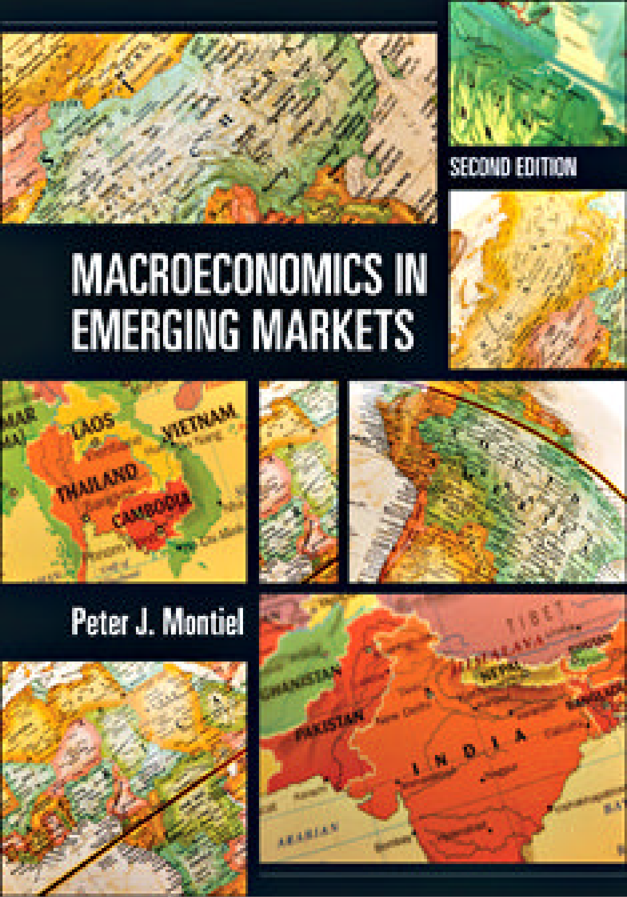
SECOND EDITION

MACROECONOMICS IN EMERGING MARKETS

Peter J. Montiel



CAMBRIDGE
www.cambridge.org/9780521514729

This page intentionally left blank

Macroeconomics in Emerging Markets

Second Edition

The macroeconomic experience of emerging and developing economies has been quite different from that of industrial countries. Compared to industrial countries, emerging and developing economies have tended to be much more unstable, with more severe boom-bust cycles, episodes of high inflation, and a variety of financial crises. This textbook describes how the standard macroeconomic models that are used in industrial countries can be modified to help understand this experience and how institutional and policy reforms in emerging and developing economies may affect their future macroeconomic performance. The second edition differs from the first in offering

- extensive new material on themes such as fiscal institutions, inflation targeting, emerging-market crises, and the Great Recession
- numerous application boxes
- end-of-chapter questions
- references for each chapter
- more diagrams, less taxonomy, and a more reader-friendly narrative
- enhanced integration of all parts of the work

Peter J. Montiel is Farleigh S. Dickinson Jr. '41 Professor of Economics at Williams College, Williamstown, Massachusetts. He formerly taught at Florida International University, Amherst College, and Oberlin College. Professor Montiel held positions as senior policy advisor at the International Monetary Fund (IMF) and as chief of the Macroeconomics and Growth Division of the Policy Research Department of the World Bank. He has served on the editorial boards of several professional journals and is a past associate editor of *World Development*. He is the author of nine books, including, most notably, the three editions of *Development Economics* (with Pierre-Richard Agénor), and of a large number of articles in professional journals. Professor Montiel has been a visiting scholar at the IMF, the World Bank, the Inter-American Development Bank, the University of Manchester, the University of Helsinki, and the Monetary Authority of Singapore.

Macroeconomics in Emerging Markets

Second Edition

PETER J. MONTIEL

Williams College, Williamstown, Massachusetts

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521733045

© Peter J. Montiel 2003, 2011

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2003
Second edition published 2011

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data

Montiel, Peter.
Macroeconomics in emerging markets / Peter J. Montiel. – 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-521-51472-9 (hardback) – ISBN 978-0-521-73304-5 (paperback)
1. Developing countries – Economic conditions. 2. Macroeconomics. I. Title.
HC59.7.M5763 2011
339–dc22 2010047125

ISBN 978-0-521-51472-9 Hardback
ISBN 978-0-521-73304-5 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

To Jana, Ruthie, and Alex

Contents

Preface

Since the publication of the first edition of *Macroeconomics in Emerging Markets* in 2003, I have used the book in my master's-level courses at Williams College and have lectured from it to a wide range of policy audiences, particularly at international financial institutions. In doing so, I became aware of several potential opportunities for improving the book that held the promise of both rendering it more user-friendly for students and substantially increasing its current policy relevance. This new edition incorporates changes that are intended to clarify and enrich the exposition in the first edition and to introduce a substantial amount of new material. The changes relative to the first edition are of six types, as follows.

I. CLEARER LANGUAGE, MORE DIAGRAMS, LESS TAXONOMY

The first edition was closely written, and frankly, many of my students found it to be terse in places. The new edition reflects a substantial amount of rewriting to make the language less terse where that seemed indicated and to develop ideas in a more gradual, step-by-step fashion. As part of this process, this new edition makes more extensive use of diagrams than the first, and the diagrams are presented more clearly (e.g., by including arrows to denote the direction of curve shifts).

An important expositional change involves the development of the baseline short-run macroeconomic model in Part 2. That model is a very rich one. It can be used to analyze the effects of shocks to the economy that are either anticipated or unanticipated and that emerge under various assumptions about the home country's degree of integration with world financial markets as well as about the alternative monetary policy regimes that the country may be operating. All of these are important considerations in the real world. Consideration of the various combinations of these factors that may arise in practice gave the analysis of the model in the first edition something of a taxonomic flavor and was sometimes confusing to students. To clarify the development of the baseline short-run model, the second edition develops the model under a single reference case (featuring anticipated shocks, imperfect capital mobility, and domestic credit targeting) that can readily be modified to consider alternative assumptions about the degree to which the domestic economy is integrated with international financial markets and about the monetary policy regime that it operates.

To clarify the model, I have also changed the description of financial market equilibrium. In the new edition, domestic financial market equilibrium is described by analyzing the domestic bond market rather than the money market. The financial-market equilibrium (FM) curve of the first edition is therefore replaced by a new bond-market equilibrium (BB) curve. Though the two approaches are equivalent, and the money market approach is undoubtedly more familiar to instructors, I have found that this change greatly simplifies the analysis for students because it allows the model to be developed without having to keep track of the effects of capital flows on the money supply, which can be complicated under imperfect capital mobility. In the new edition, the role of capital flows can be given a simple graphical interpretation. The intent of the change is to make the development of the baseline model itself much cleaner and less cluttered.

II. APPLICATION BOXES

The first edition contained empirical applications that reported on econometric research into various topics considered in the book. These have been retained, and several new ones have been added. In addition, however, the new edition contains a generous sprinkling of boxes focusing on more immediate and topical *policy* applications intended to help the material come to life for both students and instructors.

III. END-OF-CHAPTER QUESTIONS

To facilitate the use of the book as a textbook, I have included end-of-chapter questions in all but the first and last chapters of the new edition. The end-of-chapter questions are of two types: review questions and exercises. The review questions are intended to call attention to the most important concepts in each chapter. The purpose of the exercises is to induce students to think more deeply about the material by asking them to apply it in a nonmechanical way, often by analyzing hypothetical real-world situations. Answers to the exercises are available for instructors at www.cambridge.org/us.

IV. REFERENCES FOR EACH CHAPTER

In addition to end-of-chapter questions, each chapter contains suggested additional readings in "References and Further Reading" sections. These reflect a mix of policy applications, empirical work, references to the original papers from which the analysis of each chapter was developed (which are cited in the text of each chapter), and directions to more advanced work for the interested student.

V. BETTER INTEGRATION OF THE ANALYTICAL FRAMEWORK OF PART 2 WITH THE REST OF THE BOOK

The first edition did not sufficiently exploit the connection between the analytical model developed in Part 2 of the book and the rest of the material. The new edition does not lay aside the analytical model after Part 2 but rather continually revisits the model throughout the rest of the book. First, new chapters explore how the model needs to be modified to study floating exchange rates and a bank-based domestic financial system. Second, the model is explicitly used to explore some of the macroeconomic implications of the issues treated in Parts 3–7. For example, it is applied to an exploration of the macroeconomic consequences of the emergence of a sovereign risk premium (in Part 3), to an evaluation of the arguments in support of fixed or floating exchange rates (in Part 5), to an examination of how exchange rate bands work (also in Part 5), and to the study of the macroeconomic effects of the various types of emerging-market crises (in the new Part 7).

VI. NEW MATERIAL

Finally, I have used the opportunity provided by a second edition not only to update the existing material in the first edition but also to incorporate additional material into the book, some of which should have been in the first edition and some of which I have become aware of since writing that edition. The most important additions are as follows.

VI.1. The Introduction of Sticky Nominal Wages

The first edition described the nominal wage as fully flexible and explained deviations of real output from potential as the result of unanticipated shocks using a Lucas-type supply function. In the new edition, that function is treated as a special case of a more general model in which the supply wage may be sticky in the short run, allowing for the possibility of Keynesian unemployment in response to anticipated negative aggregate demand shocks.

VI.2. The Incorporation of Exchange Rate and Inflation Expectations

To understand how sustained inflation can arise in an emerging-market economy that maintains an officially determined exchange rate, Chapter 8 extends the short-run model to the medium run. To do so in an internally consistent fashion, inflation expectations are now incorporated into the GM and BB curves more explicitly than was done in the first edition of the book. To avoid new complications when these curves are first derived, however, I initially assume no inflation and then introduce it only in Chapter 8.

VI.3. Fiscal Institutions

The first edition contained a discussion of alternative means for emerging-market economies to produce a credible fiscal adjustment and to implement a sustainable fiscal policy. However, it did not deal with the political economy of fiscal policy formulation and the possible role of fiscal institutions and/or fiscal rules in securing a sustainable fiscal policy. These issues have received a substantial amount of attention in emerging-market economies in recent years, and they are included in this edition, thus rounding out the book's treatment of alternative means to achieve fiscal credibility. The new Chapter 11 is devoted to these issues.

VI.4. Monetary Transmission with Banks

Because most students from industrial countries are familiar with the intermediate-level IS-LM model, which analyzes monetary transmission in a world in which there are no banks and in which financial assets consist only of currency and interest-bearing securities, Part 2 of the book develops the benchmark model under a similar setup. However, banks are important players in most emerging-economy financial markets, and there are interesting issues posed for the channels of monetary transmission when financial intermediation is dominated by banks. The first edition contained a thorough discussion of the role of banks as financial intermediaries but did not revisit the issue of monetary transmission in the presence of banks. This edition remedies that omission by incorporating banks into the benchmark model in Chapter 23, thus providing a fairly realistic description of monetary transmission in developing and emerging economies.

VI.5. Floating Exchange Rates

Part 5 in the first edition contained an extensive discussion of exchange rate management, but the first edition did not provide a floating exchange rate version of the benchmark model. Because many emerging-market economies indeed maintain floating exchange rate regimes, the new edition provides a floating exchange rate version of the benchmark model in Chapter 17.

VI.6. Inflation Targeting

Inflation targeting has become the dominant monetary policy regime in emerging-market economies in recent years. The first edition of the book, however, did not include an extensive discussion of this topic. The new edition contains a separate chapter on the topic (Chapter 15), building on the analysis of monetary policy credibility presented in the first edition.

VI.7. Varieties of Emerging-Market Crises

A new Part 7 is devoted to various types of emerging-market financial crises. There was some material on crises in the first edition, but I have chosen to bring it together into a separate part of the book both to call attention to the importance of the topic and to give it a unified treatment. Accordingly, the second edition contains chapters on sovereign debt crises (Chapter 25), banking crises (Chapter 26), and exchange rate (currency) crises (Chapter 27) as well as a chapter on the lessons that the many crises of the last two decades have taught us about macroeconomic management in emerging and developing economies (Chapter 28).

VI.8. The Great Recession

At the time of writing (2010), the international economy is undergoing its most significant downturn since the Great Depression. Though emerging and developing economies have by no means avoided the crisis, the experience of several such economies has been more favorable this time around than in past international crises. This provides an opportunity to explore how the institutional and policy reforms explored in this book have contributed to these countries' resiliency in the face of the crisis. The book's final chapter takes up this issue.

A large number of students who have passed through the master's program at the Williams College Center for Development Economics have helped me develop the material in this book, and I am grateful to all of them. I am especially grateful for comments and excellent research assistance from George Bakradze, Mamadou Barry, Pablo Cuba, and Daniel Hernaiz.

PART 1

The Macroeconomic Framework

1

Introduction and Overview

As of 2009, the International Monetary Fund (IMF) had 192 member countries or territories. In each of these, as well as in several other countries or territories that are not members of the IMF, policy makers face the continuous need to make decisions about macroeconomic policies – decisions about fiscal policy, monetary policy, and exchange rate policy as well as about many other policies that affect the aggregate economy. The vast majority of the countries in which these decisions are made are developing countries – countries with incomes per person that are much lower than those in the advanced economies of North America, Western Europe, and East Asia. What this means is that most macroeconomic policy decisions around the world are actually made in the context of developing economies.

Though people may be the same everywhere, the economies in which they live are not. Among other things, economies differ with respect to their macroeconomic institutions, their production structures, and their economic links with the rest of the world. These factors, as well as many others that distinguish developing economies from advanced industrial economies, affect the way that economies work at the macroeconomic level. Moreover, developing countries themselves are far from homogeneous. Most important, a relatively small subgroup of such countries, typically at middle-income levels, has achieved *emerging-market* status – a term that is used to denote economies that have become closely linked financially with international capital markets. In both emerging-market economies and other developing countries, the economic environment in which macroeconomic decisions are made is quite different from the environment typically described in standard macroeconomic textbooks for advanced industrial countries.

This book is about macroeconomic analysis in emerging and developing countries. Although many of the analytical tools that we will use here are similar to those applied in industrial countries, a focus on macroeconomics in emerging and developing economies requires a change in emphasis along several directions. For example, in both emerging and developing economies, fiscal policy is often the source of macroeconomic shocks, and the solvency of the public sector (its ability to service its debt) is sometimes in question, a fact that has important macroeconomic implications. Such countries tend to be very open commercially but imperfectly (and sometimes sporadically) open financially. Similarly, only a small minority of emerging and developing economies maintains freely floating exchange rates, and the structure of financial markets in almost all of these economies is heavily dominated by commercial banks rather than securities markets. Finally, most of these countries enjoy very limited macroeconomic credibility, and their macroeconomic experience over the past several decades has been punctuated by severe instability and various types of crises, with problematic effects for their long-run growth prospects.

All these issues will be addressed in this book. As a point of departure, this chapter provides some background about the macroeconomic environment in which emerging and developing countries operate and describes some features of such economies that need to be taken into account when trying to understand how these economies work at the macroeconomic level.

I. COUNTRY CLASSIFICATIONS

When we draw distinctions, such as those referred to earlier, among different types of countries – that is, advanced industrial countries, emerging-market countries, and developing countries – to which countries are we referring exactly? One way to identify the specific countries that we might want to place in each of these groups is to make use of the classification systems employed by international organizations.

Three of the most commonly used classifications are those of the United Nations (UN), the World Bank (WB), and the IMF. The classification systems used by these institutions differ from each other because each system is designed to address the corresponding institution's specific operational needs. The UN, for example, works with economic groupings in which developing countries are classified into four main categories: least developed countries, landlocked developing countries, small-island developing countries, and transition countries (countries that recently undertook the transition from centrally planned to market economies). These categories do not exclude each other, so a country can belong to more than one category. The WB's analytical income categories, on the other hand, are based on the WB's operational lending categories, with countries divided into four groups according to their 2006 gross national income per capita: low income, lower middle income, upper middle income, and high income. The high-income category is in turn subdivided into high-income Organization for Economic Co-operation and Development (OECD) member countries and high-income non-OECD countries. Finally, the IMF's *World Economic Outlook* report divides the world into two major groups: (1) advanced economies and (2) emerging markets and developing countries. This last category is also classified according to analytical criteria that reflect the composition of

Table 1.1. *Country Classifications in Three International Organizations*

United Nations (for Developing Countries)	World Bank (Based on 2006 GNI Per Capita)	International Monetary Fund
Least developed countries	High income ($11,116 or more)	Advanced economies
Landlocked developing countries	Upper middle income ($3,596–$11,115)	Emerging-market and developing countries (subcategories by analytical group)
Small island developing states	Lower middle income ($906–$3,595)	
Transition countries [(1)]	Low income ($905 or less)	

[(1)] Countries in transition from centrally planned to market economies.

countries' export earnings, their net debtor-creditor positions, and whether they are part of the Heavily Indebted Poor Countries joint IMF-WB initiative. These classifications are summarized in Table 1.1.

We can use these classification systems to determine which specific countries fall into the advanced, emerging-market, and developing categories. Using the WB's criterion for identifying advanced countries and the IMF's criterion for separating emerging-market economies from among the remaining group of developing countries yields 34 advanced economies, 24 emerging economies, and 134 developing countries.[1] However, among the advanced economies in the WB classification, the four Asian tigers of Korea, Hong Kong, Singapore, and Taiwan are typically considered emerging economies because they reached their high-income levels relatively recently and still share many macroeconomic features with other emerging economies. Reclassifying these four countries as emerging-market economies yields the country classification presented in Table 1.2. Notice that of the 192 countries classified in the table, only 30 are advanced industrial countries. Thus the vast majority of countries (162 out of 192) are actually emerging and developing countries. Of these, 28 are emerging economies, including the large and systemically important Brazilian, Russian, Indian, and Chinese economies (collectively known as BRIC) as well as several other economies – such as Argentina, Chile, Mexico, Indonesia, Korea, and Turkey – whose recent macroeconomic histories have received a substantial amount of international attention.

II. Economic Structure and Macroeconomic Performance

Of course, the observation that macroeconomic policy decisions are primarily made in an emerging-market and developing-country context would not be very significant analytically if that context were quite similar to that for which most

[1] Emerging-market economies are developing economies that are included in the Morgan Stanley Capital International Index.

Table 1.2. *Advanced, Emerging, and Developing Economies*

Advanced Countries	Emerging Markets	Developing Countries		
Australia	Argentina	Afghanistan	Gambia, The	Niger
Austria	Brazil	Albania	Georgia	Nigeria
Belgium	Chile	Algeria	Ghana	Oman
Canada	China	Angola	Grenada	Panama
Cyprus	Colombia	Antigua and Barbuda	Guatemala	Papua New Guinea
Denmark	Czech Rep.	Armenia	Guinea	Paraguay
Finland	Egypt, Arab Rep.	Azerbaijan	Guinea-Bissau	Romania
France	Hong Kong, China	Bahrain	Guyana	Rwanda
Germany	Hungary	Bangladesh	Haiti	Samoa
Greece	India	Barbados	Honduras	Sao Tome and Principe
Iceland	Indonesia	Belarus	Iran, Islamic Rep.	Senegal
Ireland	Israel	Belize	Iraq	Serbia
Italy	Jordan	Benin	Jamaica	Seychelles
Japan	Korea, Rep.	Bolivia	Kazakhstan	Sierra Leone
Kuwait	Malaysia	Bosnia and Herzegovina	Kenya	Slovak Rep.
Luxembourg	Mexico	Botswana	Kiribati	Solomon Islands
Malta	Morocco	Bulgaria	Kyrgyz Rep.	Somalia
Netherlands	Pakistan	Burkina Faso	Lao People's Democratic Rep.	Sri Lanka
New Zealand	Peru	Burundi	Latvia	St. Kitts and Nevis
Norway	Philippines	Cambodia	Lebanon	St. Vincent and the Grenadines
Portugal	Poland	Cameroon	Lesotho	Sudan
Qatar	Russian Federation	Cape Verde	Liberia	Suriname
Slovenia	Singapore	Central African Rep.	Libya	Swaziland
Spain	South Africa	Chad	Lithuania	Syrian Arab Rep.
Sweden	Taiwan	Comoros	Madagascar	Tajikistan
Switzerland	Thailand	Congo, Democratic Rep.	Malawi	Tanzania
United Arab Rep.	Turkey	Congo, Rep.	Maldives	Togo
United Kingdom	Venezuela, Bolivian Rep.	Costa Rica	Mali	Tonga
United States		Côte d'Ivoire	Marshall Islands	Trinidad and Tobago
		Croatia	Mauritania	Tunisia
		Djibouti	Mauritius	Turkmenistan
		Dominica	Micronesia	Uganda
		Dominican Rep.	Moldova	Ukraine
		Ecuador	Mongolia	Uruguay
		El Salvador	Montenegro	Uzbekistan
		Equatorial Guinea	Mozambique	Vanuatu
		Estonia	Myanmar	Vietnam
		Ethiopia	Namibia	Yemen
		Fiji	Nepal	Zambia
		Gabon	Nicaragua	Zimbabwe

Source: Rogoff et al. (2004).

advanced-country macroeconomic models are designed. That turns out not to be the case. In general, economic theory gives us good reason to expect macroeconomic performance to be affected by a large number of country characteristics that differ significantly among advanced, emerging, and developing economies. Such characteristics include, for example, income per capita, economic size, the structure of production and trade, the composition of balance of payments flows, the degree of integration with international financial markets, the structure of the domestic financial system, the mechanism for formulating fiscal and monetary policies, the exchange rate regime, and the structure of labor markets.[2] In the rest of this section, we examine why each of these may matter.

1. Income Per Capita

The most obvious difference between advanced economies and the others is, of course, the definitional one that the former have much higher levels of per capita income than the latter. Indeed, the per capita income gap between the world's richest and poorer countries amounts to a factor of about 40! Differences in income per capita affect macroeconomic performance in a variety of ways.

First, differences in income per capita are associated with very large differences in institutional quality across countries (Easterly 1999). Many development economists in fact perceive a causal relationship between institutional quality and income per capita and attribute a large share of existing cross-country differences in income per capita precisely to differences in institutional quality (Acemoglu et al. 2004). Though these differences apply to broad institutions, such as the role of democracy versus autocracy, property rights, the rule of law, and the prevalence of corruption, they also apply to more narrowly macroeconomic institutions such as the quality and credibility of the central bank, the efficacy of budgetary institutions, and the efficiency of the civil service in administering the tax system and enforcing economic regulations.

Second, differences in income per capita are systematically related to differences in important macroeconomic relative prices such as the price of foreign goods in terms of domestic goods (the *real exchange rate*) or of today's goods in terms of future goods (the *real interest rate*).

Third, differences in income per capita should also be expected to affect the spending behavior of the different types of agents that macroeconomists typically study. Consider, for example, the behavior of the largest component of aggregate demand: private consumption. At low levels of income per capita, where households may have limited ability to save, private consumption is much more likely to track current household income than would be true at higher income levels, where households may have more discretion about when they consume.

2 For an advanced treatment of this issue, see Agenor and Montiel (2008).

Finally, there are systematic differences across income levels in the macroeconomic role of government. In low-income countries, for example, government tends to play a much larger role in production and a much smaller role in income redistribution than it does in advanced countries. This may have implications for how the government budget responds to fluctuations in the prices of the country's export commodities or in the level of domestic economic activity.

2. Economic Size

A second important characteristic is economic size. Among both advanced and developing economies, most countries (excluding the United States, the European Union, China, and possibly Japan) are sufficiently small that they must essentially take the international economic environment as given to them and not as susceptible to being influenced by their own actions. However, the vast majority of emerging and developing countries are quite small, even by these small-country standards. Small economic size means two things:

1. Small countries tend to be quite specialized in their structures of production.
2. They also tend to be relatively open commercially in the sense that they trade extensively with the rest of the world; that is, the sums of their exports and imports are large relative to the sizes of their economies. Box 1.1 shows that by this measure, emerging and developing countries have typically been more open than advanced countries.

Specialized production and openness to trade make emerging and developing economies particularly susceptible to external shocks in the form of variations in external demand for their products, which often show up in the form of fluctuations in their *terms of trade* (the relative price of their exports in terms of their imports) that are much more substantial than those typically faced by more advanced and diversified economies.

3. Structure of Production and Trade

In addition to specialization in production and commercial openness, the specific structure of production (i.e., the nature of what is produced) in emerging-market and developing countries also differs from that in advanced industrial countries. The service sector, for example, tends to be smaller in emerging-market and developing countries, and because services are typically nontraded, this means that a larger proportion of domestic output in such countries tends to be exposed to international competition.

The composition of exports also tends to be quite different. Though many emerging-market economies have become exporters of manufactured goods, such goods are typically less sophisticated than those exported by advanced industrial

Box 1.1. Real Openness in Developing and Emerging Economies

Openness to the exchange of goods and services with the rest of the world is an important feature of developing economies. Table 1.3 uses a standard measure of openness to demonstrate this: the ratio of the sum of a country's exports and imports of goods and services to its total gross domestic product (GDP). The first five rows present the average openness ratio for emerging and developing economies by region during the 1980s, the 1990s, and the period 2000–2006, whereas the sixth row presents the OECD average.

Table 1.3. *Ratios of External Trade to GDP for Five Developing-Country Regions and the OECD*

	1980–1989	1990–1999	2000–2006
East Asia and Pacific	36.5	56.0	76.0
Europe and central Asia		60.0	76.6
Sub-Saharan Africa	54.2	56.3	66.4
Middle East and North Africa	51.8	55.4	60.9
Latin America and Caribbean	27.9	35.1	45.1
OECD	35.1	36.8	43.7

Source: World Development Indicators, World Bank.

As you can see, the openness ratio for emerging and developing countries is higher than the OECD average in every period and every region, except for Latin America and the Caribbean. But by 2000–2006, that region also exceeded the OECD in its openness measure.

countries (see Hausmann and Klinger 2006), and many lower-income developing countries still have a high concentration of their exports in primary commodities. Moreover, both emerging-market economies and developing countries do much of their trade with industrial countries, implying that economic developments in such countries have a large impact on the economies of emerging-market and developing countries.

4. Balance of Payments Flows

The structure of balance of payments flows also tends to be different in emerging-market and developing countries from the structure in advanced countries. Flows such as *concessional loans* (loans offered on terms that are more generous than those generally available in the market) from bilateral and multilateral sources and receipts of workers' remittances loom much larger in the balance of payments of emerging-market and developing countries than they do in most advanced industrial economies. This again ties macroeconomic stability in such countries to external developments, because both workers' remittances and budgetary allocations for foreign assistance tend to be sensitive to economic conditions in the advanced economies.

5. Effective Degree of Financial Integration

However, a much more important difference in the balance of payments accounts of emerging-market and developing countries is the role of private capital flows. Advanced industrial countries have become highly integrated with each other financially over the past several decades, and private financial flows tend to dwarf trade flows in the balance of payments of such countries. The picture is much more mixed for emerging-market and developing countries. Standard economic theory suggests that because low-income countries are capital poor, the return to capital should be relatively high there, and such countries should be large recipients of capital inflows from advanced economies. Yet, although emerging-market economies, by definition, are active participants in private international capital markets, their participation in such markets has been sporadic in recent decades, characterized by periods of easy access to private external financing and periods when such access disappears in "sudden stops" of capital inflows, often resulting in macroeconomic crises in the countries that stop receiving such flows. The poorest developing countries, by contrast, have only very limited – if any – access to international private financial markets. Box 1.2 describes some contrasts in effective financial integration among advanced, emerging, and developing economies.

Because it has long been known that a country's effective degree of integration with world capital markets has important implications for how its economy works, this is an important difference between emerging-market and developing countries, on one hand, and advanced economies, on the other.

6. Institutional Framework for Finance and Structure of the Domestic Financial System

One of the reasons that international financial integration is so imperfect among developing countries is that the domestic institutional environment for financial intermediation is weak. Property rights are often poorly defined and weakly enforced, the legal system is unreliable (making it difficult to enforce contracts), accounting and disclosure standards are rudimentary, creditor rights are not clearly defined, and the macroeconomic environment is unpredictable. The upshot is that domestic financial intermediation is costly, and the absence of efficient intermediaries between domestic borrowers and foreign lenders restricts the domestic economy's access to private foreign capital.

Independent of its effects on the country's international financial integration, the limited development of the domestic financial system is a separate factor that affects the macroeconomics of emerging-market and developing economies. Though the

Box 1.2. Financial Integration among Advanced, Emerging, and Developing Economies

Measuring a country's effective degree of financial integration with international capital markets is not an easy task. Because financial integration involves two-way trade in financial assets between the domestic economy and the rest of the world, one way to do so is based on the extent to which domestic residents hold financial claims on the rest of the world (the domestic economy's *financial assets*) and foreign residents hold financial claims on the domestic economy (the domestic economy's *financial liabilities*). Scaling the sum of the domestic economy's financial assets and liabilities by the size of the economy provides one indication of the extent of such two-way financial trade and thus of the economy's degree of financial integration with world capital markets.

Unfortunately, this measure is imprecise, especially for poor countries, because such countries often borrow externally from nonmarket sources and because a large share of external lending by such countries is done by official agencies, especially their central banks. A refined measure of financial integration should exclude nonmarket financial transactions of these types. Such a measure of financial integration has recently been constructed by Dhungana (2008). Figure 1.1 shows his results. As is evident from Figure 1.1, ratios of private financial assets and liabilities to GDP tend to be much higher among advanced countries than among emerging and developing countries.

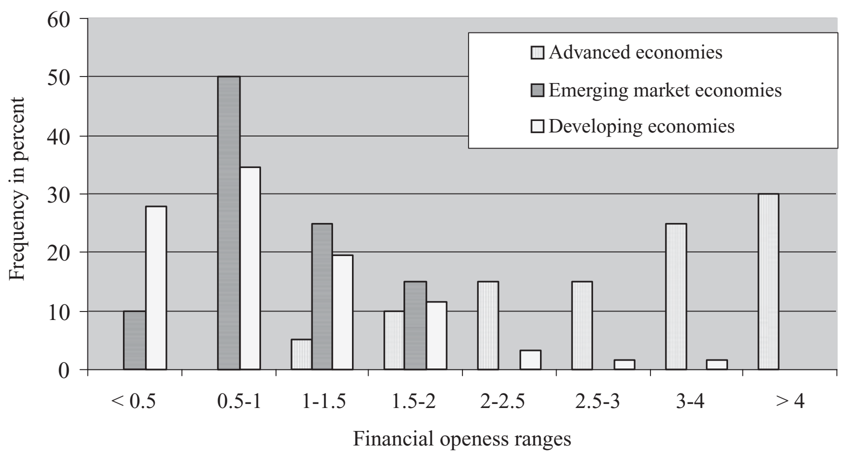


Figure 1.1. Financial openness in advanced, emerging, and developing economies. ***Source*: Dhungana (2008)**

financial systems of many emerging-market economies are relatively sophisticated, resembling those of advanced countries in having well-functioning banks, securities markets, and equity markets, those of developing countries are quite different. Many developing countries, for example, tend to have very limited securities markets, and their financial systems are dominated by banks. Policies adopted toward the banking systems in these countries often severely constrain the behavior of banks, in a policy syndrome referred to as *financial repression*. In many other cases, the banking systems are dominated by government banks.

7. Fiscal Policy

Emerging-market and developing countries have quite different fiscal structures from those of advanced industrial economies. First, as mentioned previously, the government has traditionally had a much more active role in production in such countries than in advanced industrial countries, with state-owned enterprises playing a relatively large role in the economy. Second, many emerging-market and developing countries face severe challenges in tax administration and are characterized by widespread tax evasion and persistent difficulties in raising revenue. Third, government revenue structures tend to be very different in such countries from those in advanced industrial economies, with government deriving a substantial amount of revenue from nontax sources, such as mineral royalties, and a relatively smaller amount from personal and corporate income taxes.[3] This makes tax revenues more sensitive to the terms of trade and less responsive to fluctuations in general economic activity than tends to be true in advanced industrial countries. Fourth, government budgetary institutions tend to have a larger pro-deficit bias because of the relative weakness of checks and balances in the budgetary process. Fifth, emerging-market and developing countries tend to rely more heavily on *seignorage revenue* (printing money to finance fiscal deficits) than do advanced industrial economies. Sixth – in part, for all these reasons – many emerging-market and developing countries often do not enjoy the confidence of their private creditors. This prevents such countries from being able to sustain levels of public debt (relative to the size of their economies) that are as large as those in advanced industrial economies. Finally, in part because of this lack of creditor confidence, fiscal policy often tends to be pro-cyclical in emerging-market and developing economies, with governments adopting more expansionary fiscal policies during good times and more contractionary policies in bad times.

8. Monetary Policy

Central banks in emerging-market and developing countries often have very little independence from the finance ministry, and monetary policy decisions therefore have to be subservient to fiscal ends, a situation known as *fiscal dominance*. Because of the limited development of the domestic financial system, monetary policy in many developing countries is conducted through the direct supply of credit by the central bank to the domestic banking system, and policy is often conducted

[3] For a recent investigation of the role of commodity revenues in the fiscal accounts of several Latin American countries, see Vladkova-Hollar and Zettelmeyer (2008).

by targeting specific monetary aggregates, such as the money supply or stock of domestic credit, rather than – as is more typical in advanced industrial countries – a specific short-term domestic interest rate.

9. The Exchange Rate Regime

In addition to formulating monetary policy, central banks are charged with the conduct of exchange rate policy, which involves the purchase and sale of foreign currency in exchange for domestic currency by the central bank to affect the relative price of the two currencies (the *nominal exchange rate*) in the foreign exchange market. The rules that the central bank follows in conducting these transactions are referred to as the *exchange rate regime*. In most advanced industrial countries, central banks tend to operate *floating exchange rate regimes*, which means that they limit their interventions in the foreign exchange market, leaving exchange rates to be primarily determined by the market. In many emerging-market and developing economies, however, central banks tend to play much more active roles in foreign exchange markets and to pursue much more well-defined targets for the exchange rate. In many cases, referred to as *managed exchange rates*, these targets are relatively loose and are not explicitly announced. In other cases, referred to as *fixed exchange rates*, the central bank adopts an officially announced target for the exchange rate and stands ready to buy or sell foreign exchange in unlimited amounts (at least in principle) at that price to ensure that the market exchange rate matches its official target.

The classification of countries according to their exchange rate regime is a challenging task because the exchange rate regime officially reported by a given country to the IMF (the country's de jure regime) often differs from the policy that the country actually implements (its de facto regime). Many economists have recently attempted to identify the exchange rate regimes that countries have adopted by examining what these countries actually do rather than what they say they do (e.g., Ghosh et al. 2002; Levy-Yeyati and Sturzenegger 2005; Reinhart and Rogoff 2004). It is useful for the purposes of this book to see how these de facto regimes differ among advanced countries, emerging-market economies, and developing countries.

To do so, consider the classification system of Reinhart and Rogoff (2004), shown in Table 1.4. Reinhart and Rogoff construct an exchange rate regime classification consisting of 14 regime categories, but they also present a less disaggregated "coarse" classification consisting of five types of regimes. The aggregation of the fine classification into the coarse is shown in the table, with the coarse classification in the first column and the fine in the second. The first three coarse regimes listed are essentially versions of fixed exchange rate regimes, including managed floating, which differs from the other officially determined exchange rate regimes only in that the central bank's exchange rate target is not officially announced. Floating

Table 1.4. *Exchange Rate Regimes in Reinhart and Rogoff (2004)*

1	No separate legal tender
	Preannounced peg or currency board arrangement
	Preannounced horizontal band that is narrower than or equal to ±2%
	De facto peg
2	Preannounced crawling peg
	Preannounced crawling band that is narrower than or equal to ±2%
	De facto crawling peg
	De facto crawling band that is narrower than or equal to ±2%
3	Preannounced crawling band that is wider than or equal to ±2%
	De facto crawling band that is narrower than or equal to ±5%
	Moving band that is narrower than or equal to ±2%
	Managed floating
4	Freely floating
5	Freely falling

regimes are in the fourth coarse category, whereas the fifth category is a minor category that essentially includes only high-inflation countries.

It is instructive to examine how the various types of countries we identified in the last section fit into these different types of regimes. Tables 1.5–1.7 present the classification of the advanced, emerging, and developing countries of Table 1.2 into the five coarse Reinhart-Rogoff classifications as of 2006. These tables show that as of 2006, advanced countries largely operated with floating exchange rates (category 4), whereas emerging and developing countries tended to adopt some type of fixed exchange rate regime (categories 1–3). Because it has long been known that the exchange rate regime has an important influence on how the macro economy responds to shocks, this is an important difference between advanced industrial countries, on one hand, and emerging-market and developing countries, on the other.

10. The Structure of Labor Markets

Finally, the economy's short-run supply behavior has long been one of the most contentious issues in macroeconomics, with a central question being the extent to which wages and prices respond to shocks in the short run. If prices and wages do not fully adjust to shocks in the short run, then domestic real output will tend to respond to changes in the aggregate demand for domestic goods. If they do, then changes in domestic real output will largely reflect shocks to the supply side of the economy. It is widely recognized that this is not an either-or proposition and that an economy's degree of short-run wage-price flexibility depends on a variety of institutional characteristics such as its degree of centralized wage bargaining, the extent of unionization, the prevalence of price indexation in wage setting, and so on.

Table 1.5. *Advanced Countries: De Facto Exchange Rate Regimes, 2000*

Reinhart-Rogoff "Coarse" Exchange Regime Category				
1	**2**	**3**	**4**	**5**
Cyprus			Australia	
Qatar			Austria*	
Slovenia			Belgium*	
United Arab Emirates			Canada	
			Denmark*	
			Finland	
			France*	
			Germany*	
			Greece*	
			Iceland	
			Ireland*	
			Italy*	
			Japan	
			Kuwait	
			Luxembourg*	
			Malta*	
			Netherlands*	
			New Zealand	
			Norway*	
			Portugal*	
			Spain*	
			Sweden	
			Switzerland	
			United Kingdom	
			United States	

* Countries that have adopted the euro.

Table 1.6. *Emerging Markets: De Facto Exchange Rate Regimes, 2000*

Reinhart-Rogoff "Coarse" Exchange Regime Category				
1	**2**	**3**	**4**	**5**
China	Argentina	Brazil	Korea, Rep.	
Egypt, Arab Rep.	Czech Rep.	Chile	South Africa	
Hong Kong, China	Hungary	Colombia	Taiwan	
Jordan	India	Indonesia	Turkey	
	Malaysia	Israel		
	Morocco	Mexico		
	Pakistan	Poland		
	Peru	Singapore		
	Philippines	Thailand		
	Russian Federation			
	Venezuela, Bolivian Rep.			

Table 1.7. *Developing Countries: De Facto Exchange Rate Regimes, 2000*

Reinhart-Rogoff "Coarse" Exchange Regime Category				
1	2	3	4	5
Antigua and Barbuda	Albania	Armenia	Congo, Democratic Rep.	
Barbados	Algeria	Botswana	Zambia	
Belize	Azerbaijan	Dominican Rep.		
Benin	Bangladesh	Fiji		
Bosnia and Herzegovina	Belarus	Madagascar		
Bulgaria	Bolivia	Nigeria		
Burkina Faso	Burundi	Paraguay		
Cameroon	Cape Verde	Romania		
Chad	Costa Rica	Samoa		
Congo, Rep.	Ethiopia	Syrian Arab Rep.		
Côte d'Ivoire	Gambia, The	Tanzania		
Dominica	Georgia			
Ecuador	Ghana			
El Salvador	Guatemala			
Equatorial Guinea	Guyana			
Estonia	Honduras			
Gabon	Jamaica			
Grenada	Kazakhstan			
Guinea-Bissau	Kenya			
Lebanon	Kyrgyz Rep.			
Lesotho	Latvia			
Maldives	Lithuania			
Mali	Malawi			
Niger	Mauritius			
Panama	Moldova			
Senegal	Mozambique			
Sierra Leone	Nepal			
St. Lucia	Nicaragua			
St. Vincent and the Grenadines	Papua New Guinea			
Suriname	Rwanda			
Swaziland	Seychelles, Rep.			
Togo	Slovak Rep.			
Ukraine	Sri Lanka			
	Sudan			
	Tajikistan			
	Tonga			
	Trinidad and Tobago			
	Tunisia			
	Uganda			
	Uruguay			
	Vietnam			

These characteristics are likely to differ from one economy to the other, and there is no reason to expect them to be the same in emerging and developing countries as they are in advanced industrial countries or even among countries within the separate groups of advanced, emerging, and developing economies. However, the smaller size of the industrial sector, wider prevalence of informal labor markets, and less effective enforcement of labor market regulations all have created a presumption among many economists that nominal wage-price stickiness is likely to be less prevalent among emerging-market and developing countries than it is among advanced industrial countries, although there is little hard evidence on this issue.

III. THE PLAN OF THE BOOK

As shown in the previous section, emerging-market and developing-country economies differ from those of advanced industrial countries in ways that matter for macroeconomic performance. The rest of the book investigates just how these characteristics affect macroeconomic performance. Our point of departure will be a review of the macroeconomic accounts for an open economy in Chapter 2. These definitional relationships have to hold no matter what the underlying structure of the economy, and therefore they are the same for advanced, emerging-market, and developing economies. Chapter 3 concludes the overview material by considering how the issue to be analyzed in the rest of the book, namely, the economy's short-run macroeconomic performance, affects the challenge of promoting long-run growth in emerging and developing economies.

In Part 2 of the book, consisting of Chapters 4–8, we build a benchmark model of a small, open economy that exhibits some of the characteristics identified earlier as being typical of emerging-market and developing economies – specifically, we will consider a small open economy that is imperfectly linked with international financial markets, that maintains an officially determined exchange rate, and that may enjoy substantial wage-price flexibility. We will explore how such economies work – that is, how an economy of that type responds to a variety of macroeconomic shocks – in both the short run (Chapters 4–7) and the medium run (Chapter 8). Understanding how such economies work provides us with an analytical framework that we can use to think systematically about some of the other emerging-market and developing-country macroeconomic features described in the last section.

We move to the first of these in Part 3, where we will explore a set of issues related to the formulation of fiscal policy in emerging and developing economies. Chapters 9–13 will examine the public sector's intertemporal budget constraint and the question of fiscal credibility. The issues we will consider there are the determination of how much and on what terms the public sector of an emerging or developing economy can borrow, how reduced confidence in the public sector's ability to meet its debt service obligations affects the domestic economy, and how the public sector can achieve credibility in the eyes of its potential creditors, thereby retaining the ability to borrow and avoiding either defaulting on its debts or printing money to service them. Among the measures to achieve fiscal credibility that we will consider are the adoption of fiscal rules, the design of budgetary institutions, and redefining the role of the government by privatizing state-owned enterprises. The section concludes with an analysis of what happens when the public sector finances fiscal deficits by printing money and how the result – high inflation – affects domestic macroeconomic performance.

In Part 4, we examine specific institutional reforms that make this outcome less likely, such as granting independence to the central bank and adopting a monetary policy regime that supports the central bank in achieving a low rate of inflation. We will consider both the analytical basis for believing that an independent central bank may produce lower inflation and what the evidence has to say about this issue. We will also examine the role of inflation targeting as a monetary policy regime.

Part 5, consisting of Chapters 16–19, takes up the issue of exchange rate management. The point of departure is an investigation of the notion of the equilibrium real exchange rate (the real exchange rate that is simultaneously compatible with full employment and a sustainable position in the country's external transactions) in Chapter 16, which builds on the analysis of medium-run equilibrium in Chapter 8. Chapter 17 revisits the short-run model of Part 2, showing how it would need to be modified to be applied to countries that maintain floating exchange rates. Having separately considered both officially determined and floating exchange rates, Chapter 18 discusses the choice of an optimal exchange rate regime. Given that most developing countries maintain an officially determined exchange rate, this part concludes with an analysis of how to manage an officially determined exchange rate in Chapter 19.

In Part 6, we turn to an area that has become of increased importance during recent years: the macroeconomic role of the domestic financial sector. In Chapters 20–22, we will study how the financial sector influences the economy's macroeconomic performance in so-called normal (noncrisis) times, how policies toward the financial sector can affect how well that sector performs its economic functions, and how macroeconomic considerations may affect the manner in which financial-sector reforms should be implemented in emerging and developing economies. With this expanded view of the financial system, Chapter 23 considers how monetary policy operates in an emerging-market or developing economy that is at least imperfectly linked with international financial markets when the domestic financial system is dominated by banks. This represents an extension of the benchmark short-run model of Part 2 to take into account the different, more bank-oriented structure of the financial system that typically exists in low-income countries. Finally, this part of the book concludes with an examination, in Chapter 24, of how such countries can cope with sudden large capital inflows (a phenomenon that has emerged at least three times in the last several decades (1974–1981, 1988–1996, and 2002–2008)).

Part 7 analyzes episodes of extreme macroeconomic instability in emerging and developing countries, dealing in Chapters 25–27 with debt, banking, and currency crises as well as with the interactions among these different types of crises. Chapter 28 considers the macroeconomic lessons that can be inferred from the frequent crises that emerging-market and developing economies experienced in the 1990s and 2000s. The book concludes with an examination, in Chapter 29, of how emerging and developing economies have been affected by the current international financial crisis (the Great Recession) and what that experience teaches us about the topics covered in this book.

IV. SUMMARY

This chapter has provided an overview of the issues that we will be exploring in this book. The main points are that most macroeconomic decision making in the world takes place in the context of emerging-market and developing economies and that these economies differ from advanced industrial economies in ways that, according to standard macroeconomic theory, should be expected to affect the ways they behave macroeconomically.

The chapter then laid out the plan of action for the rest of the book: we will begin by developing a model that describes how small open economies behave when they are imperfectly integrated with world financial markets, maintain an officially determined exchange rate, and are characterized by substantial domestic wage-price flexibility, three features that seem useful to incorporate into a benchmark model suitable for the analysis of macroeconomic issues in many emerging-market and developing countries. After developing this basic analytical framework, we will go on to probe in greater depth into fiscal, monetary, exchange rate, and financial-sector management issues in emerging-market and developing countries. All these tools will be brought together in an exploration of the various types of financial crises that these economies have experienced over the past two decades. The book closes with an analysis of the lessons that can be learned from these experiences for short-run macroeconomic management in emerging and developing economies.

REFERENCES AND FURTHER READING

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2004), "Institutions as the Fundamental Cause of Long-Run Growth," working paper 10481, National Bureau of Economic Research.

Agenor, Pierre-Richard, and Peter J. Montiel (2008), *Development Macroeconomics*, Princeton, NJ: Princeton University Press.

Dhungana, Sandesh (2008), "Capital Account Liberalization and Growth Volatility," unpublished manuscript, Williams College.

Easterly, William (1999), "Life during Growth," *Journal of Economic Growth*, vol. 4, pp. 239–276.

Ghosh, Atish, Anna-Marie Gulde, and Holger Wolf (2002), *Exchange Rate Regimes: Choices and Consequences*, Cambridge, MA: MIT Press.

Hausmann, Ricardo, and Bailey Klinger (2006), "Structural Transformation and Patterns of Comparative Advantage in the Product Space," working paper 128, Center for International Development.

Levy Yeyati, Eduardo, and Federico Sturzenegger (2005), "Classifying Exchange Rate Regimes: Deeds vs. Words," *European Economic Review*, vol. 49, pp. 1603–1635.

Obstfeld, Maurice (2004), "Globalization, Macroeconomic Performance, and the Exchange Rates of Emerging Economies," *Monetary and Economic Studies*, December.

Reinhart, Carmen, and Kenneth S. Rogoff (2004), "The Modern History of Exchange Rate Arrangements: A Reinterpretation," *Quarterly Journal of Economics*, vol. 119, pp. 1–48.

Rogoff, Kenneth, Aasim M. Husain, Ashoka Mody, Robin Brooks, and Nienke Oomes (2004), "Evolution and Performance of Exchange Rate Regimes," occasional paper 229, International Monetary Fund.

Vladkova-Hollar, Ivanna, and Jeromin Zettelmeyer (2008), "Fiscal Positions in Latin America: Have They Really Improved?" working paper WP/08/137, International Monetary Fund.

2

Concepts and Definitions: The Macroeconomic Accounts

The first step in constructing any macroeconomic model is to define the concepts that the model is intended to explain and to identify the definitional relationships that link those concepts to each other. Those definitional relationships are referred to as *identities* because since they simply define some variables in terms of others, they must always hold true. This chapter defines some key macroeconomic concepts that we will use throughout the book and introduces the definitional identities among them that comprise a country's *macroeconomic accounts*. The model that we will develop in Part 2 of the book will be based on this system of accounts.

The structure of the chapter is as follows: we begin by reviewing some basic conceptual underpinnings for macroeconomic models in Section I. Sections II and III describe the national income and balance of payments accounts, emphasizing the relevance of various concepts for economic welfare. Sections IV and V consider the relationship between the two types of accounts at the aggregate and sectoral levels, respectively, and Section VI summarizes.

I. CONCEPTUAL FOUNDATIONS

1. Stocks and Flows, National and Domestic

A fundamental distinction in macroeconomics is between magnitudes that are measured as stocks and those that are measured as flows. A *stock* is an economic magnitude that is measured at an instant of time. By contrast, a *flow* is measured over a period of time. An example of a stock is a country's proven reserves of some natural resource such as oil. An example of a flow is the rate at which those reserves change over time as the result of new discoveries or depletion. The stock of reserves is given at any moment of time, but the magnitude of the change in reserves depends on the period of time over which that change is measured.

Another important distinction in open economies is between the terms *national* and *domestic*. The term *national* refers to the permanent residents of a country and the factors of production they own, even if the people are abroad temporarily or the factors of production are employed abroad. The term *domestic* refers to *current* residents of a country and the factors of production that are employed there, even if the people are permanent residents of another country and the factors of production are owned by foreign residents.

2. Income and Wealth

The term *national income* refers to the amounts of goods and services produced during a period of time by factors of production owned by the permanent residents of a particular country, whether those factors of production are actually employed in their home country or abroad, whereas *domestic income* refers to the amounts of goods and services produced by factors of production located within a particular country, whether they are owned by domestic or foreign residents. Notice that both types of income refer to goods and services produced during a specific period of time and thus are *flow* magnitudes. Notice also that income refers to the value of *production*, not of consumption.

The term *national wealth* refers to the amounts of productive factors owned by the permanent residents of a particular country. Productive factors are those that can be used to generate income. Because national wealth does not have a time dimension – that is, it is measured at a point in time – it is a stock. A nation's wealth consists of the value of its stock of productive knowledge, the human capital (skills) embodied in its people, its physical capital, its natural resources, and its financial claims on the rest of the world.

The relationship between income and wealth is straightforward: in measuring a nation's income, we are essentially measuring the flow of economic benefits generated by its wealth. As we shall see later in the next chapter, a country's income is clearly influenced by its wealth, but is not completely determined by it: how efficiently that wealth is utilized to produce income is also an important factor. Indeed, several economists have investigated the extent to which differences in living standards (measured as income per person) across countries can be accounted for by differences in wealth rather than by how productively that wealth is used, and they have found that differences in productivity tend to be much more important than differences in wealth. Box 2.1 summarizes the results of a well-known study on this topic by Hall and Jones (1999).

II. THE NATIONAL INCOME ACCOUNTS

Our next task is to consider how the concept of a nation's income can be made operational, in the sense that it can actually be measured.

Box 2.1. Does Wealth Explain Income across Countries?

Income per capita differs dramatically across countries. Hall and Jones (1999) note, for example, that income per worker was 35 times higher in the United States in 1988 than in Niger and that it was 31.7 times higher on average in the five richest countries of the world than in the five poorest ones. How much of this difference in income levels across countries can be explained by differences in wealth?

Using an aggregate production function (see Chapter 4), Hall and Jones calculated that differences in wealth explain a surprisingly small part of such differences. Their results suggest that differences in physical and human capital per worker contribute factors of 1.8 and 2.2, respectively, to differences in output per worker between the five richest and five poorest countries, meaning that differences in wealth could explain approximately a fourfold difference in income levels between the two groups of countries. The implication is that by far, the most important factor in explaining income differences is how productively wealth is used to generate income. Hall and Jones estimated that productivity differences contribute a factor of 8.3 to differences in output per worker between the five richest and five poorest countries.

This finding has stood up well over time. For a more recent study that reaches a similar conclusion, see Hulten and Isaksson (2007).

1. Gross Domestic Product

The most common measure of a country's income is its *gross domestic product* (*GDP*). *GDP* is defined as the value, calculated at market prices, of the final goods and services produced in the economy during a given period. Notice that what matters is whether goods and services are *produced* during the relevant period (not whether they are *sold* in that period), and whether they have reached their final economic destination for the period, that is, they are not being used as inputs into the production of something else during the period. It is also worth noticing that by this definition, the influence on *GDP* of an increase by one unit in the production of any item depends on the market price attached to that item. Higher-valued items make a larger contribution to *GDP* than do items that sell for lower prices. This is as it should be because market prices presumably reflect the relative value placed on products by consumers.

The word *gross* refers to the fact that in calculating the total value of production, no deduction is made for the depreciation of the durable factors of production used in generating output of goods and services. If we think of a country's income as the amount that it can afford to consume without depleting its wealth, *GDP* tends to exaggerate a true measure of income because when wealth depreciates, some share of current production has to be devoted to replenishing it and thus is not available to be consumed. A measure of aggregate income that deducts the depreciation of physical capital from *GDP* is *net domestic product* (NDP).

GDP can be measured in three ways:

1. Because *GDP* does not include goods and services that are used as inputs during the period, it is equivalent to the aggregate *value added* (the difference between the value of output and the value of intermediate goods used in production) generated by domestic firms.
2. *GDP* also equals the sum of all *incomes* earned, including compensation of employees (wages and salaries plus benefits), operating surpluses of firms plus depreciation, and indirect taxes (i.e., taxes collected directly from the proceeds of final sales before income is collected, e.g., sales taxes).
3. Finally, note that current output is produced either by firms or government and that firms' output is either sold or accumulated in inventories, whereas government output is given away (e.g., public goods such as defense, police, and fire protection), sold, or accumulated in inventories. Treating the government output that is given away as sold to the government itself, *GDP* can also be measured on the *expenditure side* as the sum of *sales of currently produced goods and services plus inventory accumulation.*[1]

Aggregate expenditures are classified into the following types:

1. **Household consumption (*C*).** This refers to household purchases of consumer durables, nondurables, and services. Household consumption is typically by far the largest component of expenditures, often accounting for 60–65 percent of *GDP*.
2. **Gross private domestic investment (*I*).** This consists of purchases of newly produced plants and equipment, plus inventory accumulation, by private domestic firms and households. A typical share of *GDP* would be in the range of 20–25 percent.
3. **Government purchases of newly produced goods and services (*G*).** This includes spending by all levels of government (federal, state, or provincial, and local or municipal), whether for consumption or investment purposes. It is important to note, however, that it does not include all government spending. Specifically, it excludes payments that are not made in exchange for a currently produced good or service (called transfer payments) such as spending on income-maintenance programs and government debt services. Government spending is often around 10–15 percent of *GDP*.

[1] The term *sales* here has to be interpreted broadly because if the expenditure measure is to accurately capture all current production, it needs to include goods and services that are produced during the period in question but are not necessarily sold in the marketplace. This includes not just the accumulation of inventories as well as the goods and services that are provided for free by the government but also those that are consumed by the producing unit itself (e.g., owner-occupied housing, consumption of home-grown product on farms, and so on).

4. **Net exports** (*NX*)**.** These are sales of goods and services by domestic firms to foreigners, net of the portion of consumption, investment, and government spending that are devoted to imports. Net exports can be positive or negative and are usually in the range of ±5 percent of *GDP*.

Table 2.1 provides data on the share of these expenditure categories in *GDP* for 20 emerging and developing economies.

Using these symbols, we can therefore write the expenditure side of the national income and product accounts (NIPA) as

$$\begin{aligned} Y &= C + I + G + NX \\ &= C + I + G + X - IM \end{aligned} \tag{2.1}$$

The term $X - IM$ represents the excess of exports of goods and services (X) over imports of goods and services (IM). It is also called the *balance on goods and services*, or more informally, the *trade balance*. A country's openness to trade in goods and services with the rest of the world (sometimes referred to as real openness or goods-market openness) is often measured by its ratio of exports and/or imports to *GDP*, as illustrated in the preceding chapter.

Because *GDP* is a measure of aggregate income, international comparisons of living standards are often based on comparisons of *GDP* per capita across countries. However, such comparisons are fraught with difficulty because countries typically use different currencies, because the purchasing power of such currencies over goods and services tends to change over time, and because the relative prices of different types of goods and services tend to differ across countries. Box 2.2 describes how such comparisons are constructed.

2. Gross National Product

Some of the income produced in the domestic economy is produced by foreign capital that has been invested there or by domestic capital that has been financed by borrowing from foreigners. This generates profits and interest income that accrue to foreigners and therefore are not part of the income of domestic residents. Similarly, domestic residents may own capital that is engaged in production in other countries and may have made loans to residents of other countries that generate interest for domestic residents. To get an accurate picture of the income of domestic residents, we therefore would have to add to *GDP* the profits and interest earned by domestic funds invested abroad and subtract from it the profits and interest earned by foreign funds invested in the domestic economy. The difference between these flows is called *net investment income* (*NINV*), and when this is added to *GDP*, the result is called *gross national product* (GNP):

$$GNP = GDP + NINV$$

Table 2.1. *Expenditure Shares in GDP, 20 Countries (2007, in percentages)*

	Private Consumption	**Public Consumption**	**Total Investment**	**Private Investment**	**Public Investment** [(1)]	**Exports of Goods and Services**	**Imports of Goods and Services**
Argentina	58.6	12.9	24.2	na	na	24.6	20.3
Brazil	60.8	19.9	17.7	na	na	13.7	12.1
Chile	54.7	10.5	21.1	18.7	2.4	47.1	33.3
China	35.9	13.9	42.8	na	na	39.7	30.6
Colombia	62.0	17.0	23.0	17.6	5.4	21.6	23.8
India	55.0	10.1	37.6	36.1	1.5	21.2	24.7
Indonesia	63.6	8.3	24.9	22.0	2.9	29.4	25.4
Korea	54.4	14.7	29.4	25.7	3.7	41.9	40.4
Malaysia	45.6	12.2	21.9	10.0	11.9	110.2	89.9
Mexico	65.3	10.2	25.9	21.4	4.5	28.2	29.9
Pakistan	75.1	9.1	22.9	18.6	4.3	14.1	21.2
Russia	49.3	17.4	24.5	19.9	4.6	30.7	22.0
Thailand	53.6	12.6	26.8	na	na	73.3	65.7
Turkey	70.5	12.2	23.9	na	na	21.9	27.0
Venezuela	53.8	11.9	28.0	na	na	31.0	24.7
Average	**57.2**	**12.9**	**26.3**			**36.6**	**32.7**

[(1)] Information collected from the IMF Article IV consultation reports.
Source: IFS.

Box 2.2. International Comparisons of Living Standards

To compare living standards across countries, we need to know how many units of some *uniform* consumption basket people in different countries can afford to consume, given their income per person. The University of Pennsylvania's Center for International Comparisons has managed a long-running project to make such comparisons possible by calculating the exchange rate that would be required to equalize the price in U.S. dollars of a representative "world" consumption basket (consisting of approximately 150 different goods) in a large number of countries. This exchange rate is referred to as the *purchasing-power parity* (PPP) *exchange rate.*

To see how the PPP exchange rate makes comparisons of living standards possible, suppose that the domestic economy's income per person, measured in domestic-currency terms, is Y, and the corresponding value for the foreign country, in foreign currency, is Y^*. Then, if the price of the uniform consumption basket is P_u in domestic currency and P_u^* in foreign currency, the typical domestic resident can consume Y/P_u uniform consumption bundles, whereas the typical foreign resident can consume Y^*/P_u^* bundles. The relative standard of living is thus

$$\frac{Y^*/P_u^*}{Y/P_u} = (P_u/P_u^*)Y^*/Y = S_{\mathrm{PPP}}Y^*/Y$$

where $S_{\mathrm{PPP}} = P_u/P_u^*$ is the PPP exchange rate. This means that to compare living standards between the foreign and domestic economies, the foreign income should be converted at the PPP exchange rate S_{PPP} rather than at the prevailing market exchange rate S.

Table 2.2 provides comparisons of calculations of the U.S. dollar *gross national income* (*GNI*) per capita (defined below) for a sample of advanced, emerging, and developing countries using market-based and PPP-based exchange rates in 2006. Notice that the figures can be quite different, especially for the lowest-income countries. The reason is that in poor countries, the price of the uniform consumption bundle, measured in a common currency, tends to be much lower than in rich countries.

Thus *GDP* basically measures income generated within the national boundaries, whereas GNP measures income produced by factors of production owned by domestic residents, wherever they are located.

3. The Current Account

As shown earlier, although *GDP* is a good measure of the value of production within a country's borders, *GNP* is a better measure of the income earned by a country's residents. But another component of that income that is important in many emerging and developing economies is transfers of income from abroad such as in the form of foreign aid or workers' remittances (money sent home by workers

Table 2.2. *Dollar Values of GNI Per Capita with Market-Based and PPP-Based Exchange Rates in Selected Countries, 2006*

Country	Market-Based	PPP-Based
G-7 countries		
Canada	38,486	36,278
France	36,984	32,240
Germany	35,561	32,684
Italy	31,370	28,974
Japan	35,153	32,843
United Kingdom	39,919	33,645
United States	44,074	44,074
Emerging economies		
Argentina	5,332	11,672
Brazil	5,483	8,702
Chile	7,694	11,296
Colombia	3,236	6,129
China	2,025	4,664
Egypt, Arab Rep.	1,446	4,941
India	817	2,466
Indonesia	1,655	3,305
Korea, Rep.	18,370	22,988
Malaysia	55,195	12,156
Mexico	7,926	11,186
Pakistan	813	2,408
Philippines	1,434	3,434
Poland	8,533	14,251
Russian Federation	6,726	12,736
South Africa	5,273	8,899
Thailand	3,186	7,443
Turkey	5,513	8,410
Developing countries		
Bangladesh	423	1,230
Bolivia	1,154	3,809
Cambodia	491	1,554
Ethiopia	172	634
Guatemala	2,683	5,121
Haiti	459	1,066
Kenya	625	1,471
Nepal	326	1,006
Nigeria	696	1,407
Romania	5,483	10,154

Source: World Bank, World Development Indicators.

located abroad). The receipt of such transfers, net of transfers made by the home country to the rest of the world, is called *net unrequited transfers* (*NUT*). They are "unrequited" because they are not given in exchange for anything of market value.

Adding $NINV + NUT$ to both sides of the NIPA equation (2.1), we have

$$\begin{aligned} GNP + NUT = GNI & \\ &= Y + NINV + NUT \\ &= C + I + G + (X - IM + NINV + NUT) \end{aligned}$$

where $GNI = GNP + NUT$ is GNI, the gross income received by the country's residents. The expression in parentheses on the right-hand side of this equation is called the *current account of the balance of payments*. It is a very important concept. To see why, note that if the current account is positive, the domestic economy is exporting more, earning more investment income, and receiving more transfers from foreigners than foreigners are exporting to, earning from, and receiving from the domestic economy. In that case, the foreigners would have to pay for the difference by issuing financial claims to the domestic economy. This would represent a direct increase in the national wealth of the domestic economy. For that reason, a positive value of the current account (a current account *surplus*) is also often called *foreign investment*.

4. Real *GDP*

Since *GDP* is the sum of the *values* of production of goods and services (prices times quantities), to use changes in *GDP* as a measure of changes in the aggregate production of goods and services over time, we need to make a distinction between changes in *GDP* caused by changes in prices and those caused by changes in quantities. *Nominal GDP* is the value of *GDP* measured using *current* market prices of goods and services (i.e., the prices prevailing at the time that *GDP* is measured), whereas *real GDP* is *GDP* measured using a *reference* set (constant from year to year) of market prices. Because the latter is measured using a constant set of market prices, it is also referred to as *constant-price GDP*. Notice that because it is calculated at constant prices, real *GDP* can only change if the level of production changes. Real *GDP* is thus the conventional measure of a country's level of production.

But what set of constant prices should be used to calculate real *GDP*? There are essentially three choices: a *Laspeyres index*, which uses the prices of some past year, a *Paasche index*, which uses current-year prices, and a *Fisher index*, which uses the average of a past year's prices and current prices. These will give different measures of real *GDP* when relative prices change over time and production of different goods and services grows at different rates, because the contribution to the growth rate of real *GDP* of growth in the production of any particular good or service depends on the relative price attached to that good or service. To see the problem, notice that if some specific type of production experiences rapid productivity growth over time, and thus a declining relative price, a Laspeyres index would tend to overstate real *GDP* in recent years when that type of production is large because by relying on the relative prices that prevailed when that type of output was relatively expensive, it would be assigning too large a weight to that sector. In contrast, a Paasche index would tend to overstate real *GDP* in the past as it gives more weight to sectors that were relatively large then and have a high relative price today. Under these circumstances, use of a Laspeyres index would tend to overstate growth of real *GDP* over time, whereas a Paasche index would tend to understate it.

For this reason, the Fisher index has come into increasingly common usage. The resulting measure of aggregate production is called *chained real GDP*. It is constructed as follows:

1. For each of two consecutive years, a measure of *GDP* is constructed using the average price of each good or service in the two years.
2. The rate of growth of *GDP* is then calculated from the first year to the second. Once this is done for all years, a *time series* of real *GDP* growth rates results, showing the growth rate of real *GDP* in each year relative to the year before.
3. From the series of growth rates, an index for the *level* of real *GDP* can be constructed by setting the value of any one year (say, 1996) to unity and applying the growth rates to compute the relative value of real *GDP* in every other year.
4. Multiplying this index by the level of nominal *GDP* in 1996 yields a time series for the level of *chained* (*1996*) real *GDP*.

Though the resulting measure of real *GDP* is intended to indicate the level of production in a country, it is often treated as a measure of economic welfare. Box 2.3 provides some reasons to be cautious about using real *GDP* in that way, while Box 2.4 reviews the results of a recent study examining the relationship between real *GDP* per capita and a variety of quality-of-life indicators.

5. The Price Level

Up to now, we have been concerned with measures of aggregate production and income as well as their components. A second important type of macroeconomic variable has to do with the average level of money prices (prices measured in units of domestic currency) in the economy. This concept is measured using a variety of price indices, in which the average level of prices is expressed relative to that of some base year.

The broadest price index for an economy is the *GDP deflator*. This is simply the ratio of each year's nominal *GDP* to the same year's real *GDP*, multiplied by 100:

$$(GDP\ deflator)_t = (NGDP_t/RGDP_t) \times 100$$

where *NGDP* and *RGDP* refer to nominal and real *GDP*, respectively. Notice that the *GDP* deflator will be equal to 100 in the year that is used as the base for calculating the real *GDP* series. The *GDP* deflator measures the average price level of the goods and services produced by the economy during a given year.

Box 2.3. Real *GDP* and Economic Welfare

Real *GDP* was designed as a measure of a country's *production*. For a variety of reasons, it is at best an imperfect measure of its *economic welfare*. For example, real *GDP* ignores the following:

1. **The value of leisure consumed by the country's residents.** That leisure has economic value is confirmed by the fact that people have to be compensated to give it up. One way to value leisure is by its *opportunity cost*, that is, the market wage that could otherwise be earned during leisure time.
2. **The value of household production.** In many countries, measures of *GDP* include *imputations for owner-occupied housing* and *food consumed on farms* but not for other types of household production such as that by homemakers. This is an important item. The United Nations has estimated that including the value of household production would add 25–30 percent to the real *GDP* of industrial countries.
3. **The value of natural resources depleted and environment befouled.** These represent national wealth that is lost to future generations.
4. **Investment in intangible assets (research and development, advertising, software).** These expenditures by firms are intended to produce intangible capital that can increase the firm's future profits. However, the national income accounts treat them as spending on intermediate goods and services rather than as purchases of new, intangible capital. In other words, they are not assumed to produce a type of output that is separate from that represented by each firm's standard product line. The effect is to lower current output, investment, and corporate saving relative to a situation in which the accumulation of knowledge is treated as a separate output produced by the firm. Nakamura (2001) estimated investment in intangibles at $700 billion to $1.5 trillion for the United States in 2001. Because most research and development is done in industrial countries, this factor is likely to be less important in developing countries.
5. **That some goods produced (defense, police services, etc.) may actually be preserving utility, not adding to it.** Others (e.g., commuting costs) may actually be more appropriately treated as intermediate rather than final goods.
6. **Income distribution.** The value of real *GDP* per person may be high, yet most people in a given country may still be very poor if income is unevenly distributed.

The classic discussion of these issues is by Nordhaus and Tobin (1972), but see also Eisner (1988). A recent attempt to construct an alternative to *GDP* as a measure of income (the Genuine Progress Indicator) is described in Talberth et al. (2006).

An alternative and frequently used measure of the price level is the *consumer price index* (CPI). The CPI is an index of the money cost of purchasing a representative consumption bundle, again expressed relative to some base year. It is compiled as the cost of a given basket of goods and services, intended to represent the average spending pattern of a typical consumer. The prices of the items in the basket are

Box 2.4. Life during Growth

Though there are many reasons why real *GDP* may be an imperfect indicator of economic welfare, many economists have found that more direct measures of economic welfare actually tend to be highly correlated with real *GDP* per capita, in the sense that countries with high levels of *GDP* per capita also tend to exhibit better indicators of quality of life. Even if real *GDP* per capita is not a direct indicator of economic welfare, could this mean that increases in the level of real *GDP* per capita cause improvements in quality of life so that the link between real *GDP* per capita and economic welfare is an indirect one? The answer is maybe, but not necessarily. It could be, for example, that the cross-country association between higher levels of real *GDP* and better quality of life indicators arises because countries with better quality of life are able to achieve higher levels of real *GDP* per capita (so that causation runs from quality of life to real *GDP* per capita, rather than vice versa) or because country-specific factors simultaneously favor a high quality of life and a high level of real *GDP* per capita in some countries.

A recent comprehensive study of this issue was by Easterly (1999). Easterly looked at 81 indicators of quality of life in seven areas (individual rights and democracy, political instability and war, education, health, transport and communications, inequality across class and gender, and so-called bads) in 1960, 1970, 1980, and 1990. He examined the extent to which the behavior of each indicator across countries and over time could statistically be explained by real *GDP* per capita, after controlling for time- and country-specific factors.

Using his simplest statistical approach, Easterly found that 61 of his quality-of-life indicators showed a positive relationship with real per capita income, 12 a negative relationship, and only 8 no relationship at all. But when he used more sophisticated statistical techniques that controlled for factors specific to each country that could affect both the quality of life and income per capita, thereby accounting for the earlier association, as well as for possible reverse causation from the quality-of-life indicator to the level of income, the results became much weaker, and there were very few quality-of-life indicators that proved to be robustly linked to income per capita for all three estimation techniques. Easterly speculated that possible explanations could be based on long lags between growth and improvement in quality of life, on the role of so-called third factors driving both per capita incomes and more direct measures of well-being, or on the role of factors that operate at the global level on measures of well-being in individual countries.

typically measured on the basis of monthly surveys of retail establishments. The CPI differs from the *GDP* deflator essentially because the composition of household consumption differs from that of production in the country in several ways. For example, international specialization means that a country's typical export goods tend to carry a much larger weight in production than they do in consumption baskets within the country and that households tend to consume a large number of goods that are not produced within the country. The prices of imports would thus affect the CPI but not the *GDP* deflator. In addition, firms and governments buy a substantial share of a country's *GDP*, and the composition of what they buy (e.g., capital goods and the services of government bureaucrats) tends to be very different from that of household consumption. The CPI is an important measure of the price level because it provides an approximate measure of changes in the *cost of living*, that is, in the money cost of attaining a given level of economic welfare. It is by no means a perfect measure of the cost of living, for reasons spelled out in Box 2.5. Nonetheless, in many countries, workers base changes in their nominal wage demands on changes in the CPI, a variety of government payments and taxes are *indexed to* (automatically adjusted with) the CPI, and many central banks conduct their policies so as to achieve a desired outcome for the CPI.

6. Potential *GDP*

The proportional change in the CPI from the previous period to the current one is called the *rate of inflation*, and securing a low rate of inflation is an important goal of macroeconomic policy in all countries. It is interesting to consider why this is so. One reason that is often mentioned in popular discussion – because higher prices reduce everyone's real income (purchasing power over goods and services) – surely cannot be right because a uniform increase in all prices would increase nominal *GDP* by the same amount and thus raise everyone's income by an amount that is exactly sufficient to compensate for the increase in prices, leaving real purchasing power the same. Instead, the problem is precisely that when the average level of prices rises, all prices do *not* tend to rise at the same time. This is partly because some payments are stipulated ahead of time in nominal terms so that when higher inflation unexpectedly arrives, the *real* value of those payments falls (so the payee loses, while the payer gains), and partly because price changes tend not to be synchronized across sectors of the economy. When all prices do not move up together, arbitrary redistributions of real income take place. This lack of synchronization also distorts relative prices, on which economic decisions depend. Finally, the lack of synchronization in price changes also creates uncertainty, which increases the cost of doing business. These issues are all discussed in more detail in Chapter 3.

Fluctuations in real *GDP* over time can arise either from changes in a country's *capacity* to produce or from changes in the extent to which it is fully using its productive capacity. Countries may not fully use their productive capacity when the domestic price level is sticky and there is deficient aggregate demand for the goods that they produce. To distinguish between an economy's *actual* level of real output and its level of productive capacity, economists use the term *potential GDP* to denote productive capacity.[2] Potential *GDP* is simply the level of real *GDP* that

[2] Other terms in wide use are *full-employment GDP* and the *natural level of output*.

Box 2.5. The CPI and the Cost of Living

The CPI suffers from several biases that make it an imperfect measure of the cost of living (the money cost of achieving a given level of economic well-being). These biases include the following:

1. **Substitution bias.** The CPI assumes that when the price of a commodity goes up, consumers can maintain their level of well-being only by continuing to consume the same basket of commodities. The increase in the CPI is the extra amount of spending required to do that when prices rise. But it may cost consumers less to maintain their level of well-being by simply switching to other commodities that have not undergone price increases. In that sense, the CPI overstates the increase in the cost of living – that is, of maintaining the same level of well-being – associated with increases in the prices of specific goods.

   Another way to look at this is that as relative prices change, the shares of spending on items whose relative prices rise may decrease. This makes the weights unrepresentative. You can see that the severity of the problem would tend to depend on the degree of substitutability among different goods and services as well as on the magnitude of relative price changes.
2. **Quality bias.** Quality improvements are imperfectly measured in CPIs. However, it is unclear whether this overstates or understates the price level because some national statistical agencies make corrections to take quality changes into account.
3. **New goods bias.** The introduction of new goods may make the CPI measure unrepresentative. It may also overstate inflation because new goods often tend to experience rapid price decreases after they are initially introduced.
4. **Outlet substitution bias.** Shopping at discount stores is often not taken into account in CPI construction.
5. **List versus transaction price bias.** Similarly, sales, discounts (e.g., cars), rebates, promotions (e.g., giveaways), quantity discounts, coupons, and so on, may also be missed.

These biases have been judged on net to cause the CPI to overstate the rate of inflation. There is debate over the size of this bias. Wynne and Sigalla (1994) argued that it is probably around 1 percent per year in the United States. The 1995 Boskin Commission created by the Senate Finance Committee estimated it at about 1.1 percent per year: 0.4 percent from substitution bias, 0.6 from quality and new goods bias, and 0.1 from outlet bias. The Bureau of Labor Statistics began to address substitution bias in 2002 by doing expenditure surveys every two years and producing a chained version of the index.

an economy would produce if its rate of capacity utilization were at its so-called normal level, that is, without under- or overemployment of its productive resources.

Unlike the other concepts defined in this section, potential *GDP* cannot be observed directly. Thus it has to be estimated using statistical techniques. There is a wide variety of ways of doing so. Most of these techniques rely on the consensus among macroeconomists that economies contain automatic adjustment

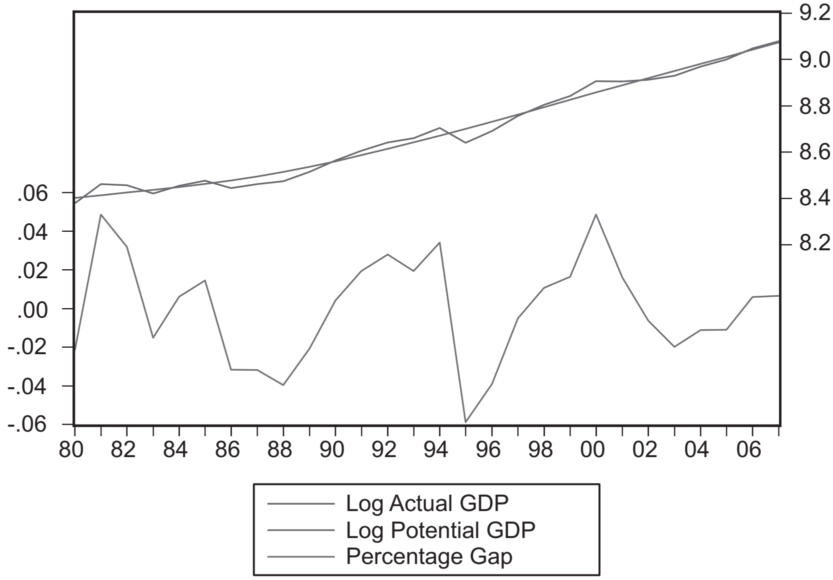


Figure 2.1. Mexico: Actual and potential *GDP*

mechanisms that cause deviations of actual from potential *GDP* to be transitory. A common technique relying on this perspective is to smooth the time series for the logarithm of real *GDP* using any of several techniques for filtering out high-frequency movements in the series such as the Hodrick-Prescott filter. Figure 2.1, for example, shows plots of (the logs of) actual real *GDP*, potential *GDP*, and the percentage gap between the former and the latter for Mexico from 1980 to 2007, constructed using a simple routine in the software Eviews. Notice that by this estimate, Mexico was producing below capacity in the second half of the 1980s, in 1995–1996, and again in 2003–2004. Later in this book, we will examine some of these episodes in Mexico's macroeconomic experience.

III. THE BALANCE OF PAYMENTS ACCOUNTS

In addition to the national income accounts, macroeconomic analysis draws on a wide range of concepts defined in a country's *balance of payments* (BOP) accounts. This is the system of accounts that keeps track of payments to and receipts from foreigners. The BOP accounts record all trade between domestic and foreign residents in goods and services as well as in financial assets. The entries in the BOP accounts, like those in the national income and product accounts, represent economic flows rather than stocks.

The BOP accounts obey the basic rule of double-entry bookkeeping: every international transaction enters the BOP accounts twice, once as a credit and once as a debit. Any transaction resulting in the receipt of something of value

Table 2.3. *Mexico: International Transactions, 2007 and 2008 (billions of U.S. dollars)*

	2007	2008
Exports of goods	271.9	291.3
Imports of goods	−281.9	−308.6
Exports of services	17.6	18.6
Imports of services	−24.1	−25.3
Income receipts	7.9	7.1
Income payments	−25.9	−24.6
Unilateral current transfers, net	26.4	25.4
Government grants	na	na
Government pensions and other transfers	na	na
Private remittances and other transfers	23.8	23.5
Unilateral capital transfers, net	0.0	0.0
Domestic-owned assets abroad, net (increase/financial outflow (−))	−40.3	−16.1
Official reserve assets, net	−10.3	−7.7
Government assets, other than official reserve assets, net	0.0	0.0
Private assets, net	−30.0	−8.4
Foreign-owned assets in Mexico, net (increase/financial inflow (+))	50.8	29.8
Foreign official assets in Mexico, net	0.0	0.0
Other foreign assets in Mexico, net	50.8	29.8
Statistical discrepancy (sum of preceding items with signs reversed)	2.3	−2.3

Source: IMF, International Financial Statistics.

by domestic residents is entered as a debit (with a negative sign), whereas any transaction resulting in the giving up of something of value by domestic residents is entered as a credit (with a positive sign). Because every transaction involves an exchange of equal (market) value, every transaction recorded in the BOP accounts generates a debit and an offsetting credit. Thus the sum of all transactions in the BOP accounts must be equal to zero.[3]

Transactions recorded in the BOP accounts are classified into six types: exports (+) and imports (−) of goods, exports (+) and imports (−) of services, receipts (+) and payments (−) for the services of factors of production (investment income), unrequited transfers received (+) and paid (−), sales (+) and purchases (−) of financial assets other than by central banks, and sales (+) and purchases (−) of financial assets by central banks. As an illustration, Table 2.3 presents the entries for each of these items for Mexico in 2007 and 2008.

Sales of financial assets to foreign residents by domestic residents are called *capital inflows*, whereas purchases of financial assets by domestic from foreign residents are called *capital outflows* (−). When a domestic resident sells a financial claim on a foreign resident to a foreign resident, this is often called *capital repatriation.* There are several subcategories of financial transactions. When claims have original maturity of less than a year, they are considered short term, and trading in such claims is referred to as *short-term capital flows.* When claims have original

[3] The exception is gifts, and to maintain the double-entry principle, the value exchanged in this case is conventionally assumed to consist of "goodwill."

maturities of more than a year, or when they have no specified maturity (e.g., equity claims), they are dubbed *long term*. When sales of equity transfer control over an enterprise to a foreigner, or increase the equity of a foreigner who already controls the enterprise, this is referred to as *foreign direct investment* (FDI; a type of long-term capital inflow) into the country where the enterprise is located.

The BOP accounts are organized into several subaccounts. Unlike the sum of all BOP transactions, the sums of transactions in these subaccounts need not be zero because, since they only contain a subset of all transactions, the balancing transaction may lie outside the included subset. For the purposes of this book, the key subaccounts are the following.

1. The Balance on Goods and Services

This consists of exports of goods and services net of imports of goods and services. We referred to this subaccount as *net exports* (NX) in Section II and expressed it as

$$NX = X - IM$$

2. The Current Account

As described in Section II, the current account (CA) consists of the balance on goods and services (NX) plus net investment income ($NINV$) and net unrequited transfers (NUT):

$$CA = NX + NINV + NUT$$

3. The Capital Account

The capital account is the difference between purchases of domestic assets by foreign (non–central bank) residents plus sales of foreign assets by domestic (non–central bank) residents (capital inflows; KI) and purchases of foreign assets by domestic (non–central bank) residents plus sales of domestic assets by foreign (non–central bank) residents (capital outflows; KO)[4]:

$$KA = KI - KO$$

4. The Official Reserve Settlements Balance

Foreign assets held by central banks are called *official international reserves*. Purchases and sales of foreign assets by central banks are called *official foreign exchange*

[4] Although the capital account has recently been renamed the *financial account* in the IMF's *Balance of Payments Manual* (which is the standard reference for compilation of balance of payments statistics), the older term remains in much wider usage.

Table 2.4. *Mexico: Subaccounts of the Balance of Payments, 2007 and 2008 (billions of U.S. dollars)*

Current account (CA)	−8.2	−16.0
Balance on goods, services, and investment income	−34.6	−41.4
Balance on goods and services (*N*)	−16.5	−23.9
Balance on merchandise trade	−10.1	−17.3
Exports of goods	271.9	291.3
Imports of goods	−281.9	−308.6
Balance on services	−6.5	−6.7
Exports of services	17.6	18.6
Imports of services	−24.1	−25.3
Net investment income (NINV)	−18.1	−17.5
Income receipts	7.9	7.1
Income payments	−25.9	−24.6
Unilateral transfers, net (NUT)	26.4	25.4
Capital account	20.8	21.5
Official reserves settlements balance (ORS)	−10.3	−7.7
Mexican official reserve assets, net	−10.3	−7.7
Foreign official assets in Mexico, net	0.0	0.0
Statistical discrepancy	2.3	−2.3

interventions. The net change in official reserve assets held by the domestic central bank is called the official reserve settlement balance. Call this *ORS*. The relationship among these concepts can be expressed through the *balance of payments identity*, presented for Mexico in Table 2.4.

Because the sum of all the transactions in the BOP accounts must (aside from statistical discrepancies, sometimes referred to as *errors and omissions*) be zero, we have

$$CA + KA + ORS = 0$$

There are two important points that can be made on the basis of this accounting.

First, from the preceding identity, a current account surplus must be offset by a deficit on the capital account plus the official reserves settlements balance, and a current account deficit must be offset by a surplus on the capital account plus the official reserves settlements balance:

$$CA = -(KA + ORS)$$

Thus, if *CA* is positive, *KA* + *ORS* must be negative and equal in magnitude; that is, a current account surplus represents an exactly equal accumulation of claims on the rest of the world. On the other hand, if the current account is negative, there is an equal accumulation of claims by the rest of the world on the domestic economy. Thus, as indicated in Section II, the current account matters because it determines the change in the economy's net international investment position (its net financial claims on the rest of the world).

Second, economists often use the term the *balance of payments* to refer to the sum of the current and capital subaccounts of the BOP accounts. Thus

$$BOP = CA + KA$$

From our previous identity, this means that BOP surpluses and deficits correspond to increases and decreases in the central bank's foreign exchange reserves, respectively. Because foreign exchange reserves are used to buy the domestic currency in the foreign exchange market when the domestic central bank deems it necessary, *BOP* deficits matter because decreases in such assets signify a reduced ability on the part of the domestic central bank to buy its own currency with foreign currency.

IV. NATIONAL INCOME AND PRODUCT ACCOUNTS IN AN OPEN ECONOMY: AGGREGATE IDENTITIES

In the preceding two sections, we examined the NIPA and the BOP accounts separately. Next we examine how these accounts are related to each other.

In a closed economy, we can write the basic NIPA expenditure identity as

$$Y = C + I + G$$

In this context, *gross national saving* (which we will call S_N) is defined as the difference between *GNP* and total domestic consumption. If we assume that all government spending is for public consumption purposes, we can write S_N as

$$S_N = GDP - C - G$$

or

$$S_N = I$$

Note that in a closed economy, wealth accumulation (saving) must take the form of the accumulation of real productive assets (investment).

By contrast, in an open economy, as we have seen, the basic NIPA identity becomes

$$\begin{aligned} GNP + NUT = GNI \\ &= (C + I + G) + (NX + NINV + NUT) \\ &= (C + I + G) + CA \end{aligned}$$

The sum $C + I + G$ is often called *domestic absorption.* It denotes total spending by domestic residents, whether on domestic or foreign goods. If we let A denote absorption, so $A = C + I + G$, we can also write that

$$CA = GNI - A$$

Thus the current account is the difference between a country's income and the goods and services that it absorbs.

Assuming once again that all government spending is devoted to consumption, in an open economy, our definition of saving must be modified to the following:

$$S_N = GNI - C - G$$

This allows us to write a very important identity relating national saving, investment, and the current account:

$$S_N = I + CA$$

As it turns out, this identity yields some important economic insights. It shows, for example, that whereas a closed economy can only accumulate wealth by building real (nonfinancial) capital, an open economy can do so by building capital or acquiring claims on the rest of the world. Domestic investment and current account surpluses are two different ways in which a country can use current output to increase its future income. Notice that wealth accumulation equals saving, no matter what form wealth accumulation takes.

How *should* an economy accumulate wealth, from a social welfare perspective? The answer is that the size of the national pie is maximized if wealth is accumulated in the form that yields the highest social return, whether it be real investment in the domestic economy or the accumulation of financial claims on the rest of the world.

The identity derived earlier can be written two other ways, both of which offer additional insights. First, writing it as

$$I = S_N - CA$$

it tells us that whereas a closed economy must finance investment by saving, an open economy can do so either by saving or by reducing its net foreign wealth (borrowing abroad). On the other hand, if we write it as

$$-CA = I - S_N$$

it says that the current account deficit ($-CA$) is the difference between investment and saving. Thus an increase in the current account deficit can be created if a country either begins to invest more without financing the extra investment with its own saving or consumes more (saves less) without cutting back on investment.

V. NATIONAL INCOME AND PRODUCT ACCOUNTS IN AN OPEN ECONOMY: SECTORAL IDENTITIES

In addition to the aggregate identities described in the preceding section, there are some useful national income accounting identities that operate at the sectoral level. For present purposes, we can confine our attention to a broad distinction between the private and public sectors. In later chapters, we will consider other sectoral breakdowns.

For simplicity, assume that the private sector owns all the factors of production and thus receives all the economy's *GNP* as factor income, while the public sector's income consists only of the net taxes paid to it by the private sector, which we will denote T. Continuing to assume that all government spending consists of public consumption rather than investment, we can define the saving of the private sector S^P and that of the public sector S^G, respectively, as follows:

$$S^P = \mathrm{GNP} - T - C$$
$$S^G = T - G$$

Adding these two identities together yields

$$\begin{aligned} S^P + S^G &= GNP - C - G \\ &= I + CA \end{aligned}$$

This permits us to write the following:

$$\begin{aligned} S^P &= I + CA - S^G \\ &= I + CA + (G - T) \end{aligned}$$

This form of the identity tells us that private saving must be used to finance domestic investment, foreign investment, or the government deficit.

We can also write this last identity as follows:

$$CA = (S^P - I) + (T - G)$$

Thus the current account surplus is equal to the excess of private saving over investment plus the fiscal surplus. This relationship is often interpreted to imply that an increase in the fiscal deficit – that is, some combination of decrease in T and/or increase in G – must be associated with an equal reduction in the current account surplus or an increase in the current account deficit. But this conclusion cannot be justified on the basis of what remains just another identity. That means that it must hold no matter what the economic relationships might be among its constituent parts. In turn, this implies that this relationship cannot tell us anything about those economic relationships. What the identity can tell us is that for this conclusion to be valid, the excess of private saving over investment must remain unchanged in response to the change in the fiscal deficit. What it cannot tell us is the conditions under which this would happen.

Finally, we can also write the sectoral saving identity as follows:

$$I = S^P + S^G - CA$$

that is, domestic investment must be financed by some combination of private saving, public saving (a fiscal surplus), and foreign saving (a current account deficit).

Box 2.6. Sources of Macro Data for Developing Countries

Several international organizations maintain comprehensive macroeconomic data sets for developing countries. The following table presents a brief description of these data sets and their locations on the Web.

Institution/ Database	Description	Institutional Web Page
International Monetary Fund: International Financial Statistics	Time series covering more than 200 countries starting in 1948; includes data on exchange rates, prices, money aggregates, national accounts, balance of payments, and a series of economic indicators	http://www.imf.org/
International Monetary Fund: World Economic Outlook	Individual country and country-group time series data for *GDP* growth, inflation, unemployment, exports, imports, external debt, capital flows, and commodity prices	http://www.imf.org/
World Bank: World Development Indicators	Data for more than 150 countries that cover more than 800 indicators in six sections: World View, People, Environment, Economy, States and Markets, and Global Links	http://www.worldbank.org/
World Bank: Global Development Finance	External debt and financial flow data for 135 countries; indicators include external debt stocks and flows, major economic aggregates, and key debt ratios	http://www.worldbank.org/
Organization for Economic Co-operation and Development	Country time series on a wide range of indicators, including production, health, education, communications and technology, and others	http://www.oecd.org/
Bank for International Settlements	International data on the global financial system	http://www.bis.org/
Penn World Tables	Purchasing power parity and national income accounts converted to international prices for 188 countries for some or all of the years 1950–2004	http://pwt.econ.upenn.edu/

This completes our overview of the main macroeconomic concepts that we will be using throughout this book. Data on the concepts we have described so far are usually available from the Web sites of the central banks of emerging and developing countries. Several international organizations also maintain macroeconomic data sets for large numbers of developing countries. Box 2.6 describes some of these data sets and where they can be found.

VI. SUMMARY

Every macroeconomic model is based on a set of definitions and accounting relationships. In this chapter, we have defined the variables that will be of primary interest to us in the rest of this book and have reviewed the main accounting relationships that link them together in each country's national income and product accounts as well as its BOP accounts. These accounting relationships not only help us define terms but also help to discipline our thinking by reminding us of key relationships that have to hold continuously among the macroeconomic variables that we are trying to understand.

However, because these relationships are just accounting identities, they have nothing to say about cause and effect among those variables. To describe cause-and-effect relationships among these variables, we need an economic model. Chapters 4–7 will develop that model, which we will be using throughout the rest of this book. Before moving to that, however, it is worth spending some time examining the relationship between the short-run macroeconomics that will be our primary concern in this book and the long-run growth of an economy's productive capacity that is the primary concern of development economists. That is the task of the next chapter.

REVIEW QUESTIONS

1. Explain the difference between national wealth and national income. What happens to each of these variables when a nonrenewable mineral resource is mined by the domestic economy and exported to the rest of the world?
2. Why is real *GDP* an imperfect measure of economic welfare? If it is indeed so imperfect, why is it so widely used?
3. Why is the current account of the balance of payments an important concept? Why is the overall balance of payments important?
4. What is the difference between the capital (financial) account of the balance of payments and the official reserves settlement balance?
5. Describe three alternative insights that can be gleaned about the differences between closed and open economies from the three ways of expressing the relationship among saving, investment, and the current account of the balance of payments.

EXERCISES

1. Suppose that a domestic resident takes out a loan in foreign exchange from a foreign bank and uses the proceeds of the loan to buy domestic currency from the domestic central bank. Explain how these transactions would affect the domestic economy's balance of payments accounts.

2. Does an increase in government spending by $1 increase an economy's current account deficit by $1? Explain why or why not.
3. In summer 2010, it was announced that Afghanistan had vast unsuspected mineral resources underground, amounting to an estimated US$1 trillion.
   a. Explain how this finding affects Afghanistan's national wealth and income in 2010.
   b. As these minerals are gradually mined and sold in the world market, explain how you would expect Afghanistan's national income accounts to be affected, and why. (Hint: there are several possible answers. What matters is that your answer be internally consistent.)
4. Journalists visiting developing countries often report levels of income in such countries, expressed in U.S. dollars, that seem impossibly low, in the sense that it would not be possible to survive in a country like the United States on earnings at such levels. Are such journalists misinformed, or is there an alternative explanation?
5. Suppose that a firm in Bolivia sells US$1 million in tin to an importer in the United States. The Bolivian firm deposits half of the proceeds of the sale in a U.S. bank and sells the remaining dollars to a Bolivian commercial bank, which exchanges it for domestic currency (pesos) at the Bolivian central bank. Explain how these transactions would affect the Bolivian balance of payments.

REFERENCES AND FURTHER READING

Easterly, William (1999), "Life during Growth," *Journal of Economic Growth*, vol. 4, pp. 239–276.

Eisner, Robert (1988), "Extended Accounts for National Income and Product," *Journal of Economic Literature*, vol. 26, pp. 1611–1684.

Hall, Robert E., and Charles I. Jones (1999), "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics*, vol. X114, pp. 83–116.

Hulten, Charles R., and Anders Isaksson (2007), "Why Development Levels Differ: The Sources of Economic Growth in a Panel of High and Low Income Countries," working paper 13469, National Bureau of Economic Research.

Nakamura, Leonard (2001), "Investing in Intangibles: Is a Trillion Dollars Missing from *GDP*?" *Philadelphia Federal Reserve Bank Business Review*, vol. Q4, pp. 27–37.

Nordhaus, William, and James Tobin (1972), "Is Growth Obsolete?" in *Economic Growth*, National Bureau of Economic Research Series 96E, New York: Columbia University Press, pp. 1–80.

Talberth, John, Clifford Cobb, and Noah Slattery (2006), *The Genuine Progress Indicator 2006: A Tool for Sustainable Development*, Oakland, CA: Redefining Progress.

Wynne, Mark A. and Fionna D. Sigalla (1994), "The Consumer Price Index," Federal Reserve Bank of Dallas *Economic Review* (Second Quarter), pp. 1–22.

3

Short-Run Macroeconomics and Long-Run Growth

Why should someone who is primarily interested in long-run growth in emerging and developing economies be concerned with the short-run macroeconomic performance of those economies? This is an important question for this book to address because the chapters that follow will all be concerned with short-run macroeconomic performance in countries whose primary aspiration is to improve their long-run growth performance. The answer is that a country's short-run macroeconomic performance can have an important effect on its long-run growth rate. Indeed, over the past two decades, a significant consensus has emerged among professional economists and policy makers in developing countries that certain aspects of short-run macroeconomic performance – getting key macroeconomic relative prices right and providing a stable and predictable domestic macroeconomic environment – play important roles in inducing the accumulation of productive factors and improvements in productivity that are the basic ingredients of long-term economic growth. A wide array of evidence is consistent with this proposition, derived both from case studies of successful and unsuccessful developing economies and from cross-country experience. The growing attention paid to macroeconomic issues by development-oriented institutions such as the World Bank and various regional development banks is one consequence of this new perception.

The proposition that we will consider in this chapter is that good short-run macroeconomic performance is conducive to higher rates of growth of productive capacity. But what exactly does this mean? We will take it to mean that *other things equal*, countries with good short-run macro performance will tend to grow faster than countries with poor performance. In other words, consider two countries that are identical, except that one has good macro policies and the other does not. The proposition suggests that the one with good policies will grow faster. It is a ceteris paribus proposition.

Why should we believe this proposition? That is what this chapter is about. In this chapter, we examine the analytical basis for the proposition as well as the evidence linking short-run macroeconomic performance to long-run growth. To develop the theory, we begin by reviewing the basic forces that drive long-run economic growth in Section I and then consider in Section II how such factors may be affected by short-run macroeconomic performance. Following a pattern that you will encounter many times in this book, Sections III and IV then briefly discuss some recent empirical research that investigates the importance of these links in practice. Section V summarizes.

I. GROWTH DETERMINANTS

Recall from Box 2.1 that a country's level of income per capita can be decomposed into the contributions made by the country's stock of productive resources (its national wealth) and by the productivity of these resources. Thus growth, in the form of sustained changes in income per capita, must be the result of the accumulation of productive factors as well as of increases in their productivity (referred to as *total factor productivity*, or TFP). If short-run macroeconomic performance indeed affects long-run growth, it must do so through one of these channels.

1. Components of TFP Growth

Since TFP is something of a catch-all variable (it is simply the portion of growth that cannot be accounted for by the accumulation of factors of production), a useful first step in analyzing the determinants of growth is to consider the factors that drive changes in TFP. To do so, we can use the familiar notion of a *production possibilities frontier*. Consider an economy that produces two goods X_1 and X_2, using two factors of production, say, capital and labor. At any moment in time, the country has a given stock of capital and a given amount of labor to allocate to the production of the two goods. The production possibilities frontier depicts the maximum amount of each good that can be produced in the economy during each period for a given level of production of the other good. This is determined by the amounts of capital and labor available to the economy as well as by the productivity of the country's technology. The more capital and labor are available to the economy, and the more output it can produce in each sector from any given amount of capital and labor employed in that sector, the more X_2 it can produce for any given level of output of X_1, and vice versa. If the two goods differ in their factor intensities – if X_2 is more capital intensive than X_1, or vice versa – then the production possibilities frontier would take the form of a concave curve in $X_1 - X_2$ space, as in Figure 3.1. Any combination of X_1 and X_2 either on this frontier or inside it is feasible for the economy to produce. Given the money prices of the two goods, say, P_{X1} and P_{X2}, the money value of domestic output for any such combination, say, at a point like C, is given by the following:

$$V_0 = P_{X1}X_{1,0} + P_{X2}X_{2,0}$$

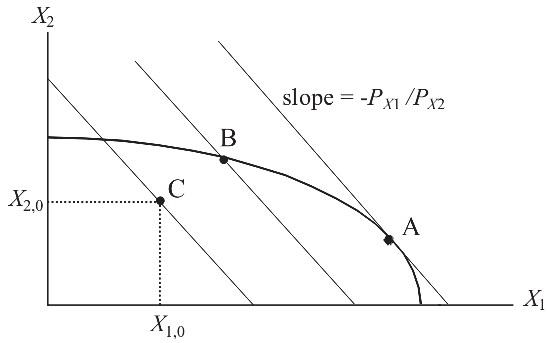


Figure 3.1. The production possibilities frontier

The value of *real* output in this economy has to be measured in units of goods – either of X_1 or X_2 or some combination of the two. For example, the real output of the economy at a point like C, measured in units of X_2, is

$$V_0/P_{X2} = (P_{X1}/P_{X2})X_{1,0} + X_{2,0}$$

Graphically, this number corresponds to the vertical intercept of a straight line with slope $-(P_{X1}/P_{X2})$ passing through the point C. Each point on this line represents a combination of outputs of X_1 and X_2 that is consistent with the same real output V_0/P_{X2}.

As is easy to see graphically, the country's maximum feasible level of real output is attained at the point where a price line with slope $-(P_{X1}/P_{X2})$ is just tangent to the production possibilities frontier. This point is labeled A in Figure 3.1. Thus, if the economy finds itself producing at a point such as C in the interior of the production possibilities frontier, or at a point on the frontier where the slope of the frontier does not equal that of the price line (such as at B), it is failing to allocate its resources efficiently, in the sense that it is not maximizing the value of domestic output, subject to its factor endowments and technology.

This simple analysis has implications for the interpretation of the determinants of TFP. Specifically, it suggests that growth in the value of domestic output can occur in two ways: by expansion of the production possibilities frontier through the gradual adoption of more productive technologies or the gradual accumulation of productive factors or by improving the efficiency of resource allocation, thus moving the economy from points like B or C to points like A. Notice that if we measure changes in TFP as the difference between the growth of output and the contributions to that growth made by the accumulation of factors of production, changes in the efficiency of resource allocation would tend to be recorded as changes in TFP. The key point is therefore that measured growth in TFP depends on changes both in the productivity of the technology available to domestic firms and in the efficiency with which the domestic economy allocates factors of production among alternative activities. Though improvements in the efficiency of resource allocation cannot yield *permanent* changes in an economy's growth rate (since once a point like A is reached, the contribution to growth of this factor would be exhausted), they may contribute to increases in TFP over what may be a prolonged transition period. We therefore conclude that the rate of growth of an economy depends on three factors:

1. the rate of accumulation of factors of production
2. the rate of change in the productivity of the technology available to the domestic economy
3. the rate of change in the efficiency of domestic resource allocation[1]

Next, consider the proximate determinants of these factors. As a preliminary matter, consider what determines the stock of effective technology (ideas about how to produce things) that is available to a specific economy at a point in time. It is useful to think about the economy's technological capacity as dependent on two factors: the position of the world technological frontier and the extent to which world technological knowledge has diffused to the economy in question. The position of the world technological frontier is typically exogenous for most emerging and developing countries because they don't tend to do the research and development that pushes that frontier outward. But the assimilation of existing technical knowledge is *not* exogenous to them. It depends, at least in part, on actions undertaken by economic agents in those economies.

Specifically, acquiring existing technology requires *learning*, and learning is an activity that absorbs resources. This can take several forms such as formal education, learning by doing, buying blueprints, leasing technology, or hiring experts. This has an important implication: because the acquisition of technological knowledge is not free, acquiring a more productive technology involves incurring a present cost for the sake of future benefits. As such, *technology transfer* (the acquisition of already-existing technology) is properly understood as a domestic *investment* activity like the accumulation of productive physical resources. Because the considerations involved are similar, therefore, we can lump together both types of investment decisions (the accumulation of reproducible capital and of technological knowledge) and consider how short-run macroeconomic performance affects productive investment in general, whether in the form of capital accumulation or the assimilation of technological knowledge. Under this approach, the three growth determinants listed earlier effectively boil down to two: investments that increase the economy's stocks of productive factors and technological know-how and improvements in the efficiency of resource allocation. To see how each of these may be affected by short-run macroeconomic performance, the next step is to examine the factors that affect each of these growth determinants.

[1] For empirical evidence of the importance of this factor in determining TFP in the manufacturing sectors of China and India, see Hsieh and Klenow (2009).

2. Investment

The decision to accumulate physical and human capital, or to increase the stock of useful technological knowledge, is essentially an investment decision because it involves incurring current costs (the cost of the new capital or new technology) for the sake of future benefits (the additional capacity to produce output that more capital or better technology can provide). In a market economy, such decisions are made both by the private sector (largely by firms, but also by households) and by the public sector (including all levels of government).

Consider the investment decision of a private firm. Private firms typically make such decisions with the objective of maximizing the value of the firm to their owners. Suppose, for example, that at any moment of time, the firm can choose from among n potential investment projects (such as the production of X_1 or X_2 in the preceding example), each of which has an uncertain future return, and denote the expected rate of return of the ith project as $\rho(i)$. Suppose that savers (domestic or foreign) dislike risk (they are *risk averse*): they are willing to fund projects with certain returns at a *real interest rate r* (the rate of return, measured in units of future goods, from investing the equivalent of one good in the present) but demand a higher return over this risk-free rate (a *risk premium*) for projects that carry more risk, to compensate them for bearing that risk. Then the firm can increase its value to its owners by undertaking an investment only if the expected return on the investment exceeds the safe interest rate plus the risk premium associated with the project. This being so, private incentives to invest will be high, and thus the economy will have a high rate of private investment, when expected private rates of return on domestic investment are high, when the risks associated with domestic investment are low, and when the real interest rate demanded by savers is also low. It is useful to probe a little more deeply into the likely determinants of each of these factors.[2]

a. Private Returns

For private returns to be high requires that investment projects generate large future profits *and* that the investing firm be able to retain those profits for itself. An important determinant of whether private returns are high is therefore the *appropriability* of those returns, that is, whether the firm whose investment generates the returns will be able to keep them. It may not always be able to do so. For example, it would not be able to do so if the returns can be seized by another private agent – say, when it operates in a country in which property rights are poorly defined and enforced – or by the government itself.

b. Uncertainty

Uncertainty may have adverse effects on investment through two separate channels. First, as discussed earlier, when a specific type of investment activity is risky,

[2] A more detailed treatment of the issues involved is provided by Hausmann et al. (2005).

risk-averse agents will demand a risk premium to undertake it. Such risk premia act as the equivalent of a tax on such activities, reducing the level of investment. A second effect arises from the *irreversible* character of much fixed capital (meaning that once it is invested, fixed capital cannot easily be converted to other uses). The irreversibility of capital investment creates an asymmetric situation: a firm that decides not to invest today retains the option to invest tomorrow, but a firm that decides to invest today does *not* have the option to reverse its decision (i.e., *not* to invest) tomorrow. When capital is irreversible, therefore, a potential investor in effect owns a valuable option *before* she makes an investment decision (the option not to invest) that she gives up once the investment is made. The value of that option is therefore part of the opportunity cost of investment. The important point is that the value of the option – and therefore the incentive *not* to invest – increases with the amount of uncertainty in the environment because the more uncertain the environment, the greater the likelihood that the firm will subsequently regret its investment decision. The implication of all this is that uncertainty is likely to have a particularly deleterious effect on domestic investment.

c. The Real Interest Rate

As we have just seen, domestic investment is likely to be a *decreasing* function of the real interest rate. Because a high real interest rate means that a dollar saved today yields a large amount of additional consumption in the future, total saving in the economy is likely to be an *increasing* function of the real interest rate. In a market economy, the equilibrium value of the real interest rate is that which equates total saving to total investment.

This means that the real interest rate has a key *intertemporal* allocative role. A high equilibrium real interest rate, for example, signals to entrepreneurs that there are many attractive investments in the economy, so only very high-return projects will be worth the social opportunity cost of the resources that such projects would absorb. A low real interest rate, by contrast, signals that even projects offering relatively low returns may be worth undertaking. In this sense, the real interest rate is a key intertemporal relative price affecting both the volume of investment and its productivity.

But real interest rates can be too low or too high, thus providing inappropriate incentives to shift production and consumption over time. For example, if an economy that has access to highly productive technologies cannot signal the need to defer consumption through a high real interest rate – that is, if the domestic real interest rate is somehow kept too low – the resources required to finance the capital accumulation required to implement those technologies will not be forthcoming (because saving may be low, and some resources may be absorbed by lower-return projects), and the investment may not be undertaken. On the other hand, if domestic investment offers relatively low future returns, then a domestic real interest rate that is too high will discourage even these moderate-return activities from taking place.

In either case, not getting the real interest rate right will imply that the economy's growth rate will be lower than it would otherwise have been.

Though this discussion applies most directly to investment decisions made by the private sector, investment decisions made by the public sector should ideally be guided by a set of considerations similar to those just described, except that the objective of public investment should be to maximize value for society at large rather than for the private owners of a particular firm. Accordingly, public investment decisions should be based on expected *social* rates of return, the *social* risks associated with specific investments, and the *social* opportunity cost of the resources that the investment would absorb.

3. Efficiency of Resource Allocation

As discussed earlier, the position of a country's production possibility frontier at a point in time depends on its factor endowments as well as on the productivity of the technologies to which it has access. But this frontier only describes the production *possibilities* open to the economy. The *actual* productivity of its factors in production depends on the extent to which their potential productivity is fully exploited by an efficient allocation of factors among competing uses. Thus, given a country's state of technology, its level of total factor productivity can be affected by policies that influence the efficiency of resource allocation, that is, that determine whether the economy produces inside or on the boundary of its production possibilities frontier.[3]

In a market economy, resource allocation decisions made by private firms are guided by the relative prices that prevail among goods and services at a given moment of time, referred to as *intratemporal relative prices*. Private agents decide where to allocate investment funds based on the relative prices associated with different activities because those prices provide the profit incentives to produce specific goods and services. A fundamental tenet of microeconomics is that for intratemporal relative prices to guide resources to their most productive uses, such prices should reflect true social scarcities.

But getting the economy's key intratemporal relative prices right on average in this sense is not enough to achieve efficient resource allocation. It is also important that the *permanent* (sustainable) values of these relative prices not be unnecessarily

[3] It is worth recalling that because changes in the efficiency of resource use affect where the economy operates on a *given* production possibility frontier, such changes affect the *level* of TFP, not its permanent rate of growth. But to change the level, the rate of growth must change at least temporarily. For countries that are below the technological frontier and that are not allocating resources efficiently, investments in knowledge acquisition as well as improvements in allocational efficiency can achieve temporary increases in TFP growth above the worldwide rate of technological advancement (catch-up), but once they attain the frontier, the rate of TFP growth will be given by the worldwide rate of technological advancement.

obfuscated, that is, to avoid unnecessary *uncertainty* about the sustainable value of relative prices. This is important because the reallocation of resources from one type of economic activity to another often involves incurring a fixed cost, for example, the costs of acquiring irreversible physical capital. In the presence of such costs, the relative prices that determine the allocation of resources are *normal* or *permanent* relative prices, not necessarily the *actual* relative prices that are observed at any given moment in time. When there is uncertainty about the future, the relevant permanent prices cannot be observed directly. Instead, they must be inferred.

II. SHORT-RUN MACROECONOMIC PERFORMANCE AND LONG-RUN GROWTH: ANALYTICS

The upshot from the previous discussion is that what we need to explain are the links between short-run macroeconomic performance and the rate of investment (broadly understood to include investment in new technology) as well as between short-run macroeconomic performance and changes in the efficiency of resource use in the domestic economy. The main aspect of short-run macroeconomic performance with which we will be concerned is the quality of the domestic macroeconomic *policy* environment because among all the exogenous variables that can affect an economy's short-run economy performance, only the policy variables can be influenced by domestic decision makers. This section will examine how so-called inappropriate macroeconomic policies (the meaning of this term will emerge from our discussion below) can both discourage investment and impair the efficiency of resource allocation.

1. Macro Policies and Investment

Inappropriate macroeconomic policies can discourage investment in several ways:

- They can *lower the expected future return* on investment.
- They can increase *uncertainty* about future returns.
- They can *increase the opportunity cost* of domestic investment by raising the return that savers demand to finance it.

Consider each of these in turn.

a. Effects on the Expected Return on Investment

Inappropriate policies can lower the expected future return that investment activities are expected to yield for private firms. Such policies can affect the expected future productivity of current investment, the prices at which its output is expected to be sold, and the appropriability of any profits that it may produce. For example,

- Inadequate public investment can result in insufficient *total* capital accumulation, not just through the obvious direct effect of reducing the growth of the public capital stock but also indirectly by discouraging private investment. Complementarities between publicly provided infrastructure and human capital, on one hand, and private capital, on the other, imply that an inappropriate allocation of public spending – one that undersupplies communication and transportation infrastructure or human capital, for example – may reduce the productivity of private capital and thus the incentive to accumulate it.
- In addition, short-run macroeconomic policies can affect expected *future* policies in a way that can potentially affect both the expected returns on private investment and the share of such returns that the private sector expects to be able to appropriate. Current macroeconomic policies, for example, may affect the expected future real exchange rate and thus the relative future profitability of investment in sectors of the economy that differ in their exposure to international competition. As another example, the government's current fiscal performance may affect expectations of future taxation, which in turn would affect the anticipated future appropriability of returns on private investment.

b. Effects on Uncertainty about Future Investment Returns

Inappropriate macro policies, in the form of policies that contribute to an unstable macroeconomic environment, can increase uncertainty about future returns. As discussed earlier, for any given expected rate of return, when investors are risk averse, increased uncertainty will reduce investment. By increasing the degree of uncertainty in the environment, short-run macroeconomic instability thus discourages the accumulation of physical capital.

But what precisely do we mean by macroeconomic instability? At a heuristic level, the answer is obvious: it refers to a situation in which the future evolution of key macroeconomic variables is difficult to predict. But how might this situation arise? Uncertainty can be created in two distinct ways: through *volatility* and *unsustainability*.

i. Volatility. A constantly changing domestic macroeconomic environment would tend to increase uncertainty. Macro instability can arise in many forms. One important distinction is between *exogenous* sources of instability (exogenous shocks) and policy-induced sources (policy shocks). Exogenous shocks are those beyond the domestic economy's control; they include technology shocks, changes in world goods-market and financial-market conditions, and unexplained changes in the behavior of private economic agents (in the form of changes in consumption or investment, in preferences between domestic and foreign goods, in preferences between domestic and foreign assets, between money and other assets, and so on). Policy shocks can arise through fluctuations in fiscal, monetary, and exchange rate policies; in the composition of taxation; or in *commercial policies* (tariffs and quotas on imports). A particularly important policy-induced source of uncertainty is high inflation. High inflation creates instability and uncertainty both because high inflation tends to be unstable inflation, increasing the uncertainty associated with intertemporal relative prices, and because different speeds of nominal price adjustment in different sectors of the economy imply that high and unstable inflation is associated with instability in intratemporal relative prices. We will consider mechanisms for keeping inflation under control in Part 4 of this book.

ii. Unsustainability. Alternatively, a superficially stable situation may not be expected to last, that is, it may be perceived as unsustainable, but it may not be known when the change will come or what form it will take. The result is once again to create uncertainty about the future. Macroeconomic unsustainability can arise in a variety of ways, but in the context of emerging and developing economies, certain specific manifestations have been of particular importance over the past several decades:

- **Prospective fiscal insolvency.** When a country's government is prospectively *insolvent* – that is, it is unable to service its debts – something has to change. Either the government will have to make an adjustment to its budget, which may involve reducing expenditures that benefit some economic activities or raising taxes on others; it may increase its reliance on *seignorage* (the resources that the government can command by printing money), thereby possibly triggering high inflation; or it may simply repudiate its debt, either de jure or de facto. Debt repudiation, in turn, may generate a variety of macroeconomic dislocations through the actions of the government's creditors. The consequences of prospective fiscal insolvency for long-term growth, including through high inflation, are the subject of Part 3 of this book, where we will also examine various means that are available to emerging and developing economies for achieving a credibly sustainable fiscal position.
- **Exchange rate misalignment.** Large and persistent real exchange rate misalignment also increases the uncertainty associated with intratemporal as well as intertemporal relative prices. When the real exchange rate is known to be far from its equilibrium value, that equilibrium value becomes unobservable and therefore uncertain. Moreover, when the country's capital account is open, the expectation of a real exchange rate adjustment affects the level of the country's equilibrium real interest rate through international financial parity conditions, as well as the potential dispersion around the expected real interest rate. We will return to this subject in Part 5.
- **Financial-sector fragility.** Finally, financial-sector fragility, in the form of a financial sector that has low net worth, has high risk exposure, and is poorly regulated, tends to magnify macroeconomic boom-bust cycles, both by generating such cycles and by amplifying them when they originate outside the

Brazil (1989): Fiscal Balance (% of GDP)		
Average 1984–1988	1989	Average 1990–1994
-0.8	-5.9	-0.3

Bolivia (1985): Inflation (%)		
Average 1980–1984	1985	Average 1986–1990
352.0	11749.6	67.8

Indonesia (1998): Nominal Exchange Rate		
Average 1993–1997	1998	Average 1999–2003
2349.6	10013.6	8885.2

Mexico (1994): Current Account Balance (% of GDP)		
Average 1989–1993	1994	Average 1995–1999
-4.54	-7.03	-1.98

Figure 3.2. Illustrative episodes of unsustainability in emerging economies

financial sector itself. Extreme swings in economic activity are likely to be associated with greater uncertainty for both intra- and intertemporal relative prices. Moreover, generalized insolvency in the financial sector is likely to be associated with potential fiscal insolvency as well, through the government's backing of the liabilities of the financial system. These issues are taken up in Part 6.

Figure 3.2 illustrates some well-known instances of unsustainability in emerging economies. Brazil's large fiscal deficit in 1989, Bolivia's inflation rate in 1985, Indonesia's exchange rate in 1997, and Mexico's current account deficit in 1994 all proved to be unsustainable ex post, as shown in the figure. Although what matters for uncertainty is ex ante unsustainability, we will see later in the book that most of these episodes were perceived as unsustainable in an ex ante sense as well.

As mentioned before, these are not the only conceivable symptoms of macroeconomic instability, but they seem to have been the most important ones in emerging economies during the past two decades. The major macroeconomic crisis that afflicted many developing countries during the decade of the 1980s, the international debt crisis, was at bottom a fiscal phenomenon, while the major crises of the 1990s, the Mexican and Asian crises, arose from interactions between inappropriate financial sector and exchange rate policies in a context of increased financial integration. The recent crisis experience in emerging and developing economies is the subject of Part 7.

c. Effects on the Opportunity Cost of Investment

Finally, inappropriate macroeconomic policies can affect the opportunity cost of investment through their effects on the return that savers (domestic and/or foreign) demand to finance domestic investment, that is, the real interest rate. As discussed earlier, real interest rates can be too high or too low, and inappropriate macroeconomic policies can contribute to making them one thing or the other. The types of policies that can have such effects and the mechanisms through which they can do so depend on whether the economy is financially closed or open as well as on the efficiency of the domestic financial system.

In a financially *closed* economy, the *natural* real interest rate is the real interest rate that is consistent with the equality of saving and investment at full employment. Excessive fiscal deficits – that is, deficits driven by inappropriately high public consumption or investment or by the government's inability to raise revenue – can increase the natural real interest rate, even if the government is confidently expected to service its debt, simply because of the need they create to mobilize additional saving.

In an open economy, the domestic interest rate is determined by the world risk-free rate plus the *country risk premium* (the risk premium that foreign creditors demand to lend to agents in the domestic economy). The country risk premium is a function, among other things, of the economy's short-run macroeconomic performance. Prospective fiscal insolvency, for example, will increase default risk for all economic agents, increasing contractual real interest rates. This effect is magnified if lenders are risk averse because then they will demand a risk premium over and above the contractual interest rate required to match lenders' opportunity cost of funds. Even with a solvent government, in an open economy, an overvalued real exchange rate would increase the domestic real interest rate.

Finally, other effects may be relevant whether the economy is closed or open. For example, policies that increase the costs of financial intermediation would increase the real interest rate for borrowers, while lowering it for savers. Similarly, excessively expansionary or contractionary monetary policy could cause the real interest rate to deviate from its natural level for extended periods of time.

2. Macro Policies and Allocational Efficiency

Short-run macroeconomic performance can also affect the economy's growth rate through its effects on the efficiency of resource allocation. Macroeconomic policies that induce an existing stock of factors to be allocated to less productive uses than they otherwise could be, given the level of technological knowledge, would tend to reduce both the level and rate of growth of TFP. Since resource allocation decisions are made by the public sector, private firms (entrepreneurs), and the financial sector, these effects could arise if short-run macro policies induce inappropriate allocation decisions by any of these actors.

a. Allocation of Resources by the Public Sector

Consider first the public sector. Inappropriate allocation decisions by the public sector concerning either the level or the composition of public investment would reduce the productivity of domestic resources. To maximize this productivity, the share of total investment undertaken by the public rather than the private sector should be such as to equalize the risk-adjusted social rates of return in the two activities. Deficient or excessive levels of public investment – that is, levels of investment that violate this rule – reduce the productivity of the resulting aggregate capital stock. Similarly, a given public investment budget should be allocated among competing public investment projects to the projects with the highest risk-adjusted social rates of return. Failure to observe this principle would again reduce the productivity of the resulting capital stock.

In practice, however, because public investment decisions are made through the political process, rather than by a benevolent planner, decisions about both the allocation of public investment among competing activities and the total level of public investment are generally made in response to political factors, which are in turn susceptible to influence from short-run macroeconomic developments.[4] This means that even in normal times, public investment decisions may be distorted by political factors, even when resources are plentiful.[5] The implication is that productive public expenditures, such as spending on infrastructure, health, and education, may tend to be shortchanged, resulting in allocational inefficiencies.

But short-run macroeconomic instability may make the resource-allocation implications of political decisions on public-sector investment more severe. In boom times, for example, the political system tends to treat transitory favorable shocks as permanent, and too much public investment may take place from an optimality perspective.[6] During times of resource scarcity, by contrast, spending cutbacks allocated by political expediency often mean that too little public investment takes place (an example is the reduction of public infrastructure

[4] E.g., public investment budgets swelled in heavily indebted developing countries during the heady days of the 1970s, featuring favorable terms of trade and negative international real interest rates, but they collapsed in the same countries in response to the subsequent debt crisis (Easterly 1989).

[5] Notice that this is an important way in which corruption can harm economic growth (by resulting in a misallocation of the public capital stock).

[6] For applications to the decades of the 1970s and 1980s, see Corden (1991) as well as Calderon et al. (2003). Commodity price booms have also often been associated with ill-advised public investment drives, especially in oil-producing countries.

investment in Latin America during times of fiscal stringency during the decade of the 1980s).

The key challenges for developing and emerging economies in this area are therefore not just how to create fiscal and macroeconomic "space" for productive public expenditures but also how to put public investment decisions in the hands of technocrats who are both honest and insulated from the political process.

b. Allocation of Resources by Private Firms and Households

Inappropriate macro policies can interfere with the appropriate allocation of investment among specific activities by private firms and households. They can do so by distorting the permanent structure of intratemporal relative prices and/or by destabilizing such prices, thereby making their "true" values difficult to infer.

i. Effects Operating through the Permanent Structure of Relative Prices. Macroeconomic policies can affect the sustainable structure of intratemporal relative prices in several ways. Consider an economy that produces a wide variety of traded and nontraded goods. In that context, the real exchange rate is generally defined as the relative price of the broad basket of traded goods in terms of that of nontraded goods. It provides the incentives that guide the allocation of resources between the two broad economic sectors. Macroeconomic policies that cause the real exchange rate to become *misaligned* (to be far away from its sustainable equilibrium value, which we will consider in Chapter 16) would tend to distort the allocation of resources between these two broad production aggregates.

Within the category of traded goods themselves, relative prices may be distorted by commercial policies (tariffs, quotas, or export subsidies). Such policies are often adopted for political reasons (to provide protection to politically powerful groups) but may also be motivated by fiscal considerations or adopted to defend a misaligned real exchange rate.

Finally, a highly distorted tax system can exert distortionary effects at an even more disaggregated level of production, through its influence on the after-tax relative returns that factors of production can earn in different activities *within* either the traded or nontraded goods sectors.

If relative prices such as these are not right – that is, they do not reflect true social scarcities – then factors of production will be guided into uses that are not as productive as they might otherwise have been, reducing the level of total factor productivity in the economy. The obvious implication is that fostering allocative efficiency involves the avoidance of real exchange rate misalignment, restrictive trade policies, and excessively distortionary taxation and/or subsidies.

ii. Effects Operating through the Efficiency of Price Signals. As discussed earlier, macroeconomic instability tends to generate uncertainty about whether the relative prices observed in the marketplace will prove to be permanent or transitory. Such volatility can persist because macroeconomic policies themselves create it (e.g., through stop-go policy cycles) or because macro policies fail to counter volatility that arises exogenously, say, because of changes in the international economic environment. In either of these forms, instability creates uncertainty about whether observed relative prices will prove to be permanent or transitory.

Effects on the growth of total factor productivity arise from two sources:

- When sustainable relative prices cannot be observed easily, the private sector has to solve an inference problem; that is, it has to generate estimates of "permanent" relative prices from the prices it can actually observe as well as any other relevant information. This activity itself absorbs resources. This has a direct effect on total factor productivity because the resources absorbed in generating and processing information are not available to be used in the production of goods and services.
- Moreover, even under the best of circumstances, the expenditure of resources in solving this inference problem will be unable to resolve all uncertainty about future relative prices. The unpredictability of relative prices may induce economic agents not to respond to relative price signals in the fear that they will prove to be transitory. This makes resources inflexible in their allocation across sectors and reduces the efficiency of resource use.

c. Allocation of Resources by the Financial Sector

The structure and observability of relative prices affect the demand for investment funds by firms. But because of the lumpy nature of investment projects, firms can seldom finance their desired investment out of their own resources. The financial sector has the job of allocating investable funds among competing demands. Financial-sector distortions may thus affect the efficiency of resource allocation through their effects on the *supply* of investment funds. Even if relative prices accurately reflect the true social scarcities of various goods and services, a poorly functioning financial system may misallocate the supply of resources among potential investment opportunities.

Macroeconomic policies can generate such distortions directly or indirectly. Direct effects are those that operate through policies directed at the financial sector itself, whereas indirect effects operate through the implications of broader macroeconomic policies for the efficiency with which the financial sector can operate. An example of the former is *financial repression* (a set of policy restrictions on the behavior of banks, often adopted for fiscal reasons, as we will see in Chapter 20). Another is a deficient regulatory environment for a liberalized financial sector, which can result in financial-sector fragility in the form of low net worth of banks coupled with high risk exposure. This distorts the incentives faced by managers of financial institutions, affecting the efficiency of resource allocation directly. Indirect effects on resource allocation may operate through various channels. For example, an unstable macro environment (e.g., as the result of stop-go policies) can induce even a well-functioning financial sector to make inappropriate allocation decisions, as we will see in Chapter 21.

III. MACROECONOMIC STABILITY AND LONG-RUN GROWTH: CROSS-COUNTRY EVIDENCE

The proposition that we considered analytically in the previous section was that inappropriate short-run macroeconomic policies (meaning policies that create distortions in macroeconomic relative prices and generate unnecessary uncertainty) are likely to be harmful for a country's long-run economic growth. We interpreted this proposition as suggesting that, other things equal, countries with good macroeconomic policies will tend to grow faster than those with poor policies. The previous section examined some analytical links between short-run macroeconomic performance and the rate of growth of the economy's productive capacity, highlighting the potential role of poor short-run macroeconomic performance in impairing both the accumulation of productive factors and the growth of total factor productivity. It also identified some specific phenomena that can be interpreted as symptoms of inappropriate macroeconomic policies. But is there any evidence that such phenomena have indeed been associated with slower growth of productive capacity in developing and emerging economies? In other words, how would we determine empirically whether the theoretical arguments of the previous section are correct?

1. Cross-Country Evidence

The first step in taking a hypothesis such as this to the data is to define it conceptually as precisely as possible. The first problem we face is to define what we mean by good macroeconomic policies and determine how to measure these empirically. As described before, "good" policies are those associated with a stable macroeconomic environment and the absence of relative price distortions. To identify a good policy environment, we could rely on macroeconomic *outcome* variables, such as the rate of inflation or the size of current account deficits, or on macroeconomic *policy* variables, such as the size of fiscal deficits, the size of government, the magnitude of fluctuations in the real exchange rate, the average real interest rate, and so on. Unfortunately, one problem with testing our proposition empirically, as we shall see later, is that none of these is really ideal.

To continue our discussion, however, let us finesse this problem by focusing on the particular indicator used earlier: the rate of inflation. Consider then, as a concrete example, how we would examine the empirical link between high inflation and the rate of growth of productive capacity. A naive way to test whether a negative association indeed exists between these two variables is to see whether changes in real gross domestic product (GDP) and changes in the price level are correlated in some sample, either for a single country over time (a *time series* sample) or across countries (a *cross section*). Notice, however, that the arguments we examined earlier refer to links between *long-run* (capacity) growth and inflation. Because they do not necessarily represent changes in capacity growth, year-to-year correlations may be meaningless as evidence. The first step, therefore, is to make sure we have an appropriate operational definition of the variables.

Inflation can be observed directly, but growth of productive capacity cannot. It must be estimated. One way to get an empirical handle on it is to exploit the observation that the *actual* growth rate of GDP in a country during a given year tends to fluctuate around the growth of capacity, either because the production function is subject to random shocks or because random demand shocks cause temporary deviations from capacity output that are gradually eliminated through nominal wage and price flexibility (this mechanism is discussed more fully in Chapter 4). From a statistical perspective, these fluctuations can be taken to be *mean-zero serially correlated* random shocks.[7] With such a statistical model in mind, we could estimate the growth of productive capacity by taking the average of several years' growth rates during which we believe the statistical model to be valid. In other words, for a given country, the mean of several years' actual growth rates can be taken as an estimate of capacity growth during those years in that country.

Thus, with an appropriately chosen sample, our best bet is to look at correlations between long-run average growth rates and long-run average inflation rates. This cannot generally be done for a single country, unless we have enough data to generate periods that are long enough for meaningful averages to be calculated. Thus a natural approach is to look at cross-country experience, that is, to use cross-sectional evidence.

To demonstrate empirically that good macroeconomics are neither necessary nor sufficient for countries to achieve high growth rates, one merely has to produce counterexamples. This is not hard to do. That strong short-run macroeconomic policies are not *necessary* to achieve high rates of growth is demonstrated by the experience of several countries. Brazil, for example, averaged an inflation rate of 28 percent during 1965–1973, yet managed an average growth rate of 7.2 percent per capita during those years. Though its inflation rate increased to 45 percent during 1973–1980, and its growth rate fell, it still registered a hefty per capita growth rate of 4.5 percent during that period. Colombia, Israel, and Turkey, among the larger emerging economies, have also experienced extended periods of high income growth coinciding with high inflation. That strong short-run macroeconomic policies are not *sufficient* to achieve high growth is demonstrated by regions like South Asia and Sub-Saharan Africa, where the macroeconomic environment has been stable for long periods of time during which growth was disappointing. In Section IV, we will look in more detail at the experience of Latin America in the 1990s,

[7] These are random shocks that cancel each other out on average but have the property that shocks in one direction (positive or negative) are likely to be followed by shocks in the same direction.

Table 3.1. *Best-Performing Countries in Growth Terms*

Country	Annual GDP Per Capita Growth (%) (1980–2006 average)	Annual Inflation (%)
Hungary	1.95	13.13
Colombia	1.42	18.62
Dominican Rep.	2.53	17.56
Nepal	1.92	8.74
Chile	3.54	13.01
Israel(1)	1.89	52.62
Indonesia	3.75	11.31
Malaysia(2)	3.64	3.22
Thailand(2)	4.70	4.65
Singapore(2)	4.53	1.96
Pakistan(2)	2.73	7.70
Sri Lanka	3.13	11.48
Egypt, Arab Rep.	2.81	11.65
Botswana(2)	5.26	10.20
India(2)	4.02	8.03

(1) One of the 15 countries with the worst inflation performance.
(2) One of the 15 countries with the best inflation performance.
Source: author's calculations.

where improved macroeconomic performance was not associated with a strong increase in the growth rate.

Tables 3.1 and 3.2 present a more systematic look at the country experience. Table 3.1 lists the 15 emerging and developing countries with the best growth performance – measured by a weighted average of the average per capita growth rate and the standard deviation of the growth rate – over the 1980–2006 period. As is evident from the table, these countries had sharply divergent inflation experiences during this period, ranging from under a 2 percent average annual inflation rate in Singapore to one of over 50 percent in Israel. Similarly, Table 3.2 ranks countries by inflation performance. Again, the growth experience of the top performers in this dimension was also diverse. Indeed, the country with the best inflation performance over this period (Saudi Arabia) registered a *negative* average growth rate in income per capita, and of the 15 best inflation performers, four (Saudi Arabia, Cote d'Ivoire, Niger, and Senegal) were among the 15 countries with the *worst* growth performance over the period.

The diversity of this experience indicates that low inflation is neither necessary nor sufficient for high growth. Yet the tables also suggest that there is something systematic in the relationship between low inflation and high growth. Notice, for example, that six countries were found among the 15 best performers on *both* criteria (Malaysia, Thailand, Singapore, Pakistan, Botswana, and India).

Table 3.2. *Best-Performing Countries in Inflation Terms*

Country	Annual GDP Per Capita Growth (%) (1980–2006 average)	Annual Inflation (%)
Saudi Arabia[1]	−1.53	0.57
Singapore[2]	4.53	1.96
Panama	1.53	1.90
Malaysia[2]	3.64	3.22
Morocco	1.94	4.91
Thailand[2]	4.70	4.65
Pakistan[2]	2.73	7.70
Burkina Faso	1.81	4.17
India[2]	4.02	8.03
Botswana[2]	5.26	10.20
Jordan	1.11	5.18
Cote d'Ivoire[1]	−2.39	5.49
Niger[1]	−1.70	3.50
Trinidad and Tobago	2.00	7.98
Senegal[1]	0.15	4.58

[1] One of the 15 countries with the worst growth performance.
[2] One of the 15 countries with the best growth performance.
Source: author's calculations

To address empirically the weaker ceteris paribus proposition that good short-run macroeconomic policies are *conducive* to improved growth performance is a little trickier. The proposition basically claims that there is a cause-and-effect relationship between short-run macroeconomic performance and long-run capacity growth. This suggests that we should be able to see an association between these two variables in the data. But how do we find out if such an association is actually present?

There is a large literature that has examined the effects of short-run macroeconomic performance on long-run growth rates using cross-country data. To illustrate some of the methodological issues involved, consider as an example a well-known study by Fischer (1993). Fischer describes various ways to examine empirically the links between macroeconomic stability and growth. One way to do it is to look at simple patterns of correlation, as we did earlier. Fischer looks at the association across regions and for the same region over time between average rates of economic growth and various macro variables that serve as indicators of stability (Fischer, 1993, Table 1). Using Asia, Latin America, and Africa as the regions, and observations averaged for 1960–1973, 1973–1980, and 1980–1988, he finds the following:

- negative correlations between growth and inflation as well as between growth and government budget deficits

- positive correlations between growth and the current account balance of the country's balance of payments as well as between output growth and export growth

But just how strong is evidence of this type? One problem with such evidence is that these correlations could simply reflect accidental outcomes. In other words, we might observe such outcomes by chance even if there were no true relationship between the variables. By using regions as our basic unit of observation, we do not have enough data to be statistically sure that this is not so, or at least to form some idea of the likelihood that the outcome is purely accidental. Suppose, for example, that countries' long-run growth experience is unrelated to their inflation rates, contrary to the preceding hypothesis. In particular, assume that whereas countries differ with respect to their long-run growth rates, their normal inflation rates are all the same but their actual inflation rates during any sample period are purely random. Then how could we be sure that the association we observe between growth rates and inflation rates does not simply reflect a lucky draw for the high-growth countries and an unlucky one for the low-growth countries?

The answer is, of course, that we cannot be sure. However, we can use statistical methods to quantify the extent of our uncertainty about the role of random factors. For example, because the hypothesis concerning the link between growth and macroeconomic stability is in any event about countries, suppose we use countries rather than regions as our units of observation and pose the question whether the "normal" inflation rate for high-growth countries is lower than that for low-growth countries. To address this question, we can classify countries according to their growth experience and examine the differences between the average inflation rates of the high-growth and low-growth country groups. The more countries that are in each group, the more precise the average inflation rate of each group will be as an estimate of the "normal" inflation rate in that group. The question, then, is what the likelihood would be that any observed difference between the average inflation rates of the two groups could emerge purely as the result of random factors if indeed there were no difference between the "normal" inflation rates of the two groups. This is precisely the kind of question that can be addressed through statistical difference of means tests.

For example, Fischer cites the work of Levine and Renelt (1992), who ranked countries by average growth rates and found statistically significant differences between fast growers (56 countries with growth above the mean) and slow growers in their sample of 109 countries (1960–1989) with regard to the ratio of investment to GDP, the ratio of government consumption to GDP, the rate of inflation, and the *black market premium* (the percentage difference between the exchange rate in the unregulated foreign exchange market and that in the official market) as indicators of macroeconomic stability.

What can we learn from tests of this sort? What these tests tell us is whether long-run growth tends to be correlated with several indicators of macroeconomic stability, taken one at a time, in the cross-country evidence. However, the well-known adage that correlation does not establish causation should give us pause about taking such evidence too seriously. To address the question of whether macroeconomic stability fosters long-run growth, we need to investigate whether the variables that we use as indicators of macroeconomic stability have a causal association with growth. A systematic (nonrandom) negative correlation between growth and inflation could have arisen in any of three ways:

1. Greater instability reduces growth (the hypothesis we are considering).
2. Some third factor causes high growth and lower instability and/or slow growth and high instability.
3. Higher growth reduces instability (reverse causation).

We need to eliminate the second and third possibilities if we are to believe the first. How can we do this?

Conceptually, think of running an experiment. Countries would be assigned randomly to two groups, and a dose of "instability" would be administered to one of the groups (the treatment group) but not to the other (the control group). This would allow us to see what happens to the rate of growth in those countries as a result. The random assignment would ensure that third factors would be expected to behave similarly between the control and treatment groups and would simultaneously ensure that the treatment was exogenous (i.e., not influenced by the country's subsequent growth performance). Thus running a controlled experiment would allow us to eliminate the possibility that any correlation that we observe in the data arises from the effects of third factors or from reverse causation.

But running such an experiment is obviously impossible. Since we cannot do so, to try to identify the *independent* (partial) correlation between each factor and the growth rate, we have to deal with the factors that complicate inference through statistical rather than experimental methods. The most commonly used tool by economists for this purpose is *regression analysis.*

2. Cross-Country Growth Regressions

Regression analysis is important in empirical economics because it provides a statistical means of controlling for the effects of "third factors" that may affect a particular economic relationship under study. To apply regression analysis, we proceed in two steps:

1. First, we need to specify the dependent variable in which we are interested as a function of a set of independent explanatory variables, using a specific functional form (typically linear) to describe the effects of the independent variables on the

dependent variable, and allowing for the possible influence of random effects arising from unobserved variables by including a stochastic error term.

2. Second, we use statistical techniques to estimate the coefficients of the explanatory variables in this *multiple regression* (a regression with more than one explanatory variable). We test our hypothesis by examining the resulting signs and statistical significance of the specific coefficients that capture the particular hypotheses in which we are interested.

In the context of this chapter, and in many other places in this book, we will be making reference to a particular application of regression analysis that has become common in the empirical study of growth determinants, usually referred to as *cross-country growth regressions.* These are typically multivariate cross-sectional regressions. The dependent variable in such regressions generally consists of the average growth rate recorded by individual countries over some extended period of time, which, as we have seen, serves as a measure of capacity growth in the country during that time. Explanatory variables are specified to control for determinants of growth other than the one(s) in which the investigator happens to be interested. The variables that are taken to be standard growth determinants differ slightly across applications, but they typically include the country's initial level of real per capita GDP, a human capital measure such as the secondary school enrollment rate, the rate of growth of the population, and a measure of political stability. Regional dummy variables, usually for Latin America and Sub-Saharan Africa, are often included as well to capture unmeasured region-specific influences on growth. The effects of particular policy variables on the long-run growth rate are then tested by adding the relevant variable to this core regression to see whether the variable exhibits a significant partial correlation with the average growth rate – that is, whether its coefficient in the estimated regression has the theoretically expected sign and is measured with sufficient statistical precision.

If it does, then further investigation, can shed light on whether the relevant variable affects the growth rate primarily by inducing greater accumulation of productive resources or by enhancing the efficiency of resource use in the economy. This is tested by adding the investment rate to the growth equation and determining whether the coefficient on the relevant variable retains its sign and statistical significance. If so, then an efficiency effect is detected because any influences operating through investment are already accounted for by the inclusion of that variable in the regression. If not, and if the variable is statistically significant in an equation similar to the core regression, but with the investment rate as the dependent variable, then its association with growth is attributed to a resource-accumulation effect.

Fischer adopted this approach as his third "cut" at investigating the effects of macroeconomic stability on long-term growth. As indicated earlier, the first step in implementing the method is to identify the full set of potential third factors that could influence the growth rate. Fischer thus began by specifying cross-sectional regressions of the type

$$\Delta Y/Y = a_0 + a_1 Y_0 + a_2(I/Y) + a_3(\Delta N/N) + \Sigma a_j X_j$$

The first four terms on the right-hand side of this equation capture the influence of standard growth determinants. Levine and Renelt (1992), for example, estimated cross-sectional growth regressions with a sample of 101 countries and data over 1960–1989. They found that initial GDP, the level of secondary school enrollment in 1960, and the rate of population growth were all independently related to growth, as was the rate of investment. Fischer built on this by adding other variables, captured in the last term on the right-hand side (the X_js). Using a sample of 73 countries with data averaged from 1970 to 1985, he found that in addition to the Levine-Renelt variables, high growth was associated with low inflation and high budget surpluses. Other variables, such as the country's level of debt in 1980, while potentially (theoretically) important, were not statistically significant in the regression, though dummy variables for Sub-Saharan African and Latin American countries proved to have negative coefficients that were statistically significant, suggesting that some potential growth determinants had been omitted from the regression.

These results suggest that if macroeconomic stability in the form of low inflation and responsible budgetary policies has independent effects on growth, these effects must operate at least partly through their influence on total factor productivity because the regression already controls for the effect of investment on growth. To investigate whether the effects might also operate through influences on investment, Fischer replaced real output growth with the rate of growth of the capital stock as the dependent variable and found that capital accumulation was negatively linked with inflation and the black market premium, though not with the other macro variables.

3. Evaluating the Evidence

What can we learn from studies such as the one just described?[8] Unfortunately, because they suffer from several methodological problems, probably the best we can hope for is to extract a set of suggestive empirical regularities from the data, and we will use studies of this type in that spirit throughout this book.

The most obvious problem with cross-country growth regressions is that heterogeneity among countries is likely to be very important. Because countries are different, what can emerge from cross-country regressions at best is an average

[8] A systematic analysis of the problems associated with the application of cross-country regressions to learn about the effects of policies on growth is given by Levine and Renelt (1992). See also Temple (1999) and Brock and Durlauf (2001).

relationship, not particularly applicable to any single country. One implication of heterogeneity, both across countries and over time, is that this empirical methodology suffers from a lack of *robustness*; that is, the variables of interest may cease to enter the regression with coefficients that bear the theoretically expected sign and are statistically significant when the sample changes, or once other reasonable variables are added to the growth regression.[9] Unfortunately, few of the variables typically included in such regressions fail to exhibit this property.

A second problem is one that is endemic in the use of statistical procedures to control for the possible influence of third factors in studying a hypothesized relationship between two economic variables. Specifically, note that while Fischer controlled for a variety of such factors that may affect growth, it is hard ever to establish that relevant ones have not been omitted, a problem known as *specification error*. For example, as mentioned earlier, that Sub-Saharan African and Latin American countries had systematically lower growth than the others, even after accounting for all the variables included in the regression, suggests that systematic growth determinants have been omitted from the regression. This is important because if relevant variables that are omitted from the regression happen to be correlated with some of the variables that are included in the growth specification, the measured impacts of the latter on the growth rate, as indicated by their coefficients in the regression, may partly tend to reflect the effect of the omitted variables, inducing an *omitted variable bias* on the estimated coefficients.

Third, even if the estimates can be shown to be robust, unless the reverse causation problem (interpretation 3) is explicitly dealt with, all that the parameters may really indicate is the strength of a partial correlation[10]; that is, the finding of a statistically significant coefficient just indicates that there is a reasonably tight relationship in the data between the explanatory variable of interest and the portion of the dependent variable not accounted for by other independent variables. It does not necessarily indicate a causal relationship running from the relevant independent variable to the dependent variable.

Finally, even if one can establish the direction of causation, the explanatory variable typically included to study the effects of policy is often not itself a policy variable but rather a performance indicator that is endogenous to policy. Thus its coefficient cannot be interpreted as revealing the long-run growth effect of a given characteristic of the policy environment.

There are ways to deal (at least in principle) with each of these statistical problems, and some of the existing cross-country empirical growth work has implemented them. Heterogeneity, for example, has been tested by including slope and/or shift dummies for regions, and robustness has been investigated through sensitivity

[9] See Levine and Renelt (1992).

[10] Notice that Fischer did not attempt to deal with the problem of potential reverse causation.

analyses in which different sets of explanatory variables are included in the regression to detect whether the coefficients on the variables of interest are affected by the inclusion or exclusion of other variables in the regression. Statistical techniques are also available to attempt to deal with the problem of reverse causation, for example, by using initial values of explanatory variables rather than average values during the sample period, or by using special statistical techniques designed for the purpose. Some of these are described in Box 3.1 and are discussed in more detail in Chapter 13. Finally, alternative proxies for policies have been tried to examine the sensitivity of results to different ways of measuring the explanatory variables of interest.

We will review more of the evidence generated by cross-country growth regressions about the links between macroeconomic policies and long-run growth – keeping in mind these methodological concerns – later in the book. For now, however, we conclude that though the evidence cannot be regarded as definitive, it is certainly suggestive of a positive relationship between good macroeconomic policies and long-run economic growth.

IV. AN APPLICATION: MACRO STABILITY AND DEVELOPING-COUNTRY GROWTH IN THE 1990S

The perceived relationship between the quality of the macroeconomic policy environment and an economy's long-run growth rate was an important factor causing many emerging and developing countries to undertake policy reforms in the late 1980s and 1990s, a process that remains ongoing. Though these reforms – many of which we will be examining in the rest of the book – proved controversial and difficult to implement, the political will to do so was created in part through their promise of improved growth performance.[11] Yet, in many parts of the world – especially in Latin America – subsequent growth performance tended to disappoint expectations. In light of the material we have covered in this chapter, it is therefore useful to ask, if macro stability is so important for growth, why was the macro stability achieved in many emerging and developing economies during the 1990s not associated with more growth?

This experience illustrates why good macro policies are not a sufficient condition for rapid growth. The question has been investigated by many authors. Montiel and Serven (2006), for example, offer a four-part answer:

1. Macroeconomic stability was not in fact achieved as widely and to the extent perceived by those who have judged the subsequent growth performance to

[11] Edwards (2008) documents the anticipated link between macroeconomic reforms and the expectation of improved growth performance in Latin America.

Box 3.1. Some Recent Evidence on the Macro Stability-Growth Relationship

Other researchers have followed Fischer in investigating the empirical importance of macroeconomic stability for long-run growth using cross-country regressions. Examples of this line of work are Barro (1995), Bruno and Easterly (1995), Sarel (1996), and Fernandez-Arias and Montiel (2001). The Barro and Bruno and Easterly studies are described in Chapter 13. This box summarizes the findings of the last two papers.

Sarel (1996)

Sarel was motivated by the observation that it was not until the 1970s and 1980s, when many countries experienced severe inflation episodes, that empirical studies systematically began to find a negative relation between inflation and growth. Because the new evidence seemed to indicate a change in the behavior of the variables during the high and persistent inflation decades, Sarel tested formally whether the inflation-growth relationship had changed over time (i.e., whether the relationship had undergone a *structural break*) and also investigated what the estimated effects of inflation on growth were both below and above the threshold (break) inflation rate.

To this end, he used a *panel* (a combination of time series and cross-sectional observations) for 87 countries during the 1970–1990 period. His basic approach was to regress the average growth rate on log (inflation), a set of control variables, and a term of the form DD[log(inflation) − log(inflation*)], where DD = 1 if inflation > inflation* and 0 otherwise, and inflation∗ is the inflation rate at which the structural break occurs. When inflation < inflation*, the effect of inflation on growth is estimated by the coefficient of log(inflation), whereas when inflation > inflation*, the effect is given by the sum of the coefficients of log(inflation) and DD[log (inflation) − log (inflation*)].

Sarel estimated this equation for different values of inflation*, selecting the one that provided the best *regression fit* (the one that explained the largest share of the variation in panel growth rates), and found a break inflation rate of 8 percent per annum. His main findings were that when inflation was low (below 8% per annum), it had no significant negative effect on economic growth, but when inflation was high (above an annual rate of 8%), the negative effect of inflation was negative, statistically significant, and robust to different specifications. In the basic regression, the annual growth rate of GDP would decrease by 1.7 percentage points when the annual inflation rate doubled.

Fernandez-Arias and Montiel (2001)

Fernandez-Arias and Montiel (2001) also used country panel data to explore the empirical relationship between several macroeconomic stability and reform variables and long-run growth. Their sample consists of data for 69 countries in the 1961–1995 period. Their approach was similar to that of Fischer, except that their basic regression included variables to measure structural reforms undertaken in their sample countries, in addition to standard growth controls. Their measures of short-run macro performance included the inflation rate, the share of government consumption in GDP, the ratio of broad money to GDP (a proxy for financial depth), the black market premium, and a conventional measure of commercial openness (exports plus imports over GDP). The set of control variables consisted of lagged educational attainment, lagged GDP per capita, and the country's international terms of trade.

Another aspect that distinguishes this work from that of Fischer (1993) is that it addressed the issue of potential reverse causality by using *instrumental variable* estimation techniques, described in Chapter 13. The authors ran specification tests to determine whether this estimation alternative affected the results and found that it did not. They found a substantial and statistically significant positive growth effect associated with lower public consumption, lower inflation, financial deepening, and exchange rate unification (a lower black market premium).

have been disappointing.[12] On the other hand, it may have been *overachieved* in some contexts:

- Certain policies intended to achieve improved macroeconomic performance in specific dimensions such as reducing inflation through the adoption of fixed exchange rates, including in the extreme forms of *currency boards* or *dollarization* (see Chapter 17) may have increased macroeconomic fragility and therefore created growth-impairing instability.
- An excessive preoccupation with fiscal deficits, over and above that required to promote stability, may in some cases have eliminated *fiscal space* for productive public expenditures.
- Similarly, an excessive preoccupation with fiscal deficits may have often prevented appropriate countercyclical use of fiscal policy. Both this point and the previous one are discussed in Chapter 10.

2. Because investment is forward looking, what matters for growth is *sustainable* macro stability. This requires not just better macroeconomic outcomes in the present but also reform of macroeconomic institutions of the macroeconomic policy *regime* to create the expectation that a stable macro environment will prevail in the future. The types of institutions involved are described in Parts 3–6 of the book. Reforms of this type have been harder to achieve than have improved macro outcomes.
3. The macroeconomic reform agenda of the 1990s was often incomplete and in some ways misguided. In particular, it often featured the following:
   - Premature financial and capital account liberalization (Chapter 23).
   - Insufficient concern with financial-sector soundness (Chapter 25).
4. Finally, the growth benefits of macroeconomic stability may have been oversold. As argued in this chapter, macroeconomic stability is not a sufficient condition for satisfactory growth performance. In particular, it is not enough to drive the growth process when other essential ingredients are lacking, that is, when

[12] See also Fernandez-Arias and Montiel (2001) as well as Edwards (2008).

other constraints on growth are binding. These constraints involve the various policies and institutions that shape the opportunities and incentives to engage in growth-enhancing activities. They include secure property rights and market-oriented *microeconomic* incentives as well as proactive government interventions to overcome informational externalities and coordination failures. It is precisely because of the importance of such factors that good short-run macroeconomic performance is not sufficient to ensure high rates of economic growth.

V. SUMMARY

This chapter has examined the basic proposition underlying this book: that good short-run macroeconomic performance, in the form of appropriate values of key macroeconomic relative prices as well as a predictable domestic macroeconomic environment, is conducive to long-term growth in emerging and developing economies. We examined the analytical channels through which appropriate relative prices and a stable macroeconomic environment are likely to foster both higher rates of factor accumulation and more rapid growth in an economy's total factor productivity. We also reviewed some important and widely cited empirical evidence on the topic. We concluded that theory supports the presence of a link between short-run macroeconomic performance and growth of productive capacity, whereas the evidence we reviewed in this chapter – while not entirely compelling – is at least consistent with the existence of such a link.

Given this link between short-run macroeconomic performance and long-run growth, we next want to address the issue of how short-run macroeconomic performance in emerging economies is influenced by domestic macroeconomic policies. The first step in doing so is to develop an analytical framework linking macroeconomic policies to macroeconomic outcomes. That is the task that we will undertake in the next four chapters.

REVIEW QUESTIONS

1. What is the difference between macroeconomic volatility and macroeconomic unsustainability? Explain how these phenomena may be related to the long-run growth of an economy's productive capacity.
2. What is total factor productivity? What determines the rate of growth of an economy's total factor productivity?
3. Explain the sense in which the adoption of new technology in emerging and developing countries can be interpreted as an investment activity.
4. What are the channels through which macroeconomic uncertainty can discourage economic growth?
5. Why did Latin American countries fail to experience the growth acceleration that was expected as a result of their improved macroeconomic policies during the 1990s?

EXERCISES

1. Is a high real interest rate good for growth because it encourages saving, or bad for growth because it discourages investment? Explain.
2. Use the production possibility frontier analysis to show how the appropriate inter- and intratemporal relative prices are both important in achieving a satisfactory rate of growth in emerging and developing economies.
3. In the 1980s, growth in Sub-Saharan Africa was both low and unstable. In the 2000s, by contrast, it was both relatively high and stable. Do you think that these outcomes reflect a cause-and-effect relationship? If so, in what direction? If not, how else could one explain them?
4. Suppose that in a sample of 50 countries with 30 years of data for each country (a total of 1,500 observations), annual changes in inflation rates were found to be positively correlated with annual changes in real GDP. What could you infer from this evidence about whether high inflation is good for capacity growth? Explain.
5. Consider the same set of data as in the previous problem. Suppose that after averaging the growth rates and inflation rates of the 50 countries over the 30 years of observation for each country, you now find that growth and inflation are *negatively* correlated in the resulting cross-sectional sample. What can you now conclude about whether high inflation is good for capacity growth? Explain.

REFERENCES AND FURTHER READING

Barro, Robert J. (1996), "Inflation and Growth," Federal Reserve Bank of St. Louis *Review* Vol. 78, No. 3 (May/June), pp. 153–169.

Brock, W. A., and S. N. Durlauf (2001), "Growth Empirics and Reality," *World Bank Economic Review*, vol. 15, pp. 229–272.

Bruno, Michael, and William Easterly (1995), "Inflation Crises and Long-Run and Growth," NBER Working Paper 5209.

Calderon, Cesar, William Easterly, and Luis Serven (2003), "Infrastructure Compression and Public Sector Solvency in Latin America," in W. Easterly and L. Serven, eds., *The Limits of Stabilization: Infrastructure, Public Deficits, and Growth in Latin America*, Washington, DC: World Bank, pp. 119–138.

Corden, Max (1991), "Macroeconomic Policy and Growth: Some Lessons of Experience," in *Proceedings of the World Bank Annual Conference on Development Economics 1990*, Washington, DC: World Bank, pp. 59–88.

Easterly, William (1989), "Fiscal Adjustment and Deficit Financing during the Debt Crisis," policy planning and research working paper 138, World Bank.

Easterly, William (1991), "How Much Do Distortions Affect Growth?" unpublished manuscript.

Easterly, William, Robert King, Ross Levine, and Sergio Rebelo (1992), "How Do National Policies Affect Long-Run Growth? A Research Agenda," discussion paper 164, World Bank.

Edwards, Sebastian (2008), "Globalization, Growth and Crises: The View from Latin America," *Australian Economic Review*, vol. 41, pp. 123–140.

Fernandez-Arias, Eduardo, and Peter J. Montiel (2001), "Reform and Growth in Latin America: All Pain, No Gain?" *IMF Staff Papers*, vol. 48, pp. 522–546.

Fischer, Stanley (1993), *Does Macroeconomic Policy Matter? Evidence from Developing Countries*, occasional paper 27, International Center for Economic Growth.

Hausmann, Ricardo, Dani Rodrik, and Andres Velasco (2005), "Growth Diagnostics," unpublished manuscript, Harvard University.

Hsieh, Chang-Tai, and Peter Klenow (2009), "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, vol. 124, pp. 1403–1448.

Levine, Ross, and David Renelt (1992), "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, vol. 82, pp. 942–963.

Montiel, Peter J. and Luis Serven (2006), "Macro Stability in Developing Countries: How Much Is Enough?" *World Bank Research Observer*, pp. 151–178.

Sarel, Michael (1996). "Non Linear Effects of Inflation on Economic Growth," *IMF Staff Papers*, vol. 43, pp. 199–215.

Temple, Jonathan (1999), "The New Growth Evidence," *Journal of Economic Literature*, vol. 37, pp. 112–156.

PART 2

A Benchmark Macroeconomic Model

4

The Aggregate Production Function, the Labor Market, and Aggregate Supply

Our next task is to develop an analytical framework – a model – that we can use to think about macroeconomic issues in emerging and developing economies. That will be our objective in Part 2, consisting of this chapter and the four that follow. We will use this model to identify the types of shocks that tend to disrupt short-run macroeconomic stability in those economies, to analyze how such shocks affect a broad range of macroeconomic variables, and to examine the set of instruments that policy makers can potentially deploy to ameliorate the effects of such shocks. We begin that task in this chapter and the next by developing a benchmark version of our model, that is, a simple version to which we will add bells and whistles later on. This chapter and the three that follow build a "short-run" version of this model, in the sense that it holds constant certain important macroeconomic variables that in the real world tend to change very slowly over time. We extend the model to a medium-run setting in Chapter 8 by studying how some of these slowly changing variables evolve over time.

The economy that we will begin describing in this chapter has certain features that are worth calling attention to at the outset. First, it is an open economy in the sense that it trades both goods and services as well as financial assets with the rest of the world. As we have seen, "real" openness – openness to the exchange of goods and services with the rest of the world – is an important feature of the vast majority of emerging and developing economies, and because as developing countries grow into emerging ones, their degree of financial integration with the rest of the world tends to increase, the effects of financial openness on domestic macroeconomic performance will be one of our main concerns throughout this book.

Second, in keeping with the "stylized facts" about emerging and developing economies described in Chapter 1, we will assume that the economy that we are describing is small, so it can have no influence on the market prices of the goods and services that it buys from the rest of the world, or on the interest rates that prevail in international financial markets.

Finally, in this part of the book, we will also assume that the central bank maintains an officially determined exchange rate regime, so the official exchange rate can be used as a policy instrument. As we have seen, this is true of many, though not all, emerging and developing economies. We will consider how we would need to modify our framework to apply it to an economy with floating exchange rates in Chapter 17.

With these preliminaries out of the way, in the rest of this chapter, we will begin building our first simple model by examining the factors that determine the aggregate supply of domestic goods – that is, the relationship between the price of domestic goods and the desired level of output of domestic firms – in the economy we are describing. The first step in doing so is to specify the technical conditions of production, that is, how aggregate output is linked to aggregate inputs through the production process. That will be the task of Section I. Because in the short run, labor will turn out to be the only variable factor of production, Section II will study how the equilibrium level of employment is determined in the labor market. In Section III, we will derive the aggregate supply of domestic goods from the aggregate production function and the equilibrium level of employment. Section IV summarizes.

I. THE AGGREGATE PRODUCTION FUNCTION

The supply side of the goods market describes the relationship between the price of domestic goods and the quantity of such goods offered for sale by domestic firms. To derive this relationship, we need two components: an *aggregate production function* and an analysis of how the equilibrium level of employment is determined. This section describes the aggregate production function, and the next section explains how equilibrium is determined in the labor market.

The first step in describing how goods are produced in a given economy is to specify how many distinct types of goods we need to consider. For simplicity, it is convenient to assume that only one type of good is produced in the domestic economy (economists refer to this as *complete specialization* in production).[1] We will let the symbol Y denote the amount of this good produced during a given period of time and will refer to it as *real output*. Notice that Y is a *real* quantity because it is measured in units of goods, not of currency, and that it has the characteristics of a *flow*; that is, precisely because it denotes production during a specific period of time, it is measured per unit of time. In the real world, Y would represent a country's real gross domestic product (GDP).

[1] We can think of this single good as a composite, possibly consisting of many individual goods. Our assumption of complete specialization just means that we will not be analyzing changes in relative prices among goods produced domestically.

The *aggregate production function* is a relationship that summarizes how the feasible level of aggregate output in an economy is influenced by the amounts used of whatever inputs are used for production in that economy as well as by the productivity of such inputs. To produce the good, we will suppose that firms in the economy employ the services of physical capital and labor. We can therefore write the aggregate production function in the form

$$Y = AF(\underset{+}{K}, \underset{+}{L}), \tag{4.1}$$

where A is a parameter that serves as an index of the total productivity of the resources employed, K is the stock of physical capital, and L denotes the level of employment.[2] The positive signs under K and L signify that independent increases in each of these variables (i.e., an increase in one, holding the other constant) increase the level of output. Notice that we can write A as $A = Y/F(K, L)$, so A is the average productivity of a combination of physical capital and labor. An increase in A means that the economy becomes more productive, in the sense that more output can be produced with the same combination of physical capital and labor. Because changes in A correspond to changes in the productivity of a combination of both factors of production, A is usually referred to as an indicator of total factor productivity (TFP), which was introduced in the preceding chapter.[3]

1. Properties of the Aggregate Production Function

To use this production function, we will need to say something about its properties. We will assume that this function has three properties that are typical of *neoclassical* production functions:

- The function will be assumed to be *twice continuously differentiable*. This means two things. First, it means that each of the factors of production can be varied continuously and that such variations will produce continuous changes in the level of output. The change in the level of output corresponding to a small increase in one of the factors, holding the other constant, is the *marginal product* of that factor. Second, it means that the marginal product of each factor itself varies continuously when the factor is changed by small amounts.
- A second property is that these marginal products are positive and decreasing (the familiar property of *diminishing marginal returns*) for both physical capital and labor. This means that we can draw the marginal products of both types

[2] We are implicitly assuming that the flow of services of capital used in production is proportional to the size of the capital stock and that the flow of labor services depends on the number of workers employed.

[3] In contrast with TFP, the *average productivities* of physical capital and labor, taken in isolation, are given by Y/K and Y/L, respectively.

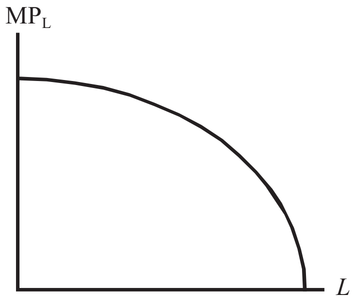


Figure 4.1. The marginal product of labor

of capital as negatively sloped curves in the positive quadrant. For example, the marginal product of labor can be depicted as in Figure 4.1. A similar picture could be drawn for the marginal product of physical capital. This property turns out to be important in short-run macroeconomic models such as the one we are beginning to build in this chapter.

- Finally, the function will be assumed to exhibit *constant returns to scale* (CRTS), which implies that if both of the factors of production are multiplied by a positive constant (i.e., if they are both changed in the same proportion), the level of output will change by a factor equal to that same constant (e.g., doubling the amounts of capital and labor used in production doubles the amount of output produced).

2. Short Run and Long Run in Macroeconomics

Over the rest of this chapter and the next, we will build a simple macroeconomic model that can be used to study the economy's short-run equilibrium. The macroeconomic *short run* is usually defined as a period of time over which total factor productivity and the physical capital stock are not perceptively affected by shocks to other variables in the economy – that is, they can be taken as exogenous. The basic intuition is that stocks of capital and the determinants of TFP tend to change very slowly compared to the pace at which several other important macroeconomic phenomena play themselves out.[4]

[4] To get a sense for this, consider the following example, which illustrates typical annual changes in a country's capital stock. Suppose the ratio of the capital stock to annual output (the capital-output ratio) is 3 and that 7% of the capital stock wears out each year. Under these circumstances, gross domestic investment of 21% of GDP would be required to keep the capital stock from changing. If an economy invests 30% of GDP (a high figure, as you can verify from Table 2.1), then the net addition to the capital stock each year would be 9% of GDP. But this is only a 3% change in the size of the capital stock. In many countries, this would barely be enough to keep up with the expansion of the effective labor force (i.e., labor force growth augmented by the change in worker productivity) so that the ratio of the capital stock to effective labor would remain unchanged. Thus achieving large changes in the capital stock relative to the size of the economy tends to be a slow process.

Given the capital stock and the level of TFP, the level of output that the economy can produce depends on how much labor it employs. Heuristically, *full employment* refers to a situation in which everyone who wants a job can get one (the next section provides a more precise definition). When total employment L is at its full-employment level, say, L_P, the resulting level of output is variously referred to as the *potential*, *natural*, or *full-employment* level of GDP, as discussed in Chapter 2. Thus full-employment GDP is given by

$$Y_P = AF(K, L_P,) \tag{4.2}$$

Short-run macroeconomics is typically concerned with stabilization of employment around its full-employment level, the determination of the average price level in the economy, and the behavior of various components of the economy's balance of payments. The long run, by contrast, is a period of time long enough that the capital stock and state of technological knowledge can change endogenously. *Long-run macroeconomics* is primarily concerned with what determines how the level of the economy's productive capacity changes over time. As we saw in the preceding chapter, increases in economic capacity are what we refer to when we use the term *economic growth.*

Notice that this means that growth does not just refer to an increase in real GDP but to an increase in productive *capacity*, whether that capacity is used or not. It is useful to clarify the distinction algebraically. Using the aggregate production function, we can approximate the change in (actual) output during any given period of time as the sum of contributions made by each of the three arguments in the production function, where the contribution of each is the change in that argument multiplied by its marginal product,

$$\Delta Y = F\Delta A + MP_K \Delta K + MP_L \Delta L$$

because MP_A, the marginal product of A, is just F. Dividing through by Y, and noting that $AF/Y = 1$,

$$\Delta Y/Y = \Delta A/A + MP_K(K/Y)\Delta K/K + MP_L(L/Y)\Delta L/L \tag{4.3}$$[5]

Recalling that for any variable X, $\Delta X/X$ is just the *rate of growth* of X, this equation states that the growth of Y depends on the rates of growth of total factor productivity, of the stock of physical capital, and of labor. An equation such as this is often used by economists to decompose growth into the contributions of factor accumulation and improvements in TFP, in an exercise referred to as *growth accounting*. The method is explained in Box 4.1.

Recall that under competitive conditions, the services of factors of production are remunerated at a rate equal to their marginal products. Thus the quantities

[5] Notice that the second term on the right-hand side of this equation has been simultaneously multiplied and divided by K, whereas the third term has been multiplied and divided by L.

Box 4.1. Growth Accounting

An important question in applied macroeconomics is how to measure TFP. This box describes how this is typically done in practice.

It is common to assume that the aggregate production function takes the Cobb-Douglas form $Y = AK^{\theta}L^{1-\theta}$. As we have seen, this gives the expression for the growth rate of real output in equation (4.4). Using this equation, we can write the growth rate of output per worker g_y as

$$g_y = (\Delta Y/Y - \Delta L/L) = \Delta A/A + \theta(\Delta K/K - \Delta L/L)$$

According to this equation, the growth rate of output per worker is the sum of the growth rate of TFP, $\Delta A/A$, and the growth rate of capital per worker multiplied by capital's income share, $\theta(\Delta K/K - \Delta L/L)$.

On the basis of this equation, the rate of growth of TFP can be estimated using a residual method; that is, TFP growth is given by

$$\Delta A/A = g_y - \theta(\Delta K/K - \Delta L/L)$$

To measure TFP, therefore, we need data on output per worker, the share of capital income, and the capital stock per worker. Output per worker and the share of capital income in GDP are both directly observable from a country's national income accounts and employment data. Estimates of a country's capital stock are typically constructed using a *perpetual-inventory method*; that is, the capital stock at time t is given by

$$K_t = I_t + (1 - \delta)K_{t-1}$$

where I_t is gross investment during period t (observable from the national income accounts) and δ is an assumed rate of depreciation of the capital stock. An initial estimate of the capital stock can be derived by making a reasonable assumption about the country's capital-output ratio at the beginning of the period (usually a number in the 2.0–3.0 range).

In practice, researchers usually work with a production function that includes human capital. For example, Bosworth and Collins (2003) assume a function of the form $Y = AK^{\theta}(LH)^{1-\theta}$, where H is a measure of educational attainment per worker. Table 4.1 presents the results of their growth accounting decomposition for seven major world regions. As the table shows, world output per worker grew on average by 2.3 percent per year during 1960–2000. Physical capital per worker and TFP contributed roughly 1 percentage point a year, while increases in human capital added about 0.3 points per year.

East Asia (excluding China) is the fastest-growing region, with output per worker increasing by 3.9 percent per year over 1960–2000. Interestingly, TFP in this region grew barely more rapidly than the overall world average. East Asia's rapid growth does not seem to be the result of strong gains in TFP but of above-average contributions from gains in human capital and, most important, from large and sustained increases in the physical capital stock.

Table 4.1. *Sources of Growth by Region*

Region and period	Growth in Output per Worker (Percentage a Year)	Contribution by Component (Percentage Points): Physical Capital per Worker	Education per Worker	Total Factor Productivity
World (84 countries) 1960–2000	2.3	1.0	0.3	0.9
Industrial countries (22) 1960–2000	2.2	0.9	0.3	1.0
China 1960–2000	4.8	1.7	0.4	2.6
East Asia except China (7) 1960–2000	3.9	2.3	0.5	1.0
Latin America (22) 1960–2000	1.1	0.6	0.4	0.2
South Asia (4) 1960–2000	2.3	1.0	0.3	1.0
Africa (19) 1960–2000	0.6	0.5	0.3	−0.1
Middle East (9) 1960–2000	2.1	1.1	0.4	0.5

Note: Regional averages are aggregated with purchasing-power-parity GDP weights
Source: Bosworth and Collins (2003)

$MP_L L/Y$ and $MP_K K/Y$ are the shares of the aggregate income generated in the economy that are received by labor and capital, respectively. Under constant returns to scale, these shares must sum to unity. To simplify matters further, let us assume that these shares are constant, and let the symbol θ denote the share of physical capital, that is, $\theta = MP_K(K/Y)$.[6] Then we can write equation (4.3) as follows:

$$\Delta Y/Y = \Delta A/A + \theta \Delta K/K + (1-\theta)\Delta L/L \tag{4.4}$$

Now doing the same thing for full-employment output, we have the following:

$$\Delta Y_P/Y_P = \Delta A/A + \theta \Delta K/K + (1-\theta)\Delta L_P/L_P \tag{4.5a}$$

Subtracting the second of these equations from the first and reorganizing,

$$\Delta Y/Y = (1-\theta)(\Delta L/L - \Delta L_P/L_P) + \Delta Y_P/Y_P \tag{4.5b}$$

[6] The assumption of constant shares means that the aggregate production function takes the *Cobb-Douglas* form $Y = AK^{\theta}L^{1-\theta}$, where θ is a positive fraction.

This equation explains how the economy's *actual* growth (in the form of year-to-year changes in output) can differ from the growth in its productive capacity. Growth of actual real GDP (given by the left-hand side of this equation) and growth of productive capacity (given by the last term on the right-hand side) are not the same thing. They differ whenever the rate of growth of actual employment differs from that of the labor force.

II. THE LABOR MARKET

As already mentioned, our model is intended to explain only *short-run* variations in the level of output. Operationally, this means that we will treat the level of total factor productivity A, the stock of capital K, and the size of the labor force L_P as exogenous variables. As we have just seen, the production function implies that for the given levels of these variables, changes in real output must be associated with changes in the actual level of employment L relative to L_P. Thus we can determine how much output firms would like to sell if we can determine how much of the available labor they would like to employ. Our next task is therefore to examine what determines the equilibrium level of employment in our economy. To do so, we need to investigate how the labor market works.[7]

1. The Aggregate Demand for Labor

The first step in doing so is to consider how much labor firms wish to hire, given their technology and capital stock. To give this question a concrete answer, we need to describe the economic environment in which firms operate and to specify what they are trying to achieve. We will assume that all the firms in our economy operate in perfectly competitive markets both for the output they sell and for the inputs they use. This means that they take market prices as given in both types of markets. We will also make the familiar assumption that their objective is to maximize profits. A standard result from introductory economics describes how profit-maximizing firms make their labor-hiring decisions under such circumstances: they choose the level of employment for which the *marginal revenue product* of labor (the extra revenue that the firm can make from selling the output produced by an extra worker) is just equal to the market wage that the firm has to pay:

$$PMP_L(\underset{-}{L^D}; \underset{+}{A}, \underset{+}{K}) = W \tag{4.6}$$

where P is the market price of the firm's output, W is the nominal wage prevailing in the labor market, L^D is the level of employment (the demand for labor), and $MP_L(\;)$ is the function denoting the marginal product of labor, which is decreasing in the

[7] For overviews of how labor markets function in developing countries, see Agenor (1996) and Behrman (1999).

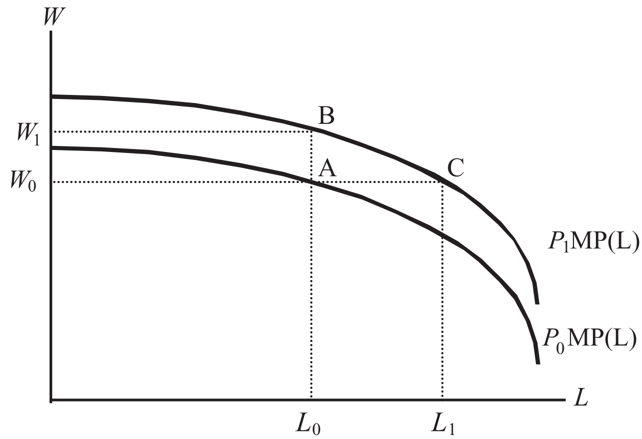


Figure 4.2. The aggregate demand for labor

level of employment. This function depends not just on the level of employment but on the level of productivity and the capital stock as well. The positive signs under A and K reflect the plausible assumption that higher TFP and a larger stock of capital with which to work both increase the productivity of an extra worker.

Equation (4.6) describes the firm's demand for labor. To see why this is so, notice that for any given values of the domestic price level P and of the nominal wage W faced by the firm in product and labor markets, the solution of this equation for the demand for labor L^D tells us how much labor the firm would have to hire to maximize profits. Similarly, the firm's desired employment at any other value of the nominal wage could be determined by solving equation (4.6) for L^D using the new value of W. Because perfect competition in product markets means that individual firms take the price of domestic output P as given, and because the property of diminishing marginal returns implies that the function $MP_L(\ )$ is decreasing in L^D, decreases in the nominal wage W will be associated with higher levels of desired employment at the firm level; that is, when plotted in (L, W) space, the individual firm's demand-for-labor curve has a negative slope.

The economy's *aggregate* demand for labor is derived by adding the amount of labor that all firms in the economy would like to hire at a given nominal wage, which is the same for all firms. Because a change in the wage rate would give rise to a similar change in the desired level of employment by all the firms in the economy, the curve in (L, W) space depicting the economy's demand for labor must also have a negative slope, albeit a much flatter one than the curve depicting the demand for labor by an individual firm (i.e., the aggregate demand for labor curve is more *elastic* than that of an individual firm), because a given change in W results in a much larger change in employment for the economy as a whole than it does for an individual firm. Otherwise, the properties of the aggregate demand for labor resemble those of the demand for labor by an individual firm. The economy's aggregate demand for labor is depicted graphically as the curve $P_0\ MP_L(L)$ in Figure 4.2.

The aggregate demand for labor has an important property of which we will make frequent use later on: when the values of the domestic price level P and nominal wage W facing firms change in the same proportion, the economy's aggregate demand for labor will be unchanged. You can verify this from equation (4.6) for each individual firm: multiplying P and W by the same constant leaves the equation satisfied at the same value of L^D. This result is shown graphically for the aggregate demand for labor in Figure 4.2. A change in the price of output causes the labor demand curve to shift vertically in the same proportion as the change in P. Thus, when P increases by $(P_1 - P_0)/P_0$ percent, the labor demand curve shifts upward by $(P_1 - P_0)/P_0$ percent, say, from $P_0MP(L)$ to $P_1MP(L)$. This means that if the nominal wage rate rises in the same proportion, that is, by $(P_1 - P_0)/P_0$ percent, say, to $W_1 = [1 + (P_1 - P_0)/P_0]\ W_0$ – the PMP_L curve and the horizontal line depicting the nominal wage would intersect at a point such as B, directly above their previous intersection at A. As a result, the demand for labor would be unchanged at its original quantity L_0.

Notice, however, if the value of W were to remain unchanged in the face of the increase in P, the $PMP_L(L)$ curve would shift vertically while the W line would remain in the same position. In this case, the two curves would intersect at a point like C, and employment would rise, say from L_0 to L_1 in the figure at the unchanged value of the nominal wage W_0.

2. The Aggregate Supply of Labor

In the model economy we are constructing, we will assume that domestic workers cannot work abroad and that foreign workers cannot work in the domestic economy. Under these circumstances, the supply of labor in the economy depends only on the behavior of domestic households. On the basis of the standard microeconomic analysis of the labor-leisure choice made by consumers, the supply of labor is typically taken to be an increasing function of the *real wage* (the nominal wage divided by the price of the typical worker's consumption basket). The dependence on the real rather than nominal wage reflects the assumption that what workers care about is the purchasing power of their wages over goods and services.

Notice that what firms should care about, by contrast, is the relationship between the nominal wage they have to pay workers and the price of their specific output, that is, the real wage measured in *units of their own output*. This is called the *product wage*. This asymmetry between the real wage concepts that are relevant to the firm and to workers has an important implication. Whereas firms can continuously observe the prices of the goods or services that they sell, workers cannot necessarily continuously observe the prices of the goods and services that they consume because they typically consume a wide variety of things, some of which they buy infrequently. Consequently, at any given moment in time, they may not be perfectly informed about the prices of all the goods and services that are relevant to their labor supply choice. This means that workers may have to make their labor supply decisions *before* they know the prices of the goods and services that they intend to buy with their wages. Under these conditions, the price level relevant for their labor supply decision is the one that they *expect* to prevail during the term of their employment at the agreed wage, which we can denote P^e. Thus the aggregate labor supply function can be written as follows:

$$L^S = L(\underset{+}{W/P^e}, \ldots)$$

where L^S is the supply of labor and the ellipses represent other unidentified variables that may affect the supply of labor, intended to capture the effects on labor supply of phenomena such as changes in the total size of the labor force, in labor force participation rates, and so on.

This labor supply function can equivalently be expressed in terms of the *expected real wage* that workers must be offered for them to supply a given amount of labor, that is,

$$W/P^e = w(\underset{+}{L^S}, \ldots)$$

or as the *nominal* wage that workers must be paid for them to willingly supply a given amount of labor:

$$W = P^e w(\underset{+}{L^S}, \ldots) \tag{4.7}$$

In the real world, the economy-wide nominal wage demanded by workers may be "sticky," in the sense that it may only gradually adjust to the desired nominal wage given by equation (4.7). For example, because not all labor contracts expire at the same time, the average nominal wage in the economy will partly reflect the wage prevailing in labor contracts that were negotiated sometime in the past and have not yet expired. We can capture this nominal wage stickiness by specifying a process of wage adjustment in which the nominal wage adjusts gradually to the level given by equation (4.7); that is,

$$\Delta W = W - W_{-1} = \sigma(P^e w(L^S, \ldots) - W_{-1})$$

where W_{-1} is last period's nominal wage and σ, a positive fraction, measures the degree of nominal wage flexibility: the larger is σ, the *smaller* the influence of past wages on current wages. When $\sigma = 1$, past wages have no influence on current wages, and you can verify from this equation that the nominal wage will be given by equation (4.7). We can refer to this case as one of *perfect wage flexibility* (no wage stickiness). On the other hand, when $\sigma = 0$, this period's nominal wage must be the same as last period's nominal wage, and current events can have no influence on the nominal wage. In this case, the nominal wage is exogenous. In the intermediate case with $0 < \sigma < 1$, the *supply price of labor*, or the *supply wage* (i.e., the nominal wage that workers require to be paid to induce them to supply a given amount of labor), can be written as

$$W = (1 - \sigma)W_{-1} + \sigma P^e w(\underset{+}{L^S}, \ldots) \tag{4.7a}$$

Graphically, equation (4.7a) can be depicted as a curve in (L, W) space, drawn for given values of last period's wage and the expected price level. The curve has the following properties:

- It has a positive slope that depends on how willing workers are to supply additional labor in response to an increase in the expected real wage. The greater the responsiveness of labor supply to a change in the expected real wage, the flatter the curve.
- Changes in W_{-1} or in P^e cause the curve to shift vertically in the same direction as the change in these variables. However, the size of the vertical shift in the labor supply curve must be *less than* in proportion to the change in W_{-1} or P^e.

Notice that the existence of nominal wage *inertia* (stickiness) has two effects on the labor supply curve:

1. The greater the degree of wage inertia (the smaller σ), the greater the share of the labor force that is willing to work at the previously prevailing wage, so the *smaller* the increase in the wage required to elicit a given amount of additional labor supply. In other words, greater wage inertia makes the labor supply curve flatter. In the limit, if $\sigma = 0$ (fixed nominal wage), the labor supply curve would be horizontal.
2. The effect of changes in the *expected* price level on the supply price of labor depends on the degree of wage inertia. When the expected price level changes, the labor supply curve shifts vertically by σw times the change in the expected price level. Thus the *smaller* the degree of wage inertia, the *larger* the vertical shift in the labor supply curve that is associated with a given change in the expected price level. In particular, when $\sigma = 1$, the labor supply curve shifts vertically in the *same* proportion as the change in the expected price level.

3. The Formation of Expectations

Because changes in the expected price level shift the labor supply curve, you can see that the particular mechanism that workers use to form their expectations about the price level will play an important role in determining how the labor market responds to any macroeconomic shocks that may be expected to affect the aggregate price level. Several mechanisms for the formation of expectations are conceivable. For example, under *static expectations*, workers always expect the current price level to be unchanged from the previous period, that is, $P^e = P_{-1}$. Under *adaptive expectations*, workers revise their price-level expectations based on the expectational errors that they made the previous period – that is, $\Delta P^e = \gamma({P^e}_{-1} - P_{-1})$, where

γ is a positive fraction (here and in the rest of the book, the symbol Δ denotes the change in a variable from one period to the next). Notice that under either of these mechanisms for forming expectations, current price-level expectations are not affected by anything that happens in the economy during the current period. That means that P^e is effectively predetermined.

We will assume instead that workers use a mechanism called *rational, or model-consistent, expectations.* This means that in forming their expectations about the price level, workers use the model itself, together with all the available information about any factors that may be affecting the current price level, to form their expectations about what the current price level will be. In that case, in the absence of any unanticipated shocks, their expectations will turn out to be correct. In the absence of unanticipated shocks, therefore, model-consistent expectations are equivalent to *perfect foresight.*

We therefore have two possibilities:

1. When shocks to the economy are unanticipated by workers, the supply wage is given by equation (4.7a), with P^e unaffected by the shocks and therefore exogenous.
2. By contrast, when shocks to the economy are correctly anticipated by workers, workers will correctly infer the value of the current price level, so $P = P^e$, and the supply wage is given by the following:

$$W = (1-\sigma)W_{-1} + \sigma P\, w(L^S, \ldots) \tag{4.7b}$$

Notice that when price-level changes are unanticipated, changes in the actual price level P have no effect on the supply wage because P^e does not change. When shocks are anticipated, however, an increase in the price level induces workers to demand a higher nominal wage, causing the labor supply curve to shift vertically upward. However, wage inertia causes the vertical displacement of the labor supply curve to be less than in proportion to the change in the price level. This means that the vertical shift in the labor supply curve must be smaller than that in the labor demand curve. Because the effects of price-level changes on the magnitude of shifts in the labor supply curve relative to shifts in the labor demand curve are therefore *qualitatively* similar (i.e., the demand curve shifts more than the supply curve), whether shocks are anticipated or not, in what follows, we will assume for concreteness that shocks are anticipated and will use the expression for the supply wage given by equation (4.7b).

4. Labor Market Equilibrium

Now that we have described both labor demand and labor supply, we can examine equilibrium in the labor market. The labor market equilibrium condition is given by the following:

$$PMP_L(L, \ldots) = (1-\sigma)W_{-1} + \sigma\, Pw(L, \ldots) \tag{4.8}$$

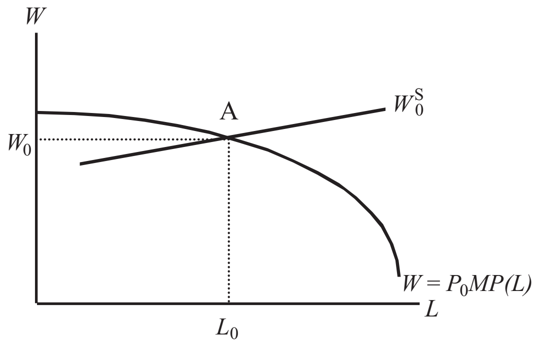


Figure 4.3. Labor market equilibrium

This equation states that the equilibrium level of employment L is that at which the wage that firms are willing to offer (the *demand wage*, given by equation (4.6)) is equal to the wage that workers require to be paid (the supply wage, given by equation (4.7b)). Our labor market model thus consists of equations (4.6), (4.7b), and (4.8). Given the equilibrium level of employment determined by equation (4.8), the equilibrium wage can be determined from equations (4.6) or (4.7b). The model is depicted graphically in Figure 4.3. Equilibrium is initially at the point A, where equation (4.8) holds, so the demand for labor is equal to the supply of labor. The equilibrium level of employment is L_P, and the equilibrium value of the nominal wage is W_0.

5. Full Employment

The labor market model that we have just described is inherently dynamic, in a very specific sense: if the current period's equilibrium wage turns out to be different from W_{-1}, which was last period's equilibrium wage, then the labor market equilibrium would have to change in the *next* period. The reason is that the current equilibrium wage W_0 would become next period's lagged wage, and because that is different from W_{-1}, the position of the labor supply curve would have to be different as well. The only labor market equilibrium that would tend to be sustained over time, in the absence of new shocks to the economy, is the one that prevails when $W = W_{-1}$. Its sustainability makes it natural to refer to the level of employment that prevails under these circumstances as the *full-employment* level of employment, which we previously denoted L_P.

From equation (4.7a), the condition $W = W_{-1}$ means that the supply wage must be given by $W = Pw(L^S, \dots)$. The full-employment level of employment must therefore satisfy the labor market equilibrium condition (demand wage equals supply wage):

$$PMP_L(L; A, K) = Pw(L, \dots)$$

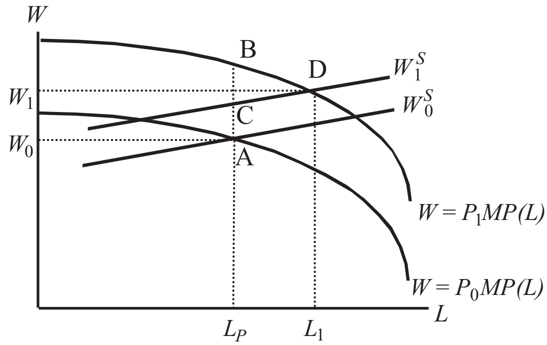


Figure 4.4. Effects of a change in the price level

or

$$MP_L(L; A, K) = w(L, \ldots)$$

Because increases in A and K would tend to increase the marginal product of labor at a given level of employment, they would increase the demand wage, thus shifting the labor demand curve upward and increasing the full-employment level of employment. In addition, any unobserved factors that shift the labor supply curve downward (upward) would also have the effect of increasing (decreasing) the full-employment level of employment. Thus we can write the full-employment level of employment as follows:

$$L_P = L_P(\underset{+}{A}, \underset{+}{K}, \ldots) \tag{4.9a}$$

where the ellipses represent unobserved shocks to the labor supply curve. Substituting this expression into our expression for the supply wage $W = Pw(L, \ldots)$, we can write the full-employment level of the real wage as follows:

$$w_P = Pw[L_P(A, K, \ldots), \ldots]/P = \omega(\underset{+}{A}, \underset{+}{K}, \ldots) \tag{4.9b}$$

6. Equilibrium Employment and the Price Level

Our next task is to examine what happens in the labor market when the aggregate price level changes. Suppose that at the initial equilibrium point A in Figure 4.4, the economy is at full employment (i.e., the current nominal wage is the same as last period's nominal wage). Now consider what happens in the labor market when there is a change in the aggregate price level.

To investigate this question, recall that the labor demand curve must shift vertically in the same proportion as the change in the price level, but the presence of wage inertia causes the labor supply curve to shift vertically *less* than in proportion to the change in the price level. In the case of an increase in the price level, for example, the labor demand curve shifts upward to a point like B in Figure 4.4, but the labor supply curve shifts vertically by a smaller amount, say, to a point like C. The market thus moves to a new equilibrium at the point D.

The increase in the price level thus results in an increase in both the equilibrium value of the nominal wage and the equilibrium level of employment. Firms are willing to hire more labor at the new equilibrium because, as you can verify from the figure, the equilibrium value of the nominal wage increases *less than in proportion* to the price level, causing the real wage to fall (recall that the labor demand curve shifts upward in proportion to the increase in the price level, but as you can see from the figure, the equilibrium nominal wage increases by less than that amount). In turn, workers are willing to supply more labor because the increase in the nominal wage exceeds what they require to compensate them for the increase in the price level (given by the point C). In the case of a decrease in the price level, a similar analysis would hold, except that because the labor demand and supply curves would both shift downward, the equilibrium level of employment would fall.

7. The Equilibrium Level of Employment and Full Employment

To conclude our analysis of the labor market, notice that by dividing both sides of equation (4.8) by the price level P, the labor market equilibrium condition can be written in *real* terms as follows:

$$MP_L(L, \ldots) = (1-\sigma)(W_{-1}/P) + \sigma\, w(L, \ldots) \tag{4.8a}$$

This equation tells us that the equilibrium level of employment depends on W_{-1}/P, the ratio of lagged nominal wage to the current price level. You can easily verify that when $W_{-1}/P = w_P$ (the full-employment real wage), this equation is satisfied with $L = L_P$ (the full-employment level of employment). Notice that when $W_{-1}/P > w_P$, for the equation to hold, we must have $L < L_P$, and when $W_{-1}/P < w_P$, we must have $L > L_P$. Therefore we can write the equilibrium level of employment as follows:

$$L = L[w_P \underset{+}{/} (W_{-1}/P)]L_P = L(Pw_p \underset{+}{/} W_{-1})L_P \tag{4.10}$$

with $L(1) = 1$; that is, if $W_{-1}/P = w_P$, then $L = L_P$. Notice that this equation makes the equilibrium level of employment an *increasing* function of the price level, as we showed earlier.

III. AGGREGATE SUPPLY

It is now easy to see what determines the level of output in this economy. Substituting the level of employment given by (4.10) into the aggregate production

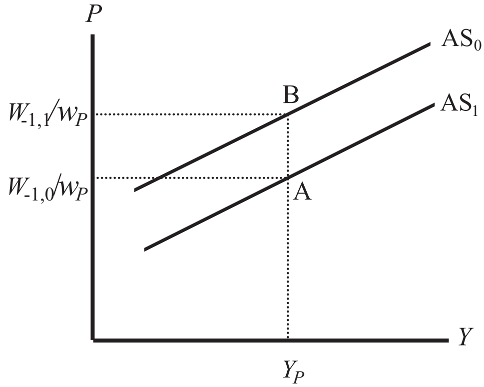


Figure 4.5. The aggregate supply curve

function (4.1), we have the following:

$$Y = AF[K, L(Pw_P/W_{-1})L_P] = Y(P; A, K, W_{-1}, \ldots) \qquad (4.11)$$

This relationship is the economy's *aggregate supply function.* It links the current level of output to the current price level as well as to the exogenous variables A and K and the predetermined variable W_{-1}.

This aggregate supply relationship is depicted in Figure 4.5, in the form of the curve labeled AS. This curve is drawn for given values of TFP, the capital stock, and the lagged nominal wage. Its properties are as follows:

- It has a positive slope because an increase in P for a given value of W_{-1} causes the labor demand curve to shift vertically upward by a larger amount than does the labor supply curve, thus increasing the equilibrium level of employment, as we have already seen. With a higher level of employment, the aggregate production function implies that firms produce a higher level of output.
- As the economy moves to the northeast along the aggregate supply curve, the nominal wage must be increasing and the real wage must be falling. Again, the reason is that increases in the price level shift the labor demand curve upward more than the labor supply curve, increasing the nominal wage less than in proportion to the increase in the price level, which causes the real wage to fall.
- Because equations (4.10) and (4.11) indicate that when $W_{-1}/P = w_P$, we must also have $Y = Y_P$, the AS curve must pass through the point $(Y_P, W_{-1}/w_P)$. This is an important observation. Its implication is that the *height* of the AS curve above the point Y_P is given by the ratio of the lagged nominal wage to the full-employment real wage. The intuition for this result is that $P = W_{-1}/w_P$ is the value of the price level that would be required to move the actual real wage – and thus the level of employment – to its full employment level. A consequence of this property is that when W_{-1} changes, the AS curve must shift vertically in the same proportion. For example, when W_{-1} increases from $W_{-1,0}$ to $W_{-1,1}$ in

Figure 4.5, the AS curve must shift upward from AS_0 to AS_1, passing through the point B.

- The elasticity of aggregate supply with respect to *P* is directly related to the elasticities of the labor demand and supply curves with respect to the nominal wage: the more elastic the labor demand and supply curves, the more elastic the aggregate supply curve. To see why this must be so, recall that a change in the price level shifts both the labor demand and supply curves vertically in proportions that are independent of how flat or steep the two curves are. But the resulting effect on the equilibrium level of employment – and therefore on the level of output – will tend to be larger the more elastic the labor demand and supply curves are because the more elastic those curves are, the larger the increase in the level of employment required to close the gap between the demand wage and supply wage created by given vertical shifts in the two curves.
- Finally, because a given change in the level of employment will have a larger effect on output the larger the marginal product of labor, holding constant the shapes of the labor demand and supply curves, the larger the marginal product of labor, the flatter the aggregate supply curve.

IV. SUMMARY

In this chapter, we have constructed the first building block of our short-run macro model: the aggregate supply function. Our point of departure was an aggregate production function that linked the level of real output to the economy's total factor productivity, its stock of capital, and the level of employment – the economy's use of labor services. Though, as we saw in the preceding chapter, improvements in total factor productivity and increases in the stock of capital are the key drivers of long-run growth, these changes tend to occur relatively slowly over time, so in the short run, labor is the only variable factor of production. This means that to explain the level of output that an economy produces during a given period, we need to explain what determines the level of employment.

This led us to an analysis of labor market equilibrium. We saw that the nature of that equilibrium depends in an important way on the factors that determine the wage that workers require to be paid to supply a given amount of labor – the supply wage. If this wage is not fully flexible in the short run – that is, if the nominal wage displays some degree of "stickiness" – then changes in the aggregate price level will be associated with changes in the equilibrium level of employment, and therefore with changes in the aggregate level of output. This relationship between the aggregate price level and the aggregate level of output that firms wish to supply in the goods market is what we mean by the economy's aggregate supply curve.

Our next tasks are to examine the demand side of the goods market and to put the supply and demand sides together to explain the short-run determination of the level of output and the aggregate price level. Chapter 5 takes up those issues.

REVIEW QUESTIONS

1. What is an aggregate production function? Describe the properties of the aggregate production function that we have been using in this chapter, including an explanation of which variables can be taken as exogenous in the short run, and why.
2. Explain the role of the labor market in determining the level of employment in an economy in the short run. Explain the effect each of the following would have on the equilibrium level of employment:
   a. an unanticipated increase in the domestic price level
   b. an increase in the expected price level, with no change in the actual price level
   c. an anticipated change in the price level
3. Explain how the full-employment level of employment and the full-employment (potential) levels of output are determined in the short run.
4. What is an economy's aggregate supply curve? Explain how the aggregate supply curve can be derived from the labor market and the aggregate production function. Explain what determines the position of the aggregate supply curve in (Y, P) space and what would cause the curve to shift.
5. Explain how the degree of short-run nominal wage inertia affects the slope of an economy's aggregate supply curve.

EXERCISES

1. Suppose that instead of the form given in equation (4.1), the aggregate production function takes the following form:

$$Y = F(K, AL)$$

   where the symbols are as defined in the chapter. What is the difference in the role of TFP between these two functions? Using the Cobb-Douglas version of this function, explain how you would estimate the rate of growth of an economy's total factor productivity in this case.
2. Two extreme cases of the labor supply function that we assumed in this chapter are the Lucas labor supply function, in which the nominal wage is fully flexible (i.e., with $\sigma = 1$), and the Keynesian labor supply function, in which the nominal wage is fixed (i.e., $\sigma = 0$). Derive the economy's aggregate supply curve (for goods) under each of these conditions and compare it to the curve that we derived in this chapter.
3. Suppose that the labor supply curve in our model is given by the following:

$$W = \omega P^{\beta}(P_{-1})^{1-\beta}$$

   where W is the nominal wage, ω is a positive constant, P is the actual (current) price level, P_{-1} is the previous period's price level, and β is a positive constant.

You will need to know that this function has the property that a 1 percent increase in P increases W by β percent, whereas a 1 percent increase in P_{-1} increases it by $1 - \beta$ percent. Derive the economy's aggregate supply (of goods) curve for each of the following three cases:

a. $\beta = 1$
b. $0 < \beta < 1$
c. $\beta = 0$

What is the relationship between these three curves? What is the economic reason for that relationship?

4. When modeling labor demand, we assumed that firms did not pay taxes on labor, but payroll taxes are common in both industrial and developing countries. To analyze the macroeconomic effects of such taxes, suppose we introduce a tax on labor such that the firm's profit maximization condition becomes the following:

$$PMP_L(A, K, \ldots) = (1 + \tau)W$$

where τ is the fixed tax rate per unit of hours worked.

a. Derive the shape of the AS curve in the presence of the tax and compare it to the standard AS curve.
b. In response to a positive productivity shock, are the effects on output larger or smaller in the economy with the labor tax compared to the standard case?

5. The aggregate supply model that we developed in this chapter assumed that firms were perfectly competitive in product markets. Suppose that they are *imperfectly* competitive instead, and that they set their prices according to the formula

$$P = (1 + m)W/MP_L(L)$$

where m is a constant markup over the marginal cost of production, given by $W/MP_L(L)$ (notice that we have implicitly assumed $m = 0$ in the chapter). Assuming that workers are willing to supply any amount of labor that firms want to hire at a fixed nominal wage, derive the aggregate supply curve for this economy. How would that curve be affected by an increase in the markup m?

REFERENCES AND FURTHER READING

Agenor, P. R. (1996), "The Labor Market and Economic Adjustment," *IMF Staff Papers*, vol. 43, pp. 261–335.

Behrman, J. (1999), "Labor Markets in Developing Countries," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, vol. 3, Amsterdam: North-Holland, pp. 2859–2939.

Bosworth, Barry, and Susan M. Collins (2003), "The Empirics of Growth: An Update," *Brookings Papers on Economic Activity* 2, pp. 113–206.

Canarella, Giorgio, and Stephen Pollard (1989), "Unanticipated Money Growth, Output, and the Price Level in Latin America," *Journal of Development Economics*, vol. 16, pp. 345–358.

Chopra, Ajai (1985), "The Speed of Adjustment of the Inflation Rate: A Cross-Country Study of the Role of Inertia," *IMF Staff Papers*, vol. X32, pp. 693–733.

Choudhary, Munir A. S., and Amer K. Parai (1991), "Anticipated Monetary Policy and Real Output: Evidence from Latin American Countries," *Applied Economics*, vol. 23, pp. 579–586.

Hausmann, Ricardo, and Michael Gavin (1996), "Securing Stability and Growth in a Shock-Prone Region: The Policy Challenge for Latin America," working paper 315, Inter-American Development Bank, Office of the Chief Economist.

Sheehy, Edmund J. (1986), "Unanticipated Inflation, Devaluation, and Output in Latin America," *World Development*, vol. X14, pp. 665–671.

5

Aggregate Demand and Goods-Market Equilibrium

Now that we have seen what determines the supply of goods produced in the domestic economy, we can turn to the other side of the market: the total demand for those goods. In this chapter, we will derive the aggregate demand for domestic goods under the simplifying assumption that the domestic interest rate is determined as an exogenous policy variable by the central bank. We will then put our description of aggregate demand together with that of aggregate supply from the preceding chapter to complete our first model of short-run macroeconomic equilibrium. We will use this model to explore how the domestic economy responds to a variety of shocks, both to the supply and the demand side of the economy, and we will trace the effects of such shocks not just on the market for domestic goods but also on the labor market and the economy's balance of payments accounts.

We begin by examining in Section I the factors that influence the aggregate demand for domestic goods by domestic as well as foreign residents. In Section II, we will put aggregate supply and demand together to show how the equilibrium values of the domestic price level and level of output are determined, as well as those of a variety of other macroeconomic variables that are of interest to policy makers in emerging and developing countries. Section III examines how all these endogenous variables respond to a wide range of both nonpolicy and policy shocks. Section IV shows how the analysis in the two preceding sections can be summarized graphically in the form of the goods-market equilibrium (GM) curve. This analysis serves as a building block for the more complete models that we will construct in the chapters that follow. Section V summarizes.

I. AGGREGATE DEMAND

Recall that we have assumed that the goods market in the economy we are describing is open to trade with the rest of the world and that the domestic economy is completely specialized in the production of a single (possibly composite) good.

To isolate the factors that determine demand for the domestically produced good, we need to specify the relationship between the domestic good and the goods produced by the rest of the world. We will assume that the domestically produced good is an *imperfect substitute* for the single (again, possibly composite) good produced by the rest of the world.[1]

1. Aggregate Expenditure

We will begin describing aggregate demand for the domestic good by examining what determines aggregate expenditure by domestic residents. Recall from the national income accounts of Chapter 2 that total spending by domestic agents can be separated into the spending done by households, firms, and government. Aggregate expenditure by domestic agents (sometimes called domestic *absorption*) can therefore be expressed as follows:

$$A = C + I + G$$

where C denotes total consumption spending by domestic households, I is desired investment spending by domestic firms, and G is total spending on goods and services by all levels of domestic government. All these quantities are measured in *real* units, that is, in units of the domestic good. To simplify things, we will draw a basic distinction between *private* absorption $C + I$, which we will denote A_P, and *public* absorption G. We will assume that public absorption, which consists of spending by the government on current goods and services, is an exogenously determined policy variable. Private absorption, however, is endogenous. We will assume that it is an increasing function of the *disposable income* of the private sector (national income Y net of taxes paid to the government by the private sector T) but a decreasing function of the domestic interest rate R. Higher disposable income increases the resources available for households to spend on consumption and encourages firms to spend more on investment, whereas a higher domestic interest rate would tend to reward saving by households and make it more costly for firms to invest.[2] We can therefore write private absorption as follows:

$$C + I = A_P = A_P(Y \underset{+}{-} T, \underset{-}{R}, \ldots)$$

[1] A production structure in which the domestic economy is completely specialized in the production of a single good that is an imperfect substitute for the rest of the world's output is the defining characteristic of the *Mundell-Fleming* model, named after Robert Mundell and J. Marcus Fleming, two economists who pioneered open-economy macroeconomic models in the early 1960s using a production structure of this type.

[2] We will assume for now that there is no ongoing inflation in the economy, so we do not need to distinguish at present between nominal and real interest rates. This distinction is introduced in Chapter 8.

The ellipses indicate that private absorption may also depend on other variables, which we leave unspecified. For now, we will assume that both the level of taxation and the domestic interest rate are exogenous policy variables. The former is a fiscal policy variable, whereas the latter will be taken to be the monetary policy variable in this chapter.[3] Finally, we will make the traditional assumption that the private sector's marginal propensity to spend out of current disposable income is less than unity, so a one-unit increase in $Y - T$ increases private absorption by less than one unit. With these assumptions, we can write aggregate domestic expenditure as follows:

$$A = A_P(\underset{+}{Y - T}, \underset{-}{R}, \ldots) + G$$

2. Aggregate Demand for Domestic Goods

Aggregate demand for domestic goods differs from aggregate domestic expenditure for two reasons: because not all spending by domestic residents is on domestic goods (some is on foreign goods, in the form of imports) and because foreigners also buy domestic goods (in the form of exports). Taking these two facts into account, we can write the aggregate demand for domestic goods as follows:

$$\begin{aligned} AD &= (A_P + G - Q \cdot IM) + X \\ &= A_P + G + (X - Q \cdot IM) \end{aligned} \tag{5.1}$$

where X denotes foreign demand for goods produced by the domestic economy (exports), measured in units of the domestic good, and IM is the domestic demand for foreign goods (imports), measured in units of the *foreign* good. Because IM is measured in units of the foreign good, we need to convert it into the equivalent number of domestic goods to compare it with the other magnitudes in equation (5.1). We make this conversion by multiplying IM by the relative price of foreign goods in terms of domestic goods, called the *real exchange rate* and denoted Q. The first term on the right-hand side of the first line in equation (5.1) is the *domestic* demand for domestic goods (total domestic spending minus the portion devoted to the purchase of foreign goods), whereas the second term captures the *foreign* demand for domestic goods. This expression for aggregate demand is rewritten in the second line as the sum of domestic expenditure and *net exports*, equal to exports minus imports.

What determines the real exchange rate Q? As the relative price of two goods, the real exchange rate is just the ratio of the nominal prices of the two goods, expressed in the same currency. Recall that the price of a domestic good in terms of domestic currency is P, and suppose that the price of the foreign good in terms of foreign currency is P^*. Let S denote the *nominal exchange rate* (the units of domestic

[3] The next chapter explains how the central bank sets the interest rate.

currency that must be given up per unit of the foreign currency). Then Q must be given by

$$Q = SP^*/P \tag{5.2}$$

Assuming that the domestic economy is small in the world market, it cannot influence foreign variables such as P^*. Thus we will take P^* to be exogenous. The determination of S, on the other hand, depends on the country's *foreign exchange regime* (the rules that govern the central bank's behavior in the foreign exchange market). As mentioned before, we will begin by assuming that S is *officially determined.* This means that the central bank determines its value by announcing its willingness to exchange the domestic currency for foreign exchange in unlimited amounts at the official price S. As we saw in Chapter 1, this is usually referred to as a *fixed exchange rate.* For our purposes, the importance of the assumption is that it makes S an exogenous policy variable. It is our model's indicator of *exchange rate policy.* With S and P^* exogenous, Q must be endogenously determined by P, with P and Q being inversely related.

3. Net Exports

Since we have already explained the determinants of aggregate expenditure, all we have left to do to complete the explanation of the determination of aggregate demand is to discuss what drives net exports $X - Q \cdot IM$. To do so, note that exports from the domestic economy reflect foreigners' demand for domestic goods, whereas imports by domestic residents reflect their demand for foreign goods. To keep things simple, we will assume that the domestic government only buys domestically produced goods. The private sector, however, buys both domestic and foreign goods. We will assume that the share of the domestic private sector's aggregate expenditure devoted to demand for domestic goods depends positively on the relative price of foreign goods (the price of foreign goods in terms of domestic goods, or the real exchange rate). Thus demand for domestic goods by the domestic private sector is given by $\varphi(Q, \ldots)A_P$, where $0 < \varphi < 1$ denotes the fraction of private absorption devoted to the purchase of domestic goods. The fraction is larger the cheaper such goods are, so it is an increasing function of Q, but may also depend on other unspecified variables that affect the relative preferences of domestic residents between domestic and foreign goods.[4] This means that spending on imports is given by

$$Q \cdot IM = (1 - \varphi\,(\underset{+}{Q}, \ldots))A_P \tag{5.3}$$

[4] Notice that because this means that total spending on imports must be a *decreasing* function of the real exchange rate, the demand for imports must implicitly be assumed to be elastic with respect to the real exchange rate. This is a stronger assumption than the more familiar Marshall-Lerner condition, which only requires the sum of export and import demand elasticities to exceed unity.

On the other hand, we will suppose that exports respond positively to the real exchange rate. The reason is, of course, that an increase in the real exchange rate represents an *increase* in the relative price of foreign goods (thus reducing domestic demand for such goods in the form of imports) but a *decrease* in the relative price of domestic goods (thus increasing foreign demand for them in the form of exports). We will also allow for the possibility that exports may depend on a host of other exogenous influences, such as partner-country incomes, terms-of-trade changes, and weather effects. Thus we can write the demand for exports from the domestic economy as follows:

$$X = X(\underset{+}{Q}, \ldots) \tag{5.4}$$

where the ellipses represent the other variables mentioned earlier.

Using equations (5.3) and (5.4), we can write net exports NX as

$$\begin{aligned} NX &= X - Q \cdot IM = X(Q, \ldots) - (1 - \varphi(Q, \ldots))A_P \\ &= NX(\underset{+}{Q}, \underset{-}{Y - T}, \underset{+}{R}, \ldots) \end{aligned} \tag{5.5}$$

Notice that the assumptions made so far imply that net exports are an increasing function of the real exchange rate and the domestic interest rate but a decreasing function of disposable income. A depreciation of the real exchange rate (an increase in Q) increases foreign demand for domestic goods and decreases domestic demand for foreign goods, thus increasing net exports, whereas increases in disposable income and *decreases* in the domestic interest rate both increase private absorption, thus increasing imports and decreasing net exports.[5]

4. The Aggregate Demand Curve

Substituting equation (5.5) into equation (5.1) and setting $AD = Y$, we have our final expression describing the aggregate demand for domestic goods:

$$Y = \varphi(SP^*/P, \ldots)A_P(Y - T, R, \ldots) + G + X(SP^*/P, \ldots) \tag{5.6}$$

Because aggregate demand is measured in units of the domestic good, we can plot it in (P) space, as we did the aggregate supply curve. The aggregate demand curve is shown in Figure 5.1 and labeled AD. Its properties are as follows:

- The AD curve has a negative slope. The reason is that a decrease in the domestic price level causes the real exchange rate to *depreciate* (i.e., causes Q to *increase*), which makes domestic goods relatively cheaper (because *more* of them have

[5] For an examination of the empirical evidence on the determinants of trade flows in developing countries, see Rose (1990).

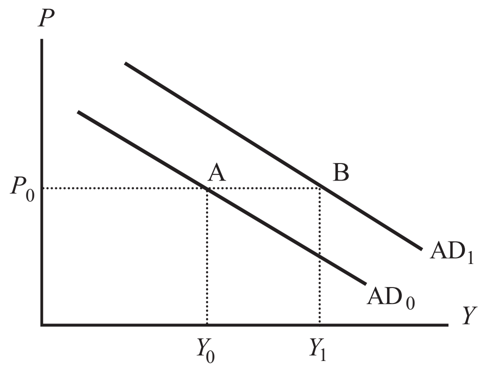


Figure 5.1. The aggregate demand curve

to be given up for each foreign good). This relative price change induces both domestic and foreign residents to switch away from buying foreign goods and toward buying domestic goods, thereby increasing the aggregate demand for domestic goods. This is referred to as the *expenditure-switching effect*. As long as the marginal propensity to spend is less than unity (so that an increase in domestic output induces a *smaller* increase in aggregate demand), an increase in the level of domestic output is required to meet this increase in demand, which gives the AD curve its negative slope. The effects of a change in the domestic price level on aggregate demand are depicted as movements *along* a given negatively sloped AD curve.

- The shape of the AD curve depends on the strength of the expenditure-switching effect. The stronger this expenditure-switching effect, the greater the effect of price changes on aggregate demand, and the flatter the AD curve.[6]
- The position of the AD curve depends on all the factors that influence the demand for domestic goods other than the domestic price level. Changes in any of these factors cause the curve to shift because they change the level of aggregate demand at any given value of the domestic price level. For example, an increase in government spending, a reduction in taxes, or an exogenous increase in exports would all increase the demand for domestic goods at any given value of the domestic price level, from a point such as A at the price level P_0 in Figure 5.1 to a point such as B. Consequently, each of these shocks would cause the AD curve to shift to the right, to a position such as AD_1 in Figure 5.1.
- An increase in the domestic-currency price of the foreign good SP^*, brought about by changes in S or in P^*, would have a similar effect because the resulting depreciation of the real exchange rate at any given initial value of the domestic price level would stimulate demand for the domestic good. In this case, however,

[6] We can also show that given the elasticities of export and import demand, the aggregate demand curve will be flatter the larger the volumes of exports and imports, i.e., the more open the domestic economy.

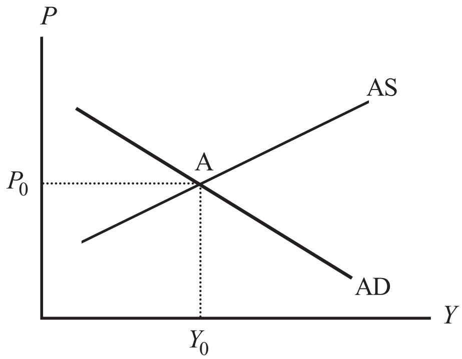


Figure 5.2. Equilibrium in the market for domestic goods

it is useful to consider the magnitude of the *vertical* shift in AD as well. To do so, suppose that at some initial level of domestic real output Y_0, the real exchange rate $SP^*/P = Q_0$ satisfies equation (5.6). Then, if there is a change in SP^*, say, to $(SP^*)'$, the new price level $P' = (SP^*)'/Q_0$ must continue to satisfy equation (5.6) at the level of output Y_0 because it would produce the same real exchange rate that prevailed before the change in SP^*. That means that the price level that satisfies equation (5.6) at $Y = Y_0$ must increase *in proportion* to the increase in SP^*; that is, the AD curve must shift *upward* in proportion to the increase in SP^*. We will make use of this result later on.

II. EQUILIBRIUM IN THE MARKET FOR DOMESTIC GOODS

We can now put the aggregate supply analysis from Chapter 4 together with the aggregate demand analysis of Section I. Setting supply equal to demand ($Y = AD$) and using the supply relationship (4.11) together with the demand relationship (5.6) yields the goods-market equilibrium condition:

$$\underset{+}{Y(P, \ldots)} = \varphi\,(\underset{+}{SP^*/P}, \ldots) A_P(\underset{+}{Y(P, \ldots)} - \underset{-}{T}, R, \ldots) + G + X(\underset{+}{SP^*/P}, \ldots) \tag{5.7}$$

where the determinants of aggregate supply other than the domestic price level have been subsumed into the ellipses to avoid cluttering up the notation. In this section, we will explore the properties of this equilibrium. Though this could be done directly from equation (5.7), we will instead proceed graphically using Figure 5.2. In this figure, equilibrium in the market for domestic goods is depicted at the point A, where the aggregate supply and aggregate demand curves intersect.

A good place to start in studying the properties of goods-market equilibrium is to identify the range of economic variables that may potentially affect this equilibrium and therefore change both the aggregate price level and the level of output in the economy. These variables are of various types:

- First, equilibrium may be affected by *nonpolicy domestic aggregate demand shocks*. This refers to the factors captured by the ellipses in the functions $\varphi(\ )$ and $A_P(\ )$. For example, domestic residents may change their preferences between domestic and foreign goods, or they may change their levels of spending in the absence of changes in incomes, taxes, or interest rates.
- Second, equilibrium may also be affected by various forms of *aggregate demand policies*. These include *fiscal policy* in the form of changes in G and/or T, *monetary policy* in the form of changes in the domestic interest rate R, and *exchange rate policy* in the form of changes in the officially determined exchange rate S.
- Third, the economy may be subjected to *external shocks*, consisting of shifts in the function $X(\ )$ (if foreign residents change their preferences for their own as opposed to the domestic economy's goods) and changes in the external price level P^*.
- Finally, it may be subjected to *supply shocks*. These may consist of changes in the level of total factor productivity A, in the size of the domestic capital stock K, or in the lagged nominal wage W_{-1}.

Our next task is to investigate how changes in all these variables affect the domestic economy. In addition to their effects on the price level and the level of output, we will also examine their effects on a variety of other variables, including the level of employment, the nominal wage, the real wage, and the level of net exports.

III. ADJUSTMENT TO SHOCKS

Because there are many possible shocks and various effects to consider, it is useful to organize this analysis in a systematic way. A key distinction is between shocks that affect the economy by shifting the aggregate demand curve (*demand shocks*, in the form of the first three types of shocks listed earlier) and those that do so by shifting the aggregate supply curve (*supply shocks*). We consider each of these in turn.[7] For simplicity, we assume that the economy is initially in a position of full employment, and for now, we consider only the effects of the shocks in the period in which they arrive.[8]

[7] There is a large literature on the role of shocks in creating volatility in emerging and developing countries. For Latin America, see Hausmann and Gavin (1995, 1996). For developing countries in general, see Easterly et al. (2000) as well as Mobarak (2005).

[8] Recall from Chapter 4 that if the economy is *not* at full employment, the nominal wage would be changing each period, which means that we would have to consider the effects of the nominal wage change together with those of the shock we are analyzing.

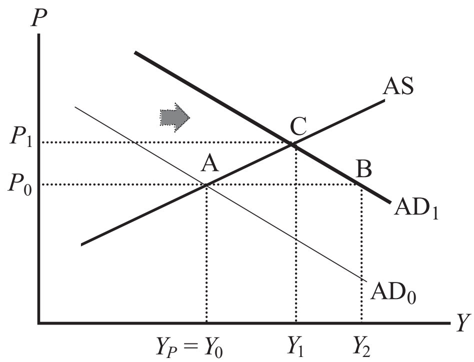


Figure 5.3. An expansionary aggregate demand shock

1. Aggregate Demand Shocks

Consider first how the economy reacts to a change in one of the exogenous variables that affects the aggregate demand for domestic goods. As we saw earlier, this could take the form of a shift in any of the functions $\varphi(\ )$, $A_P(\ )$, or $X(\ )$, a change in the government's budget (in government spending G or taxation T), a change in the domestic interest rate R set by the central bank, or a change in the domestic-currency price of foreign goods SP^* (as the result of either a change in the official exchange rate or a change in the foreign-currency price of such goods). To be concrete, let's examine the effects on the goods market of a demand shock that has the effect of increasing the demand for domestic goods at a given value of the domestic price level (a *positive* demand shock).

The implications of such a shock for the goods market are shown in Figure 5.3. Because the aggregate demand shock increases the demand for domestic goods at any given value of the domestic price level, it causes the AD curve to shift to the right, to a position such as AD_1. At the original value of the domestic price level, this would result in an excess demand for domestic goods (in the amount $Y_2 - Y_0$). This creates upward pressure on the domestic price level, which induces an increase in the level of real output.

To see how this increase in output comes about, we have to go back to the labor market. Recall that a higher domestic price level causes the labor demand curve to shift upward in proportion to the change in the price level, while the labor supply curve shifts upward less than in proportion to the price-level change, as long as the nominal wage is sticky. These shifts induce an increase in the equilibrium value of the nominal wage, albeit one that is less than proportional to the increase in the domestic price level. Because the increase in the nominal wage is proportionately less than that in the price level, the real wage falls, and firms are thereby induced to expand employment and produce more output. This explains why the economy moves to the northeast along the aggregate supply curve.

At the same time, the higher price of domestic goods causes both domestic and foreign residents to switch spending away from domestic goods and into foreign goods, so the aggregate demand for domestic goods contracts; that is, the economy moves to the northwest along AD_1. The supply expansion and demand contraction continue until the equilibrium between supply and demand is restored at the point C. This point is associated with the following macroeconomic outcomes.

First, as we have seen, it results in a higher level of domestic prices, output, and employment as well as a higher equilibrium nominal wage, but a lower real wage.

Second, the effects of shocks of this type on the real exchange rate depend on the source of the shock. Because domestic prices are higher, if the shock takes any form other than an increase in the domestic-currency price of foreign goods (SP^*), the real exchange rate must *appreciate* (i.e., Q must *fall*). However, if the shift in AD is the result of an increase in SP^*, the end result must actually be a *depreciation* of the real exchange rate. The reason is that, as we saw in the last section, when SP^* increases, AD shifts vertically by an amount that is proportional to that increase. However, as can be seen from Figure 5.3, the increase in the domestic price level from P_0 to P_1 is smaller than the vertical shift in AD. Because this means that P rises less than in proportion to the increase in SP^*, $Q = SP^*/P$ must increase.

Third, the effects of the shock on the economy's net exports also depend on the source of the shock. Recall that net exports are given by the following:

$$X(SP^*/P, \ldots) - (1 - \varphi\,(SP^*/P, \ldots)]A_P(Y - T, R, \ldots)$$

Thus expansionary shifts in demand that cause P and Y to rise while either decreasing T and R – or leaving them unchanged – would cause net exports to fall. Expansionary fiscal and monetary policies would have this effect, for example, as would exogenous nonpolicy increases in demand arising from shifts in the A_P function. However, demand expansions arising from net exports themselves (i.e., from increases in SP^* or from expansionary shifts in the $X(\ )$ or $\varphi(\ )$ functions) would tend to *increase* net exports. To see why, go back to equation (5.6). Adding $(1 - \varphi)A_P$ to φA_P, subtracting it from X, and rearranging terms, we can write the goods-market equilibrium condition as follows:

$$X - Q \cdot IM = Y - [A_P(Y - T, R, \ldots) + G] \tag{5.8}$$

This equation, which expresses net exports as the difference between domestic output and total domestic absorption, is sometimes referred to as the *absorption approach* to the explanation of how net exports are determined. Because the "marginal propensity to absorb" is less than unity, it implies that any shock (such as those just mentioned) that results in an increase in Y while leaving T, R, and G unchanged must increase net exports. Increases in SP^* and expansionary shifts in the $X(\ )$ or $\varphi(\ )$ functions would all have this effect.

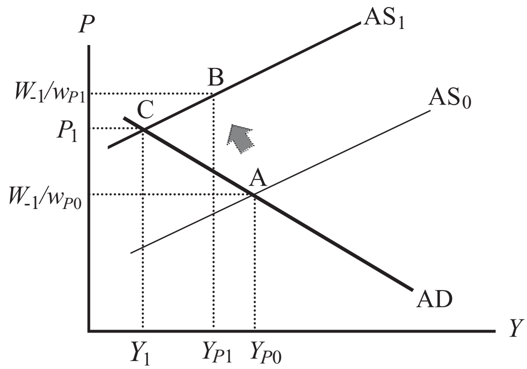


Figure 5.4. A contractionary aggregate supply shock

All these results would, of course, be exactly reversed in the event of a *contractionary* aggregate demand shock, that is, a change in an exogenous variable that causes the aggregate demand curve to shift to the *left*.

2. Aggregate Supply Shocks

Next, consider the effects on the economy of a change in one of the exogenous variables that affects aggregate supply: the level of total factor productivity A, the capital stock K, or the lagged nominal wage. Recall from the preceding chapter that the aggregate supply curve must pass through the point $(Y_p, W_{-1}/w_p)$, where $w_p = \omega(A, K), \ldots)$ and $Y_p = AF(K, L_p(A, K), \ldots))$, and that the effects of A and K on the full-employment real wage w_P as well as on the full-employment level of employment L_P are both positive. Thus we can consider two types of aggregate supply shocks: shocks that affect A or K, on one hand, and changes in W_{-1}, on the other.

Consider first shocks to A or K. For concreteness (and variety), let us consider a shock to one of these variables that reduces the full-employment level of output Y_P, that is, a *contractionary* aggregate supply shock. This could take the form of a decrease in total factor productivity A or a reduction in the capital stock K. In the labor market, either type of shock would shift the labor demand curve downward, reducing the full-employment real wage w_P as well as the full-employment level of employment. This means that such shocks cause the full-employment level of output to fall for *two* reasons: because the decrease in A or K reduces output at any given level of employment and because the full-employment level of employment itself falls. The decreases in the full-employment real wage w_P and full-employment level of output Y_P imply that contractionary shocks to A or K would *increase* W_{-1}/w_p while *reducing* Y_p. Consequently, the point $(Y_p, W_{-1}/w_p)$ that determines the position of the aggregate supply curve would move up and to the left, as shown in Figure 5.4. The aggregate supply curve would now have to pass through a point like

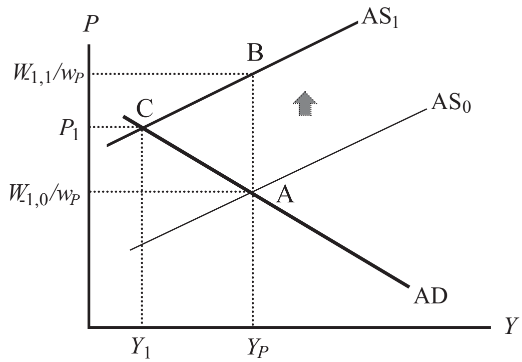


Figure 5.5. Effects of an increase in the lagged nominal wage

B rather than A. The curve would thus shift from AS_0 to AS_1, and the economy's new short-run equilibrium would move from A to C.[9] The upshot is that the equilibrium level of real output contracts, and the price level rises. Because SP^* is constant in this case, as a result of the increase in the domestic price level, the real exchange rate must appreciate.

What happens in the labor market? When the shock takes the form of a decrease in A or K, the labor demand curve shifts vertically *downward*. This is why the full-employment level of employment decreases. However, the increase in the domestic price level from the point A to the point C in Figure 5.4 causes the labor demand and supply curves to shift vertically *upward* (recall that the shift in the supply curve must be smaller than that in the demand curve). Because the final vertical shift in the labor demand curve may therefore be either downward or upward, the equilibrium level of employment may ultimately either increase or decrease relative to its starting point. The outcome depends on how large the increase in the equilibrium price level is, which in turn depends on the slope of the AD curve. Employment is more likely to increase the larger the increase in the equilibrium price level, that is, the steeper the AD curve. However, the increase in the price level cannot increase employment enough to restore the level of real output back to its original level because of the negative effect of a higher domestic price level on aggregate demand.

Now consider how the economy's short-run equilibrium would be affected by an increase in the lagged nominal wage W_{-1}. In this case, as we saw in Chapter 4, the full-employment level of output is unaffected, but because an increase in W_{-1} causes W_{-1}/w_p to increase in the same proportion, the aggregate supply curve shifts vertically upward in proportion to the change in W_{-1}. As shown in Figure 5.5, the position of the aggregate supply curve shifts so as to pass through a point like B

9 Though in Figure 5.4, the new equilibrium level of real output is depicted as being below the new lower level of full-employment output, it may be above or below Y_{P1}, depending on the slope of the aggregate demand curve.

rather than A. The economy moves from its initial position at A to a point like C, where the equilibrium level of output has fallen and the equilibrium value of the domestic price level has increased. Because the level of output has fallen with no change in TFP or the capital stock, the equilibrium level of employment must have decreased as well. For this to be true, the real wage must have increased. Note that in this case, unlike the previous one, output must fall short of its full-employment level at the new equilibrium.

In the case of either type of contractionary aggregate supply shock, net exports must fall. This outcome may seem ambiguous at first: the decrease in real output increases net exports, while the appreciation of the real exchange rate caused by the increase in the domestic price level in both cases reduces them. But from the absorption approach in equation (5.8), you can see that in the case of aggregate supply shocks, all changes in absorption arise from changes in income. Because these changes in absorption must be smaller than the changes in income that induce them, equation (5.8) implies that when aggregate supply shocks are contractionary, net exports must unambiguously fall.

IV. THE GM CURVE

The aggregate supply-aggregate demand diagram with which we have been working is a convenient and intuitive way to represent our first short-run model. However, it is not the only way of doing so, and when we move to treating the domestic interest rate as an endogenous variable in Chapter 7, it will be useful to be able to show directly on the graph that we will use to analyze the resulting model how the domestic interest rate is determined and how it responds to shocks. For that purpose, in this section, we will develop an alternative graphical representation of our model, in the form of a single curve drawn in price level-interest rate (P, R) space.

To construct such a curve, we essentially have to derive the set of all possible combinations of R and P that are consistent with equilibrium in the goods market for given configurations of the exogenous variables that affect demand and supply in that market. To identify such combinations, we need to determine what values of P would be required to maintain equilibrium in the goods market for different alternative values of R.

However, because a change in R is just an aggregate demand shock, we have essentially already done this. Recall that we analyzed the effects of aggregate demand shocks in the last section. That analysis indicated that to maintain equilibrium in the goods market, any expansionary aggregate demand shock – such as a reduction in the domestic interest rate – must be associated with an increase in the domestic price level. Geometrically, the reason is that a lower domestic interest rate causes the aggregate demand curve to shift to the right, moving the economy northeastward along the positively sloped aggregate supply curve. This results in a higher equilibrium domestic price level and a higher equilibrium level of real output.

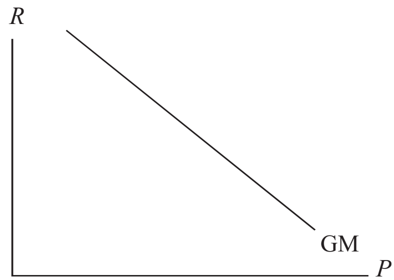


Figure 5.6. The GM curve

It is useful to review the economics involved in this outcome. A reduction in the domestic interest rate tends to increase private absorption A_P and thus gives rise to an excess demand for domestic goods at the initial value of the domestic price level. To maintain equilibrium in the domestic goods market, the adjustment in the price of domestic goods must be such as to increase the net supply of such goods to meet the excess demand caused by the lower domestic interest rate. This requires an *increase* in P, which restores equilibrium through two separate mechanisms. First, an increase in P increases the supply of domestic goods. Though this increase in output stimulates additional absorption of domestic goods, under the assumption that the marginal propensity to spend is less than unity, the supply of domestic goods increases more than the demand for such goods. Second, an increase in P reduces the demand for domestic goods by causing the real exchange rate to appreciate (i.e., by decreasing $Q = sP^*/P$), thereby making domestic goods relatively more expensive and causing both domestic and foreign residents to switch demand from domestic to foreign goods. These two effects together create the additional excess supply of domestic goods required to sustain equilibrium in the goods market in response to a decrease in the domestic interest rate.

This equilibrium relationship between the two endogenous variables R and P is illustrated in the form of the GM curve in Figure 5.6. The GM curve depicts the set of all combinations of R and P that are consistent with equilibrium in the domestic goods market.

Its properties are as follows:

- It has a negative slope, reflecting the adjustment mechanism just described.
- The slope of the GM curve depends on the sensitivity of private absorption to the domestic interest rate as well as on the effectiveness of changes in the domestic price level in restoring goods-market equilibrium when it is disturbed. The *stronger* the effect of the domestic interest rate on private absorption, and the *less* effective domestic price-level adjustments in equilibrating the goods market, the *flatter* the GM curve must be. The reason is that under these conditions, larger adjustments in the price level are required to restore goods-market equilibrium after a change in the domestic interest rate. In terms of the AD-AS diagram of Section III, the GM curve is relatively flat when changes in the domestic interest

rate induce large shifts in the AD curve, and when the aggregate demand and aggregate supply curves are themselves relatively *steep*, so large changes in the price level are required to restore equilibrium after an aggregate demand shock. In turn, the analysis of Chapter 4 suggested that the aggregate supply curve would be steep when the labor demand and supply curves are themselves steep, while Section II of this chapter indicated that the aggregate demand curve is steep when the elasticity of demand for domestic goods by both domestic and foreign residents with respect to changes in the real exchange rate is small.

- The position of the GM curve depends on the values taken by all the exogenous variables whose effects on the goods market we analyzed in the last section. As we saw there, given the domestic interest rate, shocks that create an excess demand for domestic goods (expansionary aggregate demand shocks or contractionary aggregate supply shocks) require an increase in the price of domestic goods to maintain equilibrium in the goods market. Because this means that the equilibrium value of P must be higher for a *given* value of R, such shocks must cause the GM curve to shift to the right. The magnitude of the rightward shift in the GM curve must be exactly equal to the increase in the equilibrium price level caused by such shocks in the AD-AS diagram. On the other hand, contractionary aggregate demand shocks or expansionary aggregate supply shocks call for a *decrease* in the equilibrium value of the domestic price level and consequently cause the GM curve to shift to the *left* by an amount equal to the decrease in the equilibrium price level in the AD-AS diagram.

V. SUMMARY

In this chapter, we have constructed a simple benchmark model of a small open economy by adding an aggregate demand relationship to the aggregate supply relationship of Chapter 4. For the demand side of the model, we adopted standard textbook descriptions of the determinants of the various components of aggregate demand for domestic goods.

Putting supply and demand together, we derived a simple model capable of explaining the behavior of a wide range of endogenous macroeconomic variables that policy makers in emerging and developing countries care about. These include the level of real output, the aggregate price level, the level of employment, the nominal wage, the real wage, the real exchange rate, and net exports. We examined how all of these variables are affected by a wide range of macroeconomic shocks. We studied the effects of nonpolicy domestic and foreign aggregate demand shocks as well those of fiscal, monetary, and exchange rate policies. We also examined the effects of several possible aggregate supply shocks.

In this chapter, we simply assumed that the domestic interest rate was determined as an exogenous policy variable by the central bank. But as we shall see, it is not always possible for the central bank to control the interest rate in this way under our provisional assumption that it also maintains an officially determined exchange rate. Moreover, even when this is possible, simply assuming that the central bank sets the interest rate at whatever value it chooses obviously provides no explanation as to how exactly it can manage to do that. Finally, the central bank has options for conducting monetary policy other than that of targeting the domestic interest rate. To investigate all these issues, we have to examine the requirements for equilibrium in the economy's financial markets. That is the subject of the next chapter.

REVIEW QUESTIONS

1. What is the difference between domestic absorption and the aggregate demand for domestic goods?
2. Explain why the aggregate demand for domestic goods depends inversely on the domestic price level.
3. Explain how monetary, fiscal, and exchange rate policies affect the equilibrium level of domestic real output.
4. What is a supply shock? Show how a favorable shock to total factor productivity would affect domestic real output and the domestic price level.
5. Explain how the GM curve can be derived from the AS and AD curves.

EXERCISES

1. Consider a small open economy that maintains a fixed exchange rate and in which the central bank targets the domestic interest rate. Explain what effect a decrease in the interest rate targeted by the central bank would have on the following domestic macroeconomic variables:
   a. real gross domestic product (GDP)
   b. the price level
   c. the real exchange rate
   d. the real wage
   e. the nominal wage
2. What effect would you expect a devaluation of the currency to have on aggregate demand for domestic goods in the model developed in this chapter? Explain the channels through which this effect would operate.
3. On the basis of the model developed in this chapter, what effect would you expect an increase in government spending to have on the level of private absorption in the economy, and why?
4. It is sometimes claimed that a nominal devaluation cannot affect net exports because its only effect will be to raise the domestic price level. Was this true in the model that we analyzed in this chapter? Explain what the answer depends on, and why.

5. Assuming that they have the same effects of the equilibrium value of real GDP, compare how an economy's net exports would be affected by each of the following shocks:
   a. an increase in government spending
   b. a reduction in the domestic interest rate
6. Weather-related difficulties in production in developing countries can be captured in our macro model in the form of negative changes in the total factor productivity parameter A. Assuming that the central bank responds to an unanticipated reduction in A by adopting whatever policies are required to keep the economy at full employment, explain the following:
   a. what the central bank needs to do to the domestic interest rate
   b. what would happen to the equilibrium levels of output and the price level
   c. what would happen to the equilibrium levels of employment and the real wage
7. In the first short-run macro model that we constructed in this chapter, we assumed that the domestic interest rate was set exogenously by the central bank. What tends to happen more commonly in the real world, however, is that the central bank sets the domestic interest rate according to a *policy rule*, as you will see in Chapter 15. Such rules have the effect of making the domestic interest rate an endogenous variable because they make the interest rate respond to domestic macroeconomic variables that the central bank seeks to influence. Consider, for example, the interest rate rule:

$$R = R(\underset{+}{P})$$

   This is a policy rule in which the central bank raises the domestic interest rate whenever the domestic price level increases and lowers it when the domestic price level falls. It is sometimes referred to as *price-level targeting*. Assuming that the central bank behaves in this way
   a. How would the shape of the aggregate demand curve differ from the one derived in this chapter?
   b. Would you expect shocks to aggregate *demand* (e.g., fiscal shocks) to have weaker or stronger effects on the equilibrium level of real *GDP* under this rule than when the interest rate is exogenous? Explain the economic reasoning that supports your answer.
   c. Would you expect shocks to aggregate *supply* (e.g., productivity shocks) to have weaker or stronger effects on the equilibrium level of real *GDP* under this rule than when the interest rate is exogenous? Explain the economic reasoning that supports your answer.
   d. On the basis of your answers to questions b and c, assuming that the central bank wants to stabilize real *GDP* around its full-employment level, is price-level targeting a good idea? Why or why not?

8. Suppose that the central bank sets an official value of the nominal exchange rate, just as we have assumed in this chapter, but manages it in such a way as to keep the *real* exchange rate fixed. In other words, the central bank sets the nominal exchange rate S so that

$$S = P/P^*$$

That is, it alters the nominal exchange rate whenever the domestic price level changes relative to the foreign price level so that the real exchange rate never changes. This is called *real exchange rate targeting*, and many developing countries have used this type of exchange rate rule in the past.

Assuming that the central bank engages in real exchange rate targeting

a. Derive the shape of the economy's aggregate demand curve.
b. Suppose that the government implements an increase in government spending. How do the effects of this spending increase on real output and the domestic price level differ under real exchange rate targeting from what they would have been with a "fixed but adjustable" exchange rate? Explain in intuitive terms why these differences arise. (Hint: what happens to the "crowding out" of net exports under real exchange rate targeting?)
c. What effect would an increase in the foreign price level have on the domestic economy (real *GDP* and the price level) under real exchange rate targeting?

9. Consider two economies: one in which the labor supply curve is very steep (economy A) and one in which it is very flat (economy B). Assume that both economies have identical aggregate demand curves. Suppose that the two economies are hit by identical exogenous increases in the demand for their exports. Compare the effects of this shock on each of the following variables in the two economies:

a. real *GDP* and the price level
b. the level of employment and the real wage

Do net exports increase or decrease in these economies as the result of this shock? Explain why.

10. Derive an economy's aggregate supply curve on the assumption that the price level that workers expect to prevail each period is equal to the price level that actually prevailed the period before. Using this aggregate supply curve, explain how the economy would adjust over time to an unanticipated expansion of aggregate demand (say, through an expansionary fiscal policy) that is sustained permanently.

11. Using the model that we derived in this chapter, explain how a contractionary monetary policy (an increase in the domestic interest rate) would affect each of the following macroeconomic variables in the short run:

a. the domestic price level and level of real *GDP*
b. the level of employment and the real wage

c. the real exchange rate, the level of exports, and the level of imports
d. the capital stock

12. "A devaluation of the exchange rate reduces the equilibrium value of real *GDP*, because it reduces real wages and thus causes workers to spend less. This reduction in demand for domestic goods causes real output to fall."

    Using the model that we derived in this chapter, comment on this quote. Specifically, do you agree or disagree? Explain why.

13. Many countries (industrial and developing) maintain a variety of restrictions on imports. One way to incorporate import restrictions into our model is to modify the $\varphi(\ )$ function to take the form

$$\varphi = \underset{+\ \ +}{\varphi(Q, \tau)}$$

    where τ is an index of the strength of import restrictions. Thus an increase in τ increases the share of absorption devoted to domestic rather than foreign goods, which explains the positive sign under τ in the preceding function.

    To explore the macroeconomic effects of an increase in the strength of import restrictions

    a. Derive the effects of an increase in τ on
        i. real *GDP* and the price level
        ii. the level of employment and the actual real wage
        iii. the level of exports, imports, and net exports
    b. Explain the answer you gave for the effects of the stronger restrictions on the level of imports.

14. Suppose that an economy can be described by the model that we developed in this chapter, except that the government is forced to follow a *balanced budget policy rule*. That means that any increase in government spending has to be financed with an exactly equal increase in taxes. What is the effect of an increase in government spending on the AD curve in this case? Compared to the standard case, what happens to the equilibrium level of output and prices when government spending increases? What happens to net exports?

15. As we saw in Chapter 1, in many emerging and developing economies, fiscal policy has tended to be *pro-cyclical*; that is, it becomes expansionary during good times and contractionary during bad times. This phenomenon can be captured in our model by assuming that government spending depends positively on the level of output, as given by

$$G = \underset{+}{G(Y, \ldots)}$$

    a. Does this change in the model affect the shape of the AD curve? Why and how?

b. Compared to the model with exogenous government spending, what happens to the equilibrium level of output and prices if there is an exogenous increase in the demand for the country's exports?
c. Suppose there is an exogenous positive shock to the $G(\ )$ function. What is the magnitude of the crowding out effect on net exports in this case compared to the standard one?
d. Compared to the standard case, what happens to output and the price level if the central bank decides to lower the domestic interest rate?

REFERENCES AND FURTHER READING

Easterly, William, Michael Kremer, and Lawrence H. Summers (1993), "Good Policy or Good Luck? Country Growth Performance and Temporary Shocks," *Journal of Monetary Economics*, vol. 32, pp. 459–483.

Easterly, Williams, Roumeen Islam, and Joseph Stiglitz (2000), "Explaining Growth Volatility," unpublished manuscript, World Bank.

Hausmann, Ricardo, and Michael Gavin (1995), "Overcoming Volatility in Latin America," in *Report on Economic and Social Progress in Latin America*, Washington, DC: Inter-American Development Bank, pp. 1–65.

Hausmann, Ricardo, and Michael Gavin (1996), "Securing Stability and Growth in a Shock-Prone Region: The Policy Challenge for Latin America," working paper 315, Inter-American Development Bank, Office of the Chief Economist.

Mobarak, Ahmed Mushfiq (2005), "Determinants of Volatility and Implications for Economic Development," *Review of Economics and Statistics*, vol. 87, pp. 347–361.

Rose, Andrew (1990), "Exchange Rates and the Trade Balance: Some Evidence from Developing Countries," *Economics Letters*, vol. 34, pp. 271–275.

6

Financial Markets

In the simple macroeconomic model of the preceding chapter, the domestic interest rate was an exogenous policy variable determined by the central bank. It affected aggregate demand through private absorption: a lower domestic interest rate was assumed to increase absorption and a higher one to reduce it. But the model did not contain an explanation of how the central bank set the domestic interest rate. Our task in this chapter is therefore to explain how the domestic interest rate is determined. As we will see, the central bank may not always be able to determine the domestic interest rate; it may not always choose to do so; and even when it can, it does not control the domestic interest rate directly but only indirectly through other policy instruments. This means that setting the domestic interest rate at a policy-determined level may be a tricky proposition, requiring the central bank to engage in a delicate balancing act in which it uses its policy levers to compensate for a variety of exogenous factors that would affect the interest rate in the absence of central-bank action. This chapter will explore how the equilibrium value of the domestic interest rate is affected both by the central bank's policy levers and by such exogenous factors.

In the real world, of course, there are a large number of different interest rates in any country, associated with the different types of assets that may exist at any one time in that country's economy. These assets may be *real* (physical) assets or *financial* assets. The former include land, structures, equipment, and other forms of tangible assets. The latter include both securities that are traded in organized markets and claims on financial institutions that act as intermediaries between borrowers and lenders. As we will discuss in more detail in Chapter 20, financial intermediaries perform key functions in all economies, and commercial banks are particularly important in that role. This is especially so in developing countries, where *secondary markets* for securities (markets where previously existing securities are traded) are often very limited in scope. We will discuss the role of banks and how they fit into a country's broad financial structure in Part 6 and will analyze financial-market equilibrium in the presence of banks in Chapter 23.

For now, however, we will abstract away from the existence of financial intermediaries such as banks. In this chapter, we will assume instead that borrowers and lenders interact with each other directly, rather than through financial intermediaries, by buying and selling interest-bearing securities in exchange for cash in well-organized markets – borrowers sell such securities, and lenders buy them. We begin this way for the sake of simplicity, as well as to make it easier to relate the model we are developing in this part of the book to those you may have seen before in intermediate-level macroeconomics textbooks for industrial countries. Once this framework is understood, it will be relatively straightforward to add commercial banks later on.[1]

We will assume that two types of interest-bearing securities are traded in the domestic economy: bonds issued by the domestic government and bonds issued by the foreign government. Because domestic residents can hold foreign securities, and foreign residents can hold domestic ones, we are effectively treating the economy as *financially open*. The domestic interest rate R is the interest rate paid on bonds issued by the domestic government. These bonds are denominated in units of domestic currency. We will assume that they have a one-period maturity, so the interest rate associated with these bonds can change every period. The total value of domestic bonds in existence at any one time, measured in domestic-currency terms, is denoted B. As we will see, the equilibrium value of the domestic interest rate will turn out to be that which is required to clear the domestic bond market, that is, to ensure that the total stock of domestic bonds that is in existence at any moment of time is willingly held. Foreign bonds are similarly one-period bonds, but they are denominated in foreign currency. They pay the interest rate R^*, and the total value of such bonds outstanding at any one time, measured in units of the *foreign* currency, is F^*. The domestic economy is taken to be small in the world market for foreign bonds, so it can have no influence over the interest rate paid on those bonds. We can therefore treat R^* as an exogenous variable.

In addition to the two types of bonds, there are two types of money in the model: domestic money and foreign money. Because there are no commercial banks in the economy (and therefore no *demand deposits*, or checking accounts, which are usually part of the definition of money), money in this economy consists only of currency issued by the central bank. The quantity of domestic money in the hands of the public at any moment is denoted M, and that of foreign money is M^*. We will assume that domestic money is held only by domestic residents and foreign money only by foreign residents.

[1] As you will see, the properties of financial-market equilibrium that we will derive in this chapter carry through when we incorporate banks into our model.

In this chapter, we will focus on a key characteristic of the financial side of the economy that plays a large role in determining its short-run macroeconomic behavior: the strength of the arbitrage links that exist between domestic and foreign financial markets. As we will see throughout the rest of the book, this is a vitally important issue that influences many of the key policy challenges that have confronted developing countries – especially emerging-market economies – in recent years. We will measure the strength of these links by the extent to which the expected rates of return on domestic and foreign bonds are *arbitraged* against each other, that is, the extent to which differences in rates of return induce individuals to sell the asset offering the relatively low expected return and buy the one offering the higher expected return. When the domestic and foreign bond markets are completely *segmented* (separated) – a situation to which we will refer as *financial autarky* – changes in the rate of return on either bond will fail to trigger any arbitrage flows between the two markets. When arbitrage between domestic and foreign bonds is imperfect, changes in the expected rate of return on either type of bond will trigger arbitrage flows between the two types of bonds, but those flows will fall short of what would be required to equalize the expected rates of return on the two types of bonds. We will refer to this as a situation of *imperfect capital mobility* and describe it as one in which domestic and foreign financial assets are effectively *imperfect substitutes* for each other. Finally, when arbitrage is sufficiently effective to equalize the expected rates of return on the two types of bonds continuously, we will refer to this as a situation of *perfect capital mobility*, or one in which domestic and foreign bonds are *perfect substitutes.*

We will build the financial side of our model based on the assumption that capital mobility is imperfect.[2] This approach represents a departure from the standard industrial-country textbook models of asset markets for open economies, which typically assume perfect capital mobility. Aside from its realism in describing the actual situation in most emerging-market economies (see Chapter 1), this assumption has two advantages for our purposes. First, the degree of substitutability between domestic and foreign financial assets can be treated as a measure of the strength of the links between domestic and foreign financial markets – that is, as a measure of the degree of capital mobility – and this permits us to explore how the strength of these links affects the behavior of the economy. Second, as we will see in the next chapter, when capital mobility is imperfect, the central bank has a much richer menu of monetary policy options available to it under fixed exchange rates than it has under perfect capital mobility. In the chapters that follow, we will explore how the choices made by the central bank among these options affect the behavior of the economy.

[2] For an analysis of portfolio balance models with imperfect capital mobility, see Branson and Henderson (1984).

This chapter is divided into six sections. Section I lays out the menu of assets from which domestic and foreign agents can choose and describes how they determine how much of each financial asset they wish to hold. Section II examines the determination of asset supplies. In Section III, we examine how domestic financial markets attain equilibrium, whereas Section IV explores how that equilibrium responds to a variety of shocks. This analysis is summarized in the form of the BB curve, the financial-market counterpart to the GM curve derived in the preceding chapter, in Section V. Section VI summarizes.

I. BALANCE SHEETS AND ASSET DEMANDS

It is useful to start with some definitions. The menu of financial assets that is held by each agent in the model is referred to as that agent's *financial portfolio.* The total financial wealth of each type of economic agent in our model is the sum of the values of all of the assets that agents of that type hold, net of their financial liabilities. The identities that express the financial wealth of each type of agent in the model as the sum of the assets held by all the agents of that type, net of their liabilities, are referred to as the *balance sheet constraints* facing those agents. The decisions that each agent makes about how to allocate his or her financial portfolio among the various assets and liabilities that are available to that agent are called *portfolio allocation decisions.* The outcomes of these decisions are *asset demand functions*, which express the demands for assets and liabilities of each type as functions of the variables that influence portfolio allocation decisions.

To describe the financial structure of our model and explore how the interest rate on domestic bonds (which we will continue to refer to as the domestic interest rate) is determined, we need to say something about who holds each of the assets and liabilities described previously. We will consider four different types of agents: the domestic private sector, the domestic central bank, the domestic government, and the rest of the world.

1. The Domestic Private Sector's Balance Sheet and Portfolio Allocation Decisions

The domestic private sector holds domestic money, domestic bonds, and foreign bonds. The sum of the values of all the assets held by the private sector in its financial portfolio represents its total financial wealth (W_P):

$$W_P = M + B_P + SF_P^* \tag{6.1}$$

where B_P and F_P^* are, respectively, the domestic-currency value of the domestic bonds and the foreign-currency value of the foreign bonds held by the domestic private sector, so SF_P^* is the domestic-currency value of the foreign bonds held by the domestic private sector. The change in W_P over time is given by

$$\begin{aligned}\Delta W_P &= (\Delta M + \Delta B_P + S\Delta F_P^*) + \Delta SF_P^* \\ &= [P(Y - T) + RB_P + R^* SF_P^* - PA_P] + \Delta S\ FP^* \end{aligned} \tag{6.2}$$

The right-hand side of the first line of this equation consists of two parts: the purchase of new assets $(\Delta M + \Delta B_P + S\Delta F_P^*)$ and capital gains and losses on existing assets, ΔSF_P^*, arising from exchange-rate changes. The second line indicates that the domestic private sector can finance the purchase of new assets through financial saving, that is, by maintaining a level of absorption PA_P that is smaller than its total income, consisting of factor income net of taxes $P(Y - T)$ plus interest income on the financial assets that the private sector holds, $RB_P + R^* SF_P^*$. Notice that in the absence of exchange rate changes, changes in the private sector's wealth during the current period should be very small relative to its wealth at the beginning of the period because the former depends on one period's saving, whereas the latter depends on the private sector's entire history of past saving. This means that at any moment of time, W_P can essentially be taken as given. We will therefore refer to it as a *predetermined* variable. The only exception to this statement occurs in the event of exchange rate changes, which give rise to capital gains or losses on the foreign bonds held by domestic households and could create large changes in private wealth, as we will see later.

Though the private sector cannot change its total wealth in the short run, it can change the form in which it holds its wealth – that is, it can alter the *composition* of its financial portfolio – instantaneously by trading in financial markets. As mentioned earlier, the desired amounts of the various assets that private agents want to hold, given their wealth, are referred to as the *asset demand functions* of those agents. We need to specify such functions for each of the three assets in the domestic private sector's portfolio to describe how the private sector allocates its financial portfolio among competing assets.

To do so, we will assume that the private sector essentially makes two decisions. First, it decides how much of its wealth to hold in the form of money. In determining its *real* demand for money M/P, it is influenced in the usual manner by transaction needs (captured by real output Y) and opportunity costs (given by R).[3] Because money is held strictly for transactions purposes, the demand for money does not increase if the private sector has more wealth to allocate to asset holdings

[3] Notice that we are implicitly assuming that the demand for money does *not* depend on the foreign interest rate R^*. This is conventional but is not strictly correct because both domestic and foreign bonds are alternatives to holding money. This assumption only makes a difference when R^* changes, as we will see later.

(i.e., it does not depend on W_P). Thus the *nominal* demand for money is given by the following:

$$M^D = PL(\underset{-}{R}, \underset{+}{Y}) \tag{6.3}$$

Once they have decided how much of their wealth to hold as money, private agents split the rest of it between domestic and foreign bonds, allocating a fraction b of their nonmonetary financial wealth $W_P - M$ to the former and a fraction f to the latter (so it must be the case that $b + f = 1$).[4] These fractions depend on the relative rates of return on the two bonds, expressed in a common currency. It is worth pausing to consider what this means.

The domestic-currency return on one unit of domestic currency invested in a domestic bond for one period (principal plus interest) is, of course, just given by $(1 + R)$. The domestic-currency return on the foreign bond is somewhat more complicated, however. One unit of the domestic currency can be used to purchase $1/S$ units of the foreign currency, each of which yields a return of $(1 + R^*)$ units of foreign currency when invested in foreign bonds. Thus the *foreign-currency* return on one unit of the domestic currency invested in foreign bonds is $(1 + R^*)/S$. This can be converted back into domestic currency by selling it at the expected future exchange rate $S^e{}_{+1}$, making the expected domestic-currency return on the foreign bond equal to $(1 + R^*)S^e{}_{+1}/S$. Thus the relevant comparison is between the rate of return R on domestic bonds and $(1 + R^*)S^e{}_{+1}/S - 1$ on the foreign bond. If the expected rate of exchange rate depreciation, given by $\hat{S} = S^e{}_{+1}/S - 1$, is small, the domestic-currency rate of return on the foreign bond is approximately equal to $R^* + \hat{S}$, so the relevant comparison becomes one between R and $R^* + \hat{S}$. However, if the exchange rate is fixed *and expected to remain so*, then $\hat{S} = 0$, and the expected domestic-currency rate of return on the foreign bond is just equal to the foreign-currency rate of return R^*.

Assuming for simplicity that this is the case, we can write the domestic private sector's bond demand functions as

$$B_P^D = b\,(R \underset{+}{-} R^*)(W_P - M);\ 0 < b(\ ) < 1 \tag{6.4}$$

$$SF_P^{*D} = f\,(R \underset{-}{-} R^*)(W_P - M);\ 0 < f(\ ) < 1 \tag{6.5a}$$

where b and f are functions denoting the shares of domestic and foreign bonds in the domestic private sector's nonmonetary wealth $(W_P - M)$, respectively, as mentioned earlier. The basic assumption is that holding more bonds of either type exposes people to higher levels of the particular risks associated with such bonds, so it takes a higher interest rate to induce them to do so.

[4] Notice that $f(\ )$ thus determines the *domestic-currency* value of households' demand for foreign bonds.

As discussed earlier, we will also examine the implications of alternative extreme assumptions in our model: that domestic and foreign residents do not hold each other's bonds at all (financial autarky) and that domestic and foreign bonds are held by all agents and are considered to be identical to each other (perfect substitutability).

Financial autarky is a situation in which domestic residents hold only domestic bonds and foreign residents hold only foreign bonds. Such a situation could arise, for example, if the domestic authorities place very strong restrictions on capital movements into or out of the domestic economy. To illustrate, suppose that the presence of such restrictions forces domestic residents to hold fewer foreign assets than they would have done if their ability to do so were unfettered by official impediments. Then their effective demand for foreign assets could be written as:

$$SF_P^{*D} = \lambda_1 f(R - R^*)(W_P - M) \tag{6.5b}$$

where λ_1 is an index of the severity of capital-outflow restrictions, with $\lambda_1 = 0$ if the holding of foreign assets by domestic residents is strictly prohibited and $\lambda_1 = 1$ if there are no restrictions in place. Notice that because the sum of the domestic and foreign bonds in the portfolio of the domestic private sector must add up to its nonmonetary wealth, in the presence of restrictions on capital outflows, we must have $b = (1 - \lambda_1 f)$. In the extreme case of financial autarky ($\lambda_1 = 0$), this implies that $b = 1$ because the domestic private sector must devote all its nonmonetary wealth to the accumulation of domestic bonds.

Perfect substitutability would arise if either of two conditions prevails: domestic and foreign bonds have the same risk characteristics or households do not care about risk but only about expected returns. In either case, domestic and foreign bonds would be identical with respect to all the characteristics that matter to portfolio managers. Thus, under perfect substitutability, it becomes meaningless to distinguish between B_P and SF_P because they refer to what is essentially the same asset. In this case, the total demand for bonds by the domestic private sector is given by

$$B_P^D + SF_P^{*D} = (W_P - M)$$

2. The Balance Sheet of the Government

Just as the domestic private sector has a financial balance sheet that determines its financial wealth, so do the other agents in the model. Recall from the previous chapter that the government spends an amount G and collects taxes T (both measured in real terms). The excess of government spending over the revenues it collects is called the *fiscal deficit*, and the government finances it by issuing new bonds, in the amount ΔB.[5] Setting the fiscal deficit equal to the amount of new bonds issued by the government yields the *government budget constraint*:

$$\Delta B = P(G - T - T_C) + RB \tag{6.6}$$

where interest payments on existing government bonds RB are included in the government's budget, and the government's total revenues include any transfers that it receives from the central bank, denoted T_C. We will see where these come from later in this section.

Analogous to the private sector's balance sheet constraint, the government's (nonfinancial public sector) net financial wealth W_G is given by

$$W_G = -B \tag{6.7}$$

Assuming that the stock of domestic government bonds outstanding at any given moment B is positive, the government's net wealth is *negative* and is equal to the outstanding stock of government debt. Because B is the government's net cumulative borrowing over all its past history, B will be positive if the government has been a cumulative net *borrower*, rather than lender, in the past. Because the inherited stock of domestic bonds outstanding is a legacy of the government's entire history of *past* budgetary policies, W_G is a predetermined variable, just like W_P.

Notice that because the government only issues one type of financial liability and holds no financial assets in our model, unlike the private sector, it has no portfolio allocation decisions to make and therefore no asset demand functions.

3. The Central Bank's Balance Sheet and Portfolio Allocation Decisions

The central bank holds assets in the form of foreign exchange reserves (i.e., foreign bonds) with foreign-currency value F_C^* and government bonds in the amount B_C. It effectively extends *credit* – that is, lends money – to the government by buying its bonds, on which the government pays interest to the central bank. Because the government is the only domestic agent whose liabilities the central bank purchases directly, we will refer to B_C as the bank's stock of *domestic credit* outstanding. The central bank's liabilities consist of the domestic money (currency) M that it prints to buy domestic and foreign bonds. Its balance sheet is therefore given by

$$W_C = SF_C^* + B_C - M \tag{6.8}$$

where W_C is the central bank's net financial wealth, a predetermined variable that is the cumulative value of the central bank's past interest earnings on the assets it

[5] For now, we will assume that the government only issues bonds denominated in domestic currency. We will allow for government issuance of bonds denominated in foreign currency in Chapter 9.

holds, net of any transfers that it has made to the government plus any capital gains or losses on the central bank's foreign exchange reserves arising from changes in the exchange rate:

$$\Delta W_C = R^* S F_C^* + R B_C - T_C + \Delta S F_C^* \tag{6.9}$$

Like the private sector, the central bank has to make decisions concerning the composition of its financial portfolio. But these are taken to be *policy*, not portfolio allocation, decisions. The central bank has two policy roles in our model: it conducts *exchange rate policy* and *monetary policy*.

a. Exchange Rate Policy

To conduct exchange rate policy, the central bank intervenes in the foreign exchange market by using domestic currency (that it prints for the purpose) to buy or sell foreign bonds.[6] These transactions show up in its balance sheet in the form of changes in SF_C^* that are offset by changes in M, leaving the bank's financial net wealth unaffected. As indicated in previous chapters, the general rules that govern the central bank's behavior in the foreign exchange market are referred to as the country's foreign exchange regime. Recall that we are assuming for now that the central bank operates an officially determined exchange rate regime, in which it determines the value of the exchange rate that prevails in the foreign exchange market – that is, the exchange rate it "fixes" in the market – by announcing its willingness to buy or sell unlimited amounts of foreign exchange at a given price in terms of domestic currency (this price, of course, represents the *official exchange rate*). Its behaving in this way will make the value of foreign exchange reserves F_C^* an endogenous variable in our model, with changes in F_C^* determined by the overall balance of payments. The overall balance of payments determines the excess demand for or excess supply of foreign exchange in the foreign exchange market at the official exchange rate and therefore the excess supply of or excess demand for foreign exchange from the central bank.

In Chapter 17, we will consider two alternative foreign exchange rate regimes. Under a *clean float*, the central bank never intervenes in the foreign exchange market. In that case, F_C^* never changes, and F_C^* can be taken as fixed. Indeed, the central bank may not hold foreign exchange reserves at all, in which case, $F_C^* = 0$. Under these circumstances, the exchange rate is determined in the foreign exchange market without central-bank involvement on either the supply or demand side. Under a *dirty (or managed) float*, on the other hand, the central bank allows the exchange rate to float but may intervene in the market at its discretion, without necessarily being bound to follow any specific, preset rule. In this case, F_C^* becomes an exogenous policy variable.

[6] Alternatively, you can think of the central bank as buying foreign currency and then immediately converting it into foreign bonds to hold, and selling foreign bonds by first converting them into foreign currency and then selling that currency in exchange for domestic currency.

b. Monetary Policy

The bank also intervenes in the domestic bond market, when it uses domestic currency to buy or sell domestic government bonds. If an active secondary market exists for such bonds, the central bank merely enters that market as another buyer or seller, a process known as an *open market operation*. If no such market exists, the central bank can achieve the same result – as is done in many developing countries – by holding periodic auctions of government bonds. In this case, the central bank conducts monetary policy by varying the quantity of government bonds that it sells or buys in such auctions. Through either of these institutional mechanisms, the buying or selling of government bonds is how the central bank actually implements monetary policy in our model. Monetary policy actions therefore show up in the central bank's balance sheet in the form of changes in B_C that are exactly offset by changes in M, again leaving its financial net wealth unchanged.

Just as the central bank can adopt different exchange rate regimes, it can also adopt alternative *monetary policy regimes*. These are identified by whether the central bank uses its monetary policy instrument B_C to determine a specific value of the domestic interest rate R, the stock of domestic credit B_C itself, or the domestic money supply M. These three regimes can be identified as an *interest rate rule*, a *domestic credit rule*, and a *monetary rule*, respectively. Under an interest rate rule, the domestic interest rate becomes the monetary policy variable, as in the preceding chapter, and both B_C and M are simply additional endogenous variables in our model. Under a domestic credit rule, B_C is the monetary policy variable, whereas R and M take on the roles of additional endogenous variables. Finally, under a monetary rule, M is the monetary policy variable, whereas both R and B_C are endogenous.

Under any of these monetary policy rules, the monetary policy variable itself can be either endogenous or exogenous. It is endogenous if the monetary policy variable (whether the interest rate, the stock of domestic credit, or the money supply) is set according to a *feedback rule*, in which the central bank adjusts the monetary policy variable in a systematic way as a function of specific macroeconomic outcomes that it wants to influence. In that case, the macroeconomic outcomes that determine the behavior of the monetary policy variables are referred to as the *targets* of monetary policy. Alternatively, the monetary policy variable is exogenous if it is chosen by the central bank independently of macroeconomic outcomes. In that case, we can think of the monetary policy variable itself as the target of monetary policy and consider alternatively *interest rate targeting*, *domestic credit targeting*, or *money supply targeting*. In Chapter 5, for example, the central bank was assumed to pursue a regime of interest rate targeting because the domestic interest rate was treated as an exogenous policy instrument.

c. Sterilized Intervention

The exchange rate and monetary policy regimes interact in important ways. When the central bank intervenes in the foreign exchange market, whether it buys or sells foreign bonds, its actions have monetary effects. For example, suppose that the central bank maintains a fixed exchange rate and encounters a situation in which an excess demand for foreign exchange would arise in the foreign exchange market in the absence of central-bank intervention, which would tend to drive down the value of the domestic currency (i.e., cause it to depreciate). In this case, to defend the fixed exchange rate, the central bank would be forced to sell foreign exchange and buy the domestic currency. If it does nothing else, this would show up in its balance sheet as a reduction in SF_C^* matched by an equal reduction in M. If the central bank is pursuing either an interest rate or money supply rule, this contraction in the money supply may be unwelcome, however. If so, the central bank can prevent it by using the domestic currency it purchased in the foreign exchange market to buy domestic bonds, thereby leaving M unchanged. In this case, the reduction in SF_C^* would be offset in the central bank's balance sheet by an increase in B_C rather than by a reduction in M. A change in F_C^* *that is exactly offset by a change in* B_C, so as to leave the money supply M unchanged, is referred to as *sterilized intervention* in the foreign exchange market. Changes in F_C^* that are allowed to be reflected in changes in M, because B_C is unchanged, are called *unsterilized intervention.* Notice that use of a money supply rule requires that intervention in the foreign exchange market be sterilized, whereas use of a domestic credit rule precludes it. Sterilization may or may not be required under an interest rate rule, as we will see later on.

d. Monetization

Exchange rate and monetary policies may also interact in other ways. Notice, for example, that because the central bank's foreign exchange reserves are held in the form of foreign bonds, which are denominated in foreign currency, a change in the exchange rate causes the central bank to incur a capital gain or loss on its foreign exchange reserves equal to $\Delta S F_C^*$ (recall that ΔS is the change in the exchange rate). As we will see later, the impacts of such an exchange rate change on the economy depend in part on what the central bank does with these capital gains or losses. It has two options. First, it can simply absorb such capital gains or losses into its net worth so that

$$\Delta W_C = \Delta S F_C^*$$

If it does so, the capital gains or losses have no effect on M. Alternatively, it can transfer these gains or losses to the government. If the government in turn uses these resources to buy back an equivalent amount of its debt from the central bank, we have the following:

$$\Delta B_C = -\Delta S F_C^*$$

In this case, $\Delta W_C = \Delta M = 0$; that is, just as when the central bank books these gains or losses, there is no change in the domestic money supply. However, if the government uses these resources to buy back bonds from the public, then we have:

$$-\Delta B = \Delta M = \Delta S F_C^*$$

That is, the capital gains or losses are reflected in an equal change in the domestic money supply and an offsetting change in the stock of government bonds outstanding. In this case, we say that the capital gains or losses are *monetized* (transformed into changes in the money supply). By contrast, in the two previous cases, they are *nonmonetized*. We will return to this issue when we investigate the effects of devaluation on the economy's financial markets.[7]

4. Aggregate Domestic Financial Wealth

Now that we have described the balance sheets of all three types of domestic agents, we can examine the balance sheet of the domestic economy as a whole. To derive an expression for the aggregate net financial wealth of the domestic economy, we can add together the financial wealth of all three types of domestic agents: the private sector, the government, and the central bank. The result is:

$$\begin{aligned} W_P + W_G + W_C &= (M + B_P + SF_P^*) - B + (S^* F_C^* + B_C - M) \\ &= S(F_C^* + F_P^*) - B_F \end{aligned}$$

because the Ms cancel out and $B_F = B - B_P - B_C$ (because any domestic bonds not held by domestic residents must be held by foreigners). The interpretation of this result is straightforward: because our model contains only the domestic economy and the rest of the world, the net financial wealth of the home country as a whole must be equal to its financial claims against the rest of the world, given by $S(F_C^* + F_P^*)$, net of the rest of the world's financial claims on the domestic economy, given by B_F. The domestic economy's net financial wealth is called its *international investment position*, which we will denote *IIP*. As we have just shown, it is given by

$$IIP = S(F_C^* + F_P^*) - B_F{}^{8} \tag{6.10}$$

Notice that because $IIP = W_P + W_G + W_C$, it is the sum of three predetermined variables. Thus *IIP* must be predetermined as well. To see how IIP changes over

7 The effects of monetization on the central bank's and the government's balance sheets are discussed in more detail in Appendix 6.1.

8 For data on the international investment positions of emerging and developing economies, see Lane and Milesi-Ferretti (2006).

time, notice that the change in IIP can be expressed as follows:

$$\begin{aligned}\Delta IIP &= \Delta W_P + \Delta W_G + \Delta W_C \\ &= [P(Y-T) + RB_P + R^*SF_P^* - PA_P + \Delta SF_P^*] \\ &\quad + [P(T+T_C-G) - RB] + [R^*SF_C^* + RB_C - T_C + \Delta SF_C^*] \\ &= [P(Y-A_P-G)) + R^*S(F_C^*+F_P^*) - RB_F] + \Delta S(F_P^*+F_C^*) \\ &= [P\cdot NX + R^*S(F_C^*+F_P^*) - RB_F] + \Delta S(F_P^*+F_C^*) \qquad (6.11)\end{aligned}$$

The term in square brackets on the right-hand side of this equation is the sum of the domestic-currency value of the economy's net exports, NX, and its net interest receipts from foreigners $R^*S(F_C^*+F_P^*) - RB_F$, which together represent the current account of the balance of payments, as defined in Chapter 2. The second term represents capital gains or losses on the foreign bonds held by domestic residents as the result of exchange rate changes. Notice that because $IIP = S\,IIP^*$ (the domestic-currency value of the country's net international creditor position is equal to the nominal exchange rate times the foreign-currency value of its net creditor position), we must have

$$\begin{aligned}\Delta IIP &= S\Delta IIP^* + \Delta S IIP^* \\ &= S\Delta IIP^* + \Delta S(F_P^*+F_C^*)\end{aligned}$$

so

$$S\Delta IIP^* = [P\cdot NX + R^*S(F_C^*+F_P^*) - RB_F] \qquad (6.12)$$

This has a useful implication that we will exploit later.

5. Balance Sheet and Portfolio Allocation Decisions of the Rest of the World

Finally, consider the balance sheet and the asset demand functions of the rest of the world. Private agents in the rest of the world are assumed to behave just like domestic ones. Thus their demand for their own money is given by

$$M^{*D} = P^*L^*(\underset{-}{R^*}, \underset{+}{Y^*}) \qquad (6.13)$$

where M^* is the money stock of the rest of the world and the other starred variables are the foreign versions of the corresponding domestic variables. We will assume that because the domestic economy is small, the foreign central bank does not hold any of its foreign exchange reserves in the form of claims issued by agents in the domestic economy. This being so, foreigners' total demand for domestic bonds emanates only from the foreign private sector. The rest of the world's demand for domestic bonds, expressed in *foreign-currency* terms, can therefore be expressed as:

$$B_F^D/S = b^*(\underset{+}{R - R^*})(W_F^* - M^*) \tag{6.14a}$$

where W_F^* is the financial wealth of the foreign private sector, measured in units of the foreign currency. Notice that because B_F is the *domestic-currency* value of the domestic bonds demanded by the rest of the world, it has to be divided by S to make it commensurate with the foreign-currency values W_F^* and M^*.

Just as we discussed the potential role that restrictions on capital outflows could play in determining the effective demand for foreign assets by the domestic private sector, we can now consider how restrictions on capital *inflows* could affect the effective demand for domestic bonds by foreign private agents. By analogy with what we did before, in the presence of such restrictions, we can rewrite equation (6.12) as follows:

$$B_F^D/S = \lambda_2 b^*(R - R^*)(W_F^* - M^*) \tag{6.14b}$$

where λ_2 is a measure of the intensity of restrictions on capital *inflows* in the domestic economy. Again, financial autarky would correspond to $\lambda_2 = 0$, whereas the absence of restrictions implies $\lambda_2 = 1$.

The resources that foreigners in the aggregate (private sector, central bank, and government) have available to invest in domestic bonds consist of their total financial wealth, which is the *negative* of the financial wealth of the domestic economy, – $S\ IIP^*$, plus any borrowing that they do from the domestic economy, consisting of their supply of bonds to the domestic economy, SF^{*S}. Thus the rest of the world's balance sheet can be written as follows:

$$B_F^D = -SIIP^* + SF^{*S} \tag{6.15}$$

We will consider how SF^{*S} is determined in the next section.

II. ASSET SUPPLIES

Now consider the determination of the supply of assets. Recall that domestic residents can hold three assets, consisting of domestic and foreign bonds as well as domestic money, so we have to account for the supply of each separately.

1. The Supply of Domestic Bonds

Consider first the supply of domestic bonds. Because the outstanding stock of government bonds depends on the past history of fiscal deficits, the supply of domestic bonds is predetermined (the supply curve is vertical at any moment of time) and equal to the outstanding stock of government debt B. Thus the supply curve for domestic bonds is

$$B^S = B \tag{6.16}$$

2. The Supply of Foreign Bonds

In the case of foreign bonds, because the domestic economy is assumed to be a small player in the international financial market, it can have no effect on the world interest rate. It therefore faces a perfectly elastic supply of foreign bonds at the world interest rate R^*; that is, domestic residents can acquire as many foreign bonds as they wish in international financial markets, as long as they are willing to accept the rate of return R^* for holding those bonds. Thus the supply of foreign bonds F^{*S} is indeterminate – it is whatever it has to be to satisfy the domestic economy's demand for such bonds at the prevailing world interest rate, say, R_0^*. We can therefore express the supply of foreign bonds as

$$R^* = R_0^* \tag{6.17}$$

3. The Supply of Money

What about the supply of domestic money? Note first that, assuming for the present that the central bank has no financial net worth ($W_C = 0$), the central bank's balance sheet implies the *monetary identity*

$$M^S = SF_C^* + B_C$$

Now recall that according to equation (6.13), IIP can be written as $IIP = S(F_C^* + F_P^*) - B_F$, or, in foreign-currency terms, $IIP = S\ IIP^* = S(F_C^* + F_P^* - B_F/S)$. Solving this for SF_C^*, we can write

$$SF_C^* = S(IIP^* + B_F/S - F_P^*)$$

This equation tells us the sources of the central bank's foreign exchange reserves. They arise from cumulative current account surpluses and deficits (IIP^*) plus cumulative capital inflows (B_F/S) minus cumulative capital outflows F_P^*. Substituting this expression into the monetary identity, we get the money supply relationship

$$M^S = S(IIP^* + B_F/S - F_P^*) + B_C \tag{6.18}$$

This relationship has some important implications. Notice in particular that because IIP^* is predetermined, discrete changes in F_C^* can occur only as the result of portfolio reallocations between domestic and foreign bonds by domestic and foreign residents (capital flows), which affect the quantity $(B_F/S - F_P^*)$ in equation (6.18). Other balance of payments transactions (i.e., the current account) will affect the stock of foreign exchange reserves through IIP^*, but this effect will only happen slowly over time. Thus the key implication of equation (6.18) is that short-run changes in the money supply can result from two sources: capital inflows and outflows operating through $B_F/S - F_P^*$ and changes in domestic monetary policy operating through B_C.

III. FINANCIAL-MARKET EQUILIBRIUM

Having described the demand for and supply of each of the three assets in our model, we are now ready to describe financial-market equilibrium. Generally, with three assets in our model, we would expect that financial-market equilibrium would require simultaneous equilibrium in the markets for domestic money, domestic bonds, and foreign bonds. However, balance sheet constraints imply that the three resulting equilibrium conditions are not independent. To see why, notice first that because $S\,IIP^* = W_P + W_C + W_G = W_P + W_G = W_P - B^S$ (assuming that $W_C = 0$), we can express the financial wealth of the domestic private sector as

$$W_P = S\,IIP^* + B^S \tag{6.19}$$

Then, from equations (6.1), (6.8), and (6.15), the asset demand functions of the domestic private sector, central bank, and the rest of the world have to satisfy the following:

$$\begin{aligned} M^D + B_P^D + SF_P^{*D} &= S\,IIP^* + B^S \\ B_C^D + SF_C^{*D} &= M^S \\ B_F^D &= -S\,IIP^* + SF^{*S} \end{aligned}$$

Adding these together and rearranging terms, we have

$$(M^D - M^S) + \left(B_C^D + B_P^D + B_F^D - B^S\right) + S\left(F_C^{*D} + F_P^{*D} - F^{*S}\right) = 0 \tag{6.20a}$$

In words, the excess demands for the three types of assets in our model always have to sum to zero, whether or not the asset markets are in equilibrium. This means that if any two asset markets are in equilibrium, the third must be as well; that is, the model contains only two independent asset-market equilibrium conditions.

However, things are actually even simpler than this. As we have seen, because the domestic economy is small in the market for foreign bonds, it can have no impact on the price of those bonds, so the supply of such bonds is perfectly elastic at the world interest rate R^*. This means that the supply of such bonds is *identically* equal to the domestic demand for them; that is, the supply of such bonds to the domestic economy is *determined* by the domestic economy's demand for them.[9] The implication is that $F_C^{*D} + F_P^{*D} - F^{*S}$ must always be equal to zero, so equation (6.20a) becomes

$$(M^D - M^S) + \left(B_C^D + B_P^D + B_F^D - B^S\right) = 0 \tag{6.20b}$$

Thus our model actually contains only *one* independent asset-market equilibrium condition. In other words, any value of the domestic interest rate that clears the domestic bond market must also clear the domestic money market, and vice versa. The implication is that we can express financial-market equilibrium in terms of *either* market, being assured that if that market is in equilibrium, the other must be so as well.

1. Bond-Market Equilibrium

While financial-market equilibrium is usually analyzed by considering the determination of equilibrium in the domestic money market, it turns out to be simpler and more intuitive in our case to do so by examining the domestic bond market instead. Using the domestic and foreign demand functions for domestic bonds from Section I, the bond-market equilibrium condition can be expressed as:

$$B - B_C = b(R - R^*)[W_P - PL(R, Y)] + Sb^*(R - R^*)(W_F^* - P^*L^*(R^*, Y^*))$$

where $B - B_C$ is the portion of the outstanding stock of domestic bonds available for sale to the private sector (the existing stock B minus the portion purchased by the central bank B_C), and $b(W_P - PL)$ and $Sb^*(W_F^* - P^*L^*)$ are the demands for domestic bonds by domestic and foreign residents, respectively.

We will be interested later in exploring how the equilibrium value of the domestic interest rate is affected by the domestic price level. To do so, however, we will need to take into account that real output Y, which affects the domestic private sector's demand for domestic bonds, is a function of the price level P, as determined by the economy's aggregate supply curve derived in Chapter 4. To take this relationship into account, we must express Y as a function of P in the bond-market equilibrium condition. This leads us to write that condition as follows:

$$B - B_C = b(R - R^*)[W_P - PL(R, Y(P))] + Sb^*(R - R^*)(W_F^* - P^*L^*(R^*, Y^*)) \tag{6.21}$$

To explore the determination of the domestic interest rate in the domestic bond market, we will proceed graphically. Figure 6.1 depicts this market. The left-hand side of equation (6.21) represents the supply of domestic bonds available for the

[9] This means that the supply of foreign bonds must equal the demand for foreign bonds for *all* possible values of the endogenous variables in the model, so the need to clear this market does not impose any restrictions on the values that these variables can take.

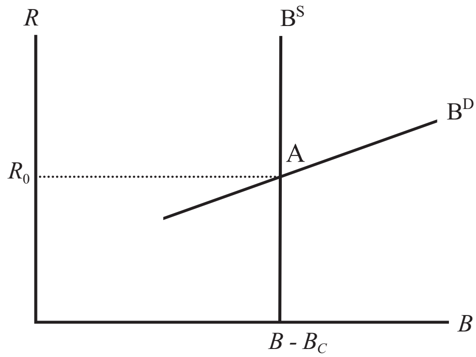


Figure 6.1 Equilibrium in the domestic bond market

private (domestic and foreign) sector to hold. Because the existing stock of such bonds is predetermined and the amount absorbed by the central bank B_C is a policy variable, for a given choice of B_C, the supply of bonds is independent of the interest rate, so we can draw it as a vertical straight line. The position of this line on the horizontal axis depends on B_C. An increase in B_C (a purchase of domestic bonds by the central bank) *reduces* the total amount of such bonds that are available for the private sector to hold and thus shifts the bond supply curve to the *left* by an amount equal to the change in B_C. A reduction in B_C, on the other hand, *increases* the amount of domestic bonds available to the private sector and thus shifts the bond supply curve to the *right* by an amount equal to the reduction in B_C.

Next, consider the demand side of the domestic bond market. An increase in the domestic interest rate, from a value such as R_0, has two effects on the portfolio allocation decisions made by domestic residents. First, they reduce their demand for money and therefore increase their demand for both domestic and foreign bonds. Second, they switch from foreign into domestic bonds because the rate of return on domestic bonds becomes relatively more attractive than that on foreign bonds. For both reasons, an increase in the domestic interest rate increases domestic residents' demand for domestic bonds. In the case of foreign residents, only the second of these effects is operative because their demand for their own money depends on the interest rate on foreign rather than on domestic bonds. Because both domestic and foreign residents increase their demands for domestic bonds in response to a higher domestic interest rate, the bond demand curve, labeled B^D, has a positive slope, as indicated in Figure 6.1.

What determines how steep or how shallow this curve is? The slope of this curve reflects how responsive the demand for domestic bonds is to changes in the domestic interest rate. We would expect that responsiveness to be greater the greater the degree of substitutability between domestic and foreign bonds because when domestic and foreign bonds are close substitutes, a given change in the domestic interest rate should trigger larger shifts in the composition of the financial portfolios of both domestic and foreign residents between the two types of bonds than when they are poor substitutes. Thus the higher the degree of capital mobility, the flatter the demand curve for domestic bonds should be. In the limit, when domestic and foreign bonds are perfect substitutes, the two types of bonds must offer the same rate of return. In that case, the demand curve for domestic bonds must be perfectly flat, at a height equal to the expected rate of return on foreign bonds.

However, it is worth noting that the bond demand curve would have a positive slope even in the case of financial autarky because even if domestic residents cannot switch between domestic and foreign bonds, and foreign residents cannot hold domestic bonds, domestic residents would still be able to change their demand for domestic bonds when the domestic interest rate changes by switching between money and domestic bonds. They would switch from money to domestic bonds when the domestic interest rate rises and the other way (from bonds to money) when it falls, causing the demand for bonds to be responsive to the domestic interest rate.

Domestic bond-market equilibrium holds at a point such as A in Figure 6.1, where the demand for domestic bonds by the domestic and foreign private sectors is exactly equal to the quantity of such bonds that the domestic central bank has made it necessary for them to hold. The domestic interest rate adjusts to ensure that the domestic bond market is always in equilibrium. When the domestic interest rate exceeds its equilibrium value R_0, the central bank finds that the demand for the quantity of domestic bonds that it is placing in the market exceeds its desired supply, and thus it can still induce the private sector to hold these bonds at a lower interest rate. Conversely, when the interest rate is less than R_0, the central bank cannot induce the private sector to absorb all the bonds that it wishes to place in the market unless it offers a higher interest rate. Thus the role of the domestic interest rate is to adjust the demand for domestic bonds to the supply of such bonds so that the quantity of domestic bonds that the private sector *wishes* to hold is exactly equal to the amount that it *must* hold, as determined by the central bank.

2. Bond-Market Equilibrium and Money-Market Equilibrium

Recall that at a point such as A, it must be true not only that the bond market is in equilibrium but also that the domestic private sector's demand for money is exactly equal to the supply of money. To see that adjustment to equilibrium in the bond market implies adjustment to equilibrium in the money market, suppose that the domestic interest rate is below its market-clearing level. As mentioned earlier, at such an interest rate, the central bank cannot place all the bonds it wishes onto the market because the private sector is unwilling to absorb them; that is, it would rather hold domestic money and foreign bonds. Because the domestic private sector can always hold its desired amount of foreign bonds, at the below-equilibrium interest rate, it must wish to hold more money (currency) than the central bank has made available to it; that is, there must initially be an excess demand for money that *exactly* offsets the excess supply of domestic bonds.

To induce the private sector to hold more domestic bonds, the central bank must offer a more attractive interest rate on such bonds. Consider what happens in the money market as it does so. When the central bank increases the interest rate on domestic bonds, the domestic private sector reduces its demand both for money and for foreign bonds, and foreigners also switch away from foreign to domestic bonds. The switch from money to domestic bonds by the domestic private sector reduces the demand for money directly, whereas the switch from foreign to domestic bonds by both the domestic and foreign private sectors causes a capital inflow into the domestic economy that increases the domestic money supply. The reason it does so is that to keep the exchange rate constant, the central bank must purchase the capital inflow by printing money. The upshot is that at the higher domestic interest rate, the demand for money decreases, while the supply of money increases. Because this reduces the excess demand for money, the money market moves toward equilibrium at the same time that the domestic bond market does so.

To see that the resulting effect on the excess demand for money must be *exactly* enough to restore equilibrium in the money market when the bond market achieves equilibrium, suppose that for every \$1 increase in demand for domestic bonds, some fraction k of that increase comes from domestic residents and that $(1 - k)$ comes from foreign residents. Suppose also that for domestic residents, some fraction s of every \$1 increase in their demand for domestic bonds reflects a reduced demand for money, while a fraction $(1 - s)$ reflects a reduced demand for foreign bonds. Then when the demand for domestic bonds increases by \$1 as a result of the central bank offering a more attractive interest rate, the domestic demand for money falls by \$$sk$, whereas the supply of money increases by \$$(1 - s)k + (1 - k)$. The excess demand for money thus changes by $-\$(sk + [(1 - s)k + (1 - k)]) = -\ \$1$, that is, by exactly the same amount that the excess supply of bonds falls.

IV. COMPARATIVE STATICS IN THE DOMESTIC BOND MARKET

We can now explore the properties of domestic bond-market equilibrium by examining what would happen to that equilibrium in the face of a variety of macroeconomic shocks.

1. Monetary Policy

Consider first the effects of an expansionary monetary policy, that is, a central-bank purchase of domestic bonds that increases B_C and M by equal amounts. As we saw earlier, an increase in B_C shifts the bond supply curve to the *left* by the same absolute amount as the change in B_C, to a position such as B_1^S in Figure 6.2.

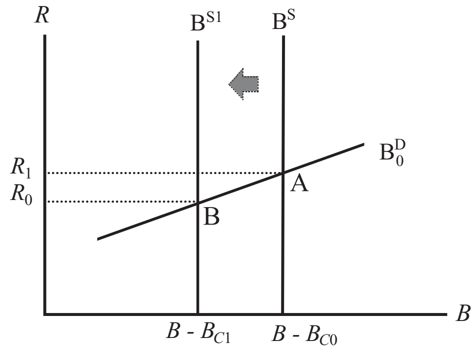


Figure 6.2. Expansionary monetary policy

Because the monetary expansion reduces the stock of domestic bonds and increases the stock of money in the hands of the private sector, at the original interest rate R_0, it creates an excess demand for bonds and an excess supply of money. To induce the private sector to willingly hold more money and fewer domestic bonds, the return on domestic bonds (the domestic interest rate) has to fall. As it does so, foreign bonds become relatively more attractive, triggering capital outflows that force the central bank to buy domestic currency (sell foreign exchange reserves) to sustain the official exchange rate. Thus the decrease in the domestic interest rate simultaneously reduces the demand for domestic bonds, increases the demand for money, and reduces the money supply. The final equilibrium is at the point B, with a lower domestic interest rate, a smaller stock of domestic bonds in the hands of the public, and a larger money supply. Notice, however, that because of the capital outflows induced in moving from the initial equilibrium at A to the new one at B, the increase in the money stock in the final equilibrium is not as large as that created by the initial monetary expansion.

Notice also that the effect of the monetary expansion on the equilibrium value of the domestic interest rate must be smaller the greater the degree of capital mobility. This is so because when capital mobility is high, the demand curve for domestic bonds is relatively flat, so a smaller reduction in the domestic interest rate is required to persuade the private sector to willingly decrease its demand for domestic bonds. The implication is that the impact of the expansionary monetary policy on the domestic money supply is smaller the higher the degree of capital mobility. In the limit, when domestic and foreign bonds are perfect substitutes, the B^D curve is perfectly flat, so the leftward shift in the B^S curve has no effect on the domestic interest rate whatsoever. Intuitively, when domestic and foreign bonds are close substitutes, a small change in the domestic interest rate causes a large switch between domestic and foreign bonds. Thus, under high capital mobility, a very small reduction in the domestic interest rate is sufficient to induce domestic and foreign residents to greatly reduce their demand for domestic bonds in favor of foreign ones, that is, to induce large capital outflows from the domestic economy. When

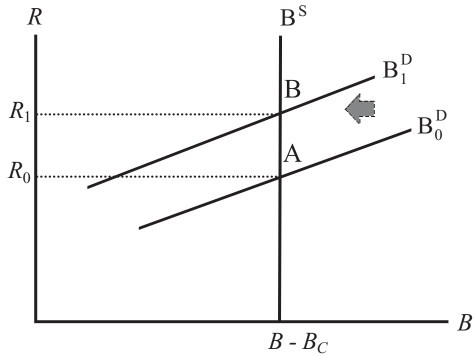


Figure 6.3. **Bond-market effects of an increase in the domestic price level**

capital mobility is perfect, no change in the domestic interest rate is required at all because since domestic and foreign interest rates are perfect substitutes, portfolio managers will gladly switch from domestic to foreign bonds without any interest rate inducement. In this case, the domestic bonds are simply replaced by foreign ones on a one-for-one basis, and the monetary expansion is therefore exactly offset by a capital outflow that leaves the domestic money supply unchanged.

2. Changes in the Domestic Price Level

Next, consider the effects on bond-market equilibrium of a change in the domestic price level. A change in the domestic price level affects the preferences of domestic residents to hold money instead of bonds. An increase in the domestic price level, for example, increases the demand for money through two channels: for a given *real* demand for money L, the *nominal* demand for money must increase in proportion to the increase in the price level, but in addition, the increase in the price level increases the real demand for money itself, through its positive effect on real output Y.[10] The upshot is that domestic residents switch from both domestic and foreign bonds to money.[11]

The effects on bond-market equilibrium are shown in Figure 6.3. The increase in the domestic price level shifts the demand curve for domestic bonds to the left. Because the quantity of domestic bonds in the hands of the private sector remains unchanged (only central-bank monetary policy can change it), the effect is to create an excess supply of domestic bonds at the initial interest rate, with a corresponding excess demand for money.[12] An increase in the domestic interest rate is therefore

[10] Recall that the demand for money is given by $PL(R, Y(P))$.

[11] The portfolio choices of foreign residents are unaffected.

[12] The increase in the domestic price level actually increases the money supply because the domestic private sector's reduced demand for foreign bonds induces a corresponding capital inflow of equal size that the central bank purchases by printing new money. But because the demand for foreign bonds only falls by a fraction f of the increase in domestic residents' demand for money, this capital

required to induce the private sector to continue to willingly hold the stock of domestic bonds that are in its hands. While it increases the demand for domestic bonds, the higher domestic interest rate reduces the demand both for money and for foreign bonds. The reduced demand for foreign bonds results in capital inflows that help restore equilibrium in the money market by increasing the domestic money supply.

Notice that because the effect of the increase in the domestic price level on the demand for money is independent of the degree of capital mobility, the size of the leftward shift in the demand curve for domestic bonds must also be independent of the degree of capital mobility. But because the demand curve is flatter the higher the degree of capital mobility, the impact of this shock on the equilibrium value of the domestic interest rate must once again be smaller the higher the degree of capital mobility.

3. A Change in the Country's International Investment Position

From equation (6.19), the financial wealth of the domestic private sector W_P is equal to the sum of the domestic economy's net international investor position, $S \cdot IIP^*$, and the stock of domestic government bonds outstanding, B. A similar expression must hold for the private sector in the rest of the world, so $W_P^* = -IIP^* + F^*$. Because both W_P and W_P^* appear in the bond-market equilibrium condition (6.21), this means that a change in IIP^* must affect the domestic bond-market equilibrium. It does so by changing the relative wealth of the domestic and foreign private sectors: when IIP^* increases, the financial wealth of the domestic private sector increases and that of the foreign private sector decreases by exactly offsetting amounts. The increase in the financial wealth of the domestic private sector increases the demand for domestic bonds, while the reduction in the financial wealth of the foreign private sector decreases it.

The net effect on the demand for domestic bonds depends on whether the desired share of domestic government bonds in the nonmonetary wealth of the domestic private sector, given by the fraction b, exceeds or falls short of the desired share of such bonds in the nonmonetary wealth of the foreign private sector, given by b^*. If the domestic and foreign private sectors were indifferent about the types of bonds they hold in their portfolios when domestic and foreign bonds pay the same rates of return, we would expect the share of domestic bonds in their portfolios to be the same, both equal to the ratio of domestic bonds to the total of domestic and foreign bonds outstanding ($b = b^* = B/(B + SF^*)$). But a stylized fact about world capital markets is that portfolio managers tend to exhibit *home bias*, that is, a preference to hold their own domestic bonds. Home bias implies that

inflow is smaller than the increase in the demand for money, with the net effect remaining the emergence of an excess demand for money.

$b > b^*$. Under this condition, the net effect of an increase in IIP^* that transfers wealth from the foreign to the domestic private sector is to transfer wealth from agents with a relatively low preference for domestic bonds to those with a relatively high preference for them. The effect is therefore to increase the total demand for domestic bonds. The resulting rightward shift in B^D would cause the equilibrium value of the domestic interest rate to fall.

In this case, the size of the rightward shift in B^D depends only on the relative *magnitudes* of b and b^*, not on the degree of capital mobility – that is, not on the sensitivity of b and b^* to the interest rate differential – because the shift is measured at a given value of the domestic interest rate. Because the size of the horizontal shift in B^D is unaffected by the degree of capital mobility, while the intersection of the new B^D curve with the stationary B^S curve will be closer to the original interest rate the flatter the B^D curve is, the effects of a change in IIP^* on the domestic interest rate must be smaller the higher the degree of capital mobility.

4. A Change in the Foreign Interest Rate

Next, consider the effects on the equilibrium value of the domestic interest rate of a change in the foreign interest rate. For concreteness, suppose that the foreign interest rate falls. A reduction in the interest rate on foreign bonds affects the demand for domestic bonds through several channels. First, domestic residents switch from foreign to domestic bonds. This clearly increases the demand for domestic bonds. For foreign residents, however, the situation is more complicated. On one hand, foreign residents also find it attractive to switch their initial nonmonetary wealth from holding foreign to holding domestic bonds (a substitution effect). On the other hand, the lower foreign interest rate also induces them to hold more of their own money and therefore reduces their nonmonetary wealth, inducing them to wish to hold *fewer* of both domestic and foreign bonds (a portfolio scale effect). The net effect on their demand for domestic bonds is therefore, in principle, ambiguous. The issue is whether, when they are induced to hold more of their own money by a lower interest rate on the foreign bond, they switch primarily out of domestic bonds or out of foreign bonds. You can see that if domestic and foreign bonds are relatively close substitutes, they would tend to switch primarily out of their own relatively less attractive bonds and therefore to demand *more* domestic bonds; that is, the substitution effect would dominate the portfolio scale effect. We will assume that this is indeed so, an assumption that is referred to as the *gross substitutes* assumption. Under the gross substitutes assumption, then, the demand for domestic bonds increases, and the equilibrium value of the domestic interest rate falls, as in Figure 6.4.

To examine the *quantitative* effect on the equilibrium interest rate on domestic bonds, let us examine the size of the *vertical* shift in the B^D curve. We can do so by investigating how much the domestic interest rate would have to fall to keep

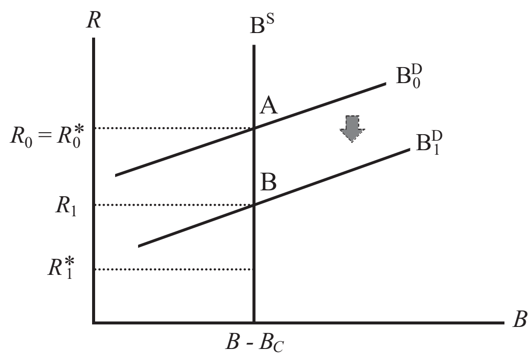


Figure 6.4. Bond-market effects of a reduction in the foreign interest rate

the total demand for domestic bonds unchanged. To answer this question, note that the portfolio allocation shares b and b^* remain unchanged when the interest rate differential $R - R^*$ is unchanged. This means that a reduction in the domestic interest rate that is exactly equal to that in the foreign interest rate would cause domestic and foreign residents to leave the allocation of their nonmonetary wealth between domestic and foreign bonds unchanged. But note also that if domestic and foreign interest rates both were to fall by equal amounts, domestic and foreign residents would *both* shift into money and out of bonds. Consequently, their demand for domestic bonds would have to fall. What this means is that to keep the demand for domestic bonds constant, the domestic interest rate *cannot* fall by as much as the foreign rate – it would have to fall by *less* than the foreign interest rate. If it were to do so, domestic and foreign residents would both still shift out of bonds and into money, but they would also shift out of foreign bonds and into domestic ones. With a large enough differential between the new domestic and foreign interest rates (in favor of the domestic interest rate), the reallocation of portfolios from foreign to domestic bonds would be just sufficient to offset that from bonds into money, leaving the demand for domestic bonds unchanged. Notice that the higher the degree of capital mobility, the smaller the size of the interest rate differential required to keep the demand for domestic bonds unchanged.

What all this means geometrically is that a decrease in the foreign interest rate must be associated with a *downward* shift in the demand curve for domestic bonds that is smaller than the size of the change in the foreign interest rate but that is closer to the size of the change in the foreign interest rate the higher the degree of capital mobility. Figure 6.4 shows the effect of a fall in the world interest rate on the domestic bond market. Assume for simplicity that the domestic and foreign interest rates are initially equal and that the foreign interest rate falls from R_0^* to R_1^*. As we have just seen, the B^D curve must consequently shift down, but by an amount that falls short of the decline in R^*. The upshot is that the equilibrium value of the domestic interest rate must fall as well, but by less than the fall in the foreign interest rate. However, the decline in the domestic interest rate will be closer to that in the foreign interest rate the higher the degree of capital mobility and will be exactly equal to that in the foreign interest rate when capital mobility is perfect because in that case, no interest rate differential can exist in the new equilibrium.

5. Changes in the Nominal Exchange Rate

Finally, consider the effects on the domestic interest rate of a change in the nominal exchange rate – say, for concreteness, of a nominal exchange rate devaluation. Recall that a devaluation creates a capital gain on the central bank's foreign exchange reserves that may or may not be monetized. We saw previously that monetization could be accomplished by transferring the capital gains to the government, which would then use those resources to buy back some of its debt from the private sector. If the capital gains on reserves are monetized, therefore, the stock of government bonds in the hands of the private sector, $B - B_C$, would fall.

However, suppose initially that capital gains on reserves are *not* monetized. In that case, as you can verify from equation (6.21), the effects of the devaluation on equilibrium in the domestic bond market operate through portfolio valuation effects. There are two such effects, on domestic and foreign residents, respectively:

1. The devaluation increases the domestic-currency value of the foreign bonds held by the domestic private sector and thus increases the domestic private sector's wealth measured in domestic currency, W_P. Because the domestic private sector devotes a fraction b of any change in its wealth to holding domestic bonds, this *increases* its demand for domestic bonds.
2. On the other hand, the devaluation of the domestic currency reduces the foreign-currency value of the domestic bonds that are held by foreigners. To maintain the same desired share of domestic bonds in their financial portfolios, therefore, foreigners would have to *increase* their demand for domestic bonds by an amount that is exactly equal to the capital loss that they suffered on those bonds, thus replenishing the foreign-currency value of the stock of domestic bonds that they hold. However, that is not the whole story, because the capital loss that foreigners suffer on domestic bonds represents a reduction in their wealth measured in foreign-currency terms, W_P^*. Because foreigners change their demand for domestic bonds by a fraction b^* of the change in their wealth, this reduction in their wealth *decreases* their demand for domestic bonds by b^* times the capital loss. However, because the effect tending to increase foreigners' demand for domestic bonds is equal to their capital loss on such bonds, while the effect tending to decrease their demand for domestic bonds is only a fraction b^* of that capital loss, the net result of these offsetting effects must nevertheless be to increase foreign demand for domestic bonds.

Because both domestic and foreign residents increase their demand for domestic bonds, the aggregate demand for domestic bonds increases. Thus the bond demand

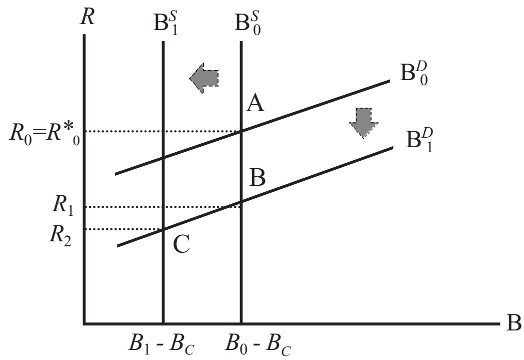


Figure 6.5. Bond-market effects of a nominal devaluation

curve shifts to the right, and the equilibrium value of the domestic interest rate falls, as at point B in Figure 6.5.[13] Notice, however, that the magnitude of this shift depends on the strength of these valuation effects, and these effects in turn depend on the initial stocks of foreign bonds in the hands of the domestic private sector as well as on the initial stocks of domestic bonds in the hands of foreigners. The smaller these initial cross-border bond holdings are, the weaker valuation effects will be, and the smaller the effect of the devaluation on the demand for domestic bonds. In the limit, if domestic and foreign residents do not initially happen to hold any of each others' bonds (even though they are not precluded from doing so by any legal restrictions), the devaluation would have no effect on the demand for domestic bonds.

How would the monetization of the capital gains on the domestic central bank's foreign exchange reserves affect this result? The answer is that because monetization would reduce the stock of domestic bonds in the hands of the private sector, the rightward shift in the B^D curve would be supplemented by a *leftward* shift in the B^S curve. As shown at point C in Figure 6.5, the result would be an even *larger* decline in the equilibrium value of the domestic interest rate.

V. THE BB CURVE

As we did for the goods market in Chapter 5, our final task in this chapter is to construct a graphical representation of the relationship between the domestic price level and the interest rate on domestic bonds that is consistent with domestic financial-market equilibrium. Recall that equilibrium in the domestic bond market implies equilibrium in the money market, and vice versa. We will eventually derive

[13] As in some of the other cases we have considered, the magnitude of the *horizontal* shift in the B^D is not affected by the degree of capital mobility, and because the B^D curve becomes flatter the greater the degree of capital mobility, the effect of the devaluation on the domestic interest rate is smaller the higher the degree of capital mobility.

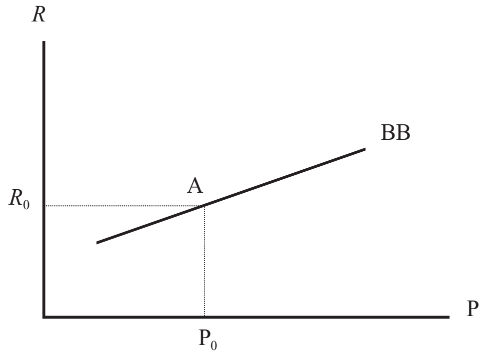


Figure 6.6. The BB curve

two such curves: one that depicts domestic financial-market equilibrium for a given value of the stock of domestic credit B_C, which we will call the BB curve, and one that does so for a given value of the money supply, which we will call the MM curve. As you will see, both curves will be useful in understanding the economy's short-run equilibrium. We will derive the BB curve in this section and turn to the MM curve in Chapter 7.

In effect, we already derived the relationship between the equilibrium value of the domestic interest rate and the domestic price level that underlies the BB curve in Section IV, where we examined the impact of a change in the price level on the value of the domestic interest rate that clears the domestic bond market. We saw there that an increase in the domestic price level shifts the demand curve for domestic bonds to the *left* in (B, R) space because the higher price level induces domestic residents to hold more money and fewer domestic (as well as foreign) bonds. As a result, the increase in the price level requires an *increase* in the domestic interest rate to induce the private sector to continue to hold the existing stock of bonds willingly. All that remains, therefore, is to plot this relationship in (P, R) space. This is done in Figure 6.6. The BB curve drawn in the figure shows that the value of the domestic interest rate required to clear the domestic bond market is an *increasing* function of the domestic price level.

Because the BB curve summarizes the conditions under which domestic financial markets are in equilibrium for a given value of B_C, its properties can be derived from the bond-market equilibrium analysis that we conducted earlier in this chapter. The most important properties are as follows:

- As we have just seen, the BB curve has a positive slope. The reason is that an increase in the domestic price level reduces the demand for domestic bonds, while leaving the supply unchanged. This means that a higher value of the domestic interest rate is required to clear the domestic bond market when domestic prices are higher.

- As we saw in Section III, the demand curve for domestic bonds is flatter the higher the degree of substitutability between domestic and foreign bonds, that is, the higher the degree of capital mobility. Because the effect of a change in the domestic price level on the *horizontal* position of the bond demand curve does not depend on how steep or flat the curve is, this means that the effect on the equilibrium value of the domestic interest rate of a change in the domestic price level must be *smaller* the higher the degree of substitutability between domestic and foreign bonds. Thus the higher the degree of substitutability between domestic and foreign bonds, the *flatter* the BB curve must be. The slope of the BB curve is steepest (but still positive) in the limiting case of zero capital mobility, while with perfect capital mobility, it is perfectly flat at the height $R = R^*$.[14]
- If domestic and foreign bonds are gross substitutes (in this case, this means that an increase in the domestic interest rate increases the demand for domestic bonds at the expense of both money and foreign bonds), movements to the upper right along the BB curve must be associated with higher values of the domestic money supply. The reason is the following: movements to the upper right along the BB curve are associated with higher values of the domestic price level as well as higher values of the domestic interest rate. As the domestic price level increases, the demand for money increases, which reduces domestic residents' demand for foreign bonds. This results in capital inflows that increase the domestic money supply. As the domestic interest rate increases, on the other hand, domestic residents reduce their demand for money and switch from foreign to domestic bonds. Though the switch from foreign to domestic bonds reduces their demand for foreign bonds, the switch from money and into bonds of both types increases that demand. The former effect dominates if domestic and foreign bonds are gross substitutes, so in that case, a higher domestic interest rate causes domestic residents' demand for foreign bonds to fall, and this effect also induces a capital inflow that expands the domestic money supply. Thus, under the gross substitutes assumption, movements to the upper right along the BB curve tend to increase the domestic money supply.
- Because the BB curve shows the value of the domestic interest rate required to achieve equilibrium in the domestic bond market for given values of the domestic price level, variables other than the domestic price level that affect the equilibrium interest rate in the domestic bond market cause the BB curve to shift. When any of these variables change, the size of the *vertical* shift in the BB curve must be exactly the same as the change in the equilibrium value of the

[14] Note that because the slope of the BB curve depends on how strongly portfolio allocations respond to interest rate changes, an intensification of capital account restrictions would tend to make it steeper.

domestic interest rate derived from the bond-market supply-demand diagrams in Section IV.

- Finally, because higher domestic interest rates increase demand for domestic bonds by both domestic and foreign residents, points above the BB curve must be associated with an excess demand for domestic bonds, and points below it with an excess supply.

VI. SUMMARY

In this chapter, we have investigated how the domestic interest rate is determined, by exploring the properties of financial-market equilibrium in a model in which the only interest-bearing financial assets are domestic and foreign government bonds. In this context, we found that the shape of the demand curve for domestic bonds depends on the nature of the domestic economy's financial links with the rest of the world.

Concerning those links, we considered three possible alternative cases: financial autarky, imperfect capital mobility, and perfect capital mobility. Financial autarky arises when the domestic economy maintains an effective prohibition against the holding of foreign assets by domestic residents and against the holding of domestic assets by foreign residents. Imperfect capital mobility exists when restrictions on capital movements are less severe, and domestic and foreign assets are imperfect substitutes in the portfolios of private agents. Perfect capital mobility represents the extreme case in which there are no restrictions on capital movements (or those that exist are ineffective), and either domestic and foreign assets share the same risk characteristics or portfolio managers do not care about risk.

We adopted as our benchmark case that of imperfect capital mobility because this provides a general framework within which the other cases can be examined as well. We investigated the determination of the equilibrium value of the domestic interest rate under these circumstances, on the assumption that the central bank targets the stock of domestic credit B_C as an exogenous policy variable. In this context, we analyzed the effects on the equilibrium value of the domestic interest rate of a variety of shocks, including changes in the stock of domestic credit, in the domestic price level, in the world interest rate, in the country's international investment position, and in the nominal exchange rate. All these results could be summarized in the form of a BB curve in (P, R) space.

In the next chapter, we will put this financial-market analysis together with the goods-market analysis of Chapter 5 to construct a complete short-run macro model of a small open economy that maintains an officially determined exchange rate and that is imperfectly integrated with international financial markets. As we have seen, this is a set of circumstances that describes many developing and emerging economies.

APPENDIX 6.1. MONETIZATION AND REVALUATION ACCOUNTS

Recall equation (6.6) describing the balance sheet of the central bank:

$$W_C = SF_C^* + B_C - M$$

What happens to W_C when the exchange rate is devalued? Suppose that initially W_C was zero, so

$$W_{C0} = 0 = S_0F_{C0}^* + B_{C0} - M_0$$

After the devaluation, we have the following:

$$\begin{aligned} W_{C1} &= S_1F_{C0}^* + B_{C0} - M_0 \\ &= (S_1 - S_0)F_{C0}^* + S_0F_{C0}^* + B_{C0} - M_0 \end{aligned}$$

Thus

$$\Delta W_C = W_{C1} - W_{C0} = (S_1 - S_0)F_{C0}^*$$

What happens to the money supply after the devaluation depends on what the central bank does with this capital gain and on what monetary policy it pursues at the same time. Consider two options, as follows.

1. No Monetization

If the central bank keeps the capital gain on its foreign exchange reserves $(S_1 - S_0)F_{C0}^*$ on its own accounts, the increase in its net worth described earlier is the end of the story. The quantity $(S_1 - S_0)F_{C0}^*$ is kept in a revaluation account that simply records the increase in the central bank's net worth. M_0 does not change, and there is no change in the money supply.

2. Monetization

Now suppose the central bank transfers these gains to the government. Because the central bank simply credits the government's account at the central bank by the amount $(S_1 - S_0)F_{C0}^*$, the initial effect is simply to reduce the amount that the government owes the central bank – that is, the stock of domestic credit – by $(S_1 - S_0)F_{C0}^*$. Now the central bank's balance sheet becomes

$$W_{C1} = S_1F_{C0}^* + B_{C1} - M_0 = 0, \quad \text{where} \quad B_{C1} = B_{C0} - (S_1 - S_0)F_{C0}^*$$

The central bank's net worth is zero because the amount $(S_1 - S_0)F_{C0}^*$ is added to $S_0F_{C0}^*$ and subtracted from B_{C0}. Notice that the money supply still has not changed.

The government is the one that experiences an increase in its net worth in this case. Its net worth becomes

$$W_G = -(B - (S_1 - S_0)F^*_{C0}) = -B + (S_1 - S_0)F^*_{C0}$$

so the government is richer.

What if the government uses its new wealth to retire debt held by the private sector? It can do so by using its credit line at the central bank, thereby restoring its debt to the central bank to the original level B_{C0} and using the proceeds in the form of $(S_1 - S_0)F^*_{C0}$ of new money to buy back debt from the private sector. The government's balance sheet is now unchanged from the one immediately preceding, but the central bank's balance sheet becomes

$$W_{C1} = S_1 F^*_{C0} + B_{C0} - M_1 = 0, \quad \text{where} \quad M_1 = M_0 + (S_1 - S_0)F^*_{C0}$$

The central bank's net worth is still zero because the amount $(S_1 - S_0)F^*_{C0}$ has been added both to $S_0 F^*_{C0}$ and to M_0.

REVIEW QUESTIONS

1. In what sense are the financial wealth of the private sector, the government, and the central bank predetermined?
2. What do we mean by *balance sheet constraints?*
3. How does the private sector's demand for domestic bonds change as domestic and foreign bonds become closer substitutes for each other?
4. Given that our financial-market model has three different assets, how was it possible to study general financial-market equilibrium by exploring how equilibrium is determined just in the domestic bond market?
5. Explain how the BB curve is determined from the analysis of equilibrium in the domestic bond market.

EXERCISES

1. In the model that we developed in Chapter 5, we assumed that the central bank determined the domestic interest rate as an exogenous policy variable, but we did not explain how it did so. Now that we have examined the market for domestic bonds, you are able to provide that explanation. This problem asks you to do so. Specifically, explain what the central bank has to do to keep the domestic interest rate fixed at its policy-determined level in response to each of the following shocks:
   a. an increase in the domestic price level
   b. an increase in the interest rate paid by foreign bonds
   c. a devaluation of the exchange rate
   d. an exogenous increase in the domestic demand for money

   e. an exogenous increase in the preferences for foreign instead of domestic bonds on the parts of both the domestic and foreign private sectors
2. How does an increase in the domestic economy's degree of integration with international financial markets affect the actions that the central bank has to take to fix the value of the domestic interest rate at its preferred value in response to such shocks? You may use any single shock to answer this question.
3. Briefly comment on the validity of the following statement: "Central banks have to choose: they cannot simultaneously fix the exchange rate and the domestic interest rate."
4. Consider a small open economy that maintains a fixed exchange rate regime and is characterized by imperfect capital mobility. Also assume that it implements a domestic credit rule that keeps the stock of domestic credit constant as its monetary policy regime. What is the effect on the money supply of an increase in the price level?
5. Assume that for unspecified reasons, domestic agents decide to hold more money at the same interest rate (exogenous shift in the money demand equation).
   a. What happens to the domestic interest rate and money supply if the central bank keeps domestic credit constant?
   b. Would the effect on the interest rate be larger or smaller if the economy had substantial capital account restrictions?
6. Following the East Asian financial crisis, many emerging economies experienced episodes of sudden capital outflows, as foreign investors decided to take their capital out of these countries.
   a. Use the model developed in this chapter to illustrate how this situation would have affected the domestic bond market, and explain what would have happened to the domestic interest rate in the aftermath of the capital outflow if the central bank had kept domestic credit constant.
   b. What would the central bank have to had done in this case to keep the domestic interest rate constant? What would have happened to the central bank's balance sheet in this case?
7. Compare how the central bank should adjust the stock of domestic credit in response to an increase in the domestic price level if it is trying to
   a. keep the domestic money supply unchanged
   b. keep the domestic interest rate unchanged

REFERENCES AND FURTHER READING

Branson, William H., and David W. Henderson (1984), "The Specification and Influence of Asset Markets," in R. Jones and P. Kenen, eds., *Handbook of International Economics*, Amsterdam: North-Holland, pp. 749–806.

Frankel, Jeffrey A. (1997), "Sterilization of Money Flows: Hard (Calvo) or Easy (Reisen)?" *Estudios de Economia*, vol. 24, pp. 263–285.

Frankel, Jeffrey A., and Chudozie Okongwu (1996), "Liberalized Portfolio Capital Inflows in Emerging Markets: Sterilization, Expectations, and the Incompleteness of Interest Rate Convergence," *International Journal of Finance Economics*, vol. 1, pp. 1–23.

Glick, Reuven, and Ramon Moreno (1994), "Capital Flows and Monetary Policy in East Asia," Pacific Basin Working Paper Series PB94-08, Federal Reserve Bank of San Francisco.

Lane, Philip R., and Gian Maria Milesi-Ferretti (2006), "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities, 1970–2004," working paper WP/06/69, International Monetary Fund.

Willett, Thomas D., Manfred W. Kiel, and Young Seok Ahn (2002), "Capital Mobility for Developing Countries May Not Be So High," *Journal of Development Economics*, vol. 68, pp. 421–434.

7

Short-Run Macroeconomic Equilibrium

In Chapter 5, we analyzed how aggregate supply and demand interact to determine the equilibrium value of the domestic price level and the level of real output. To find the equilibrium values of these variables, however, we needed to know the value of the domestic interest rate. In that chapter, we assumed that the central bank simply set the value of the domestic interest rate as a policy variable, without specifying how it managed to do so. In Chapter 6, on the other hand, we saw that in general – that is, when the central bank is not explicitly targeting the interest rate – the short-run equilibrium value of the domestic interest rate itself depends on the domestic price level, through the effects of changes in the domestic price level on equilibrium in the domestic bond market. In Chapter 6, we took the domestic price level as an exogenous influence on the domestic bond market, without specifying how the price level was itself determined.

In this chapter, we will put these two pieces of the story together to see how the domestic price level and interest rate are simultaneously determined in a general equilibrium involving the labor market, the goods market, and the financial markets. To understand how the model works, we will make use of the goods-market equilibrium curve GM from Chapter 5 and the domestic bond-market equilibrium curve BB from Chapter 6. To enrich our analysis, we will also find it useful to introduce two new curves: an iso-money curve, which we will call MM, and an iso-interest rate curve, which we will call RR.[1] We will develop both of those curves in Section I. In Section II, we will examine how the goods-market and financial-market equilibrium conditions interact to determine the economy's short-run macroeconomic equilibrium. We then turn to an analysis of how the economy adjusts to various types of shocks. Aggregate demand shocks are analyzed

[1] Recall that *iso* means "the same" in Greek, so an *iso-money curve* is simply a curve along which the domestic money stock is the same, and an *iso-interest rate curve* is just a curve along which the domestic interest rate is unchanged.

in Section III, and aggregate supply shocks are analyzed in Section IV. Section V examines how shocks of these various types have contributed to macroeconomic instability in emerging and developing economies. Section VI summarizes.

In the next chapter, we will use this short-run model to explore how domestic macroeconomic policies and other exogenous shocks affect two key macroeconomic relative prices: the real interest rate and the real exchange rate. We will also see how these two important relative prices evolve over time as the economy moves from a short-run to a medium-run equilibrium.

I. THE MM AND RR CURVES

The BB curve derived at the end of the preceding chapter was drawn for a given quantity of the stock of domestic credit in the central bank's balance sheet, B_C. As we saw in that chapter, the money supply changes along the BB curve as the result of capital inflows and outflows, with the money stock increasing as we move up and to the right along the BB curve. However, what happens to our depiction of domestic financial-market equilibrium if, instead of keeping B_C constant, the central bank adjusts it continuously to offset the effects of capital flows on the money supply, thereby keeping the money supply constant? Recall from the preceding chapter that this is referred to as *sterilizing* the central bank's intervention in the foreign exchange market. To answer this question, our next task will be to examine what domestic financial-market equilibrium would look like in (P, R) space under this alternative assumption. In other words, we want to explore what will happen to the equilibrium value of the domestic interest rate as the domestic price level changes when the central bank keeps the money supply, rather than the stock of domestic credit, constant. We will refer to the curve that shows how the equilibrium value of the domestic interest rate changes for alternative values of the domestic price level, holding constant the domestic money supply, as the MM curve.

Just to be clear about the difference between the MM and BB curves, remember that the central bank's balance sheet is given by

$$SF_C^* + B_C = M$$

Along the BB curve, the central bank holds B_C constant and allows changes in SF_C^* induced by its foreign exchange–market intervention to determine M endogenously. As we saw in the preceding chapter, M actually increases as the economy moves up and to the right along the BB curve. Along the MM curve, the central bank instead responds to changes in SF_C^* induced by its foreign exchange–market intervention by changing B_C so as to leave M unchanged. Thus, while along the BB curve, the stock of domestic credit is constant and the money supply is endogenous, along this new iso-money MM curve, the money supply is constant, and the stock of domestic credit is endogenous.

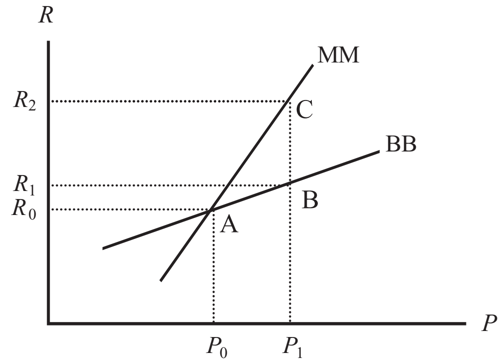


Figure 7.1. **The MM and BB curves**

1. Slope of the MM Curve

The derivation of the MM curve is illustrated in Figure 7.1, using the BB curve from the preceding chapter as a point of reference. Consider the point A on the BB curve. As we saw in the preceding chapter, an increase in the domestic price level, say, from P_0 to P_1, would increase the demand for domestic money on the part of domestic residents, while reducing their demand for both domestic and foreign bonds. Thus, with unchanged asset stocks, the increase in the domestic price level creates an excess demand for domestic money and exactly offsetting excess supplies of both domestic and foreign bonds. To restore equilibrium in the domestic money market with an unchanged value of the money supply would therefore require a change in the domestic interest rate that *reduces* the demand for money back to its original value. Because the demand for money depends negatively on the interest rate, what is required is an *increase* in the domestic interest rate. This immediately establishes that the MM curve must have a positive slope. In other words, when the domestic price level rises, a *higher* value of the domestic interest rate is needed to maintain domestic financial-market equilibrium with an unchanged value of the domestic money supply.

2. Relative Slopes of MM and BB

Next, consider the relative slopes of the BB and MM curves. It is straightforward to show that the MM curve must be steeper than the BB curve. To see how, notice that when the domestic price level rises, the increase in demand for domestic money comes partly at the expense of domestic bonds and partly at the expense of foreign bonds. It must be true, then, that the excess demand for money created by an increase in the price level must be *larger* than the excess supply of domestic bonds caused by the same increase in the price level; that is, price-level changes create a larger disequilibrium in the domestic money market than they do in the domestic bond market.

Now consider the relative effectiveness of domestic interest rate changes in restoring equilibrium in the two markets. Recall from the last paragraph that in response to an increase in the domestic price level, an *increase* in the domestic interest rate is required to move both markets back toward equilibrium. A higher interest rate reduces the demand for money while increasing the demand for domestic bonds. But when the domestic interest rate rises, domestic residents' demand for domestic bonds increases at the expense *both* of their demands for domestic money and for foreign bonds. Because the increase in their demand for domestic bonds must be exactly equal to the sum of their reduced demands for money and foreign bonds, it follows that domestic interest rate changes must have more powerful effects on the domestic bond market than they do on the domestic money market. This difference is magnified when the response of foreign residents to domestic interest rate changes is taken into account because their increased demand for domestic bonds adds to the total impact of domestic interest rate changes on the demand for domestic bonds.

In short, domestic price-level changes have relatively strong impacts on the domestic money market and relatively weak ones on the domestic bond market, whereas domestic interest rate changes have relatively strong impacts on the domestic bond market and relatively weak ones on the domestic money market. Because a given price-level change therefore creates a larger disequilibrium in the domestic money market than it does in the domestic bond market, and a given interest rate change is less powerful in restoring equilibrium in the money market than it is in the domestic bond market, it follows that for a given price-level change, a smaller increase in the domestic interest rate is required to restore equilibrium in the bond market than is required to do so in the money market. In terms of Figure 7.1, to restore equilibrium in the bond market after a price-level increase from P_0 to P_1 requires that the domestic interest rate rise from R_0 to R_1, but the domestic interest rate must rise to R_2 to restore equilibrium in the money market. It follows that the BB curve must pass through points A and B, while the MM curve must pass through points A and C. The implication is that the MM curve must be steeper than the BB curve.

What happens to the relative slopes of the two curves as capital mobility increases? Notice that the slope of the MM curve depends only on the properties of the demand for money function $PL(R, Y(P))$ and therefore does not depend on the degree of capital mobility. It is easy to see that when there is no capital mobility (i.e., under financial autarky), the MM and BB curves must coincide because any effect of a change in the domestic price level on the demand for money must be exactly offset by its effect on the demand for domestic bonds, and the same must be true for changes in domestic interest rates. Foreign bonds and foreign demand for domestic bonds would play no role in this case. However, as shown in the preceding chapter, as the degree of capital mobility increases, the BB curve becomes flatter. Thus the difference between the slopes of the MM and BB curve becomes larger as economies become more financially integrated with the rest of the world.

3. Behavior of B_C along the MM Curve

Recall from the preceding chapter that the stock of money increases as we move up and to the right along the BB curve. What happens to the stock of domestic credit B_C as we move along the MM curve? It is easy to see that B_C must be *decreasing* as we move up and to the right along the MM curve; that is, the central bank must be *selling* bonds as the economy moves up and to the right along the MM curve. To see why, compare the point A with the point C in Figure 7.1. Notice that when the price level increases from P_0 to P_1, an increase in the domestic interest rate from R_0 to R_1 keeps the domestic bond market in equilibrium with an *unchanged* value of B_C. Thus, with a further interest rate increase, say, to R_2, there must be a larger demand for domestic bonds than can be met by the existing supply. It follows that to keep the bond market in equilibrium, the central bank has to sell to the private sector the additional bonds that the private sector wishes to hold; that is, B_C must be smaller at C than at A.

4. Position of the MM Curve

Next, consider what determines the position of the MM curve. Recall that the MM curve was derived for a given value of the money supply. What would have happened to the MM curve derived in Figure 7.1 if we had chosen to hold constant a *larger* value of the money supply than the one we did? Because of the negative effect of the domestic interest rate on the demand for money, to make people willing to hold more money at a given value of the domestic price level, we would require a *lower* value of the domestic interest rate. Consequently, an increase in the money supply would cause the MM curve to shift downward. Conversely, a lower value of the money supply would cause it to shift upward.

5. Effects of a Monetary Expansion on BB and MM

Finally, consider what happens to the BB and MM curves in response to a monetary expansion. Because the central bank implements a monetary expansion by buying domestic bonds with newly created domestic money, this means that a monetary expansion causes B_C and M to increase by exactly the same amount. As we saw in the preceding chapter, the increase in B_C shifts the BB curve down and to the right, while the last paragraph showed that the increase in M shifts the MM curve down (and therefore to the right) as well. Can we compare the magnitude of the two shifts?

The answer is yes. Because a monetary expansion creates exactly equal increases in B_C and M, the expansion creates an excess demand for domestic bonds that is exactly offset by an excess supply of domestic money. However, as we have seen, changes in the domestic interest rates have more powerful effects on the domestic

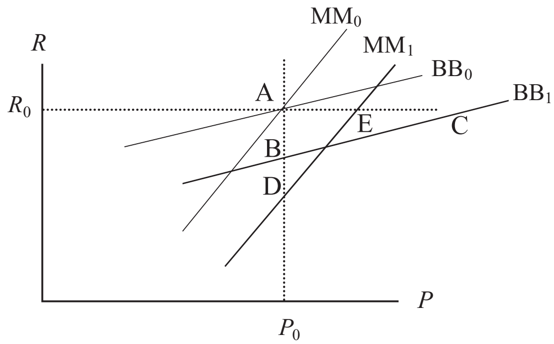


Figure 7.2. Effects of a monetary expansion on BB and MM

bond market than they do on the money market, whereas changes in the price level have more powerful effects on the money market than on the bond market. Therefore, at a given value of the domestic price level, the reduction in the domestic interest rate that would be required to induce people to hold the larger supply of domestic money created by the monetary expansion must be *larger* than that required to induce people to be satisfied to hold the smaller stock of domestic bonds that the central bank leaves available to them when it buys bonds from the market. The implication is that the *downward* shift in MM must be larger than that in BB. On the other hand, because the money market is more sensitive than the bond market to changes in the domestic price level, at a given value of the domestic interest rate, the increase in the domestic price level that would be required to induce people to hold the larger supply of domestic money created by the expansion must be *smaller* than that required to induce them to be satisfied to hold the implied smaller stock of domestic bonds. The implication in this case is that the *rightward* shift in MM must be *smaller* than that in BB.

The situation is therefore as in Figure 7.2. A monetary expansion causes MM to shift from a position such as MM$_0$ to one such as MM$_1$, whereas it causes BB to shift from BB$_0$ to BB$_1$. Notice that the downward shift in MM, to pass through a point such as D, must be larger than in BB, to pass through a point such as B. At the same time, the rightward shift in MM, to the point E, must be smaller than that in BB, to the point C.

6. The RR Curve

We conclude this section by introducing the RR curve. This is simply a curve along which the domestic interest rate is constant (that is the sense in which it is an iso-interest curve). Because the domestic interest rate is measured on the vertical axis in (P, R) space, this curve is just a horizontal straight line. It is introduced here to capture the possibility that the central bank may choose to target the domestic interest rate, as in Chapter 5. The height of the curve is determined by the interest

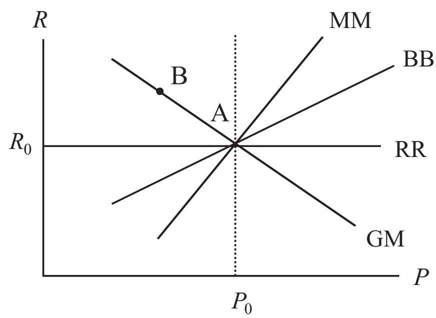


Figure 7.3. Short-run equilibrium

rate target set by the central bank. It is important to note for future reference that this target is a policy choice and therefore need not be equal to the foreign interest rate.

II. SHORT-RUN MACROECONOMIC EQUILIBRIUM

We can now assemble all the parts of our model to describe the economy's *short-run equilibrium.* Putting the GM, BB, MM, and RR curves together, as in Figure 7.3, the economy's short-run equilibrium corresponds to the point where the four curves intersect because this point corresponds to a situation in which equilibrium prevails simultaneously in both the domestic labor and goods markets, on one hand, and in the domestic money and bond markets, on the other, while being consistent with the central bank's interest rate target, if it pursues such a target. This point is labeled A in Figure 7.3. It determines the short-run equilibrium values of the domestic price level and the domestic interest rate, given by P_0 and R_0, respectively, in Figure 7.3. This equilibrium is short run in the sense that it holds for given values not only of the capital stock and technology but also of the country's international investment position (IIP^*) and stock of government bonds outstanding (B) because all of these slowly evolving variables are held constant in deriving the equilibrium conditions on which the curves are based.

How this equilibrium is reached depends on the monetary policy regime adopted by the central bank. If the central bank follows a domestic credit rule that keeps the stock of domestic credit constant (recall that in Chapter 5, we referred to this as *domestic credit targeting*), the equilibrium must be determined by the intersection of the GM and BB curves because the constancy of B_C implies that there is nothing to move the BB curve. Suppose that the MM curve does not initially pass through the point of intersection of GM and BB. For example, suppose that the MM curve passes through a point like B in Figure 7.3. At the point B, the money market would be in equilibrium (because the economy would be on MM if the curve passed through that point), but there would be an excess demand for domestic bonds (because B is above BB). Consequently, there must be an excess *supply* of foreign bonds at B. Domestic residents can immediately eliminate this excess supply by selling these bonds on the international financial market. The proceeds of those sales must be purchased by the domestic central bank to keep the exchange rate fixed, which increases the domestic money supply. This increase in the money supply shifts MM to the right. This process continues until the MM curve passes through the point A. Capital inflows would be replaced by outflows and the MM curve would shift to the *left*, if it initially passes instead through a point on the GM curve that lies *below* the BB curve. In both cases, the effects of capital inflows or outflows on the money supply cause the MM curve to adjust passively until it passes though the point where GM and BB intersect.

If the central bank instead follows a money supply rule that keeps the money supply constant (*money supply targeting*), the equilibrium is determined by the intersection of the GM and MM curves, with the BB curve adjusting passively to that point. In this case, adjustment occurs through central-bank purchases or sales of government bonds designed to offset the effects of capital inflows or outflows on the money supply. If the BB curve intersects the GM curve below the MM curve, the central bank will be led to sterilize the capital inflows described in the last paragraph, increasing the supply of domestic bonds in the hands of the private sector and causing the BB curve to shift upward. If, instead, the intersection of BB with GM is above the MM curve, the bank will have to sterilize capital *outflows* by buying back domestic bonds, and this will cause the BB curve to shift downward.

Finally, if the bank follows an interest rate rule that keeps the domestic interest rate constant (*interest rate targeting*), then it must adjust B_C so as to cause the BB curve to intersect the GM curve at the bank's desired value of the domestic interest rate on the RR curve, with the MM curve adjusting passively to this point through capital inflows or outflows, described earlier.

To see how the model works, we can now trace out the effects on the economy of a variety of policy and exogenous shocks. The initial equilibrium at A, assumed to be a full-employment equilibrium, will represent our point of departure. We will consider aggregate demand and aggregate supply shocks in the two sections that follow.

III. EFFECTS OF AGGREGATE DEMAND SHOCKS

In this section, we will analyze the effects of aggregate demand shocks in the form both of domestic policy shocks (changes in domestic monetary, fiscal, and exchange rate policies) and external shocks (both foreign financial and real shocks).

1. Domestic Policy Shocks

a. Monetary Policy

Consider first how monetary policy works in this economy. A change in monetary policy is implemented by changing the central bank's stock of domestic credit B_C. For concreteness, we will examine the effects of a credit expansion, that is,

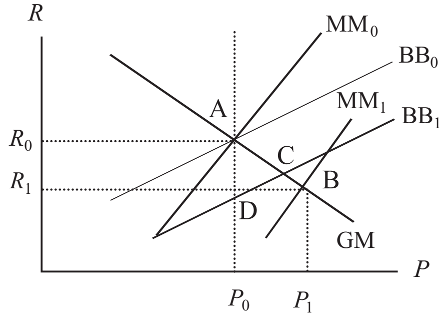


Figure 7.4. Effects of a monetary expansion

a one-time central-bank purchase of government bonds that increases its stock of domestic credit. As we saw in the preceding chapter, such a credit expansion shifts the bond supply curve to the left by the amount of the increase in the stock of credit, leaving the bond demand curve unchanged. Restoring domestic bond-market equilibrium after such a policy would therefore require a *lower* equilibrium value of the domestic interest rate, which would reduce the demand for domestic bonds by the private sector, matching their reduced supply. Because this analysis holds for any given value of the domestic price level (recall that the price level is held constant in deriving the bond demand curve), the BB curve must shift downward, as in Figure 7.4. At the initial price level, this shift is from the point A to D. As we saw in the last section, the credit expansion also increases the money supply, shifting the MM curve downward as well.

Determining the position of the new equilibrium is complicated by the fact that this policy is likely to trigger capital flows that would induce further shifts in the BB and/or MM curves. To see where the new short-run equilibrium must be, therefore, it is useful to consider first where the economy would settle in the *absence* of capital flows. How can we tell? We can use the financial autarky equilibrium as a benchmark because under financial autarky, there would indeed be no capital flows. As we have seen, under financial autarky, the MM and BB curves must coincide. Therefore, in the absence of capital flows, the economy would find a new short-run equilibrium at the point where the new MM curve intersects the GM curve – that is, at the point B in Figure 7.4 – with a higher domestic price level and lower interest rate.

What types of capital flows would be induced from this position? The answer is not obvious because at the point B, there are two offsetting effects on capital flows: a higher domestic price level induces domestic residents to want to hold more money, which causes them to repatriate capital from abroad, inducing a capital *inflow*. At the same time, however, the lower domestic interest rate causes both domestic and foreign residents to increase their demand for foreign bonds, which causes a capital *outflow* from the domestic economy. As the figure is drawn, the credit expansion must trigger a net capital outflow because the postexpansion MM curve (labeled MM_1 in the figure) intersects the GM curve below and to the right of the intersection of the postexpansion BB curve (BB_1) with the GM curve. But this outcome is not necessary. As you will see, it depends on certain characteristics of the economy – specifically, the shape of the GM curve and the degree of capital mobility.

Consider first the role of the shape of the GM curve. As the GM curve becomes flatter, it rotates around the point A in the figure. As you can verify from the figure, a relatively flat GM curve must intersect the BB_1 curve to the right of the MM_1 curve, whereas with a relatively steep GM curve, that intersection must happen to the left of the MM_1 curve. In the first case, the monetary expansion triggers a capital inflow, whereas in the second, it triggers an outflow; that is, when the GM curve is steep, it is more likely that a domestic credit expansion will result in capital outflows and reserve losses, whereas the opposite is true when the GM curve is shallow. Why the difference? The steeper the GM curve, the greater the extent to which the new short-run equilibrium features a reduction in the domestic interest rate (which induces capital outflows), and the less the extent to which it features an increase in the domestic price level (which induces capital inflows).

Next, consider the role of capital mobility. As shown in Section I, the degree of capital mobility has no effect on the shape of the MM curve or on the magnitude of its shift in response to a change in B_C. Importantly, it also has no effect on the magnitude of the *rightward* shift in the BB curve caused by the credit expansion. The reason is that the magnitude of that shift is determined by the size of the increase in the price level required to restore equilibrium in the domestic bond market *at a given domestic interest rate*. However, the degree of capital mobility does affect the slope of the BB curve, as we have seen, with a greater degree of capital mobility making the BB curve flatter. Now go back to Figure 7.4 and consider what happens as the degree of capital mobility *decreases*. As the degree of capital mobility decreases, the BB curve becomes steeper. The increased steepness of the BB_1 curve would cause it to intersect the GM curve further below and to the right than it does in Figure 7.4. For a sufficiently low degree of capital mobility, that intersection must occur to the southeast of the intersection of MM_1 and GM. In that case, the domestic credit expansion would induce a capital *inflow* rather than outflow. Similarly, as the degree of capital mobility *increases* relative to that assumed in Figure 7.4, the BB curve would become flatter, and its intersection with GM would move up and to the left along GM, implying even larger net capital outflows. Indeed, with perfect capital mobility, capital outflows would completely offset the monetary effects of the credit expansion, leaving the domestic money supply unchanged.

Now that we have examined the implications of the credit expansion for capital flows, we can identify the location of the new short-run equilibrium. That location will be determined by how the central bank responds to the capital flows induced by its expansionary monetary policy. Consider again the specific case illustrated in Figure 7.4, where the credit expansion triggers capital outflows. If the central bank is engaging in *domestic credit targeting*, its expansion of domestic credit will be a one-time event, and once the BB curve has shifted, it will remain stationary, putting the economy's new short-run equilibrium at the point C. In this case, the capital outflows caused by the credit expansion will cause the money supply to contract, and the MM curve will therefore shift to the left, until it, too, passes through the point C. However, if the central bank is engaging in either *monetary* or *interest rate* targeting, then it must have increased domestic credit because it intended for the new domestic interest rate to be R_1 and the new money supply to be that associated with the new MM curve labeled MM_1. In that case, to keep the economy on MM_1, it would have to sterilize the monetary effects of the capital outflows caused by its credit expansion to prevent the money supply from contracting endogenously. It can do so by engaging in a further expansion of domestic credit that causes the BB curve to shift downward until it passes through the point B. The upshot is that the impact of a particular monetary policy action on the economy depends on the monetary policy regime in which it is embedded because that regime determines the central bank's reaction to the capital flows induced by its actions.

To examine the other effects of the monetary expansion on the economy, let us assume for concreteness that the central bank is engaged in domestic credit targeting, so the new short-run equilibrium is at the point C. We can now trace out the effects of this shock on a variety of macroeconomic variables, including real output, labor market variables, and external variables.

Turning to real output first, note that the reduction in the domestic interest rate increases private absorption and thus the aggregate demand for domestic goods. The goods market clears despite this increase in aggregate demand because the higher domestic price level induces an increase in aggregate supply and causes a real exchange rate appreciation that switches demand from domestic to foreign goods.

In the labor market, because the level of real output increases despite unchanged total factor productivity and an unchanged capital stock, the level of employment must expand. To accommodate the increase in employment, the real wage must fall, inducing firms to hire more workers. For this to happen, the labor demand curve must have shifted upward more than the labor supply curve, and thus the nominal wage must rise less than in proportion to the increase in the price level.

Regarding the economy's external variables, because the nominal exchange rate is fixed, the higher equilibrium price level means that the real exchange rate must appreciate. Exports must therefore fall, and because domestic output is higher and the domestic interest rate is lower, both expenditure-increasing and expenditure-switching effects imply that imports must rise, so *net* exports must fall.

Finally, consider the effects of this policy on the central bank's foreign exchange reserves. Whether reserves rise or fall is not obvious simply by looking at the central bank's balance sheet because even though the lower domestic interest rate and higher price level both unambiguously increase the demand for money, it is not

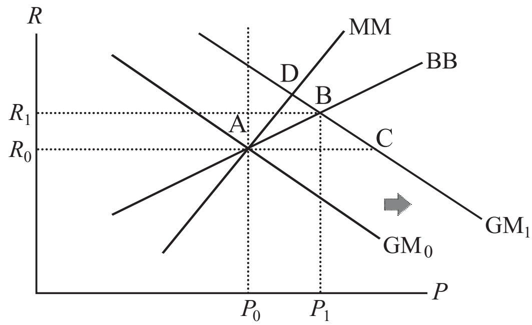


Figure 7.5. Expansionary fiscal policy

clear whether this increase in money demand is larger or smaller than the increase in the domestic component of the money supply caused by the credit expansion, thereby causing the central bank's foreign exchange reserves to increase or decrease. But as we have seen, in the case illustrated in Figure 7.4, the credit expansion triggers a capital outflow, so the central bank must have lost foreign exchange reserves. Looking at the central bank's balance sheet, its monetary liabilities have increased, with an increase in its holdings of domestic bonds more than compensating for a loss of foreign exchange reserves.

b. Fiscal Policy

Next, consider the effects of fiscal policy. As we have seen, this can take the form of a change in government spending or in taxation. For direct comparison with the case of domestic credit expansion, let us examine the case of an increase in government spending.

An increase in government spending affects only the goods market directly. Because an increase in government spending represents an increase in demand for domestic goods, this shock creates an incipient excess demand for domestic goods. This excess demand can be eliminated, and goods-market equilibrium restored, through some combination of higher domestic prices (to increase production and shift demand away from domestic toward foreign goods) and higher domestic interest rates (to reduce total spending by domestic residents, including spending on domestic goods). Because maintaining goods-market equilibrium after the increase in government spending requires a higher level of domestic prices and/or a higher domestic interest rate than it did before the shock, the shock must shift the GM curve to the right, say, from GM_0 to GM_1 in Figure 7.5. At the original price level, a higher domestic interest rate is required to maintain goods-market equilibrium, and at the original interest rate, a higher price level is required.

The location of the new short-run equilibrium once again depends on the monetary policy regime that the central bank has adopted. We can once again use the reference point given by the intersection of GM and MM, labeled D in the figure, because this is the equilibrium that would correspond to financial autarky and an unchanged stock of domestic credit. Notice that because the point D lies above BB, the economy would experience net capital inflows at D. If the central bank pursues monetary targeting, it would sterilize these flows, leaving the economy at a new short-run equilibrium at D. If it pursues domestic credit targeting, it would leave its stock of domestic credit unchanged, and the capital inflows would induce an increase in the money supply that would shift MM downward until it passes through the point B, where GM_1 and BB intersect. If it pursues interest rate targeting, as in Chapter 4, it would need to supplement the monetary effect of these capital inflows with a domestic credit expansion that would shift both MM and BB downward until they pass through the intersection of GM_1 and RR at C.

For concreteness, let us once again consider the domestic credit targeting case. As in the case of the monetary expansion, we can again say quite a bit about the domestic macroeconomic effects of this policy. The higher domestic price level at B must induce an increase in real gross domestic product (GDP). As in the case of the credit expansion, employment rises, the real wage falls, and the equilibrium value of the nominal wage must increase. Regarding the country's external accounts, because of the higher domestic price level, the real exchange rate must appreciate. Moreover, because the economy moves up and to the right along the BB curve, the domestic money supply must have increased at B. Because domestic credit is unchanged, this must be the result of reserve accumulation caused by capital inflows, as we have seen.

What happens to the country's net exports in this case? To answer this question, based on Chapter 5, we can write net exports as

$$\begin{aligned} NX &= X - Q \cdot IM \\ &= X\,(\underset{+}{SP^*/P}) - [1 - \varphi(\underset{+}{SP^*/P})]A_P[\underset{+}{Y(P) - T},\ \underset{-}{R})] \end{aligned}$$

The appreciation of the real exchange rate and increase in domestic real income associated with the increase in the domestic price level together tend to cause the trade balance to deteriorate, but the increase in the domestic interest rate causes it to improve by reducing private absorption and thus the demand for imports. In principle, the net effect on net exports is ambiguous. Net exports will fall if the composition of domestic and foreign demand for goods is very sensitive to the real exchange rate, and if domestic private absorption is *not* very sensitive to the interest rate. They will rise if these conditions are reversed.[2]

It is interesting to note in passing how these results would be affected by alternative monetary policy regimes, that is, interest rate or monetary targeting. As shown

[2] The absorption approach does not help us pin down the effect on net exports, because though output must increase relative to private absorption, the increase in government spending leaves the overall effect on net exports indeterminate.

earlier, under interest rate targeting, the central bank would have to expand the supply of domestic credit, causing the BB curve to shift to the right until it intersects the new GM curve at point C in Figure 7.5. This magnifies the effects of the expansionary fiscal policy on the price level and on real output and employment, as well as the appreciation of the real exchange rate. As we saw in Chapter 5, net exports must fall in this case because there is no increase in the domestic interest rate to moderate the increase in private absorption. If the central bank pursues monetary targeting, on the other hand, it must *contract* domestic credit to sterilize the monetary effects of the capital inflows induced by the fiscal expansion, causing the BB curve to shift upward until it passes through the point D where the original MM curve intersects the new GM curve (because at this point, the domestic money supply would be unchanged). This policy moderates the effects of the fiscal expansion on the domestic price level, while magnifying its effects on the interest rate. Consequently, net exports are more likely to increase, though the direction of change in net exports remains ambiguous.

c. Nominal Devaluation

Consider next the effects on the economy of a nominal exchange-rate devaluation that is not monetized by the central bank. It is worth noting that this shock could come about in either of two forms: it could represent an explicit policy decision by the central bank to change the value of the domestic currency against the **anchor currency** (the currency to which the domestic currency is pegged) or it could represent an exogenous shock, if the official exchange rate against the anchor currency remains unchanged but the anchor currency depreciates against the currencies of other countries with which the domestic economy trades. The latter is not an uncommon event for countries that peg their exchange rates against a single currency.

To simplify matters, assume for now that in the initial equilibrium, domestic and foreign residents do not happen to hold each others' bonds. In this case, as we saw in the preceding chapter, the devaluation would have no effect on the domestic bond market because there would be no valuation effect causing foreign residents to increase their demand for domestic bonds. As a result, the only effect of the devaluation would be to shift the GM curve to the right somewhat *less* than in proportion to the change in the exchange rate. The reason for the rightward shift is that the devaluation increases demand for domestic goods by reducing their relative price. The reason that the shift is less than proportional to the size of the devaluation is that, as we saw in Chapter 5, a devaluation causes the AD curve to shift vertically upward in proportion to the change in the exchange rate, but because the AS curve is stationary, the equilibrium price level in the goods market increases less than in proportion to the change in the exchange rate. Because this change in the equilibrium price level in the goods market is what determines the magnitude of the rightward shift in GM, the result follows.

Because the only effect of the devaluation in this case is to shift the GM curve to the right, the new short-run equilibrium in the economy would look very similar to that in Figure 7.5. Assuming that the central bank pursues domestic credit targeting, the equilibrium would be at a point like B, with a higher domestic price level and higher domestic interest rate.

Because the increase in the domestic price level is smaller than the rightward shift in the GM curve, which itself is smaller in magnitude than the change in the exchange rate, the proportional increase in the domestic price level must be smaller than that in the nominal exchange rate, which means that the real exchange rate must depreciate – albeit, because of the price offset, by less than the nominal devaluation. Moreover, the increase in the domestic interest rate will tend to curtail private absorption and, with it, the demand for imports. Both these factors are conducive to an improvement in net exports. On the other hand, the increase in the domestic price level tends to increase real output, which, through its effects on the incomes of domestic residents, increases domestic absorption and, with it, the demand for imports, inducing net exports to deteriorate. Though the overall effect on net exports would therefore appear to be ambiguous, we can use the absorption approach to show that they must actually *increase*. From the goods-market equilibrium condition,

$$Y = Ap + G + (X - Q \cdot IM)$$

you can see that because Y has risen and the marginal propensity to spend is less than unity, the increase in A_p induced by the higher level of Y must be smaller than the increase in Y itself. Moreover, the increase in the domestic interest rate caused by the devaluation would have the effect of further curtailing private spending. It follows that the goods market can only be in equilibrium if net exports $(X - Q \cdot IM)$ increase.

Finally, because the domestic interest rate and the domestic price level both increase, the devaluation triggers a capital inflow that increases the stock of foreign exchange reserves immediately. This capital inflow increases both the central bank's foreign exchange reserves and the domestic money supply. This is easy to see from Figure 7.5 because the new equilibrium point at B lies below the original MM curve, implying that the devaluation must cause the domestic money supply to increase, which can happen only through a capital inflow.

Things are a little different if domestic and foreign residents *do* initially hold each others' bonds, because in this case, as shown in the preceding chapter, a nominal devaluation will also cause the BB curve to shift down and to the right as the demand for domestic bonds increases. The magnitude of the rightward shift in BB may be larger or smaller than that in GM. The magnitude of the shift in BB depends on the size of valuation effects and will be larger the larger the initial size of cross-border bond holdings. The implication is that the change in the domestic interest rate caused by a devaluation is ambiguous under domestic credit targeting.

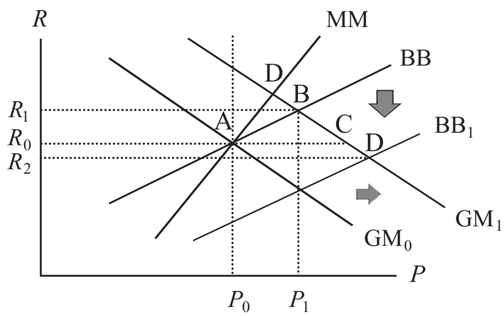


Figure 7.6. Devaluation with foreign bond holdings

For example, in Figure 7.6, the rightward shift in BB is shown larger than that in GM, corresponding to a situation where valuation effects are very large and an exchange rate devaluation has important effects on the economy's financial markets. In this case, the increase in the domestic price level is larger, but the equilibrium value of the domestic interest rate actually *falls*, under domestic credit targeting. The effect of the devaluation may be very different in this case from what we found in Chapter 5. For example, notice that because the reduction in the domestic interest rate provides an additional stimulus to aggregate demand, the increase in the price level is larger in this case than if the central bank had kept the domestic interest rate constant. Indeed, there is no reason why the proportional change in the domestic price level could not exceed that in the exchange rate, causing the real exchange rate actually to *appreciate*. If that is so, net exports would actually have to *fall* in this case because an appreciated real exchange rate, a higher level of real output, and a lower domestic interest rate would all contribute to that result.

Because monetization increases the stock of money in the hands of the private sector, while decreasing its stock of domestic bonds, it would tend to magnify the downward shift of the BB curve. The BB curve would shift down and to the right in this case, even if domestic and foreign residents do not initially hold each others' bonds. Relative to what happens when the devaluation is not monetized, there will be a smaller increase in the domestic interest rate but a larger one in the domestic price level. However, because monetization causes the money supply to increase less than in proportion to the nominal devaluation, as long as the stock of domestic credit is not increased or is increased less than in proportion to the rate of devaluation, the real exchange rate must depreciate. This is so because the increase in the domestic price level must be proportionately smaller than the change in the exchange rate.[3] When domestic and foreign residents do initially hold each others' bonds, monetization simply magnifies the effects discussed in the last paragraph, making real appreciation and net export deterioration more likely.

[3] The analysis of the effects of interest rate and monetary targeting in this case is similar to that in the case of a fiscal expansion.

2. External Shocks

Next we can turn to the effects on the domestic economy of external shocks. These can be *financial shocks*, captured in our model in the form of changes in the world interest rate, or *real* shocks, in the form of exogenous changes in demand for the country's exports. We will begin by examining the effects of financial shocks and then will turn to a shock that actually represents a combination of the two.

a. Changes in the Foreign Interest Rate

Suppose, then, that the relevant shock to the economy consists of a decrease in R^* (it is useful to consider a decrease instead of an increase to replicate the capital-inflow phenomenon that several developing countries experienced during the first part of the 1990s and again in the mid-2000s, a subject that we will take up in Chapter 24). This shock has no direct effect on the GM curve, but the BB curve shifts downward, as domestic residents repatriate capital and foreign residents shift funds into the domestic economy. Under domestic credit targeting, the economy thus moves from A to B in Figure 7.7. In the new short-run equilibrium, the domestic interest rate falls, and the price level rises. Output and employment both increase, the equilibrium real wage falls, and the equilibrium nominal wage increases. Because of the higher domestic price level, the real exchange rate appreciates. Combined with the increase in output and reduction in the domestic interest rate, this means that net exports must fall. The lower domestic interest rate and higher price level both mean that the demand for money increases, so the money supply must increase as the result of capital inflows, and the central bank's foreign exchange reserves must increase by the same amount.[4,5]

How large is the drop in the domestic interest rate in this case? The answer is that it must be smaller than that in the foreign interest rate as long as capital mobility is imperfect. To see why, recall from Chapter 6 that at a given domestic price level, the reduction in the domestic interest rate consistent with equilibrium in domestic

[4] Notice that if the central bank engages in monetary targeting (sterilizes the capital inflows induced by the lower foreign interest rate), the BB curve shifts back to intersect the MM_0 curve at A, so the shock would have no effect on the domestic short-run macroeconomic equilibrium, except to change the composition of the central bank's balance sheet: SF_C^* increases and B_C contracts by equal amounts.

[5] What difference does it make whether the response to a reduction in R^* is sterilization or capital controls? The difference shows up in the central bank's balance sheet. As noted in the previous footnote, if the fixed value of M arises from sterilization rather than capital controls, there will have to be a credit contraction that offsets capital inflows on a one-for-one basis to sustain the original value of the money supply, whereas if it arises as the result of the imposition of financial autarky through capital controls, the composition of the central bank's balance sheet would not change. For future reference, it is worth noting that because domestic bonds typically pay higher interest than do reserves, the central bank's interest income falls under sterilization but not under capital controls.

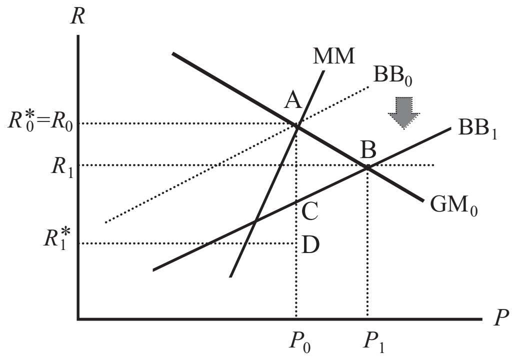


Figure 7.7. Effects of a reduction in the foreign interest rate

financial markets must be smaller than that in the foreign interest rate. This means that the downward shift in the BB curve must be smaller than the reduction in the foreign interest rate, as in Figure 7.7, where the foreign interest rate drops from R_0^* to R_1^*, whereas the interest rate required to clear the domestic bond market falls only from A to C. Both because the reduction in the domestic interest rate required to clear the domestic bond market must be less than of the foreign rate (i.e., because the point C is above D) and because, to clear the goods market, further increases in domestic interest rates and prices are required from the point C (at which the domestic money market clears with unchanged prices), the final reduction in the domestic interest rate must be smaller than that in the foreign rate.

b. Exogenous Changes in the Demand for Exports

As our final shock, consider the effects on the domestic economy of an exogenous increase in demand for its exports. To make things a little more interesting, let us assume that the change is caused by an increase in partner-country incomes. The reason this makes things more interesting is that it combines a shock to the export demand function $X(SP^*/P, \ldots)$ with a change in the rest of the world's demand for money as the result of the increase in the rest of the world's income. Thus this shock directly affects both the domestic goods and money markets.

The increase in demand for exports from the domestic economy represents an increase in aggregate demand for domestic goods and thus causes the GM curve to shift to the right, as in Figure 7.8. At the same time, if this increase in export demand is caused by an increase in partner-country incomes, their higher incomes cause foreign residents to repatriate capital to satisfy their increased demand for their own currency. The ensuing capital outflow shifts the BB curve vertically upward because the capital outflow and associated monetary contraction caused by the reduced demand for domestic bonds by the rest of the world requires an increase in the domestic interest rate to maintain equilibrium in the domestic bond market.

This shock unambiguously causes the domestic interest rate to rise, but its effect on the domestic price level (and thus on the economy's real exchange rate)

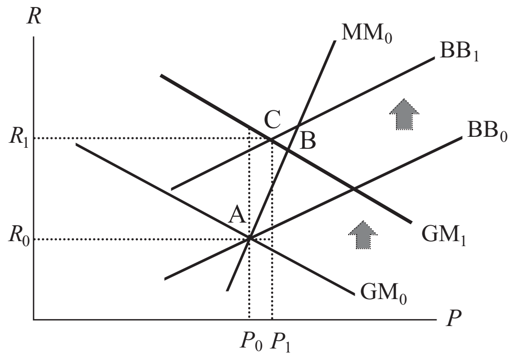


Figure 7.8. Effects of an increase in partner-country incomes

depends on the central bank's monetary policy. If the central bank does not sterilize, the domestic price level may rise or fall. The outcome depends on the relative magnitudes of the vertical shifts in the GM and BB curves. If the economy is very open commercially, and if domestic aggregate demand is not very sensitive to interest rate changes, the vertical shift in the GM curve will be large, and an expansionary outcome is more likely. It is also more likely if the domestic economy is financially relatively closed or if domestic assets represent a very small share of the financial portfolios of private agents in partner countries. Under these conditions, the vertical shift in BB will be small, and the shock will be expansionary, as in Figure 7.8. On the other hand, if the domestic central bank either targets the domestic interest rate or sterilizes the capital outflows associated with capital repatriation by foreigners, domestic prices and real economic activity will increase, as at the point B or C.

IV. EFFECTS OF AGGREGATE SUPPLY SHOCKS

The three shocks we have considered so far have affected the economy either directly or indirectly through the demand side of the goods market. Consider next how the economy's short-run equilibrium would respond to a supply shock. For concreteness, we can take the example of a discrete increase in the parameter A that measures total factor productivity in the aggregate production function. In the real world, this could correspond, for example, to an exceptionally favorable harvest or an important technological innovation. Its effect is to increase the level of real output Y corresponding to any given value of the domestic price level.

Because supply shocks affect the economy through their direct effects on the level of real output, they affect both the financial and goods markets. The increase in domestic output creates an excess supply of domestic goods, requiring a reduction in their relative price (a decrease in P) to restore goods-market equilibrium. Thus the GM curve shifts to the left in Figure 7.9. At the same time, the increase in domestic income increases the demand for money and reduces the demand for

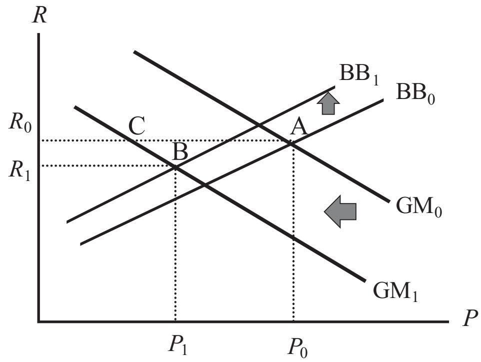


Figure 7.9. Short-run effects of a favorable aggregate supply shock

bonds, requiring an increase in the domestic interest rate to restore equilibrium in the domestic bond market, thus causing the BB and MM curves to shift upward (to keep the graph from becoming too cluttered, the MM curve is not shown in Figure 7.9).

As shown in Figure 7.9, a favorable supply shock must unambiguously lower the domestic price level, thus depreciating the domestic economy's real exchange rate. While the figure shows the domestic interest rate falling, this is not a necessary outcome. If the shift in the GM curve is smaller (i.e., if domestic and foreign goods are close substitutes), the GM curve could intersect the BB curve above the original domestic interest rate, thus causing the domestic interest rate to be higher in the new equilibrium.[6]

As shown in Chapter 5, though the price level is lower at the point C that marks the magnitude of the leftward shift of the GM curve, real output would actually be higher if the domestic interest rate were kept there by the central bank (e.g., under interest rate targeting), driven by the improvement in productivity. Would this still be true once the domestic interest rate was allowed to adjust? The answer must be yes if the equilibrium value of the domestic interest rate would actually tend to fall, as in Figure 7.9, because in that case, the reduction in the price level is smaller than it would be under interest rate targeting. But what happens if the upward shift in BB is larger so that it intersects GM above the point C, making the equilibrium interest rate rise and magnifying the drop in the price level? The answer is that real output must still increase. The reason is that if it did not, the domestic bond market could not be in equilibrium. With a lower price level and lower level of real output, the demand for money would fall, and the demand for domestic bonds would therefore rise. The higher interest rate would simply magnify the increase in the demand for domestic bonds. With a fixed supply of such bonds, the bond market could not

[6] It is also possible that the interest rate outcome could differ depending on the policies pursued by the central bank, with the domestic interest rate falling under full sterilization and rising if the bank does not sterilize.

Table 7.1. *Macroeconomic Volatility in Selected Country Groups*

	Standard Deviation (1981–2006)		
Country Group	**GDP Growth**	**Inflation**	**CA/GDP**
Advanced economies	1.10	2.38	0.45
Emerging and developing economies	1.69	34.41	1.86
Africa	2.03	10.30	2.16
Sub-Sahara	2.34	12.12	1.63
Central and eastern Europe	3.10	28.98	1.95
Commonwealth of Independent States[(1)]	6.51	375.04	4.19
Developing Asia	1.53	3.83	2.34
Middle East	2.50	4.49	8.07
Western hemisphere	2.13	129.36	1.87

[(1)] Mongolia is not a member of the Commonwealth of Independent States but is included in the group because of its geographic proximity
Source: International Monetary Fund, World Economic Outlook Database, April 2008

be in equilibrium under those circumstances. Thus the conclusions we reached in Chapter 5 remain unchanged.

V. AN APPLICATION: MACROECONOMIC INSTABILITY IN EMERGING ECONOMIES

As we have seen, the shocks we have analyzed in this chapter are capable of affecting a broad range of macroeconomic variables in emerging and developing economies. But how important are such shocks in the real world? To get a handle on this question, Table 7.1 compares macroeconomic volatility in emerging and developing economies to that in advanced economies. As shown in the table, the macroeconomic environment indeed tends to be relatively volatile in emerging and developing economies by comparison with that in advanced economies. Over the period 1981–2006, for example, the standard deviation of *GDP* growth, inflation, and the ratio of the current account surplus to *GDP* was substantially higher in emerging and developing economies as a group than in advanced economies. Although different geographic groups of emerging and developing countries exhibited varying degrees of instability in the three indicators examined in the table, every regional group had higher volatility than the advanced economies for every indicator.

The model we have been exploring in this chapter can be used to identify the sources of some of this volatility. It is useful to classify them into several categories, as follows.[7]

[7] Calderon et al. (2005) explore how vulnerability to the types of shocks examined in this section is affected by real and financial openness in emerging and developing economies.

1. Domestic Policy Shocks

In industrial-country macroeconomics, the potential stabilizing roles of fiscal and monetary policies are usually emphasized. In principle, at least, policy makers can make use of these policies to offset sources of instability arising elsewhere in the economy. Though it is often debated whether such policies can be effective in playing this stabilizing role, little attention is usually given to the possibility that domestic policy shocks, in the form of changes in fiscal and monetary policies (i.e., G, T, and B_C in our model), may themselves represent an important source of shocks. In many emerging economies, they have typically done just that, as we will see in Parts 3 and 4 of this book. Political instability, changing political coalitions, and central banks with little policy independence have often resulted in frequently changing fiscal and monetary policies, with attendant effects on instability of key macroeconomic relative prices that have disrupted the roles of such prices in the efficient allocation of resources.

2. Changes in the International Environment

As mentioned in Chapter 1, many developing economies are small, highly open to international trade, and relatively specialized in the composition of their exports. Consequently, changes in the prices of their primary export commodities, and thus in their export revenues (captured in our model through shocks in the export demand function), have large impacts on their domestic economies. High volatility in individual commodity prices thus tends to translate into high volatility of domestic macroeconomic relative prices.

Moreover, as also discussed in Chapter 1, emerging economies have been characterized by increasing financial openness, and this has also left them exposed to volatility in international financial markets. As we will discuss further in Part 6 of the book, this has manifested itself not only in the form of fluctuations in international interest rates that reflect monetary policies adopted in the industrial countries for their own domestic stabilization reasons but, perhaps even more important, in the form of fluctuations in the interest rate premia paid by specific countries over the international risk-free rate. These premia have proven to be very volatile, sometimes driven by sharp responses to domestic macroeconomic events but at other times by events happening elsewhere (a phenomenon known as *contagion*). In the context of our model, fluctuations in the interest rate at which the domestic economy can access the international financial market show up in the form of changes in R^*.

3. Supply Shocks

Even emerging economies with exports concentrated in manufactured goods tend to have large agricultural sectors, and primary production continues to loom relatively large in these economies. Consequently, weather-related supply shocks can still represent an important source of macroeconomic instability in low-income countries.[8]

4. Policy Responses to Exogenous Shocks

Finally, even in an unstable external and natural environment, macroeconomic instability can be reduced by implementing stabilizing domestic policy responses. However, in many emerging economies, not only do policies themselves represent an important source of shocks, as we have seen, but rather than play a stabilizing role in response to exogenous shocks, they often serve to *aggravate* the macroeconomic consequences of such shocks. For example, expansionary fiscal and monetary policies are often associated with export booms, and such policies tend to be tightened during cyclical downturns. In other words, policies tend to be *pro-cyclical.* We shall explore in Parts 3 and 4 of this book some reasons why this might be so.

VI. SUMMARY

In this chapter, we put the analysis of financial-market equilibrium in Chapter 6 together with that of labor- and goods-market equilibrium in Chapters 4 and 5 to study how the economy achieves a general short-run macroeconomic equilibrium. We also examined the ways in which this equilibrium can be affected by a variety of policy and exogenous shocks that tend to afflict emerging and developing economies. Though the list of shocks we examined was by no means exhaustive of such possibilities, most shocks in which we will be interested throughout this book can be interpreted as "real" shocks (shocks affecting the domestic goods market, in the form of exogenous changes in the demand for or supply of domestic goods), "nominal" shocks (shocks affecting the domestic money market, in the form of exogenous changes in the supply of or demand for domestic money), and foreign financial or real shocks. The shocks we have examined here provide representative examples of each of these.

We have seen that the domestic macroeconomic effects of such shocks depend in an important way on a broad range of factors. These include the shapes of the GM and BB curves, which summarize several behavioral characteristics of both domestic and foreign economic agents, as well as how the central bank conducts domestic monetary policy (e.g., whether it sterilizes the effects of capital flows on the domestic money supply). Though we have not explicitly considered how the effects of these shocks would differ under financial autarky or perfect capital

[8] Indeed, of more than macroeconomic instability. For the role of weather-related supply shocks in triggering civil conflict in developing countries, see Fisman and Miguel (2008).

mobility, it is a straightforward exercise to extend the analysis of this chapter to do so.

In all these cases, the period of analysis has been the short run, during which not only the domestic capital stock and state of technology but also the lagged nominal wage, the country's international investment position, and the outstanding stock of domestic government bonds have all been held constant. In the next chapter, we will move to the medium run and will consider how the analysis we conducted in this chapter needs to be modified when we allow time for some of these variables to adjust.

REVIEW QUESTIONS

1. Explain how, in an economy that pursues domestic credit (monetary) targeting, the MM (BB) curve comes to pass through the intersection of GM and BB (MM).
2. How does a situation in which the central bank targets the domestic interest rate differ from one in which capital mobility is perfect so that the domestic interest rate is pinned down by uncovered interest parity?
3. In analyzing the effects of fiscal policy, why did we not have to worry about possible effects on the BB curve of the new bonds issued by the government to finance its additional spending?
4. Describe two alternative ways of determining the effects of shocks to the economy on its level of net exports.
5. Explain in intuitive terms why a monetary expansion (an increase in the central bank's stock of domestic credit) may give rise to either a capital outflow or capital inflow in an economy with imperfect capital mobility.

EXERCISES

1. Consider the case of a small open economy with a fixed exchange rate, perfect capital mobility (i.e., interest parity holds), and complete price stability (no ongoing inflation). Explain what effect an unanticipated decrease in the world interest rate would have on the following domestic macroeconomic variables:
   a. the stock of foreign exchange reserves
   b. the money supply
   c. real GDP
   d. the price level
   e. the real exchange rate
2. It is sometimes claimed that a nominal devaluation cannot increase net exports because its only effect will be to raise the domestic price level. This was not true in the model we analyzed in this chapter. What aspects of the model prevented this outcome from happening? Explain.

3. Explain how the economy's degree of integration with world financial markets affects the effectiveness of domestic monetary policy in influencing the real exchange rate in short-run equilibrium.
4. You are an economic advisor in a developing country with a fixed exchange rate. Though your country has an open capital account, your financial markets remain imperfectly integrated with those of the rest of the world. The central bank in your country targets the stock of domestic credit. One day, an IMF team arrives and convinces the finance minister that your country's fiscal deficit is excessive. The finance minister persuades the parliament to pass a tax increase to reduce the deficit. Explain what effect you would expect this measure to have on the following:
   a. the aggregate price level and level of real *GDP*
   b. the level of employment and the real wage
   c. the level of exports, of imports, and of net exports
5. It is often said that the macroeconomic effects of devaluation depend on the nature of the accompanying macroeconomic policies. To illustrate this point, compare the effects on the real exchange rate of an unanticipated devaluation of x percent under two alternative sets of accompanying policies:
   a. no change in fiscal or monetary policies
   b. a fiscal contraction that keeps the level of real *GDP* the same as it would have been without the devaluation

   Is there a difference between the two sets of policies with respect to their effects on net exports? Explain.
6. Using the model that we developed in this chapter, describe two policies that a central bank could use if it wanted to create a short-run improvement in a country's current account balance (remember that central banks do not control government spending or taxes). Would the two policies tend to move real *GDP* in the same direction? Explain.
7. In the short-run fixed exchange rate model that we developed in this chapter, we assumed that the expected future nominal exchange rate was the same as the current nominal exchange rate, that is, $S^e_{+1} = S$. If this is *not* true, the rate of return on foreign bonds would be given by $(1 + R^*)S^e_{+1}/S - 1$ rather than just R^*. Suppose that, from a situation in which $S^e_{+1} = S$, domestic and foreign residents suddenly develop the expectation that the domestic currency will be devalued so that $S^e_{+1} > S$. Assuming that capital mobility is imperfect and the central bank targets domestic credit, explain what effect this shock would have on the following:
   a. the domestic interest rate
   b. the economy's real exchange rate
   c. domestic real *GDP*
   d. the central bank's stock of foreign exchange reserves
   e. net exports

8. Money demand shocks have become increasingly common in developing countries over the past two decades. To explore the macroeconomic implications of such shocks, suppose the demand for money is given by

$$M_D = PL(R, Y; \underset{+}{\theta_L})$$

where θ_L is a money demand shock. Assuming that capital mobility is imperfect and that the central bank follows a monetary policy regime consisting of fixing the stock of domestic credit, use diagrams to explain what would happen to the economy's equilibrium interest rate and price level if there was an unanticipated increase in θ_L.
9. Consider an economy that is at full employment but in which the government feels that the central bank's stock of foreign exchange reserves is inadequately low. Describe a set of fiscal and monetary policies that would result in an increase in the central bank's stock of foreign exchange reserves without changing the level of real *GDP*. You may assume that the central bank targets domestic credit, capital mobility is imperfect, and the policy changes are unanticipated by workers.
10. Consider a small open developing economy that maintains a fixed exchange rate and exports a large share of its output. If that economy is subjected to frequent fluctuations in the world demand for its exports (shocks to the $X(SP^*/P, \ldots)$ function in our model), would its domestic interest rate and real exchange rate be more stable if it maintained financial autarky or permitted imperfect capital mobility? Explain.
11. Consider a small open developing economy that maintains a fixed exchange rate regime and is characterized by imperfect capital mobility. Assume that the central bank increases the stock of domestic credit. Explain what effects you would expect this policy shock to have on the following:
    a. the short-run equilibrium values of the domestic interest rate, price level, and real exchange rate
    b. the equilibrium level of real output and employment
    c. the equilibrium value of the real wage
    d. net exports from the domestic economy
    e. the central bank's stock of foreign exchange reserves
12. Assuming that the central bank targets domestic credit and that capital mobility is imperfect, analyze the effects of a tax increase on each of the following macroeconomic variables:
    a. the domestic interest rate and price level
    b. real output, employment, the real wage, and the nominal wage
    c. the real exchange rate, exports, imports, and net exports
    d. the money supply and the central bank's stock of foreign exchange reserves

13. You are a resident of a developing country. Upon (successfully!) completing this course, you are immediately appointed governor of your country's central bank. Show that you are well qualified for the job by explaining how the central bank should adjust the stock of domestic credit in response to an increase in the domestic price level if it is trying to
    a. fix the domestic money supply
    b. fix the domestic interest rate
14. Again, you are a resident of a developing country. When you finish this course, much to your surprise, you discover that while you were immersed in your studies, the governor of your central bank (who did not take this course) has caused the bank to adopt an extreme form of price-level targeting (see problem 7 in Chapter 5), in which the central bank adjusts the domestic interest rate to keep the domestic price level fixed. Knowing that what people in your country care most about is stabilizing the level of real GDP, rather than the price level, and that your country is subject to frequent aggregate productivity shocks, do you conclude that this new policy is a good idea? Why or why not?
15. "An increase in the money supply causes an increase in the level of real output, and an increase in the level of real output causes an increase in the demand for money. Since an increase in the demand for money raises interest rates, an increase in the money supply must raise interest rates." Do you agree or disagree? Explain.
16. Consider two types of monetary policy regimes that central banks may choose to adopt: monetary targeting and interest rate targeting.
    a. Under what conditions would adopting one regime rather than the other cause the central bank to choose different values for the stock of domestic credit? Illustrate your answer in a money-market equilibrium diagram.
    b. Under which of the two regimes would domestic real *GDP* tend to be more stable in a country in which government spending is subject to frequent and unpredictable shocks?
17. "Central banks have to choose: they cannot simultaneously fix the exchange rate and the money supply." Is this statement
    a. always true?
    b. always false?
    c. true or false, depending on circumstances?

    Explain.
18. Suppose that a small, open economy confronts a permanent reduction in the interest rate that prevails in world financial markets. Explain how the domestic macroeconomic effects of this shock depend on whether the domestic central bank sterilizes the effects of the balance of payments on the domestic money supply.

19. Consider a small, open economy that maintains a fixed exchange rate and is *perfectly* integrated with world financial markets. Suppose that at time $t-1$, the expectation develops that this economy is going to devalue its currency at time t. At time t, the devaluation actually takes place, and no further devaluation is expected after that. Explain what happens between times $t-1$ and t to each of the following variables in this economy:
    a. the domestic interest rate and price level
    b. the level of real output
    c. the real exchange rate and private absorption
    d. net exports
20. Suppose that the domestic central bank engages in domestic credit targeting. Consider three possible shocks to the economy:
    a. a reduction in government spending
    b. a reduction in the stock of domestic credit targeted by the central bank
    c. a devaluation of the nominal exchange rate

    Explain how each of these shocks affects all of the following macro variables:
    i. the domestic price level and the domestic interest rate
    ii. the level of real *GDP* and the level of employment
    iii. the real wage and the nominal wage
    iv. the real exchange rate, the level of exports, the level of imports, and net exports
    v. the central bank's stock of foreign exchange reserves

    Please do one shock at a time, and show how each shock affects all of the variables listed. If the effect is ambiguous, explain why.

REFERENCES AND FURTHER READING

Calderon, Cesar, Norman Loayza, and Klaus Schmidt-Hebbel (2005), "Openness, Vulnerability, and Growth," unpublished manuscript, Central Bank of Chile.

Fisman, Raymond, and Edward Miguel (2008), *Economic Gangsters: Corruption, Violence, and the Poverty of Nations*, Princeton, NJ: Princeton University Press.

Hausmann, Ricardo, and Michael Gavin (1996), "Securing Stability and Growth in a Shock-Prone Region: The Policy Challenge for Latin America," working paper 315, Inter-American Development Bank, Office of the Chief Economist.

8

Medium-Run Macroeconomic Equilibrium

In the model that we developed over the last four chapters, we described the economy's macroeconomic equilibrium as a short-run one. In Chapter 4, we distinguished between the short run and the long run based on the roles of the production technology and the capital stock. In the short run, we took the production technology and aggregate capital stock as given during the period of the analysis. Our reason for holding these variables constant was that they tend to change very slowly over time, relative to the other variables in our model. But as we further developed our model in the subsequent chapters, these were not the only variables that we took as given. When we analyzed the determination of aggregate supply in Chapter 5, for example, we focused on a situation in which wage inertia caused the nominal wage to change slowly, so that the current nominal wage was partly determined by its past history. Moreover, when describing the determination of equilibrium in domestic asset markets in Chapter 6, we introduced two important financial stocks: the economy's net international investment position and the domestic government's outstanding debt (which together sum to the financial net worth of the domestic private sector). In the short run, these financial stocks were also assumed to be constant because the changes that they undergo during any given year are likely to be very small compared to the accumulated stocks that are inherited from the economy's entire past history.

As we have seen, however, shocks to the economy may result in changes in the nominal wage and in surpluses or deficits in the current account of the country's balance of payments. When these changes happen, the lagged nominal wage and the net international investment position that the economy will inherit in the next period will be different from those that it inherited this period, and that means that the next period's short-run equilibrium must be different as well, even if the economy does not experience any additional shocks. Similarly, if the government runs a deficit in its budget in the current period, it will need to issue additional bonds to finance that deficit, and if it runs a surplus, it will retire existing bonds.

In either case, the outstanding stock of government bonds that the economy inherits next period would be different from those that existed at the beginning of the current period. Again, this would tend to cause changes in the economy's short-run equilibrium. In short, because the economy's short-run equilibrium is conditioned on specific values of all three of these variables, changes in any of them will tend to alter the economy's short-run macroeconomic equilibrium over time.

This chapter will extend our model to analyze some of these dynamic effects and, in particular, will consider how the economy's macroeconomic equilibrium responds to shocks beyond the immediate short run – in particular, after adjustments to nominal wages and to the accumulation of financial assets are complete. Because we will continue to hold constant the level of total factor productivity and the capital stock, we will describe this situation as a "medium-run" equilibrium, to distinguish it both from the short-run equilibrium that we have been studying up to now and from the long-run equilibrium of growth theory in which the level of technology and the capital stock also become endogenous.

This dynamic analysis makes it possible to consider both the short-run and medium-run effects of a wide variety of exogenous and policy shocks to an emerging-market economy. In this chapter, however, we will focus on just two specific shocks: changes in domestic credit and in the exchange rate. The reason for this specific focus is that the behavior of these two nominal policy variables is crucial in determining the rate of ongoing inflation in a small open economy over the medium run. Ongoing inflation is a phenomenon that was missing from our short-run model but can be incorporated in the model by moving to the medium run. This is an important extension, not only because many emerging-market economies have indeed experienced high inflation in the past (and some still do) but also because, as we saw in Chapter 3, high inflation may have significant adverse implications for long-run capacity growth in emerging economies. We will consider the effects of changes in other (real) exogenous variables on the economy's medium-run equilibrium in Chapter 16.

Part 3 of this book will be devoted to an analysis of fiscal deficits and changes in government debt. Accordingly, in this chapter we will abstract away from changes in government debt stocks by assuming that the outstanding stock of government bonds is constant in real terms. Our focus, therefore, will be on the dynamic effects of adjustments in the nominal wage and in the country's international investment position. We will proceed as follows. In the next section, we will study the dynamic implications of nominal wage adjustment, holding constant the country's international investment position. As we will see, nominal wage adjustment will drive the economy to a full-employment equilibrium, sometimes referred to as *internal balance*, in the medium run. Section II explores the properties of this full-employment equilibrium, comparing them to those of the short-run equilibrium that we have been considering up to now. A key issue investigated in that section is how ongoing inflation is created in a full-employment setting. Section III examines how the presence of ongoing inflation in turn feeds back on the economy's full-employment equilibrium, allowing us to recast our model to incorporate this common phenomenon. Section IV then turns to wealth dynamics, considering how adjustments in the country's international investment position affect its full-employment equilibrium and what the economy's medium-run equilibrium looks like when the adjustment in its international investment position is complete, a situation known as *external balance*. Having described medium-run equilibrium, Section V analyzes how that equilibrium is affected by shocks to the two nominal policy variables on which we are focusing in this chapter. Section VI revisits the generation of sustained inflation in the full medium-run model, and Section VII summarizes.

I. NOMINAL WAGE DYNAMICS AND FULL-EMPLOYMENT EQUILIBRIUM

Because the short-run model that we analyzed in the preceding chapter had a positively sloped aggregate supply curve, shocks to the economy had the capacity to move output and employment away from their full-employment levels Y_P and L_P. The positive slope of the aggregate supply curve, in turn, resulted from nominal wage stickiness, that is, from the fact that the economywide nominal wage did not immediately adjust to the level that would have been consistent with full employment. In Chapter 4, we described the process driving nominal wage changes as

$$\Delta W = \sigma(Pw(L^S, \ldots) - W_{-1})$$

where the positive fraction σ represented the speed of nominal wage adjustment. We defined full employment as the level of employment that would prevail when nominal wage adjustment is complete, that is, when $\Delta W = 0$, so $P\, w(L^S, \ldots) = W = W_{-1}$.

Our first task in this chapter is to investigate whether the process of nominal wage adjustment just described actually brings the economy back to a full-employment equilibrium after a shock has driven it away from that equilibrium. To see that it must indeed do so, consider Figure 8.1. Suppose the economy is initially in full-employment equilibrium at a point like A, where the positively sloped AS curve of Chapter 4 intersects a negatively sloped AD curve. Recall that the AS curve must pass through the point $(Y_P, W_{-1}/w_P)$, where W_{-1} is not just the wage that prevailed in the last period but also the current-period nominal wage that would prevail in the absence of shocks. Now consider, for example, the effect of a permanent positive aggregate demand shock – that is, an aggregate demand shock that shifts AD to the right, to a position such as AD_1, and leaves it there. As we saw in Chapter 5, the economy's new short-run equilibrium would move to a point like B, with a higher level of output and a higher price level. In the labor market, this higher price level is associated with an upward shift in the labor demand curve that is

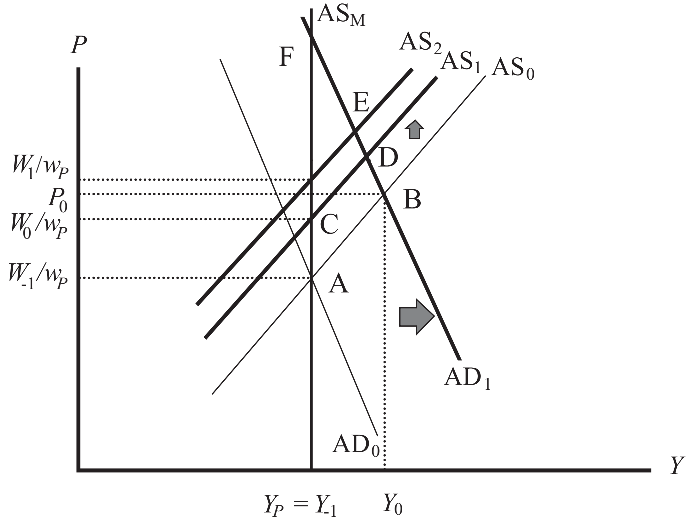


Figure 8.1. Adjustment to an expansionary aggregate demand shock

in the same proportion as the increase in the price level recorded in the goods market as well as a less-than-proportionate upward shift in the labor supply curve. The result, therefore, must be a less-than-proportionate increase in the equilibrium value of the nominal wage, from the pre-shock value of W_{-1} to a post-shock value of, say, W_0.

So far, we have simply reproduced our short-run equilibrium analysis. But consider what happens next. Because the aggregate demand shock increased the nominal wage from W_{-1} to W_0, in the next period, the lagged nominal wage becomes W_0 rather than W_{-1}. The important point is that this increase in the nominal wage must cause the short-run aggregate supply curve to shift upward. The reason is that in the period after the shock, the short-run aggregate supply curve has to pass through the point $(Y_P, W_0/w_P)$ rather than through $(Y_P, W_{-1}/w_P)$, precisely because W_0 becomes next period's lagged nominal wage. The upward shift in the short-run aggregate supply curve must therefore be in the *same proportion* as the increase in the current-period nominal wage. In Figure 8.1, this is shown as an upward shift in the short-run aggregate supply curve from AS_0 to AS_1, passing through the point C rather than the original point A. Because the nominal wage initially increased less than in proportion to the increase in the price level, C must lie below the point where a horizontal line through the new short-run equilibrium at B intersects a vertical line through Y_P.

Notice what this means. Because the AD curve stays put at AD_1 with the assumption that the shock is permanent, the economy must move to a new equilibrium at the point D, where the new short-run aggregate supply curve that passes through C, labeled AS_1 in the figure, intersects AD_1. At the point D, real output is smaller, but the price level is higher than in the initial equilibrium at B. This additional increase in the price level triggers yet another adjustment in the labor market, with the labor demand curve once again shifting vertically upward in proportion to the additional price increase and the labor supply curve shifting upward less than in proportion, resulting in another nominal wage increase. This additional increase in the nominal wage, in turn, must give rise to another upward shift in the short-run aggregate supply curve, moving it to a position such as AS_2, and moving the economy to yet a third short-run equilibrium at E.

You can see where this is going. The initial increase in the price level gives rise to an increase in the nominal wage that triggers a process of upward drift in the short-run aggregate supply curve. As the short-run aggregate supply curve drifts upward, the aggregate price level rises, and the level of real output falls back toward its full-employment level. At each stage of the process, the induced increase in the price level, and therefore in the nominal wage, becomes smaller. This process of upward drift in the short-run aggregate supply curve will stop only when the short-run aggregate supply curve has drifted upward sufficiently to intersect the aggregate demand curve directly above Y_P, where the economy has returned to full employment, but with a higher price level.

This economy therefore contains a mechanism – continuous nominal wage adjustment – that tends to automatically drive it back to a full-employment equilibrium after a shock drives it away from there. The process described above would work similarly in the case of a *contractionary* aggregate demand shock, except that full employment would be reached through continuous *downward* drift in the short-run aggregate supply curve, and it would be reached with a *lower* aggregate price level. It would also work similarly in the case of an aggregate supply shock, as we shall see later, except that the full-employment level to which the economy would converge would itself be altered by the shock. The point is that when nominal wage adjustment is complete, the economy must find itself on the vertical line labeled AS_M in Figure 8.1. Because where the economy winds up on this vertical line depends on where the vertical line is intersected by the aggregate demand curve, we can denote this vertical line as the economy's *medium-run aggregate supply curve*. The important conclusion of this analysis is that *full employment is a stable medium-run equilibrium for the economy*. As mentioned earlier, this full-employment equilibrium is often referred to as internal balance.

While the stability of the full-employment equilibrium indeed means that deviations from full employment will tend to be transitory, notice that how fast the economy converges to its internal balance configuration depends on how fast the nominal wage adjusts, and how fast the nominal wage adjusts in turn depends on how large a price increase is triggered at each stage of the adjustment process (i.e., at each upward displacement of the short-run aggregate supply curve). As you can see from Figure 8.1, price increases will be smaller the flatter the short-run aggregate supply curve. We saw in Chapter 4 that the aggregate supply curve is flatter the more inertia there is in nominal wage adjustment. It follows that deviations from full employment will tend to be more prolonged the greater the degree of nominal wage stickiness.

II. CHARACTERISTICS OF THE FULL-EMPLOYMENT EQUILIBRIUM

As we have just seen, when nominal wage adjustment is complete, the nominal wage must converge to the stable value $W = Pw_P = W_{-1}$, and the economy must return to full employment. Our next task is to explore how, after the economy has returned to full-employment equilibrium, the effects of shocks to the economy differ from what they were in the preceding chapter, that is, in the short run.

To explore how the economy behaves in full-employment equilibrium, all we have to do is add the full-employment condition $W = Pw_P = W_{-1}$ to the aggregate supply and demand equations (4.11) and (5.6), reproduced here for convenience:

$$Y = AF[K, L(Pw_P/W_{-1})L_P] = Y(P; A, K, W_{-1}, \ldots) \tag{4.11}$$

$$Y = \varphi(SP^*/P, \ldots)A_P(Y - T, R, \ldots) + G + X(SP^*/P, \ldots) \tag{5.6}$$

To analyze the resulting model, we simply replace W_{-1} by Pw_P in the aggregate supply equation (4.11). It is easy to see from equation (4.11) that in that case, we must have $Y = Y_P$ (recall that $L(1) = 1$ and $Y_P = AF(K, L_P)$). Substituting Y_P for Y in the aggregate demand equation (5.6), we can see that the goods-market equilibrium condition at full employment becomes

$$Y_P = \underset{+}{\phi(SP^*/P, \ldots)}A_P\,[\underset{+}{Y_P - T}, \underset{-}{R}, \ldots] + G + \underset{+}{X(SP^*/P, \ldots)} \tag{8.1}$$

Comparing this equation with the corresponding short-run equilibrium condition given by equation (5.7), you can see that changes in the domestic price level must have *weaker* effects on excess supply or demand in the goods market when the economy is at full employment because since the aggregate supply curve is vertical when the economy is constrained to always be at full employment, such price-level changes can no longer call forth any supply response. This means that price changes will in general have to be *larger* to maintain equilibrium in the goods market when the economy is constrained to be at full employment than when it is not. These results can be confirmed for aggregate demand and supply shocks separately.

1. Aggregate Demand Shocks

The effects of expansionary aggregate demand shocks on the full-employment equilibrium in the goods market were already illustrated in Figure 8.1. As we saw there, in contrast to what happens when wage adjustment is incomplete (at B), an expansionary aggregate demand shock results in a higher value of the domestic price level but an unchanged level of real output in the full-employment equilibrium at F. The shock's other macroeconomic implications are as follows:

- Because output is unchanged and the other variables in the production function are also unchanged, employment must be unchanged, and the equilibrium value of the real wage must be unaffected. For the real wage to be constant, the nominal wage must rise in the same proportion as the domestic price level.
- Because domestic prices are higher and the nominal exchange rate is fixed when the shock takes any form other than an increase in the domestic-currency price of foreign goods (SP^*), the real exchange rate must appreciate at full employment for all such shocks. However, the appreciation in the real exchange rate must be *larger* than in the short-run equilibrium because in the absence of a domestic supply response, real exchange rate adjustments bear the full brunt of maintaining equilibrium in the goods market. If the shift in AD is the result of an increase in SP^*, on the other hand, the end result at full employment must be an *unchanged* value of the real exchange rate. The reason is that, because AD shifts vertically by an amount that is proportional to the increase in SP^* and because the medium-run AS is vertical, the proportional change in the domestic price level must be the same as that in SP^*, leaving the real exchange rate unchanged.
- Unless the shock takes the form of an increase in SP^* or an increase in export demand, net exports must fall. The reason is that real exchange rate appreciation provides the mechanism for maintaining equilibrium in the goods market, and this means that a shock-induced expansion in other components of aggregate demand must be exactly offset by deterioration in net exports, given an unchanged supply of domestic goods. On the other hand, if the shock consists of increases in either SP^* or export demand, then net exports must remain unchanged. In the case of increases in SP^*, this is because the shock leaves all of the determinants of both exports and imports unchanged. In the case of an increase in export demand, exports must rise, but the appreciation of the real exchange rate moderates the expansion in exports and offsets their effects on *net* exports by inducing an expansion of imports.

2. Aggregate Supply Shocks

In the case of aggregate supply shocks (changes in A or K or unobserved factors affecting labor supply), the value of full-employment output will itself change, as we have seen in previous chapters. Consider the case of a contractionary productivity shock, for example. In this case, the full-employment level of output must fall, say, from Y_{P0} to Y_{P1}, and the full-employment real wage must fall as well, say, from w_{P0} to w_{P1} (Figure 8.2). Consequently, the medium-run aggregate supply curve shifts to the left, say, from AS_{M0} to AS_{M1}, and because the reduction in the full-employment real wage causes the value of W_{-1}/w_P to rise, the short-run aggregate supply curve

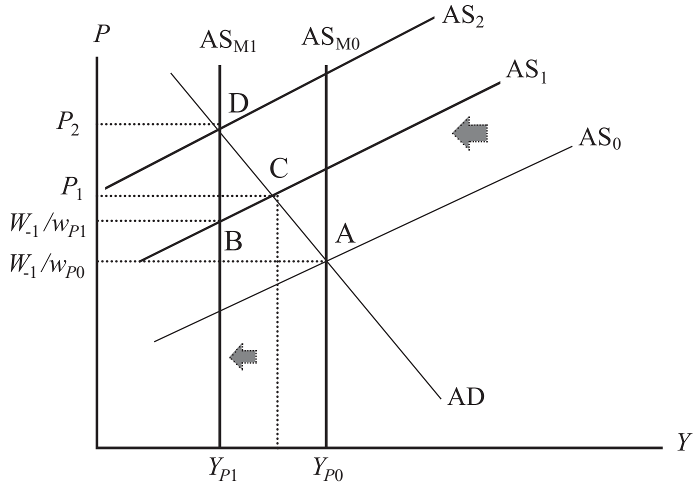


Figure 8.2. Aggregate supply shocks with full wage adjustment

shifts up and to the left, say, from AS_0 to AS_1, causing it to pass through a point like B that lies to the northwest of the original equilibrium. But because the intersection of this new short-run aggregate supply curve with the stationary aggregate demand curve at the point C causes an increase in the price level, the same process of upward drift in the short-run aggregate supply curve is set in motion as in the last section, until the economy converges to a new full-employment equilibrium at D, where the stationary aggregate demand curve intersects the new medium-run aggregate supply curve. As in the short run, output falls and the price level rises in the new full-employment equilibrium, but the decrease in output and increase in the price level are both larger after the new full-employment equilibrium is reached than they are in the short run.

At the new full-employment equilibrium, the level of employment must unambiguously fall. The reason is that because the real wage falls, the labor demand curve must shift downward relative to the labor supply curve. Because the contraction in the level of output is larger in full-employment equilibrium than it is in the short run, the absorption approach described in Chapter 2 implies that not only must net exports fall but the reduction in net exports must be larger in the new full-employment equilibrium than it is on impact.

3. The Full-Employment GM Curve

From this analysis of the properties of the goods-market equilibrium in full-employment equilibrium, it is straightforward to derive the GM curve for this case. In comparison with the GM curve derived in Chapter 5, it is easy to see that the slope of the GM curve must be *flatter* when the economy is in full-employment equilibrium, and that changes in its position in response to changes in exogenous variables must be *larger*. As we saw earlier, the reason is that under these circumstances, changes in the price level become less effective in restoring goods-market equilibrium than they are in the short run, because at full employment, changes in the price level do not affect the level of domestic real output (the AS_M curve is vertical). This means that price-level adjustments can affect the goods-market equilibrium only through their effects on the composition of domestic and foreign demand. Because changes in the price level have weaker effects on the market, *larger* price-level adjustments are required under full-employment conditions to restore market equilibrium in response to shocks.

For the purposes of this chapter, a particularly important case is that of a change in the domestic-currency price of the foreign good SP^*, brought about either by a change in the nominal exchange rate S or a change in the world price level P^*. In this case, we can be very specific about the magnitude of the shift in the full-employment GM curve. We saw earlier that an increase in SP^* would shift the AD curve upward in the same proportion, moving it vertically along the AS_M curve while leaving the economy's real equilibrium unchanged. Translated into (P, R) space, this means that in full-employment equilibrium, a change in SP^* must shift the GM curve to the right *in the same proportion* as the change in SP^*.

4. Financial-Market Equilibrium under Full-Employment Conditions

Next, consider how domestic financial-market equilibrium behaves under full-employment conditions. As we have just seen, when the nominal wage has fully adjusted, Y will remain equal to Y_P. When this is so, the bond-market equilibrium condition becomes

$$B - B_C = b(R - R^*)[W_P - PL(R, Y_P)] + Sb^*(R - R^*)(W_F^* - P^*L^*(R^*, Y^*)) \tag{8.2}$$

Just as is true in the goods market, the difference here is that price-level changes have no effect on real output: the demand for money becomes $PL(R, Y_P)$ rather than $PL(R, Y(P))$. In particular, under full-employment conditions, the demand for money changes *in proportion* to changes in the price level rather than *more* than in proportion. This means that changes in the domestic price level have *smaller* effects on the demand for domestic bonds under full-employment conditions and therefore that the horizontal shifts of the bond demand curve induced by changes in the domestic price level must be smaller as well. The implication is that changes in the domestic price level have weaker effects on the equilibrium value of the domestic interest rate under full-employment conditions than they do in the short run. Consequently, the BB curve of Chapter 6 becomes *flatter* when the economy is at full employment than when output can respond to changes in the price level.

Now consider the effect of a monetary policy change on the position of the BB curve when the economy is at full employment. Because the vertical position of the curve is determined at a *given* value of the domestic price level, whether the economy is at full employment does not affect the change in the equilibrium interest rate required to clear the domestic bond market in response to a change in the stock of domestic credit; that is, the *vertical* shift in the BB curve is the same whether or not the economy is at full employment. However, because changes in the domestic price level have weaker effects on the demand for domestic bonds when the economy is at full employment, the *horizontal* displacement of the BB curve in response to domestic credit shocks must be *larger* under full-employment conditions.[1]

Finally, consider the financial-market effects of a nominal devaluation at full employment. For simplicity, let us restrict our attention to the case in which neither domestic nor foreign agents initially hold each others' bonds, so initially, $b = 1$ and $b^* = 0$, implying that the devaluation induces no wealth effects. In this case, just as in the short-run case that we analyzed previously, the impact of the devaluation of the BB curve depends on whether capital gains on reserves are monetized:

- If the capital gains on the central bank's foreign exchange reserves created by the devaluation are not monetized, $(B - B_C)$ is unchanged, and the bond-market equilibrium condition (8.2) is undisturbed. Thus the BB curve is not affected by the devaluation.
- If the central bank's capital gains *are* monetized, the stock of domestic bonds that must be held by the public $(B - B_C)$ falls by $\Delta S \cdot F_C^* = (\Delta S/S) \cdot SF_C^*$, so restoring bond-market equilibrium requires either a reduction in the domestic interest rate or an increase in the domestic price level (i.e., the BB curve must shift down and to the right). It is easy to see that the rightward shift in the BB curve must be *less* than in proportion to the size of the devaluation, because with $b = 1$, a change in the domestic price level that is exactly proportional to the devaluation would reduce the demand for domestic bonds by $(\Delta S/S) \cdot PL$, which exceeds $(\Delta(S/S) \cdot SF_C^*$ because $SF_C^* < PL$. To preserve bond-market equilibrium at the original interest rate, therefore, the proportional increase in the domestic price level must be *smaller* than $\Delta S/S$.
- Finally, suppose that the central bank's capital gains are monetized and that, at the same time, the central bank increases the stock of domestic credit in the same proportion as the rate of devaluation. In this case, the supply of bonds in the hands of the public would fall by

$$(\Delta S/S) \cdot SF_C^* + (\Delta S/S) \cdot B_C = (\Delta S/S) \cdot (SF_C^* + B_C) = (\Delta S/S) \cdot PL$$

[1] Under interest rate targeting, the BB curve retains its horizontal shape under full-employment conditions and shifts (vertically) only when the interest rate target is modified by the central bank.

In this case, bond-market equilibrium would be preserved at the original value of the domestic interest rate only if the domestic price level increases by $\Delta S/S$ percent (i.e., in proportion to the rate of devaluation) because this would increase the demand for money by $(\Delta S/S) \cdot PL$ and therefore decrease the demand for domestic bonds by the same amount. This means that the BB curve would shift to the right in proportion to the rate of devaluation.

5. Sustained Inflation

These observations about the full-employment GM and BB curves have an important implication: they show us how sustained inflation can be generated period after period in our model under full-employment conditions.

To see how sustained inflation can emerge in our model, suppose that the central bank adopts a policy of devaluing the exchange rate by x percent each period, monetizes the associated capital gains on its foreign exchange reserves, and simultaneously increases the stock of domestic credit by x percent each period. We have just seen that under these circumstances, the full-employment GM and BB curves would both shift to the right by x percent each period. It is easy to see that the MM curve would do so as well: because the domestic-currency value of the central bank's stock of foreign exchange reserves and its stock of domestic credit both increase by x percent each period, so must the domestic money supply. For the money market to remain in equilibrium, therefore, the nominal demand for money would have to increase by x percent each period as well. But this would require the price level to increase by x percent each period, that is, it would require the MM curve to shift to the right by x percent each period. With the GM, BB, and MM curves all shifting to the right by x percent each period, the domestic price level must rise by x percent each period, and the economy therefore experiences a sustained inflation rate of x percent per period, at an unchanging value of the domestic nominal interest rate.

The interesting question, of course, is why the central bank would undertake this particular set of policies. We address that issue at the end of this chapter.

III. ANTICIPATED FUTURE DEPRECIATION AND INFLATION

Next, let us examine how the economy's full-employment equilibrium is affected by the process of ongoing inflation that we have just described. In this section, we will see that ongoing inflation requires us to make some modifications to our full-employment model. These modifications will not affect the analysis of the last section, but they will enable us to determine how the economy's other macroeconomic variables are affected by the continuous rightward shifts of the GM, BB, and MM curves that we examined there.

To see why our model needs to be modified in the presence of ongoing inflation, recall that when we introduced price-level expectations in Chapter 4, the expectations on which we were focusing were those formed by workers concerning the price level prevailing during the period over which they supply their labor to firms – which they are assumed to be unable to observe directly. But these are not the only expectations that matter in the economy. Recall from Chapter 5 that if the nominal exchange rate is expected to change by an amount $\hat{S}$ in the next period, the expected domestic-currency rate of return on foreign bonds becomes $R^* + \hat{S}$, equal to the nominal interest rate on foreign bonds plus the expected capital gain on those bonds arising from any anticipated increase in the domestic-currency value of foreign exchange (i.e., from any expected depreciation of the domestic currency). If the domestic currency is expected to depreciate, therefore, the total return on foreign bonds is *larger* than the nominal interest rate offered by those bonds, whereas if the domestic currency is expected to appreciate, the total return on foreign bonds is *smaller* than the nominal interest rate on foreign bonds because in that case, the bonds are expected to yield a capital *loss* in domestic-currency terms. For similar reasons, the foreign-currency rate of return on *domestic* bonds becomes $R - \hat{S}$.

We have been able to ignore this factor until now because we have been assuming that $\hat{S} = 0$. When there is no ongoing exchange rate depreciation in the economy, expectations that *future* changes in the exchange rate will be zero are perfectly rational, consistent with what we have assumed about the way that workers form their own expectations. But that is no longer so if we allow for constant ongoing rates of depreciation and inflation. Under these circumstances, we would expect economic agents to rationally anticipate that tomorrow's exchange rate and price level will both be different from those prevailing today. How would our full-employment model change if we allow for such expectations?

1. Real and Nominal Interest Rates

To answer this question, we can begin by drawing a distinction between nominal and real interest rates. Because domestic bonds are denominated in domestic currency, rather than in units of domestic goods, in the presence of expected inflation those bonds would be expected to lose purchasing power over goods. If the expected rate of inflation is given by π, that expected loss of purchasing power over goods is equal to π, so the *real* interest rate on domestic bonds – the rate of return on those bonds measured in units of domestic goods – is equal to the nominal interest rate that they pay minus the purchasing power that they are expected to lose. Letting r denote the real interest rate on domestic bonds, we can therefore express r as

$$r = R - \pi \tag{8.3a}$$

Alternatively, the nominal interest rate can be written as the sum of the real interest rate and the expected rate of inflation:

$$R = r + \pi \tag{8.3b}$$

Written in this form, the relationship between the nominal and real interest rates is referred to as the *Fisher relationship*, after the Yale economist Irving Fisher who first called attention to it.

2. The GG Curve

We will now introduce two modifications to the goods-market equilibrium condition (8.1). First, because it is the *real* interest rate that determines the terms on which today's goods can be converted into tomorrow's goods by savers, it is the real interest rate that affects the private sector's spending decisions. Now that we distinguish between nominal and real interest rates, therefore, the full-employment goods-market equilibrium condition given by equation (8.1) must be rewritten to reflect the dependence of private absorption on the real rather than the nominal interest rate. Second, because we want to investigate how full-employment equilibrium is affected by ongoing inflation, it is convenient to express goods-market equilibrium as a function of an endogenous value that reaches a stable value under ongoing inflation rather than one that is changing continuously. We can do this by expressing goods-market equilibrium as a function of the real exchange rate rather than the domestic price level because we saw in the last section that an ongoing depreciation of x percent would result in ongoing inflation of x percent, which would leave the real exchange rate Q unchanged. With these two modifications, equation (8.1) can be rewritten as

$$Y_P = \phi(Q, \ldots)A_P[Y_P - T, r, \ldots] + G + X(Q, \ldots) \tag{8.4a}$$

where the real exchange rate Q is used in place of its definition SP^*/P.

This equation can be solved for the value of the real interest rate required to clear the goods market as a function of all the other variables that affect that market. Notice in particular that because a depreciation of the real exchange rate creates an excess demand for domestic goods, an increase in the real interest rate would be required to clear the goods market. We can therefore rewrite (8.4a) as

$$r = r(\underset{+}{Q}, \ldots) \tag{8.4b}$$

where the ellipses represent all the other variables that affect the goods market directly such as the full-employment level of output Y_P, the fiscal policy variables G and T, and any exogenous shocks to the functions in (8.4a). Holding the real exchange rate constant, the equilibrium value of the real interest rate in the goods

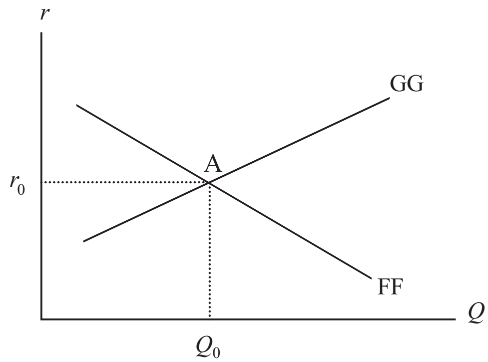


Figure 8.3. The real interest rate and real exchange rate

market would be increased by shocks to any of these variables that create an excess demand for domestic goods and decreased by shocks that create an excess supply.

The positive relationship between r and Q represented by equation (8.4b) can be depicted as a new goods-market equilibrium curve, drawn in (Q, r) space rather than in (P, R) space. This curve is labeled GG in Figure 8.3. Unlike the GM curve that we derived in Chapter 5, the GG curve must have a *positive* slope because the domestic price level, which was on the horizontal axis in the (P, R) space where the GM curve was originally derived, is *inversely* related to the real exchange rate, which is the variable on the horizontal axis in Figure 8.3. Similarly, any shocks that shift the GM curve to the right in (P, R) space will shift the GG curve to the left in the (Q, r) space of Figure 8.3.

3. The FF Curve

Anticipated future exchange rate and price-level changes have more significant impacts on asset markets. To explore these, recall from Chapter 6 that a country's aggregate financial wealth is equal to its international investment position, so $S \cdot IIP^* = W_P + W_G + W_C$. Assuming that $W_C = 0$, and recalling that $W_G = -B$, this implies that

$$W_P = S \cdot IIP^* + B$$

That is, the domestic private sector's financial wealth differs from the country's international investment position by the amount of the government's net financial liability. A similar relationship holds for the rest of the world, that is, $W_F^* = -IIP^* + F^*$. Using these relationships in equation (8.2), we can write the domestic bond-market equilibrium condition as

$$\begin{aligned} B - B_C &= b[S \cdot IIP^* + B - PL] + Sb^*[-IIP^* + F^* - P^*L^*] \\ &= (b - b^*)S \cdot IIP^* + b(B - PL) + Sb^*(F^* - P^*L^*) \end{aligned}$$

where the arguments of the various functions have been suppressed to avoid clutter. Replacing the domestic and foreign nominal interest rates in these functions by the appropriate rates of return including capital gains and losses, using the Fisher relationship, and assuming that the foreign rate of inflation is zero, we can rewrite this condition in a form that reflects the effects of anticipated future exchange rate and price-level changes:

$$\begin{aligned} B - B_C = {} & [b(r + \pi - \hat{S} - r^*) - b^*(r + \pi - \hat{S} - r^*)]S \cdot IIP^* \\ & + b(r + \pi - \hat{S} - r^*)[B - PL(r + \pi, Y_P)] \\ & + Sb^*(r + \pi - \hat{S} - r^*)(F^* - P^*L^*), \end{aligned} \tag{8.5a}$$

As we did for the goods market, we can also express domestic bond-market equilibrium as a function of the real exchange rate rather than the domestic price level. Recall that the real exchange rate is given by $Q = SP^*/P$. With foreign inflation equal to zero, we can set the price of foreign goods P^* equal to unity for simplicity, so the real exchange rate just becomes $Q = S/P$, and the domestic price level can be expressed as $P = S/Q$. Using this in (8.5a) and dividing both sides by S, we can derive a relationship between the real exchange rate and the real interest rate that is consistent with equilibrium in the domestic bond market:

$$\begin{aligned} (B - B_C)/S = {} & [b(r + \pi - \hat{S} - r^*) - b^*(r + \pi - \hat{S} - r^*)]IIP^* \\ & + b(r + \pi - \hat{S} - r^*)[B/S - Q^{-1}L(r + \pi, Y_P)] \\ & + b^*(r + \pi - \hat{S} - r^*)(F^* - L^*) \end{aligned} \tag{8.5b}$$

Because we have divided through by S, this relationship essentially expresses the bond-market equilibrium condition in foreign-currency terms.

Next, we will plot this relationship in (Q, r) space, just as we did the goods-market equilibrium condition. To do so, we need to investigate the effect of a change in the real exchange rate Q on the value of the domestic real interest rate r required to clear the bond market. You can verify from equation (8.5b) that the real exchange rate affects the bond-market equilibrium condition only through its effect on the foreign-currency value of the demand for money $Q^{-1} L$, which appears on the right-hand (demand) side of the equation. Now consider the effects on equation (8.5b) of a depreciation of the real exchange rate Q, holding constant the nominal exchange rate S, which appears elsewhere in the equation. Because a depreciation of Q with S constant can only be brought about by a reduction in the domestic price level, the effect is to *decrease* the nominal demand for money. A depreciation in Q therefore *increases* the demand for domestic bonds. Because the supply of bonds in the hands of the public is constant, other things equal, this requires a *decrease* in the real interest rate on domestic bonds to clear the domestic bond market. Equation (8.5b) can therefore be depicted as a negatively sloped FF curve in (Q, r) space, as

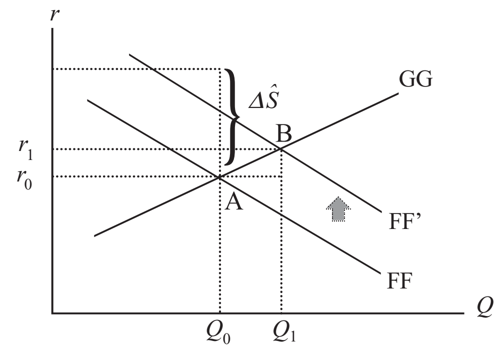


Figure 8.4. Effects of expected exchange rate depreciation

in Figure 8.3. Again, the slope is the reverse of that derived for the BB curve in the (P, R) space of Chapter 6 because of the inverse relationship between Q and P.

In effect, we have rewritten our full-employment model in terms of the domestic real interest rate and real exchange rate rather than in terms of the nominal interest rate and domestic price level as the key endogenous variables, determined by GG and FF curves that are similar to the GM and BB curves that we analyzed in (P, R) space in Chapter 7 but drawn in (Q, r) space. The advantage of doing so is that the GG and FF curves do not shift continuously to the right under ongoing inflation, as the GM and BB curves do. The economy is in full-employment equilibrium at the point A in Figure 8.3, where the GG and FF curves intersect. The point of intersection between the two curves determines the full-employment equilibrium values of the domestic real interest rate and real exchange rate.

Notice that our modified model now explicitly includes the variables $\hat{S}$ and π, denoting the ongoing rates of exchange rate depreciation and inflation. This allows us to investigate the macroeconomic effects of changes in both these variables.[2] Because the economy we are analyzing maintains an officially determined exchange rate, the (actual and expected) rate of exchange rate depreciation is an exogenous policy variable. To analyze the effects of changes in expected future inflation, we will provisionally take this variable to be exogenous as well but will investigate later on how it is determined.

4. Changes in the Expected Rate of Depreciation

The effects of an increase in $\hat{S}$ (an increase in the expected and actual rate of depreciation) are shown in Figure 8.4. Note first that a change in $\hat{S}$ has no effect on the goods-market equilibrium condition (8.4a) and therefore on the GG curve.

[2] Notice an important property of equation (8.4b): changes in the same proportion in the nominal exchange rate S, the stock of government debt outstanding B, and the stock of domestic credit B_C issued by the central bank leave the equation undisturbed. This observation will prove useful later on.

To see what happens to the FF curve, consider equation (8.5b). An increase in $\hat{S}$ increases the rate of return on foreign bonds relative to that on domestic bonds. Consequently, it induces a reduction in the demand for domestic bonds. Because the supply of domestic bonds is unchanged, maintaining bond-market equilibrium requires an *increase* in the domestic real interest rate to restore the demand for domestic bonds to its original level. The FF curve must therefore shift upward, as in Figure 8.4.

How large must the increase in the domestic real interest rate be? Under the assumption of imperfect capital mobility reflected in equation (8.5b), we can show that the required increase in r must be smaller than that in $\hat{S}$. To see why, notice that if r were to rise by exactly as much as $\hat{S}$, the differential between the rate of return on domestic and foreign bonds would be unchanged, but because the domestic real interest rate would rise while the expected rate of inflation is unchanged, the domestic *nominal* interest rate $r + \pi$ would be higher. Because this would reduce the demand for money, it would increase the demand for domestic bonds, resulting in a situation of excess demand in the domestic bond market. Thus a lower real interest rate would be required to sustain equilibrium. The upshot is that the equilibrium real interest rate in the domestic bond market must rise by *less* than the increase in $\hat{S}$. Because the FF curve must shift vertically by the amount of the change in the real interest rate that clears the domestic bond market at a given value of the real exchange rate, this means that the FF curve must shift vertically upward by less than the increase in $\hat{S}$. Because the GG curve is unchanged, this means that an increase in $\hat{S}$ must result in an increase in the domestic real interest rate that is smaller than the increase in $\hat{S}$ as well as in a depreciation of the equilibrium real exchange rate, as at the point B in Figure 8.4.[3]

5. Changes in the Expected Future Rate of Inflation

The effects on the real interest rate and real exchange rate of an increase in the expected rate of inflation are shown in Figure 8.5. Once again, because the goods-market equilibrium condition (8.4a) does not depend on the expected rate of inflation, changes in π do not affect the GG curve. To see what happens to the FF curve, notice that the expected rate of inflation enters the bond-market equilibrium

[3] It is worth pausing to note the effect of changes in capital mobility on this result. As capital mobility increases, the FF curve becomes flatter because the interest rate differential required to sustain equilibrium in the domestic bond market becomes smaller. Similarly, the size of the interest rate differential required to sustain domestic bond-market equilibrium in the face of an increase in $\hat{S}$ also becomes progressively smaller, implying larger vertical shifts in the FF curve. In the limit with perfect capital mobility, the FF curve is given by

$$r = r^* + \hat{S} - \pi \tag{8.4c}$$

You can see that this curve is flat and must shift upward by the full amount of the increase in $\hat{S}$. In this case, the domestic interest rate must increase by the full amount of the increase in $\hat{S}$.

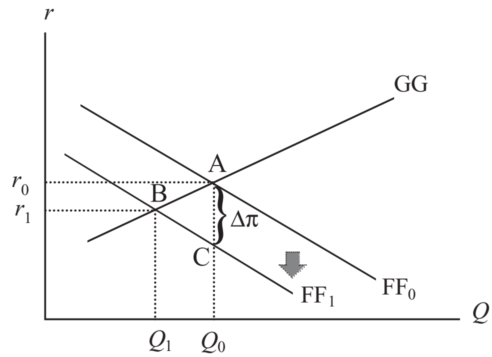


Figure 8.5. Effects of an increase in expected inflation

condition only as a component of the nominal interest rate $r + \pi$, so an increase in π that is exactly offset by a reduction in the real interest rate r at a given value of the real exchange rate would leave the nominal interest rate R unchanged and thus leave the equilibrium in the domestic bond market undisturbed. An increase in π must therefore shift the FF curve *downward* by exactly the same amount as the change in π, say, by the distance AC in Figure 8.5.

The economy's new full-employment equilibrium is at the point B, where the new FF curve intersects the GG curve. As you can verify from Figure 8.5, the increase in the expected rate of inflation reduces the domestic real interest rate by *less* than the change in the expected rate of inflation, implying that the domestic *nominal* interest rate must increase (though, again, by less than the rate of inflation). Because the real interest rate falls, increasing aggregate demand, the real exchange rate must appreciate to sustain equilibrium in the goods market.[4]

6. Offsetting Changes in Expected Depreciation and Inflation

But what we are really interested in in this section is examining the effects of the process described in Section I, in which equal rates of exchange rate devaluation and inflation recur period after period. To do so, then, consider the effects of equal increases in $\hat{S}$ and π. As we have just seen, the increase in $\hat{S}$ causes the FF curve to shift upward by less than the change in $\hat{S}$, whereas the increase in π causes it to shift downward by an amount that is exactly equal to the increase in π. On net, therefore, this combination of shocks must cause the FF curve to shift downward by *less* than the increase in π. Because the GG curve is not affected by either of these variables, it remains undisturbed. Qualitatively, therefore, the new equilibrium looks like that in Figure 8.5: the domestic real interest rate falls, but by less than the increase in the

[4] As we have seen, when capital mobility is perfect, the FF curve is flat. As you can verify from equation (8.4c), however, its downward shift in response to an increase in the expected rate of inflation will still be equal to the change in the expected rate of inflation in this case. Consequently, the domestic real interest rate will fall by the full amount of the increase in the expected rate of inflation in this case, and the effect on the equilibrium value of the real exchange rate will be magnified.

expected rate of inflation. As a result, the nominal rate rises. Because reduction in the real interest rate increases demand for domestic goods, the real exchange rate must appreciate.

The important point is that under imperfect capital mobility, the ongoing rates of exchange rate depreciation and inflation matter, that is, they have real effects on the domestic economy. In particular, in a macroeconomic equilibrium in which the GM and BB curves are continually shifting to the right in (P, R) space, and in which economic agents expect these constant rates of depreciation and inflation to continue, the constant value of the domestic nominal interest rate must be higher the higher the ongoing rates of depreciation and inflation. In other words, the continuous rightward shifts of GM and BB will cause the two curves to intersect at a higher constant value of R. At the same time, the domestic real interest rate must be lower, and because this induces an increase in demand for the domestic good, the constant value of the real exchange rate will be more appreciated the higher the ongoing rate of inflation.[5]

IV. FINANCIAL STOCKS AND MEDIUM-RUN EQUILIBRIUM

The analysis of the last two sections provides two steps toward understanding the macroeconomic effects of sustained inflation because it expands our model to take into account both full nominal wage adjustment to changes in the price level as well as the role of expectations about future exchange rate and price-level changes. As we have just seen, such expectations affect the economy's full-employment equilibrium configuration.

However, as explained in the introduction to this chapter, the full-employment equilibrium that we have just analyzed may not be sustained over time because certain financial stocks that influence that equilibrium may themselves be changing. If they do, the economy's full-employment equilibrium would have to change along with them. If the adjustments in these financial stocks are sufficiently rapid relative to those of the technology and capital stock, then it is meaningful to talk about a *medium-run equilibrium*, in which we continue to take the level of technology and capital stock as given, but in which relatively rapidly adjusting macroeconomic variables such as the nominal wage and financial stocks have reached a stable level so that the economy's macroeconomic situation would tend to persist for some time. In this section, we will consider how the analysis of full-employment equilibrium

[5] The result is different, however, when capital mobility is perfect. In this case, the upward shift in the (flat) FF curve caused by an increase in $\hat{S}$ is exactly offset by the downward shift caused by the increase in π (see equation (8.4c)), leaving the FF curve undisturbed. The full-employment equilibrium determined at the intersection of the GG and FF curves is therefore also undisturbed, and the equilibrium values of the real interest rate and real exchange rate are unchanged. In this case, ongoing depreciation and inflation simply increase the domestic nominal interest rate by an amount equal to the rate of ongoing depreciation.

is affected by the adjustment in these financial stocks, and what the economy's full medium-run equilibrium looks like when these adjustments are completed.

1. The Medium-Run Model

To see how this process works in our model, let us begin by identifying which financial stocks we actually have to consider. To that end, examine equation (8.5b) once again. According to that equation, the economy's bond-market equilibrium at any given time depends on the three predetermined financial stocks IIP^*, B/S, and F^*. Under the small-country assumption, however, the behavior of F^* is unaffected by what happens in the domestic economy, so we can take F^* as an exogenous variable and, for the sake of simplicity, assume that it is constant over time. This leaves two predetermined financial stocks that affect the bond-market equilibrium condition: IIP^* and B/S. Because IIP^* and B/S are predetermined variables that may change gradually over time even when the economy is in full-employment equilibrium, the bond-market equilibrium condition would itself change endogenously over time. If so, then the full-employment equilibrium that we have been analyzing so far in this chapter must be changing as well. What this means is that to find the economy's full medium-run equilibrium, we have to find the conditions under which IIP^* and B/S reach stable values.

To do so, we need to describe how IIP^* and B/S evolve over time. Consider IIP^* first. Recall from Chapter 2 that the domestic economy acquires net claims on the rest of the world when it runs a surplus on the current account of its balance of payments because in that case, the rest of the world can pay for the excess of its purchases from the home country over sales to the home country only by borrowing from the home country. The foreign-currency value of the home country's current account surplus CA^* is the foreign-currency value of its net exports $Q^{-1}\,NX$ (recall that we have set $P^* = 1$) plus its net receipts of interest payments from the rest of the world, consisting of its interest receipts on foreign bonds held by the private sector and the central bank, $r^*(F_P^* + F_C^*)$, minus the foreign-currency value of the domestic government's interest payments on its bonds held by foreigners, $(r+\pi)B_F/S$. Thus we can write

$$\begin{aligned}\Delta IIP^* = CA^* &= [Q^{-1}NX + r^*(F_C^* + F_P^*) - (r+\pi)B_F/S] \\ &= [Q^{-1}NX + r^*IIP^* - (r+\pi-r^*)B_F/S] \\ &= [Q^{-1}NX(Q, r, ..) + r^*IIP^* - (r+\pi-r^*) \\ &\qquad \times b^*(r+\pi-r^*)(F^* - IIP^* - P^*L^*)\end{aligned} \tag{8.6}$$

To derive a corresponding relationship for B/S, note first that

$$\begin{aligned}\Delta(B/S) &= \Delta B/S - (\Delta S/S)\cdot B/S \\ &= \Delta B/S - \hat{S}B/S\end{aligned}$$

In turn, changes in B over time are given by

$$\Delta B = P(G - T) + RB$$

This equation, which is called the *government budget constraint*, is expressed in units of domestic currency. It states that the government issues bonds (ΔB is positive) when it needs to finance an overall deficit in its budget, consisting of a *primary* (noninterest) *deficit* $P(G - T)$ plus interest payments on its outstanding debt RB. Putting these two equations together, we have

$$\Delta(B/S) = (P/S)(G - T) + RB/S - \hat{S}B/S \tag{8.7}$$

Equations (8.6) and (8.7) describe how IIP^* and B/S change over time. The current *levels* of these two stocks are the sums of all these past changes:

$$IIP_t^* = \sum_{j=0}^{\infty} CA_{t-j}^* \tag{8.8a}$$

$$B/S = \sum [(P/S)(G - T) + (R - \hat{S})B/S] \tag{8.8b}$$

IIP^* and B/S are both predetermined variables in the sense that because they are the sums of all these past changes, their current values are little affected by what happens in the economy in the current period.

To study how the economy adjusts over time to its medium-run equilibrium, let us restrict ourselves for the present to a situation in which the government runs a fiscal deficit that is just large enough to keep the foreign exchange value of its debt stock constant. This requires the government to issue new debt each period equal to the loss in foreign exchange value that its nominal debt stock experiences as the result of ongoing nominal exchange rate depreciation; that is, the fiscal deficit is equal to

$$\begin{aligned} \Delta B/S = \hat{S}B/S &= (P/S)(G - T) + RB/S \\ &= Q^{-1}(G - T) + (r + \pi)B/S \end{aligned} \tag{8.9a}$$

Under this condition, B/S will be constant in equation (8.5b). However, to make sure that this condition holds continuously, the government will have to make changes in its budget whenever there are shocks to the economy that alter any of the endogenous variables in equation (8.9a). If we assume that the variable that the government adjusts is its tax revenue T, then equation (8.9a) implies that tax revenues must be endogenous and given by

$$T = G + Q(r + \pi - \hat{S})B/S \tag{8.9b}$$

In words, the government must alter its tax revenues to cover any changes in its debt-servicing costs.

If the government behaves in this way, and if the central bank keeps the stock of domestic credit constant measured in foreign-currency terms so that

$\Delta(B_C/S) = 0$, the economy will be in a medium-run equilibrium when its international investment position IIP^*, which becomes an endogenous variable in the medium run, satisfies

$$0 = [Q^{-1}NX(Q, r, \ldots) + r^*IIP^* - (r + \pi - r^*)b^*(r + \pi - r^*) \times (F^* - IIP^* - P^*L^*) \tag{8.10}$$

This relationship is referred to as the economy's *external balance condition.*

Our complete medium-run model thus consists of the full-employment model comprising the goods-market equilibrium condition (8.4a) and the bond-market equilibrium condition (8.5b) plus the external balance condition (8.10) and the fiscal rule (8.9b). This model determines the medium-run equilibrium values of four endogenous variables: the domestic real interest rate r, the real exchange rate Q, the government's tax revenue T, and the country's international investment position IIP^*. Notice that what we have added to the model of the last section to take us from a full-employment equilibrium to a full medium-run equilibrium are the external balance condition (8.10) and fiscal rule (8.9b). The role of these two conditions is to pin down the medium-run values of the two financial stocks.

2. Solving the Model

Because the medium-run model contains four equations in four endogenous variables, to solve it graphically in two dimensions, we have to eliminate two of the endogenous variables, collapsing it into a system of two equations in two unknowns. A particularly revealing way to do so is the following.

First, substitute the fiscal rule (8.7) into the goods-market equilibrium condition (8.4a). The result is

$$Y_P = \phi(Q, \ldots)A_P[Y_P - (G + Q(r + \pi - \hat{S})B/S), r, \ldots] + G + X(Q, \ldots)$$

This equation captures the effects of the government's fiscal rule on equilibrium in the goods market. Any shocks that change the government's debt-servicing costs $(G + Q(r + \pi - \hat{S})B/S)$ require tax changes that affect goods-market equilibrium. How does this affect the properties of that equilibrium? Notice that because the marginal propensity to absorb is less than unity, it remains true that an increase in government spending G is expansionary. Moreover, because an increase in the real interest rate r would increase the government's debt-servicing costs and therefore require a tax increase, it remains true that an increase in the real interest rate is contractionary. But there are two more significant changes:

1. Because $(r + \pi - \hat{S})B/S$ is positive when the government is a net debtor and the nominal return on its debt is positive in foreign-currency terms, a

depreciation of the real exchange rate Q increases the government's debt-servicing costs and thus implies a tax increase that is contractionary. This is an important real-world problem: when the government issues domestic-currency debt, a reduction in the domestic price level increases its real debt-servicing costs and requires some kind of fiscal adjustment. The complication is that this negative effect of Q on private absorption is contractionary and therefore works in the opposite direction to the expansionary effect that real depreciation typically would have on the demand for domestic goods.

2. Aggregate demand for domestic goods now depends *negatively* on $\pi - \hat{S}$, the difference between domestic inflation and the rate of exchange rate depreciation. If we were to include these effects in the GG curve of Section III, this would complicate our analysis of the separate effects of changes in $\hat{S}$ and π on the economy's full-employment equilibrium because changes in these variables would then shift the GG curve as well as the FF curve.

However, notice that the magnitude of these effects depends on the size of the government's constant debt stock B/S. The larger that stock is, the more important these effects will be, and the smaller that stock is, the smaller they will be. Because the size of that stock is essentially arbitrary in our model, we are free to consider different possible cases. To keep things as simple as possible, we will focus on the case $B/S = 0$, in which neither of the effects mentioned earlier is present. In this case, the government's fiscal rule reduces to the simple form $T = G$.

Next, we can use the goods-market equilibrium condition (8.4b) to substitute for the real interest rate in the bond-market equilibrium condition (8.5b). The resulting equation contains two endogenous variables: the real exchange rate Q and the international investment position IIP^*. We can solve it for the real exchange rate as a function of the international investment position and of all the exogenous variables that affect the economy's full-employment equilibrium. That solution, the *reduced form* for the real exchange rate that corresponds to internal balance, takes the following form:

$$Q = S(\underset{-}{IIP^*}; \underset{+}{\hat{S}}, \underset{-}{\pi}, \ldots) \tag{8.11}$$

where all the remaining exogenous variables that affect the economy's full-employment equilibrium are subsumed into the ellipses. This is simply an expression for the real exchange rate that is determined by the intersection of the GG and FF curves. Note that increases in IIP^* cause the FF curve to shift down in (Q, r) space but leave the GG curve unchanged. This causes the full-employment equilibrium value of the real exchange rate to appreciate and accounts for the negative sign under IIP^* in equation (8.11). The effects of changes in $\hat{S}$ and π on the full-employment equilibrium value of the real exchange rate were derived in the last section and remain valid as long as $B/S = 0$.

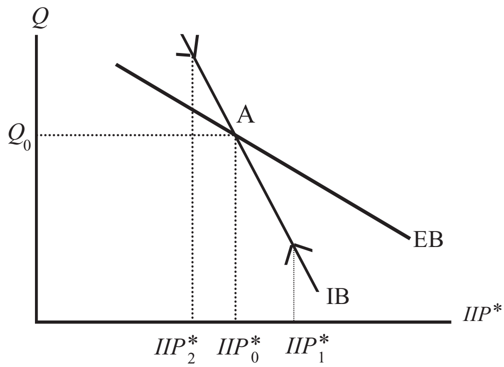


Figure 8.6. Medium-run equilibrium

Because equation (8.11) indicates that increases in IIP^* cause the full-employment equilibrium real exchange rate to appreciate, we can depict this equation graphically in (IIP^*, Q) space in the form of the negatively sloped IB (for "internal balance") curve in Figure 8.6. The positive sign under $\hat{S}$ and negative one under π in equation (8.11) indicate that this curve is shifted upward by an increase in $\hat{S}$ and *downward* by an increase in π.

Next, consider the external balance relationship (8.10). It has the following properties:

- A real exchange rate depreciation affects the foreign-currency value of the current account balance through two channels: on one hand, it increases net exports NX by inducing expenditure switching away from foreign and into domestic goods; on the other hand, if net exports measured in units of domestic goods are originally positive ($NX > 0$), it decreases their value measured in foreign currency. The first effect causes the current account balance to improve, and the second causes it to deteriorate. If domestic and foreign goods are close substitutes and net exports are initially small, the first effect will dominate. We will assume that to be the case.
- An increase in the domestic real interest rate r increases net exports through negative expenditure-reducing effects on imports but, at the same time, increases interest payments to foreigners who hold domestic bonds. The first effect improves the current account, whereas the second causes it to deteriorate, making the net effect on the current account of an increase in the domestic real interest rate ambiguous.
- If the domestic nominal interest rate exceeds the foreign rate ($r + \pi > r^*$), which is typically the case for developing countries, an improvement in the country's international investment position (an increase in IIP^*) improves the country's current account balance by increasing its net interest receipts from (or reducing its net interest payments to) the rest of the world.

Using the goods-market equilibrium condition (8.4b) once again to eliminate r from this equation, we can derive a second relationship between the real exchange rate and the economy's international investment position, in the following form:

$$0 = [Q^{-1}\mathrm{NX}(Q, r(Q), \ldots) + r^{*}IIP^{*} - (r(Q) + \pi - r^{*})b^{*}(r(Q) + \pi - r^{*}) \\ \times (F^{*} - IIP^{*} - P^{*}L^{*})$$

In this form, a depreciation of the real exchange rate affects the current account not just through its effects on net exports, as considered earlier, but also through its effects on the real interest rate required to sustain equilibrium in the goods market. Because a real depreciation requires an *increase* in the real interest rate to maintain goods-market equilibrium, the resulting expenditure-reducing effect magnifies the positive effect of real depreciation on net exports and the current account. On the other hand, a higher domestic interest rate increases interest payments to foreigners who hold domestic bonds, causing the current account to deteriorate. If we assume that the effects on net exports dominate those on interest payments, we can summarize the properties of this relationship as follows:

$$Q = m(\underset{-}{IIP^{*}}; \underset{+}{\pi}, \ldots) \tag{8.12}$$

Because an increase in IIP^{*} improves the current account, it requires an appreciation of the real exchange rate to reduce net exports and sustain a balanced position in the current account. On the other hand, because an increase in the domestic inflation rate increases the domestic nominal interest rate and therefore increases interest payments to foreigners, it causes the current account balance to deteriorate and requires a depreciation in the real exchange rate to increase net exports and restore a balanced current account. The negative effect of IIP^{*} on Q in equation (8.12) allows us to depict that equation as the negatively sloped curve EB (for "external balance") in Figure 8.6.

Notice an important property of the EB curve. Positions *above* the EB curve are associated with more depreciated values of the real exchange rate than are required to maintain a balanced current account. Thus, above the EB curve, the current account must be in *surplus*, and IIP^{*} must be *rising*. Similarly, positions *below* the EB curve are associated with more *appreciated* values of the real exchange rate than are required to maintain current account balance, so below the EB curve, the current account must be in *deficit*, and IIP^{*} must be *falling*.

This observation helps us determine the relative slopes of the IB and EB curves. When the economy is in full-employment equilibrium, it must always be on the IB curve, whatever the prevailing value of its international investment position. That means that it can only be on the EB curve where the IB curve intersects it. Thus the point of intersection of the IB and EB curves (at A in Figure 8.6) is the economy's medium-run equilibrium point, determining the medium-run equilibrium values of the real exchange rate Q_0 and the country's international investment position

IIP^*_0. Notice that if the economy is initially in a full-employment equilibrium with an international investment position such as IIP^*_1, which exceeds IIP^*_0, it can only attain its medium-run equilibrium if IIP^* falls over time. Similarly, if it starts in a position such as IIP^*_2, which is below IIP^*_0, it can only move to medium-run equilibrium if IIP^* rises over time. This means that for the medium-run equilibrium to be stable – that is, for the economy to return to it if it is moved away from it by a shock – the economy must find itself *below* MR when $IIP^* > IIP^*_0$ and *above* it when $IIP^* < IIP^*_0$>. The implication is that the IB curve must be *steeper* than the EB curve, as in Figure 8.6. Under these conditions, the economy will move along the IB curve to its medium-run equilibrium whenever it starts from an initial international investment position not equal to IIP^*_0, as indicated by the arrows in Figure 8.6.

V. DYNAMICS OF ADJUSTMENT

We are now ready to investigate how the economy adjusts to its medium-run equilibrium in response to shocks. For now, as indicated in the introduction to this chapter, we will focus on monetary and exchange rate shocks. We will consider the effects of other shocks when we study the determination of the equilibrium real exchange rate more closely in Chapter 16.

1. Changes in Domestic Credit

Consider first the effects of a one-time expansion in the stock of domestic credit B_C. As we have seen previously, an increase in B_C increases the price level in full-employment equilibrium and, given the nominal exchange rate, causes the economy's full-employment equilibrium real exchange rate to appreciate. Because the equilibrium real exchange rate appreciates at the given initial value of IIP^*, the increase in domestic credit must cause the IB curve to shift downward, as in Figure 8.7. Because B_C does not affect the external balance condition, the EB curve is undisturbed.

How does the economy respond? Because IIP^* does not change on impact, the economy jumps from its initial equilibrium at A to a point B directly below A on the new internal balance curve IB_1. Because this point lies below the EB curve, the current account must move into deficit from an initial position of equilibrium. As a result, IIP^* begins to fall over time, and the economy moves to the northwest along IB_1, with a gradually depreciating real exchange rate and a gradually decreasing current account deficit. What causes this adjustment is a gradual contraction in the domestic money supply driven by central-bank reserve losses as IIP^* falls. This process continues until a new medium-run equilibrium is reached at the point C on the EB curve, where the current account has returned to equilibrium. Notice a curious result: in this new medium-run equilibrium, the real exchange rate must

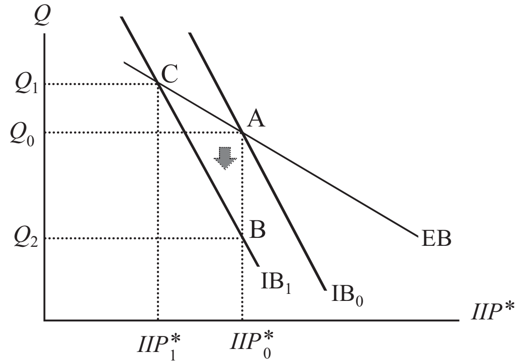


Figure 8.7. Medium-run effects of an increase in domestic credit

actually be more *depreciated* than before the shock. Because the nominal exchange rate is constant, this means that the domestic price level ultimately has to *fall*.

This should seem puzzling. Remember that an expansion of domestic credit increased the domestic price level and lowered the domestic interest rate in full-employment equilibrium. That is indeed what happens at the point B. Yet we have just found that in the medium run, these results must be reversed. Why should this be?

The answer is that, as we have seen, the entire trajectory of adjustment from B to C is characterized by a current account deficit and therefore by a continuous reduction in IIP^*. The new equilibrium, therefore, must be reached with a lower value of IIP^* than the original one. But this means that the country will have larger debt service payments to make in the new equilibrium. To be able to finance such payments, it must generate a *surplus* in its net exports. But that in turn requires some combination of *depreciation* of the real exchange rate and a higher domestic interest rate, which can come about only if the price level in the new equilibrium is actually lower than that in the original one. The important implication is that not only does the domestic credit expansion not increase the domestic price level in the medium term but it must actually *decrease* it to achieve the required depreciation in the real exchange rate.

What we have just examined are the effects of a *one-time* increase in B_C, after which the central bank is content to let the money supply adjust endogenously; that is, implicitly, the central bank is pursuing a domestic credit rather than a monetary target. But what would have happened in this case if the central bank had been pursuing a monetary rather than a domestic credit target?

In that case, the initial credit expansion would have been intended as a permanent increase in the money supply, and the central bank would have been compelled to sterilize the reserve outflows that emerge at the point B by repeated credit expansion. Because sterilization would have prevented the money supply from contracting, the domestic price level would remain unchanged after its

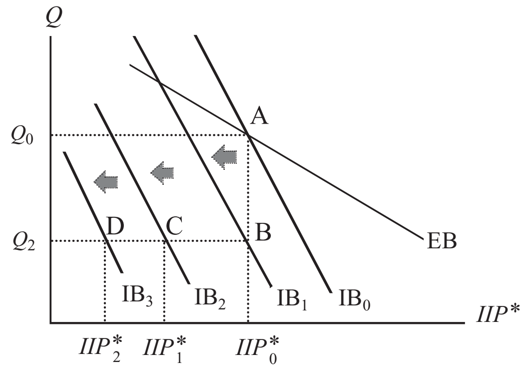


Figure 8.8. Continuous credit expansion

initial increase, and the full-employment equilibrium value of the real exchange rate would have remained at Q_2. However, it remains the case that IIP^* must be falling continuously because it is falling at B, and with the real exchange rate unchanged, the economy must be moving continuously to the left, which keeps it continuously below EB. How does this leftward movement happen? The answer is that the continuous credit expansion required to sterilize the reserve losses must keep the IB curve moving down and to the left, as shown in Figure 8.8. As the IB curve drifts continuously to the left, the economy moves through points like B, C, and D along the horizontal line through Q_2, with continuous contraction in IIP^* and reserve losses offset by continuous monetary expansion.

Notice that this situation cannot be sustained indefinitely, however. The reason is that the central bank would eventually run out of reserves, making the fixed exchange rate unsustainable. Continuous credit expansion to sterilize reserve outflows would therefore culminate in a *currency crisis* in which the central bank is ultimately forced to abandon the fixed exchange rate regime. We will come back to this topic in Chapter 27.

2. Exchange Rate Policy

Next, consider the case of a nominal devaluation. Note first from equation (8.12) that the EB curve is not affected by this shock. We saw at the end of Section I that the effect on the economy's full-employment equilibrium – and thus on the IB curve – depends on the extent to which the devaluation is monetized. Suppose it is not monetized at all. In this case, the effect of the devaluation is to shift the IB curve upward less than in proportion to the change in the exchange rate because a less-than-proportionate increase in the domestic price level partly offsets some of the effect of the devaluation on the real exchange rate. The economy jumps from A to B in Figure 8.9, and the current account moves into surplus. Over time, the current account surplus is associated with an increase in the central bank's

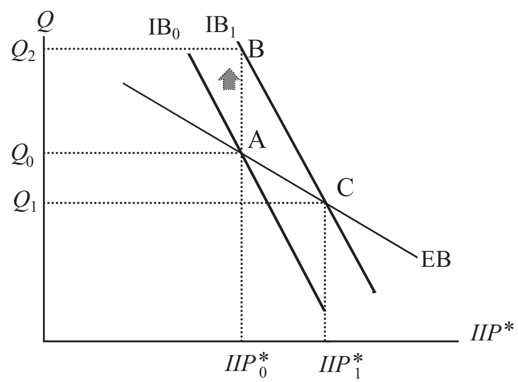


Figure 8.9. Medium-run effects of a nominal devaluation

foreign exchange reserves and, under domestic credit targeting, with a continuous increase in the money supply that drives up the domestic price level, gradually causing the real exchange rate to appreciate and the current account surplus to fall, all the while improving the economy's international investment position. In the new medium-run equilibrium, this improvement in the economy's international investment position must be associated with an *appreciated* medium-run equilibrium real exchange rate, implying that the domestic price level must ultimately increase *more than in proportion* to the change in the exchange rate. Again, the change in the economy's international investment position over time causes the medium-run equilibrium to reverse the impact effects of the shock. A similar result would follow if the devaluation were monetized but not accompanied by domestic credit expansion.

Again, it is interesting to ask what would happen if instead of just a one-time change in the exchange rate, the central bank were to decide to devalue repeatedly. Suppose, for example, that the bank adopts as its objective to keep the real exchange rate constant at a targeted value such as Q_2. In this case, because the domestic price level tends to rise at the point B, and the real exchange rate tends to appreciate, this would require continuous nominal exchange rate depreciation to keep the real exchange rate constant. A policy of continuously adjusting the nominal exchange rate to offset the effects of domestic inflation on the real exchange rate, thus keeping the real exchange rate constant, is called *real exchange rate targeting*. Notice that because the central bank has to depreciate the nominal exchange rate continuously, in this case, the IB curve drifts continuously to the *right*, as in Figure 8.10, and the economy moves along a horizontal line passing through the real exchange rate target Q_2, through points like B and C. It experiences a persistent current account surplus and a permanently higher rate of inflation. What drives domestic inflation in this case is the continuous increase in the money supply caused by continuous increases in the central bank's foreign exchange reserves arising from persistent

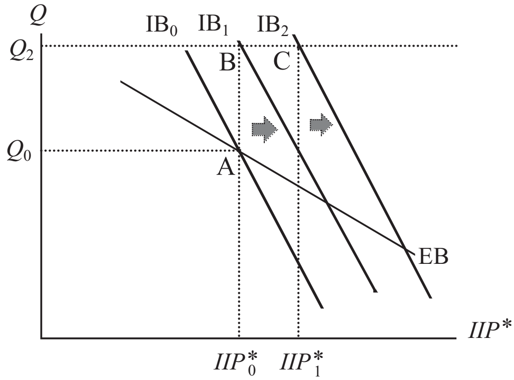


Figure 8.10. Continuous depreciation

increases in IIP^*. The resulting inflation rate will be higher the larger the current account surplus caused by the original devaluation.

VI. SUSTAINED INFLATION

Is there a limit to this process? In principle, the answer is no, because international reserves have no natural upper bound and therefore could be accumulated indefinitely. However, the economy would have to live with a permanently higher inflation rate and permanently lower level of absorption relative to income as the result of its sustained current account surplus, neither of which may be viewed as desirable. If the central bank wants to retain its real exchange rate target but avoid inflation, it would have to sterilize the effects of the sustained increase in IIP^* on the money supply, that is, it would have to continuously *decrease* B_C. This would keep the economy moving to the right along the horizontal line through Q_2 as the IB curve drifts to the right, but instead of a continuously rising price level, the economy would experience a continuous reserve buildup that is offset in the central bank's balance sheet by a contraction in its stock of domestic credit, thereby keeping the money supply constant. Although the central bank may eventually run out of government bonds to sell for sterilization purposes, it could continue to sterilize by issuing its own bonds. The limits on this policy may ultimately be fiscal ones because the interest rate on the bonds the central bank has to sell to pursue its sterilization operations are typically higher in developing countries than is the interest rate on the foreign bonds that it acquires as reserves. We will come back to this issue in the next part of the book.

To generate sustained inflation without the "problem" of reserve accumulation requires a different combination of policies. Consider in particular the policy we analyzed at the end of Section II: a monetized continuous depreciation of the domestic currency combined with an expansion of the rate of growth of domestic credit at the same rate as the rate of depreciation, so that the money supply increases each period at the same rate as the rate of depreciation. In this case, as we saw in Section II, the full-employment GM and BB curves shift out in the same proportion as the rate of depreciation, the domestic price level rises in the same proportion as well, and the full-employment equilibrium real exchange rate remains unchanged. Because the full-employment equilibrium real exchange rate is unchanged, the IB curve does not move in this case, and because the EB curve is not affected by changes in either S or B_C (as you can verify from equation (8.12)), the economy's medium-run equilibrium is undisturbed as well. This means that the current account of the balance of payments remains in balance – that is, IIP^* is constant – and there is no reserve accumulation. As indicated earlier, the money supply increases in the same proportion as the rate of depreciation in this case, but the increase in the money supply each period arises from a combination of monetization of the capital gains on a fixed stock of foreign exchange reserves and domestic credit expansion. Repeated period after period, this policy thus results in a sustainable increase in the domestic price level in every period that is exactly equal to the fixed rate of depreciation of the domestic currency and rate of credit expansion.

However, all these results are based on our maintained assumption that the expected rates of depreciation $\hat{S}$ and expected rate of inflation π are exogenous. This seems more than a little unrealistic if, for example, the economy were indeed to experience a permanently higher rate of exchange rate depreciation and inflation as the result of a policy change. Under rational expectations, for example, if the change in policy is announced and understood, we would expect that the expected rates of depreciation and inflation would adjust quickly to the new sustainable values of the actual rates of exchange rate depreciation and inflation. In other words, if, say, from a policy of keeping the nominal exchange rate and stock of domestic credit literally fixed, the central bank were to announce a change to a permanent rate of depreciation and rate of credit expansion of x percent, we would expect for $\hat{S}$ and π ultimately to adjust to their new permanent values of x percent each. How would allowing for the endogenous formation of depreciation and inflation expectations affect our previous results?

To address this issue, consider the case just described: assume that the central bank increases the rates of depreciation and credit expansion from zero to x percent, and assume that $\hat{S}$ and π *immediately* adjust to this new value when the new policy is announced and implemented. As we have seen, the policies themselves do not affect the IB and EB curves when $\hat{S}$ and π are held constant, but changes in $\hat{S}$ and π themselves do alter the positions of both curves and therefore both the economy's full-employment and medium-run equilibria.

In particular, recall from Section III that equal changes in the expected rate of depreciation of the domestic currency and the expected rate of inflation are expansionary: they cause an appreciation of the full-employment equilibrium real exchange rate and thus induce a downward shift in the IB curve, as in Figure 8.11. At the same time, while a change in $\hat{S}$ has no effect on the EB curve, an increase in

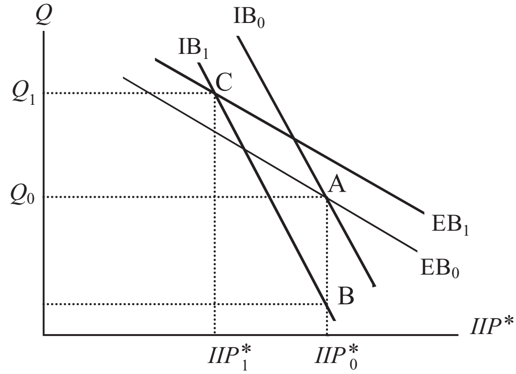


Figure 8.11. Medium-run effects of changes in expected depreciation and inflation

the expected rate of inflation increases the domestic nominal interest rate, causing interest payments to foreigners to rise and necessitating a depreciation of the real exchange rate to keep the current account in balance. The EB curve thus shifts upward in Figure 8.11.

On impact, therefore, the economy jumps from A to B, with an appreciation of the full-employment equilibrium real exchange rate and a current account deficit. Over time, the sustained current account deficit causes the economy's international investment position to deteriorate until a new medium-run equilibrium is reached at the point C, with a depreciated medium-run equilibrium real exchange rate and a reduced international investment position. At the point C, the actual and expected rates of depreciation as well as inflation are equal, and expectations are therefore correct, with a sustained rate of inflation that is equal to the rate of depreciation and credit growth selected by the central bank.

VII. SUMMARY

It appears that the only way to create ongoing inflation in this model, then, is through continuous nominal depreciation. Notice, specifically, that continuous credit expansion with an unchanged nominal exchange rate cannot do the job on its own because it would eventually result in reserve depletion and a currency crisis. How can this be reconciled with Milton Friedman's famous adage that "inflation is always and everywhere a monetary phenomenon"? There are several ways:

- First, the analysis just completed applies to an open economy. In the standard textbook closed-economy case, money growth is indeed the cause of inflation.
- Second, even in an open economy, our analysis applies only to a fixed–exchange rate regime. When the exchange rate is allowed to float, money again becomes the nominal anchor, as we will discuss in Part 5.
- More important, though, consider what would happen even in our open-economy fixed–exchange rate model if domestic credit were expanded

continuously. As we have seen, this would be associated with a continuous loss of reserves. This process can therefore only be sustained until reserves are exhausted. What is likely to happen before then, however, is that the loss of reserves will trigger a devaluation of the exchange rate to replenish the central bank's reserve stock. Thus continuous credit expansion is likely to result in continuous depreciation to preserve the reserve stock. But continuous credit expansion associated with continuous devaluation results in continuous price increases, that is, in inflation. The message is, then, that even in an open economy with a fixed exchange rate, continuous credit expansion would indeed result in increased inflation – not directly, but indirectly, through induced devaluation to prevent reserve depletion.

This leaves open the questions, of course, of how such a process would get started in the first place and what broader implications it might have for the domestic economy. The next part of the book turns to these issues.

REVIEW QUESTIONS

1. Explain how continuous nominal wage adjustment tends to restore full employment to an economy after both expansionary and contractionary shocks. What determines how quickly the economy returns to full employment?
2. Why does a response to a specific permanent shock, such as an increase in government spending, differ in the short run from what it is when the economy returns to full-employment equilibrium?
3. Why might a full-employment equilibrium not tend to be sustained over time?
4. What do we mean by a medium-run equilibrium, and what conditions are necessary for an economy to find itself in such an equilibrium?
5. How does an economy produce a sustained-equilibrium high rate of inflation in the medium run with no change in the equilibrium value of the real exchange rate or in the economy's international investment position?

EXERCISES

1. The use of devaluation as a policy instrument has always been controversial in developing countries. Some have claimed that devaluation does nothing but raise domestic prices in the same proportion, with no beneficial real effects on the economy. Can you describe a situation in which devaluation would indeed raise domestic prices in the same proportion as the change in the exchange rate, even in short-run equilibrium? Would it be true that the devaluation would have no beneficial effects in this case? Explain.
2. Assume that nominal wages are fully flexible in the short run. Would an increase in government spending have a larger (negative) impact on net exports when it

is anticipated or when it is unanticipated? Explain. (Hint: don't forget that there are two ways to assess the impact of a policy change on net exports – by looking at the expression that explains how net exports are determined and by looking at the goods-market equilibrium condition that contains net exports.)

3. Explain why permanent high inflation can only be caused by growth of the money supply.
4. The medium-run equilibrium real exchange rate is defined as the value of a country's real exchange rate that is simultaneously consistent with full employment and "external balance." On the basis of the model we developed in this chapter, can an economy's medium-run equilibrium real exchange rate be affected by fiscal policy? By exchange rate policy? By monetary policy? Explain.
5. In the model that we developed in this chapter, the central bank can directly control two nominal variables: the exchange rate and the stock of domestic credit. Assuming that capital mobility is imperfect, compare the *medium-run* impact of a permanent devaluation of the nominal exchange rate on the domestic price level to that of apermanent increase in the stock of domestic credit. Assume that the exchange rate devaluation is not monetized.
6. Are there any conditions under which a reduction in the money supply could reduce the price level without reducing real output at the same time? What are they?
7. Suppose a new technology that increases total factor productivity becomes freely available. In a world where agents have model-consistent (rational) expectations and nominal wages are fully flexible, explain the short-run effects of this shock on the following variables:
   a. the domestic price level and the interest rate
   b. the level of real output
   c. the real wage, the nominal wage, and the level of employment
   d. private absorption and net exports
8. Consider a small open economy that maintains a fixed exchange rate and is characterized by perfect capital mobility. Suppose the economy has had a stable price level (zero ongoing inflation), but its central bank suddenly undertakes rates of devaluation and domestic credit expansion that are expected to produce an ongoing inflation rate of 10 percent per year. Explain how capital flows in that economy would respond to such a change in central-bank policy.
9. China is sometimes accused by the United States and the European Union of targeting its real exchange rate at a very depreciated level to sustain an unfair competitive advantage in international markets. China indeed maintains an officially determined exchange rate and runs large current account surpluses. What would you expect to happen to the Chinese inflation rate if China indeed keeps its real exchange rate at a very depreciated level? On what does this outcome depend?

PART 3

Public Finance and Macroeconomic Performance

9

The Intertemporal Budget Constraint of the Public Sector

In the previous chapter, we saw that a central-bank policy of continuous credit expansion and exchange rate depreciation would result in ongoing inflation in the medium run. The questions that naturally arise in this context are what would lead the central bank to undertake such a policy and what the benefits and costs associated with it might be for the economy as a whole. This chapter will take up the first of these questions, leaving the second for the chapters that follow. The answer we will give to why the central bank might behave in that way is that credit expansion coupled with monetized exchange rate depreciation – monetary expansion, for short – allows it to finance a portion of the government's fiscal deficit.

But of course, monetary emission is not the only option available to the government for financing fiscal deficits. It can also borrow, both from domestic and foreign private sources. Thus, to understand what drives monetary emission, we will also need to consider what determines how much the government can borrow. To do so, we will analyze the government's *intertemporal budget constraint* (the constraint that must be satisfied by the government's fiscal choices over time) and develop the important concept of *fiscal solvency* (the perceived ability of the government to honor its financial commitments), which lies at the heart of all the issues to be discussed in the second part of this book.

The analysis of fiscal solvency allows us to consider several important related issues. For example, in this chapter, we will examine the calculation of sustainable fiscal deficits, the budgetary consequences of postponing fiscal adjustment, the issue of "unpleasant monetarist arithmetic," and the fiscal consequences of sterilized intervention in foreign exchange markets. In subsequent chapters, we will consider why the requirements of solvency often force fiscal policy to be procyclical in developing countries, in contrast to the countercyclical stance advocated for industrial countries in intermediate macroeconomics textbooks.

This chapter is divided into six sections. In the next section, we derive the relationship between the fiscal deficit and the rate of growth of the money supply. That relationship is incorporated into the derivation of the government's intertemporal budget constraint in Section II. Section III examines how the intertemporal budget constraint affects the fiscal choices that the government must make, and how the choices that it actually does make affect the evolution of the public sector's debt over time. Section IV considers the implications of those fiscal choices for the rate of money growth. That analysis is used in Section V to examine the fiscal implications of inflation stabilization. Section VI summarizes.

I. BUDGET DEFICITS AND MONEY GROWTH

We saw at the end of the preceding chapter that the view that money growth causes inflation is compatible with the assertion that in a small, open economy operating an officially determined exchange rate, inflation can only be caused by exchange rate depreciation. The logical chain was the following: domestic credit expansion causes reserve depletion, reserve depletion causes exchange rate adjustments, and exchange rate adjustments cause inflation. Ongoing inflation is the result of credit expansion accompanied by monetized exchange rate depreciation, which together result in continuous monetary expansion with no change in foreign exchange reserves. Thus so far we have learned *how* ongoing inflation happens in an emerging economy. Now we need to determine *why* inflation happens. The answer we will give in this section is that *inflation happens because the government induces the central bank to print money to finance a fiscal deficit.*

1. The Budget Constraint of the Consolidated Public Sector

To show the link between fiscal deficits and money growth, we begin by reproducing the balance sheets of the government and central bank from Chapter 6:

$$\begin{aligned} W_G &= -B && \text{(government)} \\ 0 &= S \cdot F_C^* + B_C - M && \text{(central bank)} \end{aligned}$$

In our analysis of domestic financial markets in Chapter 6, we assumed that the government only issued bonds denominated in domestic currency. In many developing countries, however, government borrowing from the rest of the world is primarily conducted in foreign currency.[1] Suppose that the stock of government foreign-currency debt, measured in foreign-currency units, is denoted B^*. Then the government's balance sheet can be amended to

$$W_G = -(B + S \cdot B^*)$$

[1] In fact, some economists have coined the term *original sin* to refer to the fact that emerging-market economies – just by virtue of being emerging-market economies – have traditionally been unable to borrow from foreigners in their own currencies. See Eichengreen and Hausmann (1999).

We will assume that the central bank does not hold the government's foreign-currency debt, so the central bank's balance sheet remains unchanged.[2]

Recall from Chapter 6 that the change in the government's outstanding debt arises from its need to finance a budget deficit. Taking into account foreign currency–denominated government debt, and allowing for any transfers that may take place from the central bank to the government (including, e.g., revenues from the monetization of devaluation-induced capital gains), we can write the government's budget constraint as

$$(\Delta B - \Delta B_C) + S \cdot \Delta B^* + \Delta B_C = P \cdot (G - T) - T_c + R \cdot B + R' \cdot S \cdot B^*$$

where T_c denotes the nominal value of transfers received by the government from the central bank, R' is the interest rate on foreign currency–denominated government debt, and the change in the government's total domestic-currency debt ΔB has been separated into the portion purchased by the private sector $(\Delta B - \Delta B_C)$ and the portion purchased by the central bank (ΔB_C). Decomposing transfers to the government from the central bank into a portion that comes from capital gains on reserves (monetization, given by $\Delta S \cdot F_C^*$) and a portion $(T_c - \Delta S \cdot F_C^*)$ that is financed from other sources, the government's demand for monetization revenues and credit from the central bank can thus be expressed as

$$\begin{aligned}\Delta S \cdot F_C^* + \Delta B_C = {} & [P \cdot (G - T) - (T_c - \Delta S \cdot F_C^*) + R \cdot B + R' \cdot S \cdot B^*] \\ & - [(\Delta B - \Delta B_C) + S \cdot \Delta B^*] \qquad (9.1)\end{aligned}$$

That is, the government demands resources from the central bank to cover the excess of its deficit (the first term in square brackets on the right-hand side of equation (9.1)) over its borrowing from the private sector (the second term in square brackets).

Now, just as we did for the government, we can write the budget constraint for the central bank:

$$\Delta B_C + S \cdot \Delta F_C^* = (R^* \cdot S \cdot F_C^* + R \cdot B_C - T_c) + \Delta M \qquad (9.2)$$

[2] What effect would the introduction of the new debt instrument have on the model of Part 2? The answer is that there would now be three assets for portfolio managers to choose among: domestic government bonds denominated in domestic currency, domestic government bonds denominated in foreign currency, and foreign bonds. How that affects the model's behavior depends on the substitutability relationships that prevail among the three types of assets. On the assumption that we will make later – that government domestic- and foreign-currency bonds are perfect substitutes – the only difference would arise through the valuation effects of exchange rate changes on the nominal value of government debt.

This identity states that the central bank can acquire assets by saving (given by the difference between its income in the form of interest earnings and its transfers to the government) or by printing money.[3]

An important question is how decisions are made about the amount of resources that the central bank makes available to the government through monetization and credit expansion, that is, about $\Delta S \cdot F_C^* + \Delta B_C$. In this chapter, we shall assume that the central bank is not *independent* of the finance ministry. This means that decisions about $\Delta S \cdot F_C^* + \Delta B_C$ are made by the finance ministry, not by the central bank.[4] This situation is referred to as one of *fiscal dominance*. In that case, $\Delta S \cdot F_C^* + \Delta B_C$ will be determined by the government's fiscal needs. To see the implications of this situation for the central bank's decisions about monetary emission, we can substitute equation (9.1) into equation (9.2), which yields the following:

$$\Delta M = P \cdot (G - T) + R \cdot (B - B_C) + R' \cdot S \cdot B^* - R^* \cdot S \cdot F_C^* \\ - [(\Delta B - \Delta B_C) + S(\Delta B^* - \Delta F_C^*)] \tag{9.3}$$

Because this substitution in effect consolidates the budget accounts of the government with those of the central bank, equation (9.3) is referred to as the budget constraint of the *consolidated public sector*. It states that the public sector's overall deficit, consisting of the *primary* (*noninterest*) *deficit* $P \cdot (G - T)$ and net interest payments $(R \cdot B + R' \cdot S \cdot B^* - R^* \cdot S \cdot F_C^*)$, must be financed by net borrowing from the private sector in either domestic $(\Delta B - \Delta B_C)$ or foreign currency $S(\Delta B^* - \Delta F_C^*)$ or by printing money ΔM. When the central bank is not independent, the right-hand side of equation (9.3) will determine the left-hand side; that is, *the nonindependence of the central bank means that the level of monetary emission is determined by the government's fiscal needs.*

2. Interest Rate Differentials

So far, this has all been accounting. Here is the first bit of economic substance. Notice that there are three interest rates in the budget constraint of the consolidated public sector – the interest rate on foreign bonds and the two interest rates on bonds issued by the domestic public sector. Differences among these three interest rates mean that *public-sector debt policy* – that is, changes in the composition of the public sector's financial portfolio – affect the fiscal accounts, as explained in Box 9.1. This

[3] Notice that this means that the change in the money supply is given by

$$\Delta M = S\Delta F_C^* + (\Delta B_C + T_c) - R^* S F_C^* + R B_C)$$

i.e., the central bank expands the money supply when the sum of its reserve acquisition and the resources it transfers to the government exceeds its interest earnings.

[4] Central-bank independence is considered at length in Chapter 14.

Box 9.1. Fiscal Implications of Public-Sector Debt Policy

Equation (9.3) has a number of interesting applications regarding the fiscal implications of the composition of the public sector's financial portfolio, that is, of public-sector debt policy.

1. Debt versus Reserve Financing

We can ask what difference it would make, for example, if the public sector were to borrow in foreign currency to finance a deficit or instead use foreign exchange reserves – in other words, if ΔB^* is positive or ΔF_C^* is negative. The difference depends on the relationship between R' and R^*. If the two interest rates are equal, then external borrowing and reserve depletion are equivalent; that is, from the perspective of the public sector's budget, it does not matter whether the public sector borrows abroad to finance a deficit or spends its foreign exchange reserves. This is so because B^* and F_C^* would enter the budget constraint only in the form $B^* - F_C^*$. But if the domestic public sector has to pay a premium above what the public sector in the rest of the world has to pay ($R' > R^*$), then reserve depletion is a cheaper form of finance.

2. Fiscal Implications of Sterilization

Recall that when balance of payments surpluses are sterilized by the central bank, increases in foreign exchange reserves are offset by central-bank sales of government bonds, so B_C contracts to offset increases in $S \cdot F_C^*$. Because the offset must be dollar-for-dollar for the money supply to remain unaffected, it follows that if $R > R^*$, this policy will tend to increase the overall deficit of the consolidated public sector because the assets acquired by the central bank, in the form of foreign exchange reserves, yield a lower return than the domestic-currency bonds that it has sold to the public. The impact of sterilizing the stock of foreign exchange reserves on the public sector's cash flow is given by $-(R - R^*)S \cdot F_C^*$.

However, notice that the true effect on the fiscal accounts is not given by the difference between R and R^* only, because foreign exchange reserves increase in value in the presence of depreciation, whereas domestic bonds do not. Thus the true comparison is between R and $R^* + \Delta S/S$. To see what this involves, note that under the conditions of equation (9.4), $R = R' + \Delta S/S$. Because this makes the previous comparison one between $R' + \Delta S/S$ and $R^* + \Delta S/S$, this means that sterilization of balance of payments surpluses will be fiscally costly if $R' > R^*$, that is, if the domestic government has to pay a premium on its foreign-currency borrowing over the interest rate that it earns on reserves. This is the same condition that determines whether reserve depletion is cheaper than external borrowing.

raises the question, just what is the relationship among the domestic-currency rates of return for these three assets?

As a first step in answering this question, notice that unlike the interest rate R on domestic currency–denominated bonds, the interest rates R' and R^* correspond to assets denominated in foreign currency. The *domestic-currency* rate of return from holding these assets is approximately $R' + \Delta S/S$ and $R^* + \Delta S/S$, respectively, where $\Delta S/S$ is the rate of depreciation of the domestic currency.[5] To make sure that we are comparing apples to apples, we should therefore ask how these three domestic-currency rates of return are likely to be related.

There are two characteristics that distinguish the three assets to which these rates of return correspond: the identity of the borrower and the currency of denomination. Assets that represent financial claims on different economic agents may carry a different probability of nonpayment (credit risk). Those that are denominated in different currencies, in turn, may carry different amounts of exchange rate risk – that is, the risk associated with the possibility that currency values may change. The existence of two different types of risk and the likelihood that creditors may demand compensation for bearing risk mean that the three assets may, in principle, offer very different domestic-currency rates of return. Therefore, to abstract away from some of these complications and sharpen our analysis, we will make two assumptions:

1. Future exchange rate movements are known with certainty. This means that even though $\Delta S/S$ may not be zero, there is no currency risk.
2. All assets issued by the domestic public sector bear the same credit risk. This means that neither domestic currency–denominated nor foreign currency–denominated bonds are "senior" (have a prior claim on the public sector's resources).

With these two assumptions, domestic currency–denominated and foreign currency–denominated bonds of the public sector must yield the same domestic-currency rate of return. In other words, we can impose the no-arbitrage condition:

$$R = R' + \Delta S/S \tag{9.4}$$

This leaves the relationship between the domestic-currency rate of return on the foreign bond $R^* + \Delta S/S$ and that on the domestic bond to be determined. This difference will depend on differences in credit risk. We will assume that the foreign bond carries no credit risk; that is, the domestic-currency rate of return on the foreign bond is the rate of return on a safe asset. To simplify notation, we will call this rate of return R_S^*, so $R_S^* = R^* + \Delta S/S$. The remaining question is what determines the differential credit risk attached to lending to the domestic government. We will come back to this question in the next chapter.

[5] Notice that this means that if it devalues, the domestic government can expect to reap a capital gain on its foreign exchange reserves, as we have discussed previously, but also to incur a capital loss on its foreign exchange–denominated debt. The latter has at times created severe fiscal problems in some emerging-market economies, as we will see in Part 7.

3. Determinants of Public-Sector Borrowing and Money Supply Growth

Returning to the thread of our argument, notice that condition (9.4) allows us to simplify equation (9.3) considerably. Let $D = (B - B_C) + S(B^* - F_C^*)$ denote the domestic public sector's net debt, in other words, its total stock of borrowing from the domestic and foreign private sectors minus its stock of foreign exchange reserves. Noting that the change in D can be written as $\Delta D = (\Delta B - \Delta B_C) + S(\Delta B^* - \Delta F_C^*) + \Delta S(B^* - F_C^*)$, we can use equation (9.4) to write equation (9.3) in the form

$$\Delta D + \Delta M = P(G - T) + (R - R^* - \Delta S/S)S \cdot F_C^* + R \cdot D \qquad (9.5)$$

As pointed out in Box 9.1, the term $(R - R^* - \Delta S/S)\ S \cdot F_C^*$ is essentially the cost to the public sector of sterilizing reserve accumulation.

In what follows, we will be interested in exploring what determines the domestic public sector's ability to borrow. An important factor will be the resources available to the public sector to service debt. Because the sterilization of foreign exchange reserves will absorb some of these resources, it will be useful to define an *adjusted primary deficit*, denoted PD', which takes into account the costs of reserve sterilization. We can express the adjusted primary deficit as follows:

$$PD' = P(G - T) + (R - R^* - \Delta S/S)S \cdot F_C^*$$

That is, the adjusted primary deficit is the primary deficit plus the cost of foreign exchange reserve sterilization. Using this notation, we can now write the budget constraint of the consolidated public sector in the following simpler form:

$$\Delta D + \Delta M = PD' + R \cdot D$$

or

$$\Delta D = PD' + R \cdot D - \Delta M \qquad (9.6)$$

This equation describes what determines the public sector's demand for new borrowing from the private sector: the public sector seeks to borrow to cover the excess of its adjusted fiscal deficit $PD' + R \cdot D$ over the resources that it can obtain by printing new money, given by ΔM. The latter is often referred to as *seignorage revenue*. The question is whether the public sector will be able to satisfy its demand for new borrowing, that is, whether the funds demanded by the public sector will be forthcoming.[6]

[6] Equation (9.6) can also be used to explain the determinants of money growth. Box 9.2 shows how.

Box 9.2. Sources of Money Growth

Equation (9.6) enables us to address directly the question posed at the beginning of this chapter: why does the central bank engage in sustained monetization of devaluation-induced capital gains on its foreign exchange reserves and sustained credit expansion? Monetization and credit expansion are two ways in which the central bank increases the money supply. Why does it do so? Solving equation (9.6) for ΔM, we have

$$\Delta M = PD' + R \cdot D - \Delta D$$

In the context of fiscal dominance, the right-hand side of this equation drives the left. Thus *monetary expansion arises from public-sector deficits not financed by net public-sector borrowing.*

Notice that this does not rule out other sources of monetary expansion that are frequently noted, for example, those arising from balance of payments surpluses or from central-bank lending to commercial banks. Reserve accumulation shows up in the form of a lower value of ΔD, which would tend to increase ΔM, unless the reserve accumulation is sterilized, which would leave ΔD unchanged, as gross debt increases one-for-one with reserve accumulation. Though we have not included commercial banks in our model so far, it is easy to see that any domestic lending by the public sector – such as to commercial banks – would also simply reduce its *net* borrowing from the private sector ΔD and thus increase ΔM, other things equal.

The relationship between money expansion and fiscal deficits is also consistent with the public sector using external borrowing to finance domestic spending. To see this, suppose that the public sector borrows an amount $S\Delta B^*$ externally to finance $P\Delta G$ of domestic spending. Because the domestic currency required for such spending is acquired by depositing the proceeds of the loan at the central bank, the bank's foreign exchange reserves rise by the same amount that the public sector's external liabilities do, that is, by $S\Delta B^*$. This means that the adjusted primary deficit increases by $P\Delta G$ (the nominal value of the increase in spending), while net public-sector borrowing is unchanged, leaving a net effect on monetary expansion of $\Delta M = P\Delta G$. In effect, then, the government has printed money to finance its increased spending.

II. THE PUBLIC SECTOR'S INTERTEMPORAL BUDGET CONSTRAINT[7]

In this section, we will show that the public sector's option to borrow may in fact be limited. This means that borrowing may not be an option to finance fiscal deficits, in which case, deficits must be either eliminated or financed through monetary expansion. As you will see, it also means that a permanent reduction in the rate of monetary expansion, such as would be involved in a permanent decrease in the rate of inflation, requires a fiscal adjustment, that is, a reduction in PD'. Why should this be so? From equation (9.6), one might think that all one needs to do to effect a

[7] For overviews of the material covered in this section, see Buiter (1985, 1990) as well as Easterly and Fischer (1990).

reduction in the rate of monetary expansion is to change the composition of deficit financing from issuing money to issuing debt. The problem, as with an increase in the deficit itself, is that creditors may not be prepared to accommodate such a shift. The basic question in both instances, then, is, how do we know when creditors will be willing to extend loans to the public sector?

In general, the answer to this question depends on who the creditors are assumed to be. Bilateral and multilateral lenders may have a range of noneconomic (usually political) motives for lending to developing-country governments. Such motives will influence both the amount of lending they do and the terms on which they make such lending available. Market creditors, on the other hand, are likely to be motivated by return and risk considerations only. We will focus here on the behavior of market creditors.

Such creditors always have the option, instead of lending to the public sector, of investing their funds elsewhere. Market creditors will therefore be willing to extend loans to the public sector – or to any other debtor, for that matter – if and only if they expect to receive a competitive rate of return, that is, a return comparable to what they could earn elsewhere. Because payments for the loan will be made at some point in the future, in practice, this means that the creditors must expect to receive in return for their funds a stream of future payments that is at least equal *in present value* to the loans that they have extended, where the relevant interest rate for discounting to present values is the creditors' opportunity cost of funds. Because creditors care about both return and risk, this opportunity cost should be the return applying to assets with the same risk characteristics as public-sector debt. For now, we will assume that this is the risk-free interest rate. When the public sector is expected to be able to make future payments to its creditors with a present value that is at least equal to the face value of the public sector's outstanding debt, creditors will willingly hold this debt because they can do no better by putting their funds elsewhere. In this case, the public sector is said to be *solvent.* When the public sector can remain solvent even if it acquires additional debt, then it will be able to borrow.

To evaluate the solvency of the public sector, we therefore have to examine what determines the payments that the public sector can credibly promise to make to its creditors in the future. These payments are typically classified into two categories: interest and amortization payments. Because amortization payments are payments that reduce the debt outstanding, we can denote total net amortization payments made by the public sector during a given period t as $-\Delta D(t)$, that is, the *negative* of the change in the debt over that period (if the debt is increasing, amortization payments are effectively negative, and if it decreases, they must be positive). Total debt service payments made during period t are therefore $R(t)D(t) - \Delta D(t)$. This means that the condition for the public sector to be solvent can be expressed as follows:

$$PV[R(t)D(t) - \Delta D(t);\ R_S^*(t)] \geq D(0)$$

where $PV[\ ]$ denotes the present value of a stream of expected future payments beginning one period from now and potentially extending to infinity, with a representative time-t payment of $R(t)D(t) - \Delta D(t)$ and discount factor $R_S^*(t)$. $D(0)$ is the stock of debt currently in existence. As discussed earlier, $R_S^*(t)$ is the domestic-currency risk-free rate relevant for discounting payments made at time t back to the present.

This solvency condition imposes a constraint on the components of the public sector's budget. To see how, rewrite the public sector's budget constraint as follows:

$$PS' + \Delta M = R \cdot D - \Delta D$$

where, for ease of notation, we have defined the *adjusted primary surplus* of the public sector (just the negative of the adjusted primary deficit) as $PS' = -PD'$. The right-hand side of this expression is, as we have seen, the flow of debt service payments (interest plus amortization) made by the consolidated public sector each period. What this form of the identity tells us is that the resources available to service debt each period are equal to the sum of the adjusted primary surplus of the public sector (PS') and new monetary emission (seignorage revenue). This identity allows us to write the solvency condition as

$$PV[PS'(t) + \Delta M(t);\ R_S^*(t)] \ \geq\ D(0) \tag{9.7}$$

This is also referred to as the consolidated public sector's *intertemporal budget constraint.*

Notice what it means. The right-hand side of the inequality (the existing stock of debt) is given at any instant, so this relationship imposes a constraint on the future paths that can be followed by the sum $PS' + \Delta M$ (hence the name), in the form

$$PV[PS'(t) + \Delta M(t);\ R_S^*(t)]_{\min} = D(0)$$

That is, at a minimum, the government must be expected to raise enough resources in the future to service its existing debt on market terms.[8]

What determines the resources that the public sector will be able to generate to service debt in the future? It is convenient to split the relevant factors into two components, conceptually similar to the familiar decomposition of tax revenue into the product of a tax *rate* and a tax *base.* First, growth in the size of the domestic economy and increases in the domestic price level are likely to be important factors.

[8] Notice that because this condition would be satisfied by debt service payments equal to $-PD' + \Delta M = R_S^*\ D$, which would merely involve making interest payments of $R_S^*\ D$ each period, the condition does not actually require the government to *repay* its debt, just to service it on market terms. As long as it can do so, it can roll it over, if not with an existing creditor, then with a new one, because creditors cannot do any better than to lend to the government.

The former will affect the magnitude of real tax revenues and the demands for real government spending, and the latter will influence both the nominal value of the real primary surplus and the demand for money. In effect, nominal gross domestic product (*GDP*) is the "tax base" for the primary surplus and seignorage revenue. In other words, in an economy that is growing and in which the price level is increasing over time, there will tend to be forces causing PS' and ΔM, both of which are expressed in nominal absolute terms, to change in the future, even with an unchanged fiscal effort (i.e., with no changes in tax laws, tax administration, or the usual procedures for determining government expenditures). A second set of factors can be considered to be those that reflect special fiscal efforts, that is, those that determine the magnitude of fiscal resources extracted from an economy of a given size: the "tax rate."

To capture this decomposition in our solvency condition, we can express the values of the variables in equation (9.7) as ratios to nominal *GDP* rather than in absolute terms. Letting $ps' = PS'/PY$ and $m = M/PY$, we can rewrite equation (9.7) as

$$PV\{[ps'(t) + (\Delta M(t)/M(t))m(t)]P(t)Y(t);\; R_S^*\} \geq D(0)$$

Assume that the rate of growth of real *GDP* (denoted n) and the rate of inflation π are both constant. In that case, the preceding equation becomes

$$PV\{[ps'(t) + (\Delta M(t)/M(t))m(t)][P(0)(1+\pi)^t][Y(0)(1+n)^t];\; R_s^*\} \geq D(0)$$

or

$$PV\{[ps'(t) + (\Delta M(t)/M(t))m(t)][P(0)Y(0)[(1+n)(1+\pi]^t;\; R_S^*\} \geq D(0)$$

Suppose, for simplicity, that $ps'(t)$ and $m(t)$ are set at constant values ps' and m (notice that these are now not functions of t). Because the constancy of the ratio of the money supply to nominal *GDP* (m) means that the rate of growth of the money supply ($\Delta M/M$) must be equal to the rate of growth of nominal *GDP* ($\pi + n$), replacing ($\Delta M/M$) by ($\pi + n$) and dividing both sides of the preceding equation by $P(0)Y(0)$ permits us to write the following:

$$PV\{[ps' + (\pi + n)m][(1+n)(1+\pi)]^t;\; R_S^*\} \geq d(0) \tag{9.8}$$

where $d(0) = D(0)/P(0)Y(0)$ is the initial ratio of debt to nominal GDP. Notice that holding the ratios ps' and m constant in a growing economy means that payments to creditors, measured in nominal absolute terms, grow at the rate of growth of nominal *GDP*.

Finally, to summarize the modified solvency condition in its most compact form, notice that the tth term of the present value expression looks like the following:

$$\frac{[ps' + (\pi + n)m][(1+n)(1+\pi)]^t}{(1+R_S^*)^t}$$

which can be written as

$$\frac{ps' + (\pi + n)m}{[(1 + R_S^*)/(1 + n)(1 + \pi)]^t}$$

However, because the denominator is approximately equal to $(1 + R_S^* - \pi - n)^t$, letting $r = R_S^* - \pi$ denote the *real* interest rate, we can write equation (9.8) in the equivalent form[9]

$$PV\{[ps' + (\pi + n)m];\ r - n\} \geq d(0) \tag{9.9}$$

In other words, discounting a stream of payments $[ps' + (\pi + n)m][(1 + n)(1 + \pi)]^t$ that is growing at the constant rate $\pi + n$ each period using the constant nominal interest rate R_S^* is equivalent to discounting the constant stream of payments $[ps' + (\pi + n)m]$ at the constant rate $r - n$. Because debt service payments are growing at a constant rate (the rate of growth of nominal *GDP*), their present value is equivalent to that of a constant level of debt service payments discounted at the lower rate $r - n$ rather than R_S^*.[10]

Because $ps' + (\pi + n)m$ and $r - n$ are both constant, as long as $r - n > 0$, the solvency condition in equation (9.9) can be expressed as follows:

$$PV(ps' + (\pi + n)m;\ r - n) = (ps' + (\pi + n)m)/(r - n) \geq d(0)$$

This allows us to write the minimum value of the resources that the public sector must be able to devote to debt service as

$$ps' + (\pi + n)m = (r - n)d(0) \tag{9.10a}$$

The left-hand side of equation (9.10a) is the sustainable "permanent" value of the sum of the adjusted primary surplus and seignorage revenue (expressed as ratios to *GDP*). According to this equation, payments each period can fall short of the accrued real interest only to the extent that economic growth allows the debt-to-*GDP* ratio to remain unchanged under new borrowing.

[9] To show the equivalence, note that $(1 + y)/(1 + x)$ is approximately $1 + y - x$. This can be shown by doing the long division of $1 + x$ into $1 + y$ until the remainder consists only of second-order terms.

[10] This form of the identity can also be written as

$$ps' + (\Delta M/M)m = [R_S^* - (n + \pi)]d(0)$$

This says that the stream of payments can be *less* than would be required to stabilize the nominal debt stock; i.e., the nominal debt stock can grow at the rate $(\pi + n)$ over time, without violating the government's solvency. How can this be? The reason is that because the stream of potential payments is growing over time, it can support (have a present value equal to) a growing stock of debt over time. But the only way the latter can emerge is if the payment ratio is less than the rate of interest.

To see what restrictions the solvency constraint imposes on fiscal policy, write equation (9.10a) as

$$ps' = (r - n)\, d(0) - (\pi + n)m \tag{9.10b}$$

This tells us what the permanent value of the adjusted primary surplus must be. According to equation (9.10b), the public sector must generate a larger value of the permanent adjusted primary surplus; that is, it must generate more tax revenues, spend less, or accumulate fewer reserves whenever

- its initial net debt stock $d(0)$ is high
- the market interest rate available to its creditors (r) is high
- the growth rate of the economy n is low
- seignorage revenue $(\pi + n)m$ is low (given the growth rate of the economy, this would happen when the domestic rate of inflation is low and/or the demand for money is low)

III. DEBT DYNAMICS

What does the solvency condition imply about the evolution of the debt-to-*GDP* ratio over time? To answer this question, suppose that creditors confidently expect the government's future fiscal plans to satisfy the solvency condition, and consequently, that the public sector can borrow at the safe rate of interest. Notice that the budget constraint of the public sector (equation (9.6)) can be written as

$$\Delta D(t) = -PS(t)' + R'D(t) - \Delta M(t)$$

Dividing both sides by nominal *GDP* and using the (constant) ratios we defined before, we can rewrite this as

$$\Delta D(t)/P(t)Y(t) = -(ps' + (\pi + n)m) + R_S^* d(t)$$

Using $R_S^* = r + \pi$, subtracting $n\, d(t)$ from both sides, and rearranging terms, leaves

$$\Delta D(t)/P(t)Y(t) - (\pi + n)d(t) = -(ps' + (\pi + n)m) + (r - n)d(t)$$

Finally, because $\Delta D(t)/P(t)Y(t) - (\pi + n)d(t)$ is approximately equal to $\Delta d(t)$, we can write the budget constraint as

$$\Delta d(t) = -(ps' + (\pi + n)m) + (r - n)d(t) \tag{9.11}$$

We can illustrate the financial choices available to the government by plotting the budget constraint (9.11) in $(d, \Delta d)$ space, as in Figure 9.1. It is easy to see that assuming that $r - n > 0$, the curve depicting the budget constraint (9.11) is a straight line in $(d, \Delta d)$ space, with a positive slope equal to the effective interest rate $r - n$ and a vertical intercept at $-(ps' + sgn)$, where $sgn = (\pi + n)m$ is the public

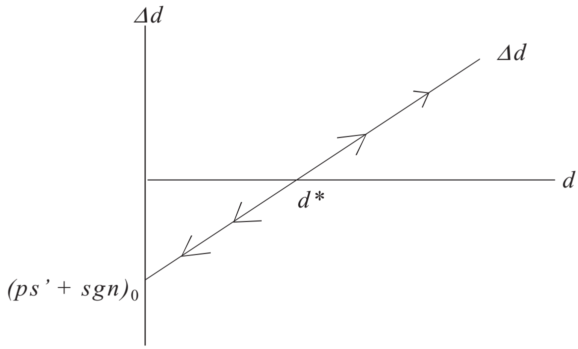


Figure 9.1. The intertemporal budget constraint of the public sector

sector's seignorage revenue. This curve can be used to analyze the debt dynamics implied by the public sector's budgetary choices.

As you can see from equation (9.11), the public sector's choice of ps' and sgn determines the height of the Δd curve. The initial value of the debt-GDP ratio, say, $d(0)$, determines where on the horizontal axis the economy is located initially. Notice that the point of intersection of the Δd curve with the horizontal axis, labeled d^* in the figure, is where $\Delta d = 0$, that is, where the debt-*GDP* ratio is neither increasing nor decreasing. The dynamics of the debt stock depend on where $d(0)$ lies relative to d^*. If $d < d^*$, then $\Delta d < 0$, and the debt-*GDP* ratio must be falling; if $d(0) = d^*$, the debt-*GDP* ratio is stable; finally, if $d(0) > d^*$, the debt-GDP ratio must be rising.

We can use this information to analyze the constraints that solvency imposes on the public sector's budget. In particular, suppose that the quantity $ps' + sgn$ is to be kept constant, so the position of the Δd line is fixed. Notice the following:

- If $d(0) = d^*$, $\Delta d = 0$. From equation (9.11), this implies that $ps' + sgn = (r - n)d(0)$. In this case, the value of $ps' + sgn$ is exactly sufficient to service the existing debt $d(0)$ on market terms, and d^* thus has the property that $(ps' + sgn)/(r - n) = d^*$. If $d(0) = d^*$, then, the necessary condition for public-sector solvency is met as an equality. This result demonstrates that following the policy in equation (9.10a) is equivalent to stabilizing the ratio of debt to *GDP*.[11]
- If $d(0)$ is to the left of d^*, on the other hand, $\Delta d < 0$. This means that $ps' + sgn > (r - n)\ d(0)$. Thus the public sector is making payments to creditors that are *more* than enough to pay the interest on the debt while keeping the debt-*GDP* ratio constant, so the ratio must be falling. In that case, the necessary condition for public-sector solvency is met as the inequality $(ps' + sgn)/(r - n) > d(0)$. Thus, given the value of $ps' + sgn$, the public sector is solvent as long as $d(0) \leq d^*$.

[11] Notice the implication for the behavior of the debt stock: not only does debt not have to be repaid, as we showed before, but it can actually grow over time at the rate of growth of the economy.

- However, for values of $d(0) > d^*$, it must be the case that $\Delta d > 0$, so $ps' + sgn < (r - n)d(0)$. In this case, given the value of $ps' + sgn$, the resources that the public sector is transferring to its creditors are not enough to service the debt on market terms. Because the shortfall in payments just adds to the debt due, if the initial ratio of debt to *GDP* $d(0)$ is to the right of d^*, the stock of debt must be rising *faster* than *GDP*. In this case, *if* ps$'$ + *sgn remains unchanged*, the public sector would be insolvent. Because an insolvent government is not able to borrow at all, the shape of its Δd curve would not be as in Figure 9.1 because positive values of Δd would be ruled out to the right of d^*. Instead, the Δd curve would coincide with the horizontal axis to the right of d^*.

Does this mean that a government that finds itself at a position on the Δd curve where $d(0) > d^*$ is necessarily insolvent? The answer is not necessarily. Given $(r - n)$, this would be true only if the government's fiscal plans call for $ps' + sgn$ to remain unchanged *permanently*. To explore the constraints that solvency places on the behavior of the public sector, suppose instead that the quantity $ps' + sgn$ can be changed discretely at some future moment of time.

Under this assumption, the key observation is that the solvency condition (9.9) can be satisfied, even though the public sector's debt-*GDP* ratio is *currently* increasing, *as long as its creditors expect it to make the fiscal adjustment required to stabilize its debt-GDP ratio at some point in the future.* This is so because, as long as this adjustment is eventually made, any extra debt that the government accumulates between the present and the time when the adjustment is made will eventually be serviced on market terms; that is, it will generate a present value of future debt service payments that is equal to the increase in the value of the debt. To see what this means in terms of Δd line, recall that the position of the line is determined by the vertical intercept $-(ps' + sgn)$. Thus a fiscal adjustment that increases ps' and/or sgn would shift the curve downward. The preceding adjustment requires increasing $ps' + sgn$ at some point T in the future by an amount that is just sufficient to shift the Δd curve far enough downward so that it intersects the horizontal axis at $d(T)$.

Under what conditions will it be possible to do so? To answer this question, let $(ps' + sgn)_{max}$ denote the *maximum* values of ps' and sgn that the government can generate to service debt under normal conditions (i.e., without extraordinary legal and/or political actions by creditors).[12] Then the maximum amount of debt that the government can service successfully is given by

$$d_{max} = PV[(ps' + \mathrm{sgn})_{\mathrm{max}};\ r - n] \tag{9.12}$$

Graphically, d_{max} can be derived as the point at which a line with slope $r - n$ starting from the point $(ps' + sgn)_{max}$ on the vertical Δd axis intersects the horizontal d axis.

[12] This is discussed further in Chapter 10.

Table 9.1. *External Debt Ratios at the Time of Adverse Credit Event, 1970–2001*

	Initial Year of Credit Event	External Debt-to-GNP Ratio in Initial Year	External Debt-to-Exports Ratio in Initial Year
Albania	1990	45.8	616.3
Argentina	1982	55.1	447.3
	2001	53.3	458.1
Bolivia	1980	92.5	246.4
Brazil	1983	50.1	393.6
Bulgaria	1990	57.1	154.0
Chile	1972	31.1	na
	1983	96.4	358.6
Costa Rica	1981	136.9	267.0
Dominican Rep.	1982	31.8	183.4
Ecuador	1982	60.1	281.8
	1999	89.2	239.3
Egypt	1984	112.0	282.6
Guyana	1982	214.3	337.7
Honduras	1981	61.5	182.8
Iran	1992	42.5	77.7
Jamaica	1978	48.5	103.9
Jordan	1989	179.5	234.2
Mexico	1982	46.7	279.3
Morocco	1983	87.0	305.6
Panama	1983	88.1	162.0
Peru	1978	80.9	388.5
	1984	62.0	288.9
Philippines	1983	70.6	278.1
Poland	1981	na	108.1
Romania	1982	na	73.1
Russian Federation	1991	12.5	na
	1998	58.5	179.9
Trinidad and Tobago	1989	48.1	112.8
Turkey	1978	21.0	374.2
Uruguay	1983	63.7	204.0
Venezuela	1982	48.6	220.9
	1995	44.1	147.2
Average		**70.6**	**254.3**

Source: Reinhart et al. (2003).

Notice that d_{max} is likely to differ from country to country, depending on each country's ability to generate resources, in the form of $[(ps' + sgn)_{max}$, to service its debt, as well as on the country's growth rate n. As shown in Table 9.1, the maximum sustainable debt level does indeed seem to vary substantially across countries because emerging and developing countries have encountered debt-servicing difficulties at very different ratios of debt to *GDP*.

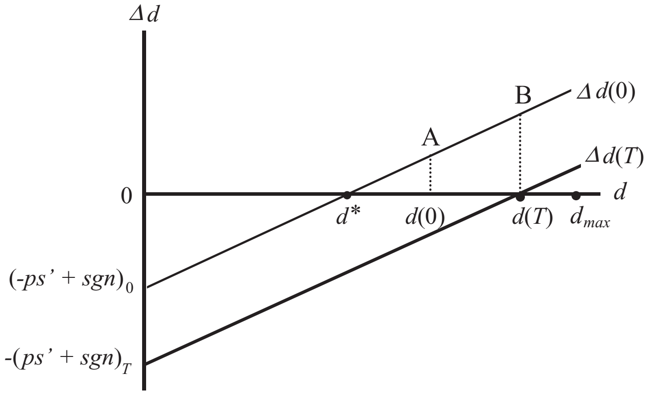


Figure 9.2. Sustainable debt accumulation

With d_{max} defined in this way, the necessary condition for public-sector solvency becomes $d(0) \leq d_{max}$. The preceding argument therefore implies that points on the Δd line with $d(0) > d^*$ can be sustained (i.e., the public sector will be able to borrow the amount $\Delta d(0)$ corresponding to such a point) only insofar as its creditors believe that at some point T in the future, the government will make the adjustments to $ps'(T)$ and/or $sgn(T)$ required to stabilize $d(T)$ before it reaches d_{max}. Notice that if $d(0) \geq d_{max}$, creditors could not have such an expectation because no additional fiscal adjustment is possible under these conditions, by definition.

We can summarize these arguments by saying that for the government to be able to borrow when $d(0) > d^*$, it is *necessary*, but not *sufficient*, that $d(0) < d_{max}$. The additional condition required is that the government must be able to credibly commit to stabilizing its debt-*GDP* ratio at a value short of d_{max} at some point in the future. A fiscal plan that meets these conditions – that is, that implies $d(t) \leq d_{max}$ for all values of t – can be said to be *feasible* in the sense that creditors would be willing to finance it. Any fiscal plan that would ultimately imply $d(t) > d_{max}$ for some future value of t is *infeasible*. Notice that a government will become insolvent not when $d(t) > d_{max}$ but rather as soon as it becomes apparent to creditors that the government will be unable to implement a feasible fiscal plan.

To illustrate these points, consider Figure 9.2. Suppose that the current stock of debt $d(0)$ is greater than d^* but falls short of d_{max}. Thus creditors will be willing to extend additional financing at this point, as long as the government convinces them that it will make a fiscal adjustment at some point in the future that will stabilize the debt-*GDP* ratio short of d_{max}. Thus Δd can indeed take on the positive values suggested by the Δd curve for values of d above d^* such as the initial point A. Because Δd is positive at A, the economy cannot stay there because the positive value of Δd means that the public sector's net stock of debt is increasing over time. This causes d to drift to the right away from $d(0)$, along the horizontal axis. As it does so, the economy moves to the northeast along the Δd curve. Suppose that the debt is to be stabilized at the value $d(T)$ that prevails at the time T, so the government makes its fiscal adjustment at that time. When it does so, it must increase ps' or sgn. This causes the Δd curve to shift downward. To stabilize the debt ratio at the value $d(T)$, the government must shift the curve downward by an amount that is exactly sufficient to have it intersect the horizontal axis at $d(T)$. The new Δd curve is labeled $\Delta d(T)$ in Figure (9.2), and the new value of $-(ps' + sgn)$ is labeled $-(ps' + sgn)_T$ in the figure.

IV. FISCAL DEFICITS AND MONEY GROWTH

With this apparatus in hand, we can now go back to the relationship between fiscal deficits and money growth. Suppose that from an initial situation where the primary surplus and seignorage revenue are ps'_0 and sgn_0, respectively, and the debt stock is d^*, the government *permanently* reduces its primary surplus, say, to ps'_1. What are its financing options? One possibility is for it to immediately print money to finance the deficit dollar for dollar. In that case, the budget constraint curve Δd remains undisturbed, but with $ps'_1 + sgn_1 = ps'_0 + sgn_0$ and $ps'_1 < ps'_0$, this means that $sgn_1 > sgn_0$, which implies a higher rate of monetary growth and inflation. Alternatively, if $d(0) < d_{max}$, the government may be able to borrow to finance the higher deficit, if it is solvent in the initial situation. However, as we have just seen, it cannot *permanently* finance the deficit in this way if $d(0) > d^*$ because if it tried to do so, its debt-to-*GDP* ratio would expand continuously, which would violate the solvency condition (9.9); that is, such a plan would not be feasible. This means that at some point before d reaches d_{max}, the government would have to shift the budget constraint line back down, stabilizing the debt ratio at some value $d < d_{max}$.

How would it do this? Because, by assumption, the increase in the primary deficit is permanent, it would have no choice but to do so by printing money. Notice also that once debt has begun to accumulate, the downward shift in the budget constraint line must be *larger* than its initial upward shift. Thus the longer the government waits to turn to money financing, the larger its reliance on seignorage revenue will eventually have to be. The reason is, of course, that it will need to service a larger stock of debt. The upshot is that a permanently higher fiscal deficit will sooner or later have to be financed by printing money. It is worth reemphasizing that even the initial debt financing would not be feasible unless creditors were convinced that the eventual turn to money financing would take place. If creditors believe that the government has no intention of eventually stabilizing its debt stock short of d_{max}, then they would not provide the temporary financing required for the government to avoid printing money initially because their new loans would not be repaid on terms comparable to those they could get elsewhere.

Now consider an alternative exercise, one in which the government wants to lower the rate of inflation. Suppose that the condition $d(0) < d_{max}$ is satisfied and the public sector's current fiscal plans are consistent with the intertemporal budget constraint (9.9), but with a relatively high value of sgn. Suppose further that with an unchanged value of ps', the public sector decides to borrow in the present to avoid the inflationary consequences of the relatively high level of money financing (why it might want to do so is the subject of Chapter 13). Then, given the ratio of debt to *GDP* currently in existence, $d(0)$, the requirement that the intertemporal budget constraint (9.9) must be satisfied means that lower monetary growth today must be accompanied by *faster* anticipated monetary growth in the future (because, given $d(0)$, if the present value of ps' is unchanged, that of *sgn* must be as well, so a lower *sgn* in the short run must mean a higher one later on). This "unpleasant monetarist arithmetic" thus implies that with an unchanged adjusted primary deficit, the government can achieve lower inflation today only at the cost of higher inflation in the future.

But what if the government seeks to achieve a *permanent* reduction in monetary growth? The key point is that the government's intertemporal budget constraint (9.9) implies that any reduction in present and future monetary growth resulting in a decreased value of $PV(sgn; r-n)$, such as would be associated with a permanently reduced rate of monetary growth intended to slow inflation, must be associated with a *higher* value of $PV(ps'; r-n)$; that is, a permanent reduction in monetary growth requires a fiscal adjustment.

When must this fiscal adjustment take place? Suppose the solvency condition (9.9) holds, so the consolidated public sector is solvent, and that once again, $d(0) < d_{max}$, so the public sector has unused borrowing capacity. Now suppose that *sgn* is reduced. Does this mean that ps' must rise right away? The answer is no. We will now show that if the consolidated public sector is solvent, it can indeed borrow for a while, so no fiscal adjustment is necessary right away. But as it does so, its ratio of debt to *GDP* will begin to rise, and the solvency condition can continue to hold only if $PV(ps'; r-n)$ rises at the same time. What this means is that in an inflation stabilization program, *postponing fiscal adjustment now must imply a larger fiscal adjustment in the future, to service the accumulated debt.*

The analysis is conducted in Figure 9.3. Suppose that *sgn* falls. Because $ps' + sgn$ would decrease in this case, the Δd curve shifts up. The government now faces two choices:

1. Adjust ps' immediately to keep the debt ratio at $d(0) = d^*$. This is the immediate fiscal adjustment option.
2. Finance the reduction in monetary expansion by borrowing. In this case, the economy moves to the point A. At A, debt starts to rise.

If no adjustment in ps' is intended ever to be made, debt would rise forever, and since creditors would not be repaid on market terms under these circumstances, this path is not feasible because the loans would not be forthcoming. On a feasible path, the public sector would have to commit itself to stabilizing d again in the future because only along such a path would all the debt accumulated be continually serviced on market terms.

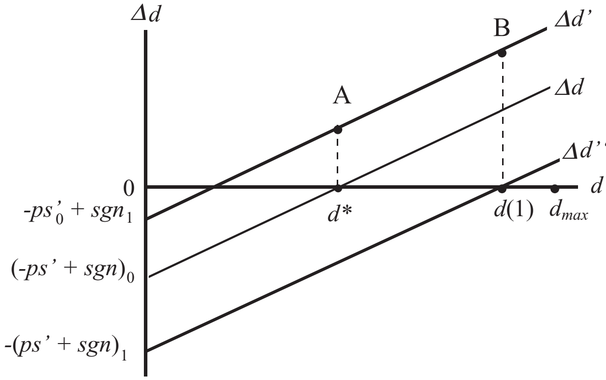


Figure 9.3. Fiscal aspects of monetary stabilization

Suppose, then, that debt has risen to $d(1)$, with the economy at the point B. To stabilize the debt at the new level, with the lower level of *sgn*, *ps′* would have to rise to have the Δd line intersect the horizontal axis at $d(1)$. Notice that in this case, just as in the previous section, the adjustment in *ps′* would have to be larger than it would have been had it been undertaken immediately, because now interest on the accumulated debt has to be paid as well. The path followed by $ps' + sgn$ can be described as in Figure 9.4.

The implication of this analysis is that reducing money growth requires fiscal adjustment, either in the present or in the future. Why not just put the adjustment off? There are two reasons:

1. To postpone adjustment requires borrowing, and for borrowing to be possible, creditors have to be convinced that adjustment will indeed take place, so this may require some up-front painful changes to achieve credibility.
2. Postponing the adjustment, moreover, means that the future adjustment must be bigger.

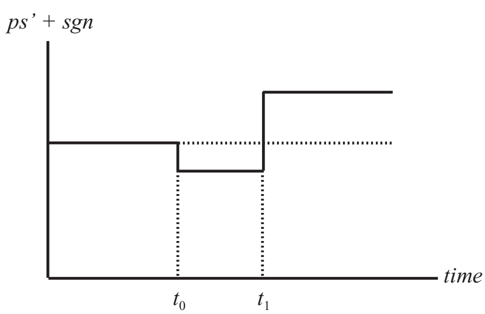


Figure 9.4. Fiscal trade-offs in monetary stabilization

V. THE SIZE OF THE REQUIRED FISCAL ADJUSTMENT

So far, we have shown that achieving a permanent drop in the rate of inflation requires a fiscal adjustment, either now or later. Suppose we make the adjustment now. In this section, we will examine how big that adjustment has to be – that is, how large is the permanent fiscal deficit required to achieve a given rate of inflation?[13]

Recall that the solvency condition told us the size of the permanent adjusted primary surplus consistent with solvency. It is given by equation (9.10b), which is reproduced here for convenience:

$$ps' = (r - n)\, d(0) - (\pi + n)m \tag{9.10b}$$

However, this equation does not directly yield a useful measure of how the permanent adjusted primary surplus is affected by a change in the long-run level of inflation. The complication is that because the demand for money depends on the nominal interest rate, the ratio of the money stock to *GDP* (m, in our preceding notation) itself depends on the rate of inflation. Taking this dependence into account, and using equation (9.10b), we can express the *sustainable value of the adjusted primary surplus* as

$$ps' = (r - n)d(0) - (\pi + n)m(\pi) \tag{9.13}$$

This equation reveals the relationship between the permanent rate of inflation and the sustainable value of the adjusted primary surplus. Consequently, it permits us to determine the magnitude of the permanent fiscal adjustment required to achieve a given permanent reduction in the rate of inflation. From equation (9.13), it is clear that the answer depends on what happens to $(\pi + n)m(\pi)$.

The complication here is that there are offsetting effects operating on seignorage revenue $(\pi + n)m(\pi)$ as the rate of inflation rises. On one hand, a higher rate of inflation increases $(\pi + n)$, but on the other hand, it also decreases m. It is easy to see that a reduction in the rate of inflation will be more likely to reduce seignorage revenue when the first effect is strong relative to the second, that is, when the demand for money does not respond strongly to changes in the rate of inflation. To see what the net effect depends on more formally, write the change in $(\pi + n)m(\pi)$ as

$$\begin{aligned}
\Delta(\pi + n)m(\pi) &= \Delta\pi\, m + (\pi + n)\,\Delta m \\
&= \Delta\pi\, m + (\pi + n)\,\Delta m \\
&= \pi m\,(\Delta\pi/\pi + (1 + n/\pi)\,\Delta m/m) \\
&= \pi m\,(1 - (1 + n/\pi)\,\eta_{m\pi})\,\Delta\pi/\pi)
\end{aligned}$$

[13] For an application of the issues addressed in this section to the case of Turkey, see Anand and van Wijnbergen (1989).

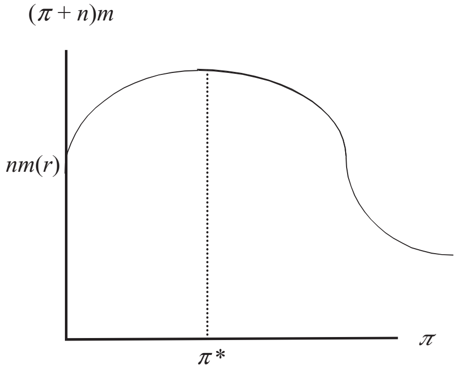


Figure 9.5. Inflation and seignorage revenue

where $\eta_{m\pi} = -(\Delta m/m)/(\Delta\pi/\pi)$ is the *elasticity of the demand for money with respect to the inflation rate.* This expression will be positive when $\Delta\pi/\pi$ is positive only when the demand for money is not very responsive to changes in the rate of inflation, that is, when $\eta_{m\pi}$ is sufficiently small. When the growth rate is zero, for example, this condition requires that the elasticity of the demand for money with respect to the inflation rate be less than unity.

To illustrate these points more concretely, we can give a specific form to the demand for money $m(\ )$. Suppose, for example, that $m(\pi)$ takes the semilog form

$$m(r + \pi) = ke^{-\alpha(r+\pi)}$$

Now let us plot $(\pi + n)\ m(r + \pi)$ against π. To see what shape the curve has, differentiate $(\pi + n)\ m(r + \pi)$ with respect to π. The result is

$$\begin{aligned} dm(r + \pi)/d\pi &= ke^{-\alpha(r+\pi)} - \alpha(\pi + n)ke^{-\alpha(r+\pi)} \\ &= ke^{-\alpha(r+\pi)}[1 - \alpha(\pi + n)] \end{aligned}$$

This implies that the slope will be positive up to some critical value $\pi^* = 1/\alpha - n$ and then turn negative for rates of inflation above π^*, as illustrated in Figure 9.5.

What does this imply about the fiscal adjustment required to achieve a given permanent reduction in the rate of inflation? In general, there is no good reason why an economy would be to the right of the critical point. If it is to the *left* ($\pi < \pi^*$), then it follows that reducing the rate of inflation reduces seignorage, and the sustainable adjusted primary surplus must *increase.* Notice that an increase in n reduces the required primary surplus, so if n responds positively to a reduction in π, the required fiscal adjustment would be smaller than the preceding discussion would indicate.

How large does seignorage revenue tend to be in practice? Easterly and Schmidt-Hebbel (1994) estimated the average amount of annual seignorage revenue collected by a 50-country sample of Organization for Economic Co-operation and Development (OECD) countries and emerging economies (there were 15 OECD countries in the sample and 35 emerging economies) over the period 1970–1988. They estimated annual seignorage revenue by taking the ratio to *GDP* of the yearly sum of deflated monthly changes in the monetary base (i.e., they added together the monthly values of $\Delta M/P$ and divided the sum by real *GDP*). They found that the average value of seignorage revenue for the OECD countries in their sample over the period amounted to about 1 percent of *GDP*, whereas for the emerging economies, the average was 2.1 percent of *GDP*. Among all the countries in their sample, the largest amount of seignorage collected over the period was by Zaire (now once again known as the Congo), where it was 4.4% of GDP.

Two other observations from this study are worth noting. First, on the basis of cross-country data, they estimated the revenue-maximizing annual inflation rate for their sample of countries (π^* in Figure 9.5) at about 68 percent.[14] Second, they noted that while the *inflation tax* (given by $\pi\ m$, the portion of the government's seignorage revenue that arises due to the public's need to replenish the inflationary erosion of the real value of its money stock) did not tend to provide large amounts of revenue for most countries over extended periods, the ratio of seignorage revenue over *GDP* tended to be highly variable in their sample, and that seignorage tended to be a large source of revenue during times of crisis, amounting in some observations to as much as 13 percent of GDP.

VI. SUMMARY

In this chapter, we have seen that a central-bank policy of continuous credit expansion-cum-devaluation may at bottom be a response to a fiscal need. In a context in which the central bank's policies are dictated by the finance ministry, the government is likely to raise funds in this way when it cannot borrow to finance a large fiscal deficit, which is likely to be the case when the government's solvency is in question, that is, when its debt obligations are already at or near the maximum that it can reasonably expect to service on market terms, given the prospective future resources that it will have available for debt service.

When a government is perceived by its creditors to be potentially insolvent, it is said to have a *debt overhang*, that is, to have more debt than it can feasibly service. What will happen under these circumstances? In principle, four outcomes are possible. The most obvious one is that creditors may be prepared to forgive and (it is hoped) forget the excess debt. However, as we will discuss in more detail in the next chapter, this is not an outcome that is likely to materialize spontaneously as the result of voluntary actions by the government's creditors. Failing some form of organized debt forgiveness, then, the government can make one of three remaining

[14] Actually, this result was sensitive to the inclusion of Argentina in the 50-country sample. The 68% figure was derived when Argentina was excluded. When that country was retained in the sample, the revenue-maximizing rate jumped to 160%.

choices: specifically, it can make a fiscal adjustment by reducing G and/or increasing T; it can accelerate the printing presses (increase $\Delta M/M$); and/or it can repudiate part or all of its debt.

Deficit reduction is seldom costless or easy. Not only do reductions in spending and increases in taxation reduce aggregate demand but the implementation of such measures is usually hampered by every economic agent's incentive to free ride, that is, to let others bear the burden of specific expenditure cuts and tax increases. We will discuss this further in Chapter 11. The upshot is that governments are frequently tempted by the two remaining options in dealing with prospective insolvency: money financing and debt repudiation. We will examine the macroeconomic consequences of money financing and the resulting high inflation in Chapter 13 and of the nonpayment of debt in Chapter 25.

REVIEW QUESTIONS

1. What is fiscal dominance? Explain how fiscal dominance determines the rate of money creation by the central bank.
2. Explain why a solvent public sector is not necessarily one that can repay its debt.
3. Is a government solvent as long as its debt-*GDP* ratio falls short of d_{max}? Explain why or why not.
4. Are solvency and debt sustainability the same thing? Explain why or why not.
5. Explain the relationship between seignorage and the inflation rate. Does an increase in inflation always generate higher seignorage revenue?

EXERCISES

1. Many economists have noted that developing countries often undergo debt crises (episodes of fiscal insolvency) at debt-*GDP* ratios that are much lower than those exhibited by several industrial countries whose governments are able to continue to borrow at essentially risk-free interest rates, even though they have very high debt-*GDP* ratios. This phenomenon has been called *debt intolerance*. Using the tools you have learned in this chapter, can you explain why debt intolerance might exist?
2. "A permanent reduction in the rate of inflation requires a fiscal adjustment (an increase in the adjusted primary surplus) either now or in the future."
   a. Do you agree or disagree? Explain.
   b. Compare the required size of the fiscal adjustment if the adjustment is made now to what it would need to be if it is postponed to some time in the future.
3. Explain why a financial-sector reform that results in a reduction of the ratio of base money to *GDP* would require a fiscal adjustment.
4. In the fiscal sustainability model that we developed in this chapter, we assumed that the economy's growth rate n and its ratio of money to *GDP* (m) were both

independent of the rate of inflation π. Show graphically how an increase in the rate of inflation π would affect the sustainable level of debt for given values of ps' and r if both the economy's growth rate and its ratio of money to GDP are instead *decreasing* functions of the domestic inflation rate, that is, if the following conditions hold:

$$n = n(\underset{-}{\pi}) \text{ and } m = m(\underset{-}{\pi})$$

5. "Inflation is always and everywhere a fiscal phenomenon."
   a. What does this statement mean?
   b. Do you agree with the statement? Explain.
   c. Is the statement inconsistent with the view that inflation is caused by money supply growth? Explain.

REFERENCES AND FURTHER READING

Anand, Ritu, and Sweder van Wijnbergen (1989), "Inflation and the Financing of Government Expenditure: An Introductory Analysis with an Application to Turkey," *World Bank Economic Review*, vol. 3, pp. 17–38.

Buiter, Willem (1985), "A Guide to Public Sector Debt and Deficits," *Economic Policy*, vol. 1, pp. 13–79.

Buiter, Willem (1990), "The Arithmetic of Solvency," in W. Buiter, ed., *Principles of Budgetary and Financial Policy*, Cambridge, MA: MIT Press, pp. 145–162.

Easterly, William, and Stanley Fischer (1990), "The Economics of the Government Budget Constraint," *World Bank Research Observer*, vol. 7, pp. 127–142.

Easterly, William, and Klaus Schmidt-Hebbel (1994), "Fiscal Adjustment and Macroeconomic Performance: A Synthesis," in W. Easterly and K. Schmidt-Hebbel, eds., *Public Sector Deficits and Macroeconomic Performance*, New York: Oxford University Press, pp. 15–78.

Eichengreen, Barry, and Ricardo Hausmann (1999), "Exchange Rates and Financial Fragility," working paper 7418, National Bureau of Economic Research.

Perotti, Roberto (2007), "Fiscal Policy in Emerging Market Economies: A Framework and Some Questions," unpublished manuscript, Universita Bocconi.

Reinhart, Carmen, Kenneth Rogoff, and Miguel Savastano (2003), "Debt Intolerance," NBER Working Paper No. 9908 (August).

10

Sovereign Risk Premia

As shown in the preceding chapter, the solvency of the public sector in an emerging or developing economy depends on the perceptions of creditors about the government's ability and willingness to service its debt on market terms. Because judgments about the government's solvency are made in the present on the basis of projections of the resources that will be available to the government for servicing debt in the future, solvency assessments are inherently forward-looking exercises and, as such, are intrinsically uncertain. When a government's expected debt-servicing capacity far exceeds its existing debt obligations, there may be little doubt on the part of creditors about the likelihood of repayment. But when the stock of debt is large relative to the government's projected debt-servicing capacity, lending to the government becomes a risky proposition for creditors. It is therefore useful to investigate how our solvency analysis is affected by the emergence of such credit risk.

This chapter undertakes that task. In the first section, we will examine how our solvency analysis needs to be modified to incorporate credit risk. As we will see, the key point is that in the presence of credit risk, the government's creditors will no longer be willing to lend to the government at the risk-free interest rate – they will demand a higher interest rate to compensate them for the possibility of nonpayment. The difference between this interest rate and the risk-free rate is referred to as the *sovereign risk premium.* We will explore how our solvency analysis is affected by the presence of such a premium. Section II will then consider how the existence of such a premium affects the economy's short-run macroeconomic equilibrium and will show why the emergence of such a premium is likely to have contractionary effects on the economy and lead to capital outflows. Because credit risk exists in the eye of the beholder – in this case, of the creditors – Section III will look at recent empirical evidence on the determinants of perceived credit risk, in part to examine whether the factors that our model suggests *should* matter for perceptions of credit risk actually do so in practice. Finally, we will consider in

Section IV what the government can do to mitigate credit risk by signaling to creditors its intentions to fully meet its financial obligations in the future. Section V summarizes and concludes.

I. SOVEREIGN CREDIT RISK

How may the public sector's fiscal choices be affected by the fact that in determining the terms on which they will lend to the public sector, creditors take into account the possibility that the public sector may not be able to repay? In general, we would expect that creditors would demand to be compensated for absorbing this risk by requiring a higher interest rate on government debt the higher this risk is perceived to be. A natural measure of this risk is the relationship between the actual stock of debt and the maximum stock of debt that the government would be able to service at the risk-free interest rates under "normal" circumstances, that is, in the absence of special measures by its creditors. This suggests that the terms on which the government can borrow may depend on the relationship between its existing stock of debt and its maximum perceived debt-servicing capacity. Written as ratios to gross domestic product (*GDP*), this suggests that the interest rate that the public sector faces at time t should depend on $d(t)/d_{max}$; that is, we should have:

$$r = r_s + \underset{+}{r}(d(t)/d_{max}) \tag{10.1}$$

where r_S is the safe (risk-free) interest rate. The positive sign under the function r indicates that the real interest rate that the government must pay on its debt is an increasing function of the ratio of its debt outstanding to its maximum debt-servicing capacity.[1] As mentioned earlier, the difference between the interest rate that the government must pay and the risk-free rate, captured by the function $r(\;)$, is referred to as the sovereign risk premium, that is, the risk premium on government debt.

How does the presence of a risk premium affect the public sector's debt dynamics? Going back to the public sector's budget constraint given by equation (9.12), we can now write it as follows:

$$\Delta d(t) = -(ps' + \text{sgn}) + [r_s + r(d(t)/d_{max}) - n]d(t) \tag{10.2}$$

The new budget constraint is depicted in Figure 10.1. Notice that because r is now an *increasing* function of the stock of debt, the $\Delta d(t)$ curve must have an increasing rather than a constant slope, and the new $\Delta d(t)$ curve must lie everywhere above the old one from Chapter 9, indicated by the dotted line.

[1] It is important to note here that the discount rate used in calculating d_{max} remains the *safe* interest rate rather than the interest rate charged to the government by its creditors because it is the safe interest rate that captures the opportunity cost of funds to the creditors.

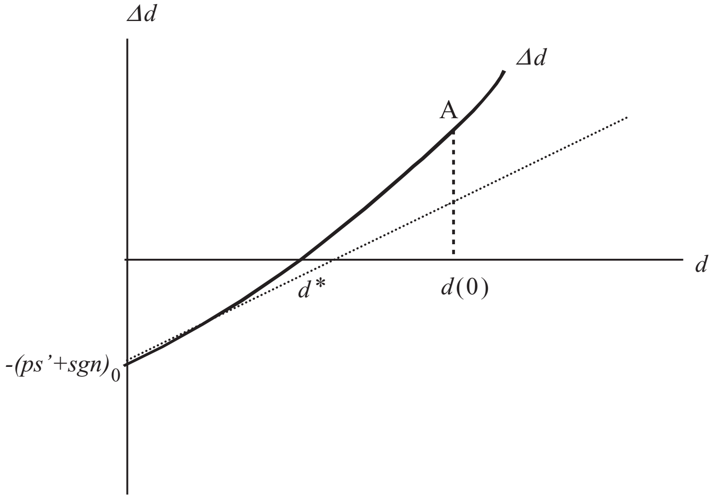


Figure 10.1. The intertemporal budget constraint with credit risk

To see the implications of the risk premium, suppose that the economy finds itself at a point like A, with an initial debt-*GDP* ratio equal to $d(0)$. The curvature of the Δd curve has two implications:

1. If the government does not make a fiscal adjustment, its debt will increase faster than it would have done in the absence of a risk premium. The reason is that as the government continues to borrow, its debt-service obligations increase, not just because it needs to service additional debt but also because the higher risk premium implies larger debt-service payments on the entire stock of *existing* debt.
2. The fiscal challenge faced by the government becomes even more acute as debt accumulates; that is, if the government is to avoid defaulting on its debt, a fiscal expansion in the present requires an even *larger* fiscal contraction in the future, and a reduction in seignorage revenue will require an even larger fiscal adjustment. The reason is that the curvature of the Δd curve causes its height to increase at any given value of the debt-GDP ratio (such as $d(0)$), requiring a larger downward shift in the curve (i.e., a larger adjustment in $ps' + sgn$) to stabilize the debt-GDP ratio.

This formulation has yet a third interesting implication: creditors' views of the government's fiscal capacity may actually have a self-fulfilling element. These views are captured in the variable d_{max}. Governments with larger values of d_{max} have larger fiscal capacities. But if creditors believe that a given government has a large fiscal capacity in this sense, equation (10.1) suggests that because that government will face a smaller risk premium, for a given fiscal stance and initial debt stock, it will accumulate less debt, making it easier for that government to carry out its fiscal plans. Geometrically, an increase in d_{max} causes the point at which the Δd

curve begins to curve upward to move up and to the right along the dotted line in Figure 10.1.

II. MACROECONOMIC IMPLICATIONS OF CREDIT RISK

The previous section considered the *fiscal* implications of an increase in credit risk on government debt – but what are its *macroeconomic* implications? We can analyze this issue by using the model of Part 2 in a fairly direct way. As we have seen, an increase in credit risk causes the government's creditors to demand a higher interest rate to hold a given amount of the government's debt. In the context of our macroeconomic model, we can capture this effect by introducing a shift parameter σ into the functions $b(\ )$ and $b^*(\ )$, representing the shares of domestic government bonds that domestic and foreign residents wish to hold in their financial portfolios. The variable σ serves as an indicator of the amount of credit risk that domestic and foreign agents perceive in the domestic government's debt. An increase in their perception of risk would increase σ and *decrease* their demand for domestic government bonds at given values of the domestic and foreign interest rates R and R^*. Using equation (6.21), we can thus write the equilibrium condition in the domestic bond market as

$$B - B_c = b(R - \underset{+}{R^*} - \sigma)[W_p - PL(\underset{-}{R}, Y(\underset{+}{P}))] + Sb^*(R - \underset{+}{R^*} - \sigma)(W_F^* - P^*L^*(\underset{-}{R^*}, \underset{+}{Y^*})) \tag{10.3}$$

where the relevant interest rate differential that determines portfolio allocations between domestic and foreign bonds is now the *risk-adjusted* one, $R - (R^* + \sigma) = R - R^* - \sigma$. Note that other things equal, an increase in the perceived credit risk on the bonds issued by the domestic government reduces the shares of their portfolios that both domestic and foreign residents choose to devote to domestic bonds.

Consider now the macroeconomic effects of a change in creditors' perceptions of the risk associated with domestic government debt, in the form of an increase in σ. Figure 10.2 shows what happens in the domestic bond market. Because the increase in risk reduces creditor demand for domestic government bonds at any given value of the domestic interest rate, it causes the bond demand curve to shift to the left, from a position such as B_0^D to one such as B_1^D. The implication is that the interest rate on domestic bonds must increase to clear the bond market, say, from R_0 to R_1. In turn, the higher interest rate required to clear the bond market causes the BB curve to shift upward by the amount $(R_1 - R_0)$ in (P, R) space, as in Figure 10.3. The economy's new short-run equilibrium is at the point B, with a higher domestic interest rate and lower domestic price level. The increased perception of credit risk is therefore contractionary in the short run.

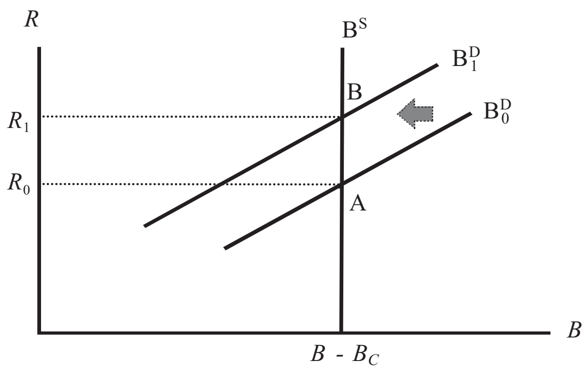


Figure 10.2. Bond-market implications of an increase in sovereign credit risk

Because the domestic interest rate is higher and the price level is lower, the demand for money must have decreased in the domestic economy, and because the central bank has not accommodated the reduced demand for money by contracting domestic credit, the higher perception of credit risk must be associated with a capital *outflow* from the domestic economy and a reduction in the central bank's stock of foreign exchange rate reserves. Sustaining the original value of the domestic interest rate under these circumstances would require a more expansionary stance of monetary policy than when credit risk is low; that is, the domestic central bank will have to absorb more of the government's debt itself when that debt becomes less attractive for the private sector to hold. But if capital mobility is high, so that a monetary expansion results in a reserve outflow (see Chapter 7), this would imply an even larger capital outflow and thus a larger loss of foreign exchange reserves.

III. DETERMINANTS OF SOVEREIGN CREDIT RISK

Fluctuating perceptions of sovereign credit risk can therefore have important short-run macroeconomic impacts on emerging and developing economies. The issue of what determines such risk perceptions is therefore an important one. This section examines some recent evidence on this issue.

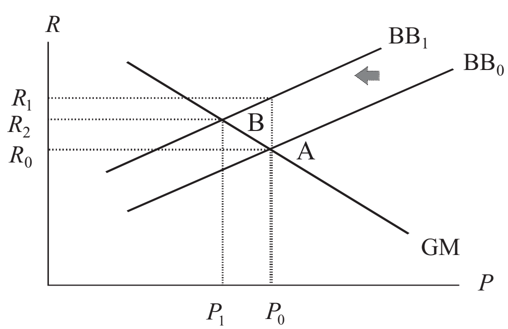


Figure 10.3. Short-run implications of an increase in sovereign credit risk

We saw earlier that credit risk would tend to rise as the government's debt-*GDP* ratio d approaches its maximum sustainable value d_{max}. Recall also that d_{max} itself depends on $(ps' + sgn)_{max}$ as well as on the slope of the Δd curve, given by the growth-adjusted interest rate $r - n$. Thus we would expect the sovereign risk premium to depend positively on the ratio of public-sector debt to GDP. It should also be a decreasing function of the factors that determine creditors' perceptions of $(ps' + sgn)_{\max}$ but an increasing function of the growth-adjusted interest rate $r - n$ because an increase in $r - n$ makes the Δd curve steeper and decreases d_{max}.

There is a large literature that investigates these propositions empirically. This literature has taken several forms:

- One approach has been to examine the factors that affect sustainable values of the debt-*GDP* ratio across countries. This essentially amounts to investigating the determinants of d_{max}.
- A second approach consists of examining the determinants of default risk on sovereign debt as perceived by credit-rating agencies such as Standard & Poor's or Moody's. These agencies rate default risk on bonds issued by a wide range of borrowers, including governments in emerging and developing countries, and classify such bonds into standardized risk categories.
- Finally, a third approach has been to examine the factors that affect risk premia on bonds issued by emerging-market and developing-country governments directly.

In the rest of this section, we will examine recent evidence about the determinants of sovereign credit risk following each of these approaches in turn.

1. Sustainable Debt

Interest in assessing sustainable public-sector debt levels has arisen from the observation that countries have tended to encounter debt-servicing difficulties at widely differing debt-*GDP* ratios. Some industrial countries (e.g., Belgium, Italy, and Japan) have long been able to sustain ratios of public debt to *GDP* in excess of 100 percent without panicking their creditors, while many emerging and developing economies have found themselves in debt crises at debt-*GDP* levels that are less than half those values. We saw in Table 9.1 that even among emerging and developing economies, there was great variation in the debt-*GDP* ratios at which governments defaulted or experienced other adverse events associated with public debt. The question is why some countries are able to sustain much higher levels of debt than others.

Reinhart et al. (2003; hereinafter referred to as RRS) provided a prominent recent investigation of this issue. They examined a sample of 53 advanced, emerging, and developing countries for which they could obtain measures of credit risk and public-sector debt over the period 1979–2002. Their procedure was as follows. First, they

Table 10.1. *Average Sovereign Credit Ratings, 1996–2003*

Country	Rating	Country	Rating	Country	Rating
Argentina	BB	Hong Kong SAR	A	Philippines	BB+
Australia	AA+	Hungary	BBB	Poland	BBB
Austria	AAA	Iceland	A+	Portugal	AA
Belgium	AA+	India	BB	Russia	BB−
Brazil	B+	Indonesia	BB	Singapore	AAA
Bulgaria	BB−	Ireland	AA+	Slovenia	A
Canada	AA+	Israel	A−	South Africa	BBB−
Chile	A−	Italy	AA	Spain	AA+
China	BBB	Japan	AA	Sweden	AA+
Colombia	BB+	Korea	A−	Switzerland	AAA
Croatia	BBB−	Lithuania	BBB	Taiwan, China	AA
Cyprus	A+	Malaysia	BBB+	Thailand	BBB
Czech Rep.	A−	Mexico	BB+	Turkey	B
Denmark	AAA	Netherlands	AAA	United Kingdom	AAA
Finland	AA+	New Zealand	AA+	United States	AAA
France	AAA	Norway	AAA	Venezuela	B
Germany	AAA	Pakistan	B		
Greece	A−	Peru	BB−		

Note: Average of end-year mean foreign-currency ratings of Moody's and Standard & Poor's. Ratings shown correspond to the notation used by Standard & Poor's. Ratings of countries less than CCC are not included in the sample. Not all countries have ratings for all years.
Source: Borio and Packer (2004).

classified the countries in their sample into three groups, according to the average value over the sample period of their *Institutional Investor* credit ratings (IIR), which ranged from a highest credit score (lowest credit risk) of 100 to a lowest score (highest risk) of 0.[2] The low–credit risk group consisted of countries with average IIR greater than 67.7, the intermediate group of countries with IIR greater than 24.2 but less than 67.7, and the high-risk group of countries with IIR below 24.2. Next, the intermediate group was further divided into four subgroups ranging from least to most "debt intolerant," using the average IIR and the debt-*GDP* ratio as criteria. Subgroup I, the least debt-intolerant group, had IIR > 45.9 (the overall group mean for all countries) and a debt-*GDP* ratio below 35 percent, group II had IIR > 45.9 but a debt-*GDP* ratio in excess of 35 percent, group III had IIR < 45.9 but a debt-*GDP* ratio < 35 percent, and group IV had IIR < 45.9 and a debt-*GDP* ratio in excess of 35 percent.

RRS then ran cross-sectional regressions explaining the average IIR for each country as a function of the country's debt-*GDP* ratio, several variables capturing the country's past default history, and a variable indicating the number of 12-month periods since 1948 during which the country experienced an average inflation rate

[2] The publication *Institutional Investor* assigns a sovereign credit risk ranking ranging from 100 (lowest credit risk) to 0 (highest credit risk) to each country twice a year, based on surveys of economists and securities analysts at leading international banks and securities firms.

over 40 percent. They found (1) that this small set of variables proved to be sufficient to account for a substantial fraction (about 75%) of the cross-country variation in IIR and (2) that the sign of the debt-*GDP* ratio, while negative for countries in the intermediate- and high-risk groups, was actually *positive* for countries in the low-risk group. This pattern justified their referring to the former as "debt intolerant"; that is, for those countries, a higher debt-*GDP* ratio was associated with greater perceived default risk. In other words, those countries were on the nonlinear portion of their Δd curves.

RRS used the coefficients from their estimated regressions, together with the actual values of the nondebt explanatory variables, to estimate IIR scores as a function of alternative values of the debt-*GDP* ratio for each country. This allowed them to calculate the debt thresholds required for each country in the intermediate-risk group to move from a relatively safe subgroup to a riskier one. Their key finding was that these thresholds (the one between the intermediate-risk group and the high-risk group, for example, would approximate our d_{max}) were much lower for countries with weak prior payment histories and/or past episodes of high inflation. Their conclusion was that when it comes to perceptions of credit risk, history matters; specifically, countries with past records of fiscal probity, in the form of timely debt payment and limited reliance on seignorage, are able to sustain higher debt-*GDP* ratios than countries that have fallen short on one of those dimensions.

2. Credit Ratings

As we will discuss in Chapter 22, bond rating agencies play an important role in overcoming information asymmetries that plague securities markets, including the markets for government bonds. The most prominent rating agencies – Moody's and Standard & Poor's – provide credit ratings by classifying the default risk associated with a particular bond into one of 17 categories, ranked by degree of default risk. Each agency uses its own label for each of these categories, ranging from AAA to CCC+ for Standard & Poor's and from Aaa to Caa 1 for Moody's. Table 10.1, taken from Borio and Packer (2004), shows the average from 1996 to 2003 of the ratings that the two agencies have assigned to the sovereign bonds of a variety of advanced, emerging, and developing economies, using the classification system of Standard & Poor's. As is evident from the table, most emerging and developing economies (Chile, Singapore, and Taiwan being exceptions) receive substantially lower credit ratings than the average advanced economy. The question in this section, of course, is why this is so.

Borio and Packer (2004) attempted to answer this question empirically by taking the average credit rating assigned to a country by the two agencies, transforming it into a number (ranging from 1 to 17) and trying to explain this number for 52 countries in an annual panel from 1996 to 2003 on the basis of macroeconomic fundamentals. Their specific interest was in testing the debt-intolerance view of RRS as well as whether credit risk is affected by a country's inability to issue bonds

Table 10.2. *Explanatory Variables in Borio and Packer (2004)*

Categories	Variables
Macroeconomic	Log per capita GDP Log inflation Real GDP growth (year on year) Investment/GDP Saving/GDP Current account/GDP
Debt burden	Net debt/GDP External debt/exports Short-term external debt/FX reserves Short-term external debt/total external debt Short-term external debt/GDP FX reserves/imports
Government finance	Public debt/GDP Fiscal balance/GDP
Political, socioeconomic variables	Corruption Political risk Central bank independence
History	Dummy = 0 if no default in past 25 years Years since foreign-currency default Percentage time over 40% inflation in past 25 years
Size	Log real GDP Log real GDP (PPP terms)
Financial development	Domestic credit to private sector/GDP Market capitalization of stock market/GDP Credit plus stock market capitalization/GDP FX derivatives turnover/GDP FX spot and derivatives turnover/GDP
Original sin and mismatch variables	OSIN2, OSIN3 MISMATCH AECM

denominated in its own currency ("original sin") and by the presence of currency mismatches in the domestic economy, two issues that will play a prominent role in our analysis of optimal exchange rate regimes in Chapter 18. To examine the independent influence of these effects, however, they needed to control for a broad range of other variables that could affect credit risk. To that end, they used more than 30 variables in their regression analysis, including macroeconomic variables (such as inflation and growth, debt measures, and fiscal measures) and institutional variables (such as measures of corruption and political risk). Table 10.2 contains the complete list, which is typical of studies of this type and calls attention to the broad information set that creditors may potentially use to assess sovereign credit risk.

Borio and Packer (2004) found that per capita *GDP*, measures of macroeconomic performance such as inflation and *GDP* growth, variables proxying for political risk and corruption, and the country's default and inflation history all performed well in their panel regressions, in the sense that their coefficients were of the expected sign and statistically significant. Of these, per capita *GDP* and the institutional variables were quantitatively the most important. In a result reminiscent of RRS, debt ratios helped to explain credit ratings for emerging and developing countries but not for advanced economies. For the former countries, the debt-intolerance variables did not make credit ratings more sensitive to debt ratios, but measures of the currency composition of debt and the incidence of domestic-currency mismatches played a role in that an increase in the share of debt denominated in domestic currency, or a reduction in the incidence of currency mismatches, both improved credit ratings.

3. Determinants of Risk Premia

A more direct study of the determinants of risk premia on emerging-market government bonds was recently conducted by Dailami et al. (2005; hereinafter referred to as DMP). DMP were primarily interested in the effects of U.S. interest rates on emerging-market bond risk premia, measured as the spreads of the yields on such bonds over U.S. treasury bills. But to address that issue, like Borio and Packer, they found it necessary to control for other determinants of spreads, making their empirical results useful in assessing the broader determinants of risk premia on emerging-market bonds during recent years. They examined the determinants of spreads on government bonds issued by 17 emerging-market countries, and their empirical work was based on monthly spreads from 1991 to June 2004.

DMP distinguished between long-run and short-run influences on spreads, and they took into account likely nonlinearities in the determination of risk premia. For example, the effect of an increase in the world interest rate r on the risk premium may be very different if a country is close to insolvency than if it has a large margin between d and d_{max}. They also posited that risk premia are likely to be affected by several characteristics of the international financial environment. In the long run, these included the U.S. interest rate, a measure of investor appetite for risk (the yield spread between high- and low-risk U.S. corporate bonds), and the number of other countries in the sample that experienced a speculative attack on their currencies during each period. The latter was intended to capture possible *contagion* (spillover) effects through creditor psychology. Transitory changes in risk premia were also taken to be affected by forecasts of future U.S. monetary policy to capture creditor expectations of the effects that future changes in international interest rates could have on public-sector solvency in the emerging-market economies. Risk premia were also taken to be affected in the long run by several characteristics of the issuing economy: the debt-*GDP* ratio, the ratio of short-term debt to total debt, the ratio of reserves to debt, and the economy's degree of trade openness.

DMP distinguished between the determinants of the risk premium during periods when each country was itself in crisis and when it was not. For present purposes, their noncrisis results are most relevant. They found that domestic variables dominated the determination of risk premia: such premia increased with the debt-*GDP* ratio and decreased with the ratio of reserves to debt as well as with the economy's degree of trade openness. Surprisingly, they also decreased with the ratio of short-term debt to total debt. They also found that the U.S. interest rate entered their regression significantly and with a positive sign when interacting with the country's debt-*GDP* ratio and that the presence of crises in other countries increased the risk premium.

Overall, then, the literature is quite consistent with the model of the previous section. Both the debt-*GDP* ratio and factors that affect d_{max} appear to affect the sovereign risk premium in the ways one would expect. An important concern for policy makers, however, is that because d_{max} reflects creditors' expectations of the government's capacity and willingness to service debt under normal conditions, the range of factors that can potentially influence d_{max} is broad indeed. The question, then, is how governments can attempt to influence creditors' perceptions of d_{max} favorably to lower the government's borrowing costs and safeguard its solvency. The next section turns to that issue.

IV. MITIGATING CREDIT RISK: ACHIEVING FISCAL CREDIBILITY

As just discussed, credibility in the eyes of prospective creditors is of vital importance to the government because it plays the central role in determining not just the cost of public-sector debt but potentially the very solvency of the public sector and thus its ability to access bond markets on any terms at all. Creditors will not lend to the government unless they perceive that it has a credible plan to service its outstanding debt on market terms in the future. Moreover, as we will see in Chapter 25, when creditors' perception of the government's ability to pay falls very far short of the outstanding stock of debt, a debt "overhang" emerges that can have serious macroeconomic consequences.

Fiscal credibility in the eyes of creditors matters not just when the government considers changing the size of the fiscal deficit but also when it tries to change how it finances a fiscal deficit. If the government has been servicing its debt in part by relying on the seignorage revenue associated with a high rate of inflation, and the high economic costs of doing so cause it to attempt to stabilize, one of the most important determinants of its success in doing so is its fiscal credibility. We saw in Chapter 8 that slowing inflation requires reducing the rate of money growth and in Chapter 9 that reducing the rate of money growth requires a fiscal adjustment. But though slowing inflation requires an immediate reduction in the rate of money growth, it does not require an *immediate* fiscal adjustment. Whether the fiscal adjustment can be smoothed over time may help to influence the success of the stabilization program. In turn, the ability to postpone fiscal adjustment depends on whether creditors are willing to provide the resources required to replace the seignorage revenue that the government forgoes when it stabilizes. If creditors find the government's intention to reform its finances to be credible, then the path of fiscal adjustment can be smoothed by borrowing.

All this raises the question of how fiscal credibility can be achieved. In the rest of this chapter and in the two chapters that follow, we will consider a number of mechanisms for doing so. In this section, we will consider several ways that a government that intends to service its debt fully, but faces a credibility problem nonetheless, can attempt to convince creditors of its future fiscal intentions by *signaling* (taking actions that reveal its intentions). In the next two chapters, we will consider how the government can alternatively try to achieve credibility by taking actions in the present that "lock in" future fiscal adjustments. We will examine two ways that it can do so: by implementing institutional reforms affecting the budget process and through the privatization of state enterprises.

1. The Role of Signaling

How can the government convince its prospective creditors that its fiscal intentions are honorable, that is, that it means what it says when it promises to service its debt on the agreed terms? The first and most obvious question to ask in this context is why the government could not just simply announce its intentions and be believed. The answer is, of course, that the situation is complicated by the fact that the government can benefit from people's beliefs. We have seen that if creditors believe that the government will indeed be fiscally responsible in the future, the government will be able to borrow at relatively low cost in the present, but if creditors do not hold that belief, the government will face an increase in its borrowing costs or even lose access to private capital markets entirely. What this means is that even governments that have no particularly strong intentions of following through on their promises will have strong incentives to make proclamations of future fiscal responsibility. Because both governments that mean it and those that do not will make the same policy announcements, the private sector cannot rely on the announcement itself to gauge the government's future policy intentions.

How can they do it, then? Because what is at issue is essentially a psychological mechanism concerning how people form expectations about the behavior of others, economists have no particularly distinctive insights to offer on this issue and have found no safe and mechanical way that the credibility of government policy announcements can be assured. Instead, the basic principles governing credibility-enhancing mechanisms are the same as would apply in any other field of human activity. What is different, of course, is how they are implemented in the context of achieving fiscal credibility. Three approaches are worth mentioning.

a. Establishing a Track Record

Possibly the most reliable way for a government to achieve fiscal credibility is the hard way, that is, to establish a past track record of prudent fiscal behavior, as emphasized by Reinhart et al. (2003). Following the maxim that actions speak louder than words, a government that wants people to believe what it says about its future policy intentions would be well advised to repeatedly do what it says, that is, to be bound in its present actions by its previous policy announcements. Repeated instances in which the public observes that the government's previous announcements bind its current behavior certainly make it more likely that the public will believe that its present announcements will constrain its future behavior. RRS provide evidence that this is indeed the case.

Though this is little more than common sense, it is important to note that this mechanism for establishing credibility is not always available when it is most needed. Specifically, it is impossible for the government to establish a track record when the government in question is a new government or when what is intended even for a government that is not new is a change in regime. In these cases, the government's goal is precisely to convince the public that the past is *not* a good guide to the future. What are the options for establishing credibility under these circumstances?

b. Breaking Fiscal Taboos

If it has no recourse to an established track record, then the government can try to establish credibility in several other ways. One such way is to accompany its policy announcement with actions that only a government that gives a very high priority to fiscal solvency would undertake. For example, the government could attempt to signal the seriousness of its fiscal intentions to the private sector by implementing a fiscal action that has long been perceived as politically very difficult to undertake. Even if such an action did not remedy a structural problem in the public sector's budget, and thus did not itself contribute to improving the government's future finances, actions of this type can enhance the government's credibility by demonstrating the government's willingness to make difficult decisions and pay the political cost of doing so. The purpose would be to signal the government's intention to make discretionary changes in its budget in the future when it becomes necessary to do so in the interest of fiscal rectitude, even if the political costs turn out to be high. In a sense, by breaking long-standing political taboos, the government can indicate that it no longer intends to conduct fiscal business as usual.

The specific actions that can be taken to demonstrate such intent are likely to vary from country to country because they depend on country-specific political circumstances. In countries with a strong and politically powerful landlord class, for example, this may involve implementing or increasing agricultural taxation, whereas in countries with powerful urban unions, it may involve the removal of subsidies that favor urban workers at the expense of the agricultural sector. A particularly common taboo is the elimination of subsidies on domestic fuel prices in oil-exporting countries. Such subsidies are sometimes perceived as a birthright, and their reduction usually generates substantial political opposition. Putting the elimination of such subsidies into play can therefore be a powerful signal of the government's intention to abandon business-as-usual practices in fiscal matters.

c. Tying One's Hands

Another type of signal is for the government to accompany the announcement of its responsible future fiscal intentions with actions that it would arguably never have taken if it were not fully committed to following through on that announcement. Those actions would be intended to signal to the public the strength of the government's commitment since, having taken them, the costs to the government of not following through on its commitments would be much larger.

An example is the adoption of institutional mechanisms that deprive the government of access to the inflation tax. If the government accompanies its announcement of future fiscal rectitude with the implementation of such a mechanism, then if it fails to act in a fiscally responsible manner in the future, its inability to avail itself of the inflation tax will force it to finance its spending in a politically more painful manner, for example, through explicit and formal taxation. Knowing this, the public is more likely to believe that the government will adhere to its announced fiscal intentions in the future because it would never "tie its hands" (limit its options) in this way if it anticipated requiring recourse to the inflation tax in the future.

An interesting question concerns the nature of the institutional mechanisms through which the government can deprive itself of future inflation tax revenues. It can do this in several ways. Ranging from the most to the least extreme, these include the following:

1. Abandoning the domestic currency and replacing it with a foreign currency as means of payment. This is called *currency unification* if done multilaterally and is typically referred to as *dollarization* if done unilaterally, though foreign currencies other than the U.S. dollar may be adopted as the means of exchange. In this case, seignorage revenue arising from the domestic demand for base money accrues to the foreign issuer of the currency that is newly used in the domestic economy, and the government can finance fiscal deficits only by borrowing. Panama and Liberia have long used the U.S. dollar as a means of exchange domestically, and both Ecuador and El Salvador have begun doing so more recently.
2. Adoption of a *currency board*. This is an arrangement whereby the central bank must hold \$1 of foreign exchange for every \$1 of domestic currency it issues. The implication of this restriction is that the central bank is unable to issue

credit, either to the government or to the domestic banking system. Unlike what happens under dollarization, the domestic public sector continues to receive seignorage under a currency board because the right to print the country's legal tender allows it to purchase something of value, that is, interest-bearing foreign currency–denominated bonds that it holds as foreign exchange reserves. However, the seignorage available to the government is limited to the interest earnings on the currency board's stock of foreign exchange reserves, and most important, the government has no discretion over the amount of this revenue. Among the major developing countries, Hong Kong has had a currency board since 1983, Argentina adopted one in 1991 (but abandoned it at the end of 2001), and Estonia and Bulgaria did so more recently. Currency boards have been advocated at various times in recent years for Mexico, Indonesia, and Russia in postcrisis situations.

3. A less extreme option than either of these is a *constitutional restriction on central-bank lending to the government*, including possibly an absolute prohibition of such lending. In this case, the central bank may still be able to lend to the commercial banks at its discretion, or its total extension of credit may be constitutionally restricted.
4. Finally, least extreme is the granting of *legal independence* to the central bank. This means that the central-bank governor does not report directly to the executive, and specifically, that the governor is not subordinate to the finance minister. Typically, this is coupled with a constitutional mandate for the central bank to seek to promote price stability, at the sacrifice of all other economic goals. As we will see in Chapter 14, this particular institutional innovation has become increasingly common among emerging economies. For example, the central bank of Chile was made independent in 1989, those of Venezuela and the Philippines in 1993, and the Mexican central bank in 1994.

Implementing any of these institutional mechanisms that restrict the government's access to central-bank financing is intended to affect the government's credibility through the simple mechanism described earlier – in essence, the implementation of one of these measures signals the seriousness of the government's fiscal intentions by depriving it of a source of revenue that would be valuable to it if it contemplated running large fiscal deficits in the future. That it is willing to tie its hands and cut itself off from the possibility of having access to such sources of revenue in the future must indicate that it means what it says when it promises to mend its fiscal ways.

Remember that the government would want to do this to reap the benefits of fiscal credibility – easier access to credit through enhanced perceptions of solvency. But it is not obvious that the adoption of such mechanisms would indeed be successful in achieving the ultimate objectives sought by the government, even if the intended gain in fiscal credibility is itself attained.

Consider, for example, the effects of the government's eschewing its discretion over seignorage revenue on its perceived solvency. The reasons that such mechanisms may not enhance the government's perceived solvency is that when the government signals the seriousness of its fiscal intentions by tying its hands in this manner, it is simultaneously depriving itself of a discretionary source of revenue (the inflation tax) that it can use to service its debt. Thus, though creditors may be gratified by the government's honorable fiscal intentions, they will be concerned by the implications of the loss of control over seignorage revenue for the government's ability to service its debt. It is not clear, for example, that the more extreme forms in which the government could tie its hands – dollarization or the adoption of a currency board – will necessarily be associated with a lower risk premium on the government's foreign currency–denominated debt.[3] Because this issue concerns the benefits and costs of alternative exchange rate arrangements, we will come back to it in Chapter 18.

When it comes to effects on inflation, we can distinguish between mechanisms that operate through a fixed exchange rate (preceding points 1 and 2) and those that operate by creating "firewalls" between the government and the central bank (preceding points 3 and 4). Governments may dollarize or adopt currency boards to signal their fiscal intentions, but the government that does so cannot bind its successors to similar fiscal performance. We have seen previously that sustained inflation requires sustained devaluation of a pegged currency. This is impossible under either dollarization or a currency board, which imposes limits on the seignorage revenue that the government can hope to receive. But what happens if a future government's "demand" for seignorage revenue exceeds the "supply" that would be forthcoming under either of these arrangements? How do we know that if this happens, it is the government's fiscal accounts that will adjust rather than the exchange rate arrangement being modified to a form that provides larger seignorage revenues? The issue is whether exchange rate arrangements of this type can effectively discipline governments with a weaker commitment to fiscal rectitude than the one that implemented them. Again, because mechanisms 1 and 2 involve extreme exchange rate arrangements, we will defer discussion of issues of this type to Chapter 18.

2. "Locking In" Future Fiscal Adjustments

The last mechanism mentioned previously operated on the government's future accounts from the financing side. By depriving itself of a less painful source of financing in the future, the government can try to convince people of its intention to adjust its accounts in the future so as not to require such financing. Alternatively,

[3] In other words, such measures could have the effect of *reducing* d_{max} if the implied reduction in sgn_{max} is not offset by an increase in $p\,s'_{max}$.

the government could take actions *in the present* that would directly affect its future fiscal accounts directly, in the sense that they would be perceived as increasing the likelihood that the government would restrain its spending or increase its revenues in the future by making it easier for it to do so.

Such actions could affect either the expenditure or the revenue side of the government's budget, or both. On the expenditure side, an important measure is the liquidation or privatization of state enterprises. Because these have often been a significant drain on the government's budget, closing them down or privatizing them could – at least in principle – "lock in" a reduction in future public-sector spending. This topic is the subject of Chapter 12, which will consider the conditions under which this effect would materialize. On the revenue side, the government could implement tax reform, in the form of an overhaul of the tax laws and the system of tax administration. Replacing the existing system by one that generates more revenue at given levels of economic activity, and/or generates the same revenue at less economic cost, would be intended to lock in improved future fiscal performance from the revenue side of the government's budget. Finally, the government could implement measures in the present that provide the prospect of producing closer matches between expenditures and revenues in the future, thus locking in improved future fiscal performance. Such measures could take the form of "fiscal rules" that bind future budgetary outcomes or of reforms of the budget process that weaken the political factors that contribute to poor fiscal outcomes. Such reforms are the subject of the next chapter.

V. SUMMARY

In this chapter, we have reviewed the implications of creditors' perception that the government may not be able to service its debt as contractually agreed, that is, of sovereign credit risk. As we have seen, sovereign credit risk exacerbates the fiscal challenge faced by the government because creditors demand higher interest rates on government bonds to compensate them for the risk they have to bear, thereby increasing the government's debt-servicing costs. The reduced demand for domestic government bonds resulting from credit risk also has a contractionary effect on the domestic economy.

The empirical evidence suggests that increasing public-sector debt-*GDP* ratios are indeed associated with higher risk premia on sovereign debt, as assumed in our model. It also suggests that creditors may look at a wide range of indicators to assess the likelihood that the government will be able and willing to honor its debts, key among them the government's previous payment history and its record of responsible fiscal policy.

This raises the question of whether there is anything a government can do in the short run – that is, short of establishing a long track record of responsible fiscal behavior – to establish its fiscal credibility and reduce its borrowing costs. We reviewed alternative ways that the government can seek to achieve fiscal credibility and focused on one of them: "tying its hands" to restrict its discretionary access to the "cookie jar" of seignorage revenue. We saw that this could be achieved in a variety of ways, some of which involve the adoption of extreme exchange rate regimes. Others, however, consisting of constitutional restrictions on central-bank lending to the government and of the granting of legal independence to the central bank, do not similarly restrict the country's exchange rate regime. However, "signaling" mechanisms of this sort may or may not be successful in enhancing the government's credibility because they operate by restricting the resources on which the government can draw to service its debt.

In the chapters that follow, we will consider two other ways of achieving fiscal credibility in the short run, both of which operate by locking in future fiscal adjustments. Chapter 11 will consider the role of fiscal rules and reforms of the budgetary process, whereas Chapter 12 will examine that of the privatization of state enterprises.

REVIEW QUESTIONS

1. How does the presence of credit risk affect the government's intertemporal budget constraint?
2. In what sense can credit risk be self-fulfilling; that is, in what sense can expectations of default make default more likely?
3. What are the macroeconomic implications of increased credit risk? How can policy makers who seek to stabilize the economy respond to this situation?
4. Describe three approaches that economists have used to explore the determinants of credit risk empirically.
5. How can governments in emerging and developing economies attempt to influence creditors' perceptions of credit risks?

EXERCISES

1. Explain why the size of the fiscal adjustment that a government has to make to stabilize its debt-*GDP* ratio may depend on how likely the government's creditors believe it to be that the government will actually make that adjustment.
2. Assume that the interest rate that the government pays on its debt is subject to a risk premium so that $r = r(d(0)/d_{max})$, where r increases as $d(0)$ approaches d_{max}. Show how a loss of confidence on the part of creditors (a reduction in d_{max}) could result in government insolvency even though d_{max} never falls below $d(0)$.
3. Consider a small, open economy that is imperfectly integrated with international financial markets and maintains a fixed exchange rate. Suppose that upcoming elections cause the government's creditors to fear that a new administration may

not be fully committed to the full servicing of the government's debt, preferring instead to devote the government's fiscal resources to other goals. What effects would you expect this expectation to have on the economy today? What would the central bank have to do to counter these effects?

4. Compare the effects of an increase in the sovereign risk premium on the domestic economy when (1) the central bank targets domestic credit and (2) when it targets the money supply.
5. In the absence of fiscal credibility, the government will be tempted to finance its budget deficit via seignorage. Indicate five mechanisms to eliminate the government's temptation to print money.
6. "Expectations of public-sector default can be self-fulfilling. If the government's creditors develop the expectation that they may not be paid, it is more likely that they indeed won't be paid, whereas if they are confident that they will be paid, it is more likely that they will be." Explain.

REFERENCES AND FURTHER READING

Borio, Claudio, and Frank Packer (2004), "Assessing New Perspectives on Country Risk," *BIS Quarterly Review*, pp. 47–65.

Dailami, Mansoor, Paul R. Masson, and Jean Jose Padou (2005), "Global Monetary Conditions versus Country-Specific Factors in the Determination of Emerging Market Debt Spreads," Policy Research Working Paper 3626, World Bank.

Eichengreen, Barry, and Ashoka Mody (1998), "What Explains Changing Spreads on Emerging-Market Debt: Fundamentals or Market Sentiment?" working paper 6408, National Bureau of Economic Research.

Mauro, Paolo, Nathan Sussman, and Yishay Yafeh (2002), "Emerging Market Spreads: Then versus Now," *Quarterly Journal of Economics*, vol. 117, pp. 695–733.

Min, Hong-Ghi, Duk-Hee Lee, Changi Nam, Myeong-Cheol Park, and Sang-Ho Nam (2003), "Determinants of Emerging Market Bond Spreads: Cross-Country Evidence," *Global Finance Journal*, vol. 14, pp. 271–286.

Reinhart, Carmen, Kenneth Rogoff, and Miguel Savastano (2003), "Debt Intolerance," *Brookings Papers on Economic Activity*, vol. 1, pp. 1–62.

11

Fiscal Institutions

We saw in the preceding chapter that as the government's debt approaches its maximum borrowing capacity, prospective fiscal insolvency can cause the government's creditors to demand high risk premia to hold the government's bonds. Such risk premia would tend to result in high domestic interest rates, depressing aggregate demand and reducing economic activity. When the government's debt increases to the point where it is perceived as having completely exhausted its borrowing capacity, it will be unable to borrow at *any* interest rate. As we will see in this chapter, such a complete loss of access to credit markets by the government, because it limits the government's fiscal flexibility, has adverse economic effects that are additional to those of high interest rates. In Chapter 25, we will examine the even more destructive effects that arise when, beyond having merely exhausted its borrowing capacity, the government is perceived as having far *exceeded* it, triggering a sovereign (i.e., government) debt crisis. Given all these negative effects, it is easy to see why avoiding perceived risks of government insolvency – achieving *fiscal credibility* – would be a high priority for fiscal policy makers in emerging and developing economies.

In the preceding chapter, we began a discussion of how fiscal credibility can be achieved by considering various mechanisms for *signaling* the government's fiscal intentions to its creditors. In this chapter, we will consider a more substantive way to do so: by adopting fiscal institutions that make it more likely that the government will indeed manage its fiscal affairs responsibly. We will proceed in several steps.

First, we will ask what kinds of fiscal performance we would like "good" fiscal institutions to deliver. Our objective is to derive a benchmark of performance against which fiscal outcomes can be judged. In popular discussion, balanced budgets or fiscal surpluses are commonly perceived as good things. But it is not necessarily the case that a benevolent social planner should always seek to make the fiscal surplus as large as possible because fiscal credibility is not the only criterion by which to judge fiscal policy. Once other criteria are considered, you will see that it will sometimes be optimal for the government to run fiscal deficits. Thus good fiscal institutions should deliver not just credible fiscal solvency but also the *flexibility* to run larger-than-sustainable fiscal deficits when these are called for. We will therefore need to explore the optimal setting of fiscal policy in more detail before we can evaluate the proper role of fiscal adjustment in securing government solvency. This will be our task in Section I.

Next, we will consider why fiscal policy outcomes often fail to attain this standard of performance. If it were easy to make fiscal adjustment effective – that is, to make it happen and make it widely perceived as sustainable – it would be hard to understand why prospective fiscal insolvency is such a recurring problem in emerging-market and developing economies in the first place. Thus our second task in this chapter, which is the subject of Section II, is to explore why *pro-deficit bias* seems to be such a persistent policy problem. Understanding the forces that contribute to poor fiscal outcomes is an indispensable step in designing corrective institutional mechanisms to produce better fiscal outcomes.

With this analysis in hand, we turn in Section III to the design of institutional mechanisms to keep this bias under control and put into effect a credibly sustainable fiscal policy *regime*. The questions to be addressed in that section are, what kinds of institutional mechanisms are available to safeguard government solvency while at the same time maintaining fiscal flexibility, and what are the advantages and disadvantages of alternative fiscal institutions? Section IV summarizes and concludes.

I. OPTIMAL FISCAL DEFICITS

As a point of departure, we need an *optimality benchmark* for fiscal policy, that is, a way of assessing the kind of performance that we want fiscal policy to deliver. To derive one, in this section we will consider what we might mean by "optimal" fiscal policy in the context of development, which is the primary policy objective in most emerging and developing economies. We will then examine the implications of that analysis for the behavior of the fiscal deficit.

1. Fiscal Policy and Development

Fiscal policy can have both direct and indirect effects on development.

a. Direct (Microeconomic) Effects

Both the spending and revenue sides of the government's budget can affect economic growth through their effects on the rate of accumulation of productive capital as well as on the efficiency of its allocation. To promote growth, the government should do the following.

i. Seek the level and composition of public investment that maximizes productivity (a supply-side goal). To achieve this objective, the composition of public investment should be such as to equalize the social rates of return on all public investment projects, given the size of the public investment budget. The size of the public investment budget, in turn, should be such as to equalize the social rates of return of public and private investment.

ii. Maintain a low tax rate applying to the broadest possible tax base, with an average tax rate that does not fluctuate over time. The low tax rate and broad tax base achieve the objective of not distorting the *intratemporal* allocation of economic activity, while the constant tax rates over time achieve two objectives:

1. They avoid distorting decisions about the *intertemporal* allocation of productive activities.
2. If the distortions introduced by taxes are a convex function of the tax rate, fluctuating tax rates increase the excess burden of taxation, so constant tax rates minimize that excess burden.

b. Indirect (Macroeconomic) Effects

Fiscal policy can also affect growth indirectly through *macroeconomic* channels, that is, effects on intertemporal relative prices and the amount of uncertainty in the macroeconomic environment. Fiscal policy that is conducive to growth should respect two principles

i. Do no harm. Avoid prospective (and, of course, actual!) fiscal insolvency so that fiscal policy does not contribute to high levels of domestic interest rates and itself become a source of macroeconomic instability and crisis.

ii. Contribute to short-run macroeconomic stabilization. Promote and use fiscal flexibility, allowing the fiscal deficit to behave countercyclically to help stabilize the economy in the face of exogenous shocks.

2. Implications for Fiscal Policy

What do these principles imply about the optimal conduct of fiscal policy from a development perspective? Recall from our analysis of fiscal solvency in Chapter 8 that, setting all variables at their permanent values, the fiscal plans of a solvent government would satisfy the present value condition[1]:

$$ps'^{p} = (r^{p} - n^{p})d(0) - (\pi^{p} + n^{p})m(\pi^{p}) \tag{11.1}$$

Treating the real interest rate and growth rate as exogenous, and assuming that the government has chosen the desired sustainable inflation rate, this equation determines the sustainable value of the ratio of the adjusted primary surplus to gross domestic product (*GDP*) that is consistent with fiscal solvency, given the existing debt-*GDP* ratio. Indeed, if all the variables on the right-hand side of equation (11.1) remain at their permanent values, setting the adjusted primary surplus in this way would stabilize the debt-*GDP* ratio at its current level of $d(0)$. But as we saw in Chapter 9, this is simply the value that the adjusted primary surplus must take on average from now into the indefinite future for the solvency condition to be satisfied; the adjusted primary surplus does *not* have to satisfy this condition in every period. The question, then, is when should the government deviate from the value of ps'^{P} implied by this condition?

One reason that the government may want to deviate from the value of the adjusted primary surplus that stabilizes the debt-*GDP* ratio is to change its net worth or change the composition of its financial portfolio. For example, if the government judges that its outstanding stock of debt is too high (i.e., too close to d_{max}), it may choose to run a higher adjusted primary surplus for a period of time, paying down its debt until it has moved its outstanding stock of debt sufficiently below d_{max}. It may be motivated to do this to lower the risk premium that it is charged by its creditors, or it may do so for macroeconomic reasons, for example, increasing its own saving to increase the national saving rate.

These examples suggest reasons why the optimal adjusted primary surplus may deviate from that which is required to stabilize the debt-*GDP* ratio, perhaps even for a prolonged period of time. There are other reasons, however, why it might be optimal for the government to allow transitory deviations of ps' from even this medium-term path.

a. Intergenerational Equity

Some government expenditures – in particular, productive government investments – yield benefits over long periods of time. That means that those benefits will be enjoyed not just by the generation that is currently alive but also by future generations. An equitable distribution of the burden of paying for those expenditures would require that all who receive benefits should bear some of the costs. This suggests that productive government investments should not be financed through taxation because doing so would place all the burden of payment on the current

[1] See equation (9.10b). The superscript *P* has been introduced in equation (11.1) to emphasize that these are "permanent" values of the variables in question.

generation. The result is the so-called *golden rule* of public financing: current government expenditures should be financed through taxation, but public investments should be financed by issuing debt. The implication for ps' is that it should be low when the investment component of government spending is high and high when the investment component of public spending is low.

b. Intertemporal Tax Smoothing

The traditional neoclassical view of public debt argues that government borrowing policy should be determined by the objective of stabilizing tax rates over time. As explained earlier, the reason this is desirable is that if tax rates are expected to differ over time, those differences would introduce distortions into the intertemporal allocation decisions of economic agents: they would be induced to move productive activities into periods when these activities are lightly taxed and away from periods when they are heavily taxed. But that would mean that the *social* returns on these activities would differ at different periods of time, which would violate standard Pareto efficiency conditions. To keep intertemporal tax rates smooth, governments should not adjust tax rates to finance fluctuations in their current levels of expenditure: instead, they should keep tax rates stable and borrow when expenditures are temporarily high, while repaying debt when expenditures are temporarily low.

What is the implication of this principle for the behavior of the fiscal deficit? As we have seen, from a development perspective, optimal fiscal policy involves setting the permanent level of productive government spending according to the microeconomic principles discussed earlier and maintaining a constant tax rate. But at what level should the constant tax rate be set? The answer is that it should be set to achieve the *macroeconomic* objective of sustaining fiscal solvency, that is, satisfying the government's intertemporal budget constraint.

To see what this means for the behavior of the fiscal deficit, assume for simplicity that the rate of return on reserves is equal to the public sector's borrowing costs, so $ps' = ps = t - g$. Thus the public sector's intertemporal budget constraint will be satisfied at any time j if

$$d(j) = [ps^p + (\pi^p + n^p)m^p]/(r^p - n^p) \tag{11.2}$$

The government should therefore set

$$t = t^p = g^p + (r^p - n^p)d(j) - (\pi^p + n^p)m^p \tag{11.3}$$

which Buiter (2003) refers to as the *permanent balance rule.*

To see what this implies for the behavior of the *actual* ratio of the conventional fiscal deficit to *GDP* (*def*), write the deficit as

$$def = g - t + Rd$$

Substituting from equation (11.3),

$$\begin{aligned} def &= g - [g^p + (r^p - n^p)\,d - (\pi^p + n^p)m^p] + Rd \\ &= (g - g^p) + (R - R^p)d + (\pi^p + n^p)(d + m^p) \end{aligned} \tag{11.4}$$

where $R^P = r^P + \pi^P$. Notice some implications of this result:

- If g, r, π, and n are at their permanent levels, $def = (\pi^P + n^P)(d + m^P)$, as in Chapter 9.
- The deficit should *accommodate* temporary increases in g and R arising, say, from lumpy public expenditures, from temporary increases in the costs of debt service due to a higher risk premium faced by the domestic government, or from an increase in expected inflation that is incorporated into the nominal interest rate.

This policy rule has several important virtues:

- It ensures solvency of the government and sustainability of its fiscal-monetary program.
- It is "inflation and real growth corrected," so it does not arbitrarily impose a balanced budget during normal times.
- It respects intertemporal tax smoothing.

3. Stabilization of Aggregate Demand

Keeping the level of government spending and the tax rate constant when shocks to aggregate demand create fluctuations in real economic activity would cause the primary surplus to fall when aggregate demand and economic activity are low (during recessions) and to rise when aggregate demand and economic activity are high (during booms). Because reduced tax revenues help to cushion private disposable income during recessions and to restrain it during booms, this allows fiscal policy to function as an *automatic stabilizer* for private disposable income. To the extent that actual private disposable income affects private absorption, as we assumed in the model of Part 2, unchanged fiscal policy therefore also helps to stabilize aggregate demand automatically. Fiscal policy can play a further stabilizing role, however, if *discretionary* changes in policy can further increase fiscal deficits during recessions and reduce them during booms. The conflict between maintaining a perception of fiscal solvency and the stabilization of aggregate demand is well illustrated by the controversy that surrounded the International Monetary Fund's (IMF's) policy advice to crisis-stricken economies in East Asia in 1997–1998, discussed in Box 11.1.[2]

[2] For a more complete discussion, see Ghosh et al. (2002).

Box 11.1. IMF Fiscal Advice during the Asian Crisis

The Asian financial crisis of 1997–1998 affected countries that had previously experienced high growth and had strong fiscal positions, with low public debt–GDP ratios. The crisis threw these countries into deep recessions, and several governments in the affected countries had to borrow heavily to provide lump-sum resources to their banking systems, which had become insolvent. How should this additional public debt have affected fiscal deficits in the afflicted countries during the crisis?

This proved to be a controversial issue. The IMF initially recommended significant fiscal tightening, in part to compensate for the additional service due on the newly issued debt and thus to sustain creditor confidence. But did this fiscal adjustment really have to be made to ensure the governments' solvency? Critics of the IMF's policy advice, such as Joseph Stiglitz, the chief economist of the World Bank, argued that because of these countries' strong fiscal positions, it did not. If Stiglitz was right, tightening fiscal policy in the midst of a recession would have been poor fiscal policy because it would have been pro-cyclical, violating point 3 above. Indeed, the IMF quickly realized that the recession in the region would be much worse than it had expected and revised its policy advice to accommodate larger fiscal deficits in several countries.

II. THE POLITICAL ECONOMY OF FISCAL DEFICITS

In the real world, fiscal policy does not behave in the ideal way just described. In both developing countries and emerging-market economies, it is instead too often characterized by two properties:

1. pro-deficit bias (in the absence of constraints, fiscal deficits tend to be suboptimally high)
2. pro-cyclicality (fiscal policy tends to be expansionary during upswings in economic activity and contractionary during downswings)[3]

Can we address these problems through the appropriate design of fiscal institutions? The challenge is to design a set of fiscal institutions that can at once enhance fiscal credibility (avoid the pro-deficit bias) and maintain the flexibility for fiscal policy to react appropriately (in a countercyclical fashion) to shocks.

Before we can prescribe an institutional remedy for these problems, however, we have to diagnose the malady. Specifically, we have to ask why fiscal policy is not set optimally in the first place. In particular, why does fiscal policy often tend to display a pro-deficit bias and behave pro-cyclically?

The short answer is that budgetary outcomes are not chosen by a social welfare–maximizing social planner. Instead, they are the result of a political process of budgeting that appears to have a pro-deficit bias. The term "political process" here

[3] For early evidence on pro-cyclicality in Latin America, see Gavin et al. (1996). More recent evidence is provided by Talvi and Vegh (2005).

refers both to the procedures for formulating, approving, and implementing the budget and to the political forces that determine the outcome of those procedures. What are these political forces, and how can the problems of pro-deficit bias and pro-cyclicality be addressed? In this section, we will consider theories of three types[4]:

- fiscal illusion
- strategic interactions between governments at different moments of time
- strategic interactions between two groups that may influence policy at the same time

1. Fiscal Illusion

Fiscal illusion is the notion that voters internalize the benefits of current spending and the costs of current taxes, but not the costs of debt, because they do not understand the intertemporal budget constraint of the government. They consequently underestimate the burden of the future tax increases or future spending cuts that will be required to finance an increase in current spending. Opportunistic politicians take advantage of this illusion to get elected by using deficit-financed spending or tax cuts. This hypothesis can thus, in principle, explain both a large size of government and large debt accumulation and prospective fiscal insolvency. It can also explain two other phenomena

- **Political business cycles.** Political business cycles are systematic fluctuations in fiscal policy over the course of an electoral cycle. Typically, these cycles are associated with expansionary fiscal policies (in the form of increased fiscal deficits) just before elections and tighter fiscal policies after elections. Models to explain this pattern are often based on fiscal illusion stories in which politicians care more about voter response at times of elections and less at other times because voters are assumed to have short memories and politicians are opportunistic. Thus politicians increase spending and cut taxes just before elections but pay down debt the rest of the time. Notice that models of this type can explain fluctuations of spending and taxes around elections but not systematic trends in debt-*GDP* ratios.
- **Asymmetric stabilization in response to shocks.** Models based on fiscal illusion are also capable of explaining another commonly observed phenomenon: that governments often have an easier time increasing spending or cutting taxes in recessions than they do in cutting spending or raising taxes in expansions. Because fiscal illusion suggests that voters always reward spending increases and penalize tax increases, it becomes politically possible to implement spending increases and tax cuts in response to adverse shocks but not spending reductions and tax increases in response to positive shocks. “Illuded” voters do not punish this asymmetry.

[4] The discussion is based on Alesina and Perotti (1995).

Though this theory thus seems to account for several real-world phenomena, it has some important empirical problems. The most significant of these is that there is substantial variation in both fiscal deficit-*GDP* ratios and public debt-*GDP* ratios in the international experience. The "fiscal illusion" hypothesis cannot explain this time-series or cross-sectional variation in fiscal deficits or the debt ratio because it does not offer an explanation of variations in fiscal illusion. Even when it has been supplemented by hypotheses that attempt to do so – for example, hypotheses that attribute differences in fiscal illusion to differences in the complexity and transparency of the tax structure – the theory has not performed well in explaining such variation. Separately, the available evidence for the United States casts doubt on whether voters actually reward big spenders, as the theory suggests they should.[5]

Suppose, then, that voters actually do understand the intertemporal budget constraint faced by the government. How else could a pro-deficit bias arise?

2. Interactions between Governments at Different Moments of Time

A second set of hypotheses explains such a bias as the outcome of strategic interactions between governments that are in power at different moments of time. Consider two versions of this approach.

a. Intergenerational Redistribution

The first version explains pro-deficit bias as the result of an attempt by the current generation to distribute resources to itself from future generations. It can do so by relying on debt financing for government spending or tax cuts because that debt will be serviced through taxes levied on future generations.

Why would the current generation do that? One explanation relies on the absence of intergenerational altruism. Because only the current generation votes, if it does not care about the economic well-being of future generations, the current generation may simply use debt to transfer income to itself from future generations. A second explanation is somewhat less cynical. It assumes that intergenerational altruism exists but that the current generation is populated by rich and poor families that have different preferences about the direction that intergenerational transfers should take. Rich families expect to be better off than their descendants, so they plan to leave positive bequests. They do not care about government debt policies because they can always undo the effects of those policies on their descendants by leaving larger or smaller bequests. Poor families, on the other hand, expect their descendants to be better off than they are, so they would like to leave *negative* bequests. Because the latter is not possible through private means, poor families can only do this by inducing the government to finance its spending by borrowing rather than by levying current taxes. Because poor families therefore favor debt financing, while the rich do not care whether the government taxes or issues debt,

[5] See Peltzman (1992).

social choice results in public debt as the only way for poor families to effect negative bequests.

Unfortunately, theories of this type have both empirical and analytical problems. Empirically, these theories, like fiscal illusion theories, do not provide satisfactory explanations for why debt problems would arise at different times in the same country, or in certain countries and not others, unless altruism differs implausibly among countries and over time in the same country. They also find it difficult to explain why high debt is often accumulated and repaid within a single generation. Analytically, a serious problem with such theories is that they do not explain why future generations would not be expected just to default on the debt that they inherit. If they were expected to do so, of course, the current generation would be unable to borrow to finance larger deficits. One reason that could be offered is that the future generation (the young) may care about the older generation, which holds the debt that the young are called on to pay. But this would imply an odd one-sided generational altruism (the young care about the old, but the old do not care about the young).

b. Debt as a Constraint on Future Governments

An alternative approach focuses not on interactions between generations that succeed each other in time but between political parties that may succeed each other in power. Suppose that there are two political parties with different preferences over the composition of government spending – say, one party prefers to spend on national defense, while the other prefers social welfare spending. Suppose that the party that likes defense spending is currently in power, but there is uncertainty over who will gain power at the next election. Under these circumstances, the party that likes defense could make the average composition of spending over time more to its liking by spending heavily on defense while it is in power, and financing that spending by issuing debt, thereby imposing a tighter budget constraint on the future actions of the party that likes social welfare. A similar story could be told if one party is a high spender and the other a low taxer (i.e., if the parties have different preferences over private and public goods). Both high and low spenders would find it in their interest to issue debt – the former by spending more money, the latter by lowering taxes.

This theory has the attractive feature that unlike those discussed previously, it can, in principle, generate time-series and cross-sectional variation in pro-deficit bias. Under this story, today's government borrows more the greater the disagreement between the two parties and the more unlikely it is that today's government will be reelected. Thus variations across space and time in political polarization as well as in government fragility would be the key variables to explain variations in debt accumulation across countries and over time.

There are two analytical complications, however, that diminish the appeal of this approach. First, the amount of debt accumulation that one can explain through this mechanism may be very limited because the strategy will not be feasible (the current government will not be able to issue debt) if the future government is expected to default, and the future government indeed has an incentive to make it known that it will do so, precisely to block the current government's strategy. If there are costs of default, then the current government may still be able to borrow, but the maximum amount of debt that it can issue will be limited to that which just makes the future government indifferent between paying and defaulting. Second, the theory takes the probability of reelection as exogenous. But governments may actually issue debt precisely to influence that probability. For example, right-wing governments may issue debt to make more people debt holders, thus making them oppose left-wing governments who may default. This would imply a different theory that would be based on the political identity of the incumbent.

3. Interactions among Multiple Groups That May Influence Policy at the Same Time

Yet a third general approach interprets pro-deficit bias as the outcome of a "tragedy of the commons" based on self-interested atomistic behavior on the part of individual interest groups. The idea here is that because public-sector benefits are provided through the political process, rather than through the market, recipients do not tend to pay the full cost for them. Public spending that benefits a particular interest group will always be supported by that interest group, even if that spending is entirely tax financed, because the group receives the full benefit of the spending but pays only a small share of the total cost in the form of taxes. The smaller this share (the larger the number of groups among which the tax burden is distributed), the greater the incentive for each group acting alone to support increased spending.

In principle, this is a story about large government rather than about pro-deficit bias because voters would tend to support more spending even if it were fully tax financed, expecting someone else to pay the majority of the tax cost. However, it can become a story about pro-deficit bias if intergenerational altruism is limited, because in that case, the tax burden can be transferred not just to other contemporaneous groups but also to future generations. This makes the share of the taxes paid by the group that benefits from the spending even smaller than it would be under full tax financing. Thus the "commons" problem combined with limited generational altruism can explain a combination of large government with deficit spending. The case for pro-deficit bias becomes even stronger when one recognizes that deficit reduction (stabilization) is a public good because a stable macroeconomic environment (one with a solvent government) has the two classic attributes of a public good: it is both *nonexcludable* and *nonrivalrous*. This being so, every individual interest group has an incentive to "free ride" by enjoying the benefits of macroeconomic stability without paying *any* part of its tax costs. The upshot is that every interest group has an incentive to support public spending that benefits that group and to oppose tax increases.

Self-interested behavior on the part of individual interest groups can also explain how both favorable and unfavorable shocks to the government's budget constraint could lead to debt accumulation. When there are favorable shocks to the government's resources, every interest group will want to capture more than its proportional share of the additional resources, not internalizing the reduced spending implied for others. This *voracity effect* induces overspending of favorable transitory shocks to the government's budget. If an adverse shock hits that creates a need for a tax increase, on the other hand, this approach explains why that adjustment may be delayed: the shock leads to distributional struggles as every group seeks to place the tax burden on someone else. The upshot is that adverse shocks to the government's budget may also lead to debt accumulation.

The latter is possible, of course, only if creditors are willing to absorb the additional debt during bad times. But the so-called voracity effects that emerge during good times may actually prevent this from happening, resulting in pro-cyclicality. The reasoning is as follows: if voracity effects lead to overspending and/or tax cuts – and consequently, to excessive debt accumulation – during good times, then during bad times, creditors will cut financing below normal to impose a debt path on the government that is consistent with fiscal solvency. This will force spending to be lower or taxes to be higher during bad times than they would have been in the absence of voracity effects, thereby making fiscal policy pro-cyclical during both upswings and downswings.

In addition to explaining pro-cyclicality, models based on self-interested behavior by atomistic agents are at least in principle capable of explaining time-series and cross-sectional variation in pro-deficit bias. They relate the accumulation of debt to fragmentation of governments and to the degree of political cohesion. They suggest that countries with larger numbers of disparate interests represented in the political process would tend to have larger public debt because the larger the number of interests represented, the larger the share of the costs of government spending that can be transferred to others. Thus political, institutional, and electoral laws leading to multiparty coalition governments should be associated with higher deficits.[6]

Specifically, this approach makes predictions about the effects of different political systems on the degree of pro-deficit bias. In parliamentary systems of representation, the key characteristics are electoral formulas and district magnitudes. Electoral formulas can be "first past the post" (plurality), in which one

[6] These models are, in principle, capable of explaining, e.g., why a deficit bias emerged among several Western European countries after the 1973–1974 oil crisis because they can explain why adjustment to the oil price increase was delayed. Different fiscal institutions among these countries explain the different responses of fiscal deficits to this common shock.

representative (the one with the most votes) is elected per district or in which all seats in a district go to the party that receives the most votes; "proportional," in which multiple representatives are elected per district, allotted in proportion to the vote for each party; or mixed (a combination of the two). The district magnitude consists of the average number of representatives per district. Plurality systems tend to produce two-party systems, majority parties, and disproportionality between representation and voting shares (i.e., a party that wins small majorities in many districts may have a much larger share in the parliament than its share of the votes cast). Proportional systems with large district magnitudes, on the other hand, tend to encourage multiparty systems and coalition governments. Because the number of interests represented in the parliament is therefore likely to be larger under proportional representation with large district magnitudes, these should be more prone to pro-deficit bias than plurality systems. In presidential systems, the key characteristic is the number of rounds in presidential elections. The number of rounds affects the number of parties, with two-round systems encouraging more parties that can form coalitions in the second round of presidential elections.

The evidence is indeed consistent with these predictions. Specifically, it suggests both that minority governments in general have larger deficits and that proportional systems do so as well (see Alesina and Perotti 1995).

III. BUDGETARY INSTITUTIONS[7]

The key problem that budgetary institutions need to address, then, is that government programs tend to create benefits that are concentrated either geographically or sectorally but that are typically financed from a common pool of resources. Those who benefit from such programs thus fail to internalize their full costs because most of those costs are borne by others. Because most agents involved in the budget process represent geographical or sectoral interests, this creates a commons problem that can manifest itself both within the cabinet and within the legislature.

Budget institutions affect the rules of the game within which these agents interact, either by placing constraints on the whole process or by defining the distribution of power and responsibility among the agents. There are therefore two types of budgetary institutions: *target-oriented* and *procedure-oriented*.[8] Target-oriented institutions, usually referred to as *fiscal rules*, impose binding quantitative targets on specific (perhaps multiple) fiscal outcomes. By contrast, procedure-oriented institutions impose specific rules on the process by which the budget is formulated and approved by the legislature. Note that target-oriented institutions can affect

[7] This section draws on a follow-up paper by Alesina and Perotti (1996).
[8] See von Hagen and Harden (1995).

budgetary outcomes only if they are more difficult to change than the budget law itself, while procedure-oriented institutions can make a difference only if the actual procedures followed affect the distribution of power within the cabinet, between the cabinet and the legislature, and/or within the legislature in such a way as to increase the influence of the agents who are accountable to the public for the *total* budgetary outcome rather than for any specific part of the budget. In the remainder of this section, we will examine these two types of fiscal institutions in more detail.

1. Target-Oriented Institutions (Fiscal Rules)

Quantitative targets impose *numerical constraints* on fiscal outcomes. Their purpose is to depoliticize the macroeconomic framework. There are many possible types of fiscal rules, differing from each other in various ways.

a. Budget Rule or Debt Rule

The quantitative constraint imposed by the rule can apply to a budget (flow) concept or to a debt (stock) concept. In turn, budget rules can apply to a variety of budget concepts, including the *overall* fiscal balance, the *current* (noninvestment) balance only (a concept referred to as the *golden rule*, as we saw earlier), the *nonoil* balance (for oil-exporting countries), or the *primary* balance. The rule can specify a maximum for the deficit on any of these balances or a minimum for the surplus.

b. Presence or Absence of Additional Rules

Any of the types of rules described previously can be supplemented by additional rules that affect the composition of the budget and may facilitate observance of the first rule, for example, formal quantitative restrictions on total public expenditures, on primary expenditures, on the public-sector wage bill, on public debt service, and so on.

c. Coverage

Rules may be applied to the fiscal performance of the general government (central government, social security administration, plus subnational governments) or separately to the central government or subnational governments. In the United States, for example, the federal government is subject to a congressionally legislated debt ceiling, while many state governments have constitutionally mandated balanced-budget rules.

d. Nature of Escape Clauses

A budget rule that allowed for no contingencies under which the rule could be violated would not be credible because exceptional circumstances can be expected to create politically irresistible pressures for suspension of the rule. Thus fiscal rules often contain contingency clauses that allow the rule to be violated without violating the law. Fiscal rules can also differ with respect to the nature of these clauses.

Standard clauses are for war or natural disasters. Other escape clauses can take the form of a multiyear definition of the rule, cyclical adjustments to the budgetary outcome to which the rule applies, or provision for a contingency or stabilization fund. For example, a multiyear definition of a balanced-budget rule would require the budget to be balanced only on average over the number of years over which the rule applies. This would add flexibility because it would permit fiscal deficits during any given year, as long as they were made up for in other years. Allowing for cyclical adjustments means correcting the budgetary concept to which the rule applies for the effects of booms and busts on the government's revenues and expenditures, which is a way of preventing the rule from mandating pro-cyclical fiscal policy (see the example of Chile's budget rule in Box 11.2). Finally, a contingency fund is a fund that absorbs resources when the government runs a fiscal surplus and releases them under specified conditions to finance a deficit. This is a contingency clause in the sense that it allows the government to run a deficit under a specific set of conditions, that is, if it has previously run cumulative surpluses of sufficient magnitudes so as to have accumulated sufficient funds to finance the proposed deficit.

e. Statute

An additional dimension with respect to which rules may differ concerns their legal basis. Rules can be enshrined in the country's constitution, be a legally enacted (legislated) provision, or just be adopted as a policy guideline.

f. Sanctions

Finally, the sanctions that are applied for violations of the rule can also differ among rules. Rules may affect fiscal outcomes just through a so-called negative spotlight effect that calls attention to undesirable fiscal behavior. Applicable sanctions may thus be purely reputational. Alternatively, there may be formal judicial sanctions, allowing the government to be sued for violation of the rule and providing for dismissal of the accountable officials, financial penalties, or even criminal penalties.

The evidence suggests that fiscal rules can potentially affect fiscal outcomes. Balanced-budget rules, for example, seem to affect fiscal deficits. However, rules have been criticized on various grounds:

- Depending on how they are specified, they may be too inflexible. For example, a limit on the budget deficit can induce pro-cyclical behavior on fiscal policy – causing spending to be cut or taxes to be raised – in response to negative shocks to the economy that reduce the government's tax revenues. At such times, balanced-budget rules may also jeopardize public investment because politically, this is often the easiest item to cut in the short run. Inflexible rules

Box 11.2. Chile's Structural Surplus Rule

A policy guideline adopted by Chile in 2000 commits the government to run a minimum overall structural surplus of 1 percent of *GDP*. The term *structural* refers to the fact that the surplus measure in question corrects the actual surplus for copper revenues arising from abnormally high or low copper prices and for revenue shortfalls or surpluses arising from abnormally high or low levels of economic activity. In other words, the surplus is corrected for terms-of-trade and business-cycle effects. It is based on the use of a long-term price of copper to estimate government revenues from copper (an industry that remains publicly owned in Chile) and of a full-employment measure of *GDP* to estimate other revenues. The structural surplus (*SSURP*) is thus defined as follows:

$$SSURP_t = tP_tY_{Pt} + P_P^C Q_t^C - P_tG_t - R_tD_t$$

where t is the tax ratio, P_t is the price level, Y_P is the full-employment level of real *GDP*, P_P^C is the permanent price of copper, Q_t^C is copper output, G_t is real government spending, R_t is the interest rate on public debt, and D_t is the stock of public debt. The rule states that

$$SSURP_t \geq 0.01P_tY_t$$

Because the actual surplus is given by

$$SURP_t = tP_tY_t + P^C Q_t^C - P_tG_t - R_tD_t$$

we can derive the implications of the rule for the actual surplus as

$$\begin{aligned} SURP_t &= (SURP_t - SSURP_t) + SSURP_t \\ &= tP_t(Y_t - Y_{pt}) + (P^C Pp^C)Q_t^C + SSURP_t \\ &\geq tP_t(Y_t - Y_{pt}) + (P^C - P_P^C)Q_t^C + 0.01P_tY_t \end{aligned}$$

The implications for the actual surplus are as follows:

1. When copper prices are abnormally high ($P^C > P_P^C$), the excess revenues have to be saved because though the *actual* surplus rises, the structural surplus is not affected.
2. When economic activity is low ($Y_t < Y_{Pt}$), the government can run a countercyclical deficit because though the *actual* surplus falls, the structural surplus is not affected.

may thus imply a gain of credibility in exchange for a loss of flexibility. However, rules that are too inflexible may actually lose credibility because they will very likely be violated. For rules to be credible, the cost of lost flexibility cannot be too high.

- They may create incentives for creative accounting (e.g., through the adoption of excessively rosy scenarios for future revenues, by reclassifying expenditure items so that they are no longer subject to the rule, by transferring spending from the budget to the quasi-fiscal operations of the central bank, or through the adoption of measures that do not alter the present discounted value of spending

Box 11.3. Brazil's Fiscal Responsibility Law

After struggling with high fiscal deficits for many years, Brazil adopted a fiscal rule in May 2000 called the Fiscal Responsibility Law. This law applied to all levels of government in Brazil (federal, provincial, and municipalities). It requires each government to do the following:

1. maintain *current* fiscal balance (i.e., limit all *current* fiscal spending to the levels that can be financed out of current revenue) – this allows only public investment to be financed by borrowing, in accordance with the *golden rule*
2. limit its stock of indebtedness to a specified percentage of its current revenues determined by the federal senate in response to a proposal by the president
3. limit payroll expenditures to a specified percentage of tax revenues

Each level of government must also meet a primary balance target that is consistent with the target for indebtedness.

Noncompliance with the law is subject to corrective action and possible sanctions. Public officials in noncomplying governments are actually subject to criminal prosecution under the Fiscal Crime Law, which could include the payment of fines, job loss, ineligibility for public office, and even imprisonment.

and revenues but simply reallocate them over time). Creative accounting reduces transparency in the budgetary process and makes it more difficult for the private sector to monitor the government's fiscal policy (see Easterly 1999; Milesi-Ferretti 2003).
- They may be difficult to enforce.

These criticisms suggest that for rules to be effective, they must be designed correctly. Ideally, they should

- be constitutionally grounded so that they are costly to amend
- include contingency clauses so unusual circumstances can be accommodated without breaking the rule
- be based on ex post rather than ex ante accounting to avoid the temptation to adopt assumptions about economic performance that are designed primarily to allow the rule to be complied with
- include provisions for transparency to allow the public to more effectively monitor the government's compliance with the rule
- be enforced by an open and politically independent review panel with significant sanctions for violations

2. Procedure-Oriented Institutions

Procedure-oriented rules are, in principle, superior to target-oriented ones because they can strike a better balance between credibility and flexibility by addressing the source of pro-deficit bias directly without tying the hands of the government. There are three broad types of procedure-oriented institutions. In descending order of the degree to which they constrain the government's budgetary discretion, these are delegation, procedural rules, and transparency.

a. Delegation

Delegation involves appointing an autonomous entity (say, a national fiscal council) to set the constraints on overall fiscal outcomes. The national fiscal council should consist of a group of expert individuals who are insulated from the political process and thus have little incentive to deviate from the medium-term optimum for fiscal outcomes. This politically independent agency could not be overruled by the legislature. Because the council would set the constraints on overall fiscal outcomes using its discretion, rather than on the basis of a previously specified rule, delegation represents an attempt to optimize the credibility versus flexibility trade-off. The government would be just as constrained as it would be under a fiscal rule, but the council would retain the flexibility to modify these constraints if circumstances require it.

Wyplosz (2003) discusses some of the principles that a national fiscal council could follow. It would be composed of a group of recognized experts, appointed for long terms by the highest political authority and not subject to removal unless they were to violate the law. The ultimate objective of the council is to maintain a government debt-*GDP* ratio that is consistent with fiscal solvency. Because any ratio that satisfies $d \leq d_{\mathrm{MAX}}$ complies with that requirement, the specific targeted debt-*GDP* ratio should be decided ex ante by the political authorities. The council's role could take two forms.

A strong version would endow the council with the legal authority to determine the government's primary balance. In this case, the council's responsibility would be to establish a path for the primary surplus that is consistent with achieving the debt target over some prespecified horizon, conditional on specific assumptions about economic activity as well as other variables that affect the overall fiscal outcome (e.g., interest rates or the price of some key primary commodity that has important effects on government revenues, such as copper in Chile). The links among the primary balance path, the assumptions about exogenous variables, and the outcome of the debt-*GDP* ratio that are assumed by the council should be clearly articulated and publicized in a regularly published debt sustainability report, and the members of the council would be held accountable to the legislature for meeting the latter's announced debt target by presenting testimony on a regular basis about the path of the debt ratio.

A weak form would give the council no legal authority to determine the path of the primary balance but would require it only to present publicly its expert opinion on the desirable path of the primary balance, thereby creating pressure on the government to explain any deviations from that path.

b. Procedural Rules

Procedural rules leave the fiscal outcome unconstrained from the perspective of the government but impose restrictions on the *process* through which that outcome is brought about. The budgetary process has three stages, involving in turn the formulation, approval, and implementation of the budget. The formulation stage involves the negotiation of the proposed budget within the cabinet, the approval stage involves the passage of the budget by the legislature, and the implementation stage involves the actual collection of revenues and the carrying out of spending by the bureaucracy. Procedural rules concern the process that determines how each of these stages is carried out.

The common-pool interpretation of pro-deficit bias suggests that the key issue at the formulation stage is the roles played by agents representing specific constituents (the line ministries) versus those representing the population as a whole (the finance minister). At the approval stage, the key procedural issues concern the rules of agenda setting and of amendment in the legislative process, that is, who holds agenda-setting power in the legislature and what types of amendments are possible on the legislative floor. At each stage, arrangements may be *hierarchical* or *collegial.* Hierarchical rules favor the finance minister in the drafting stage and restrict the power of the legislature to modify the budget presented by the cabinet, whereas collegial rules do not give the finance minister disproportionate power in the budget formulation process and allow the legislature wide range to amend the budget after it gets to the legislative floor.

Procedure-based rules are based on the view that procedures that limit universalism and reciprocity tend to encourage fiscal restraint. They therefore attempt to deal with the common-pool problem by making both the budget formulation and budget approval process more hierarchical. The objective of these rules is to give agenda-setting powers in the budget preparation and execution process to the agents that can best internalize the cost of funds and are most accountable for the total budgetary outcome. In the budget formulation stage, this would be the finance minister rather than the spending ministers. In the budget approval stage, it would be the executive rather than the legislature. In effect, procedural rules are a special case of delegation, in which the agent to which the decision on the overall budgetary outcome is entrusted is in this case *inside* the government rather than outside it, but is the agent within the government that is most likely to make overall budgetary decisions that reflect the *general* interest rather than parochial ones.

c. Transparency

Transparency is the least constraining of all the potential procedure-oriented institutional mechanisms for influencing fiscal outcomes because it constrains neither the outcome nor the process but instead requires that the process be *observable* by the public. The justification for the view that transparency can influence fiscal outcomes is based on the perception that the formulation of fiscal policy can be interpreted as a *principal-agent problem* in which the government acts as the agent of the public, which is therefore the principal.[9] The specific problem is that the agents in the government who have the responsibility for formulating fiscal policy (i.e., the ministers and the legislature) have incentives to conduct policies that benefit themselves (either financially or politically) rather than promote the public good. Informational deficiencies in the budgetary process (e.g., through hidden resources, miss-estimated assumptions, extrabudgetary items, hidden liabilities, and contingent liabilities) worsen agency problems because they make it more difficult for the public to observe and therefore to monitor the government's true fiscal behavior. By improving the public's ability to monitor the government's fiscal performance, transparency of the process can therefore result in better fiscal outcomes.

d. Evidence

How successful have institutional reforms been in affecting fiscal outcomes? Stein et al. (1998) find that budget procedures that include constraints on the deficit, introduce hierarchical elements into the budget process, and are relatively transparent produce smaller deficits. Alesina and Perotti (1996) also find that more hierarchical procedures produce smaller deficits. A more recent survey of the evidence by Kirchgassner (2001) confirms these results. Differences in budgetary institutions therefore seem capable of explaining cross-country differences in outcomes; however, unless the effects of shocks depend on budgetary institutions, it is harder for such differences to explain variations in pro-deficit bias over time in specific countries because such institutions tend to change slowly over time. Moreover, there is no evidence that strong fiscal institutions can neutralize the effects of very fragmented governments such as are likely to arise under proportional representation with large district magnitudes.

IV. SUMMARY

This chapter has considered the role that reform of fiscal institutions can play as an up-front measure to enhance a government's fiscal credibility. Institutional reforms implemented in the present that promise to deliver responsible fiscal performance in the future can result in immediate credibility dividends, if the reforms are well designed. The main challenge for the design of well-functioning fiscal institutions is to strike an appropriate balance between credibility and flexibility because fiscal policy needs not just to avoid destabilizing the economy on its own through (prospective or actual) government insolvency but also to be deployed flexibly to counter the effects of shocks arising elsewhere in the economy. Both roles

[9] The term *principal-agent problem* refers to a situation in which one economic actor, the *agent*, is supposed to act on behalf of another, the *principal*, but which is rife with opportunities for the agent to seek his or her own benefit at the expense of the principal.

are important if fiscal policy is to contribute to a stable domestic macroeconomic environment.

Designing fiscal institutions to attain this goal is a challenging task. Among other things, it requires understanding the political-economy source of pro-deficit bias. We considered three common explanations for pro-deficit bias in this chapter – fiscal illusion, intertemporal interactions between governments, and intratemporal interactions among competing interest groups – and found that the last of these seems best able to explain the substantial variation that appears to exist in pro-deficit bias both across countries and over time.

Institutional responses to the political-economy challenge represented by pro-deficit bias can take two main forms: the implementation of budget rules that act as a constraint on the budgetary process from outside or the redesign of the budgetary process itself to make it less susceptible to pro-deficit bias. We considered the pros and cons of both types of institutional reforms and concluded that the evidence is consistent with the view that such reforms matter: well-designed fiscal rules and movement to more hierarchical rather than collegial procedures for budget formulation can indeed improve fiscal performance.

In the next chapter, we will consider another way to try to achieve fiscal credibility by locking in improved future fiscal performance: through the privatization or liquidation of loss-making state enterprises.

REVIEW QUESTIONS

1. Describe three alternative explanations for pro-deficit bias in fiscal policy in developing and emerging economies.
2. Explain how fiscal illusion could explain both a large size of government and a pro-deficit bias.
3. Describe what hierarchical rules for budget formulation and approval are and how they can address common-pool reasons for pro-deficit bias.
4. Describe four institutional measures that a new finance minister could implement to enhance his or her government's fiscal credibility.
5. Describe some pros and cons of balanced-budget rules.

EXERCISES

1. Describe Chile's Structural Balance Rule. What would the advantages and disadvantages of having such a rule be under circumstances in which copper prices (a main export and a key source of government revenue) and real output both fall sharply?
2. Suppose that the government increases spending temporarily whenever real *GDP* falls below potential, and reduces it when real *GDP* rises above potential,

but otherwise follows Buiter's permanent balance fiscal rule. How would the fiscal deficit behave over the course of the business cycle under this assumption?

3. Suppose that the government follows Buiter's permanent balance rule, except that tax rates are set according to

$$t = t_p + \varepsilon$$

where ε is a transitory deviation of t from the permanent tax rate t_P. What does the principle of tax smoothing suggest about the optimal behavior of ε? What would countercyclical fiscal policy suggest?

4. The Stability and Growth Pact (SGP) of the European Union stipulates, among other things, that member countries that adopt the euro as their currency should exhibit fiscal deficits of no more than 3 percent of *GDP* and public debt–*GDP* ratios of no more than 60 percent.
   a. Explain how these two components of the SGP might be related to each other.
   b. On the basis of the material in this chapter, provide at least three criticisms of the SGP.

REFERENCES AND FURTHER READING

Alesina, Alberto, and Roberto Perotti (1995), "The Political Economy of Budget Deficits," *IMF Staff Papers*, vol. 42, pp. 1–30.

Alesina, A., and R. Perotti (1996), "Budget Deficits and Budget Institutions," working paper 5556, National Bureau of Economic Research.

Buiter, Willem (2003), "Fiscal Sustainability," paper prepared for the Conference on Sustainability, Efficiency and Public Expenditures in Egypt, Egyptian Center for Economic Studies, October 19–20.

Easterly, William (1999), "When Is Fiscal Adjustment an Illusion?" *Economic Policy*, vol. 14, pp. 55–86.

Eslava, M. (2006), "The Political Economy of Fiscal Policy: A Survey," RES working paper 4487, Inter-American Development Bank.

Fatas, A., and I. Mihov (2003), "The Case for Restricting Fiscal Policy Discretion," *Quarterly Journal of Economics*, vol. 18, pp. 1419–1448.

Gavin, M., R. Hausmann, R. Perotti, and E. Talvi (1996), "Managing Fiscal Policy in Latin America: Volatility, Procyclicality and Limited Creditworthiness," OCE working paper 326, Inter-American Development Bank.

Ghosh, Atish, Timothy Lane, Marianne Schulze-Gattas, Ales Bulir, Javier Hamann, and Alex Mourmouras (2002), "IMF-Supported Programs in Capital-Account Crises," occasional paper 210, International Monetary Fund.

Kirchgassner, Gebhard G. (2001), "The Effects of Fiscal Institutions on Public Finance: A Survey of the Empirical Evidence," discussion paper 2001-15, Universität St. Gallen.

Kopits, George (2001), "Fiscal Rules: Useful Policy Framework or Unnecessary Ornament?" working paper WP/01/145, International Monetary Fund.

Milesi-Ferretti, Gian Maria (2003), "Good, Bad, or Ugly? On the Effects of Fiscal Rules with Creative Accounting," *Journal of Public Economics*, vol. 88, pp. 377–394.

Peltzman, S. (1992), "The Voters as Fiscal Conservatives," *Quarterly Journal of Economics*, vol. 107, pp. 327–361.

Stein, Ernesto, Ernesto Talvi, and Alejandro Grisanti (1998), "Institutional Arrangements and Fiscal Performance: The Latin American Experience," working paper 6358, National Bureau of Economic Research.

Talvi, Ernesto, and Carlos Vegh (2005), "Tax Base Variability and Procyclical Fiscal Policy," *Journal of Development Economics*, vol. 78, pp. 156–190.

Von Hagen, Jurgen, and Ian J. Harden (1995), "Budget Processes and Commitment to Fiscal Discipline," *European Economic Review*, vol. 39, pp. 771–779.

Wyplosz, Charles (2003), "Institutions for Fiscal Sustainability in Brazil," unpublished manuscript, Graduate Institute of International Studies.

12

Privatization

As we saw in Chapter 10, governments can seek to achieve fiscal credibility by signaling their future fiscal intentions. They commonly do so by cutting themselves off from access to certain financing sources, something that they would be unlikely to do if they anticipated requiring recourse to those sources of finance in the future. Yet we also saw in Chapter 10 that while this may help to signal the seriousness of the government's fiscal intentions, and may even help to discipline future governments not similarly inclined to fiscal rectitude as the one that institutes the measure (see Chapter 18), the impact of such measures on the public sector's perceived solvency is problematic because actual and potential creditors will correctly perceive that the government has deprived itself of a source of funds with which to service debt.

An alternative approach to enhancing the credibility of fiscal adjustment is to lock in the adjustment in present value terms by undertaking measures in the present that are irreversible – or at least, that are very costly to reverse – and that can be expected to exert favorable effects on the government's budget over the indefinite future. Chapter 11 described how the implementation of fiscal rules and reforms of the budgetary process could help to commit the government to responsible future fiscal behavior. These reforms operate by constraining the fiscal actions of future governments and, in that sense, function as commitment devices.

Privatization or reform of loss-making state-owned enterprises represents yet a third potential route to fiscal credibility. In this case, the prospect of improved future fiscal performance is enhanced not by imposing constraints on the behavior of future governments but by reducing a source of pressure for future budgetary outlays. The privatization of state-owned enterprises has been widely undertaken in emerging and developing countries, and it has often been given a fiscal justification.

This chapter explores how such measures can potentially be used by the government as mechanisms to enhance the credibility of fiscal adjustment by locking in that adjustment in the present. It also describes the economic arguments for privatization and analyzes the implications of privatization for the government's intertemporal budget constraint, emphasizing that the effects are not limited to the revenue received from the sale of such enterprises and therefore that the fiscal implications of privatization are more complicated than is often imagined. In particular, the fiscal effects of privatization are not necessarily favorable. Finally, the experience with privatization and reform of state-owned enterprises in developing countries is reviewed.

I. ECONOMIC ROLE OF STATE-OWNED ENTERPRISES

In many developing countries, state-owned enterprises (SOEs, defined as government-owned or -controlled firms that generate most of their revenue from selling goods and services) have been a significant drain on the government budget. Divesting itself of such enterprises by liquidating them or privatizing them is thus one way that the government can achieve a credible fiscal adjustment. Credibility in this case comes from the fact that the source of the deficit was the location of such enterprises in the public sector, and renationalizing them would be economically and politically costly. Thus the government can in principle lock in a (present and) future fiscal adjustment by divesting itself of loss-making state-owned enterprises.

Alternatively, the government can retain these enterprises in the public sector but implement reforms in the way that they are managed that are intended to permanently improve their economic performance. However, because these measures may be less costly to reverse, or may subsequently be undermined by new developments, reform of SOEs, no matter how successful in the short run, may be less effective as a mechanism for locking in future fiscal adjustments than divestiture.

State-owned enterprises have been a common postindependence phenomenon in a large number of emerging and developing countries. In many such countries, the state has been the driving force behind the country's development efforts, in part because of the absence of similarly well-organized private-sector institutions and in part because of the perceived productivity of public goods (such as infrastructure, health, and education) in development. Beyond the provision of public goods, a specific role for the state in *production* was perceived to be important in many such countries in their early stages of development, for a variety of reasons. These included, of course, ideological motivations favoring the public sector over the private sector as well as production by nationals as opposed to direct foreign investment. They also included motivations of a purely political sort because an expanded public role in production increased the scale and scope of the economic dispensations under the control of the ruling party. Beyond these, however, there were more strictly economic reasons for favoring direct public production in some spheres of economic activity. Specifically, private firms had limited access to external finance, and domestic financial underdevelopment – not to mention financial repression (see Chapter 21) – limited the access of such firms to domestic sources of finance as well.

During the 1970s, the combination of favorable terms of trade for many countries and low international real interest rates gave a huge boost to debt-financed public investment in emerging and developing economies. The accumulation of debt during that decade that we will examine in Chapter 25 was in fact associated with an increase in investment in many heavily indebted countries, and this investment often took the form of the creation or expansion of state-owned enterprises.

An interesting question is why such enterprises have widely become drains on public-sector budgets. In part, the answer is simply that they represented poor investments. The collapse of international terms of trade for many emerging economies during the decade of the 1980s implied that many investments made during the previous decades turned out to be poor ones ex post. Beyond this, however, such enterprises have widely suffered from structural problems brought on by *soft budget constraints* (essentially fiscal subsidies) and political interference. The absence of market discipline associated with soft budget constraints resulted in mismanagement, and political interference in the form of the imposition of nonmarket objectives on the management of the public enterprises also impaired their performance.

Thus divestiture through liquidation or privatization of such enterprises has become an attractive way to help cope with fiscal deficits. Arguments for privatization, however, are of two types.

1. Efficiency Arguments

Economic efficiency is ultimately the most important consideration in deciding whether to privatize a state-owned enterprise. Structural problems such as those mentioned earlier may make it better to run many firms privately, even if *in principle* public ownership can be compatible with economic efficiency. Moreover, even if a public firm makes substantial profits, and is therefore not a drain on the public-sector budget, the economy may be better off moving the enterprise to the private sector, if the latter can produce an even higher return.

2. Fiscal Arguments

The fiscal argument for privatization has already been made above. However, the fiscal effects of privatization are often measured in the popular press by the contribution that privatization proceeds make to fiscal revenues, reducing the fiscal deficit in the year that privatization takes place and revenues from sales of public enterprises are received. But this is incorrect, as we will see later. The fiscal effects of privatization are not generally equal to privatization revenues for two reasons:

1. This measure neglects the fact that the privatization, as an asset sale, is a one-time event.

2. Measuring the fiscal effects of privatization in this way fails to take into account the (positive or negative) effects on the government's budget of *keeping* the enterprise.

In the next section, we will examine the fiscal effects of privatization in more detail, taking both these factors into account.

II. THE FISCAL EFFECTS OF PRIVATIZATION

As a point of departure, recall the condition that determined the sustainable value of the public sector's adjusted primary fiscal surplus (equation (9.13)):

$$ps' = (r - n)d(0) - (\pi + n)m(\pi)$$

To identify the impact of public enterprises on the public sector's budget, we can break down the adjusted primary surplus into two components: the portion contributed by the state-owned enterprise sector and the rest of the primary surplus,

$$ps' = p + (r_G - n)\,k$$

where p is the nonenterprise component of the adjusted primary surplus; k is the public-sector capital stock valued at replacement cost, expressed as a ratio to gross domestic product (GDP); and r_G is the ratio of the net income (profits minus depreciation) received from public-sector enterprises to the value of the public-sector capital stock. Thus $(r_G - n)$ is the *net* revenue that the public sector receives on a permanent basis from maintaining a public enterprise capital-output ratio equal to k (because each period, it has to plow the amount nk back into public-enterprise investment, the budget receives only the excess $(r_G - n)k$). With this notation, the sustainable value of the public sector's primary surplus can be written as follows:

$$p + (r_G - n)\,k = (r - n)d(0) - (\pi + n)m(\pi) \tag{12.1}$$

State-owned enterprises are a net drain on the public sector's budget if $(r_G - n)k$ is negative. The drain on the budget is larger, of course, the smaller $(r_G - n)$. We can see that because the sustainable value of the primary surplus is given by the right-hand side of equation (12.1), this budgetary drain puts pressure on p, which has to be larger to compensate for any revenues lost through negative values of $(r_G - n)k$.

How do we know whether privatizing state enterprises provides a permanent improvement to the public sector's budget? We can answer this question by investigating in what direction the nonenterprise adjusted primary surplus p would have to change for the preceding equation to continue to hold if the government sells off the public-sector capital stock k to the private sector.

As a first pass, we may want to consider what would have to happen to p if k were simply given away. In that case, there would be no privatization revenue, but privatization would nevertheless have a fiscal effect. The reason is because the elimination of the term $(r_G - n)k$ would require an adjustment in p to maintain the solvency condition in effect. If $(r_G - n)k$ is negative (state-owned enterprises are a budgetary drain), p could afford to fall if k were set to zero through the liquidation of state enterprises, even if that liquidation brought in no revenue. It is because the sustainable value of its primary surplus can afford to be smaller under these conditions, while still retaining the public sector's solvency, that we can say that liquidation of unprofitable state-owned enterprises can enhance the government's fiscal credibility. Removing this drain on the budget simply makes it easier (less fiscal adjustment is required) for the public sector to service its debt in the future. This exercise makes the point, then, that the fiscal implications of privatization are not limited to those of the privatization revenues.

To examine how the sale of state-owned enterprises affects the public sector's fiscal accounts in the more realistic case in which public enterprises are sold, suppose the private return from operating these enterprises would be $(r_P - n)$. Then the private sector would be willing to pay an amount given by

$$Q = (r_P - n)/(r - n)$$

for each unit of k if it is offered for sale to the private sector. Notice, by the way, that if $(r_P - n) < 0$, then no one will buy the state-owned firms, and privatization may not be feasible (e.g., this may have been the case for many public enterprises in the former Soviet Union). Now write the public-sector solvency condition in stock terms as follows:

$$\frac{p + (r_G - n)\,k + (\pi + n)m(\pi)}{r - n} = d(0) \tag{12.2}$$

Note that if the private sector pays Q per unit of public-sector capital, then the effect of the sale is to replace $(r_G - n)k/(r - n)$ (the present value of net revenues from state-owned enterprises) on the left-hand side of equation (12.2) by the sales revenue Qk. Thus the fiscal impact of the sale is given by[1]

$$\frac{\Delta p}{r - n} = \left[Q - \frac{(r_G - n)}{r - n}\right] k \tag{12.3}$$

This means that the government's fiscal position will be eased (the required adjustment in p will be *negative*) as long as $r_P > r_G$ because this will make the term inside the square brackets positive.

[1] Profit taxes do not affect these calculations. If the government collects taxes on the profits earned in the private sector by the enterprises sold, then this would reduce the sale price by the present value of those taxes, but the government would exactly recoup this loss in the form of higher tax revenues.

The lessons to be drawn about the fiscal implications of privatization are the following:

1. The state-owned enterprise can be sold to the private sector only if it can make money in private hands.
2. The government may be (fiscally) better off selling a state-owned enterprise, even if the enterprise is making money in public hands.
3. The fiscal benefit to the government from the sale will be
   - less than the sale price if $(r_G - n) > 0$ (i.e., if the enterprise was profitable in public hands)
   - negative if $r_P < r_G$ (if the enterprise was more profitable in public hands than it will be in private hands)
   - equal to the sale price if $(r_G - n) = 0$ (if the enterprise was neither a net drain on nor a net contributor to the public sector's resources)
   - more than the sale price if $(r_G - n) < 0$ (if the enterprise was a net drain on the budget but will be profitable in private hands)

Going back to the credibility issue, then, the important point is that by selling a loss-making state-owned enterprise, the government can lock in a future fiscal adjustment. It can increase the primary surplus in the present and the future at the same time, so creditors do not have to rely on promises of future fiscal actions that may not be kept by the same or a different government; that is, the public sector can generate a permanent fiscal adjustment (in a present value sense) *today*.

Finally, it is worth making two additional points. First, it is important that the actual impact of privatization on the budget be measured correctly. The key question is by how much the sustainable nonenterprise primary surplus is affected by privatization. The answer was given in stock terms in equation (12.3). In flow terms, it is given by

$$\Delta p = (r - n)\left[Q - \frac{(r_G - n)}{r - n}\right]k \tag{12.4}$$

That is, the present-value fiscal benefit has to be amortized to calculate the impact of privatization on the sustainable primary surplus.

Second, whether privatization is or is not a good idea does not depend on its fiscal implications; rather it depends on whether the resources involved yield a greater *social* return in public- or private-sector use.[2] But wherever their return is greater, the decision to keep the resources in the public sector or transfer them to

[2] Though there are serious methodological challenges in determining whether privatization has actually improved efficiency in practice, the empirical evidence strongly suggests that improvements in enterprise performance after privatization have been the rule rather than the exception in most countries around the world (see Megginson and Netter 2001).

the private sector will invariably have fiscal implications, and these have to be taken into account to retain the public sector's solvency.

III. PRIVATIZATION AND REFORM OF STATE-OWNED ENTERPRISES IN PRACTICE

As a result of the wave of market-oriented reform that swept emerging and developing economies during the last part of the decade of the 1980s and the decade of the 1990s, privatization has received a substantial amount of attention, and many countries have privatized at least some subset of their state-owned enterprises. Consequently, a substantial amount of experience has accumulated not just with the operation of state-owned enterprises but also with the process of reforming and privatizing them in emerging, developing, and transition economies. A World Bank study prepared at the height of privatization activity (World Bank 1995) summarized both types of experiences.[3]

1. Evidence on State-Owned Enterprise Performance

With respect to the performance of state-owned enterprises, the World Bank study summarized the a priori arguments to expect them to perform inefficiently as follows: because no one really owns these firms, no one has a clear stake in maximizing their economic returns. Moreover, the existence of soft budget constraints means that these enterprises operate under no threat of bankruptcy to serve as a check on inefficiency. Finally, their placement in the public sector subjects them to multiple objectives and multiple constraints, which tend to increase transaction costs, distort incentives, and decrease managerial efforts.

The question, of course, is whether the evidence supports these a priori arguments. The World Bank study reviewed some of the literature that attempts to test these propositions. Early studies attempted to address this issue by comparing the performance of state and private firms in the same sector, across sectors, and over time. But these studies found it difficult to control for potential influences on the performance of these firms other than the type of ownership that characterized them. More recent studies looked instead at the performance of individual firms before and after divestiture. Though this type of evidence is subject to its own shortcomings (in particular, the failure to control for other changes in the economic environment in which these firms operate), this literature tends to find that privatized firms operating in competitive markets performed better than they

[3] For more recent reviews of the international experience with privatization, see Davis et al. (2000), Shirley and Walsh (2000), and Megginson and Netter (2001).

had in the public sector.[4] Results, however, tended to be less clear-cut for firms that operated in noncompetitive markets.

The World Bank study thus concluded that state-owned firms tended to hinder economic growth both because of microeconomic inefficiencies, such as those mentioned previously, and because of the financing implications that have been emphasized in this chapter: subsidies to state-owned enterprises put pressure on nonenterprise primary fiscal surpluses, thereby crowding out more productive government spending on infrastructure, health, and education, and to the extent that this pressure resulted in money financing, it led to inflation.[5]

2. Evidence on Privatization and Reform

a. Privatization

Turning to the process of privatization, the World Bank study found that despite much discussion of privatization and much actual divestiture during 1987–1993, state-owned enterprises accounted for nearly as large a share of economic activity in developing countries in the mid-1990s as they did 20 years previously, about 11 percent of *GDP*. In lower-income developing countries, the state-owned enterprise sector was found to be even larger, accounting for about 14 percent of *GDP*. The shares of state-owned enterprises in total employment in these economies tended to be similar to their shares in *GDP*.

The World Bank also found that divestiture tended to be concentrated in a few countries. Over the relatively heavy-divestiture period of 1988–1993, for example, five countries (Argentina, Brazil, Hungary, Mexico, Poland) accounted for 30 percent of developing-country transactions and 60 percent of the value of state-owned enterprises sold.

b. Enterprise Reform

With respect to reform of firms in the state enterprise sector, the World Bank study found that such attempts were not always very successful. It concluded that for reform of SOEs to be successful, it had to be politically desirable both to the leadership and to its constituency. This can happen when the government changes in such a way that those who would lose from reform are no longer a significant

[4] The caveat that factors other than changes in ownership need to be controlled for in such comparisons is particularly germane in such circumstances because privatization tends to occur precisely in the context of market-oriented reform that entails substantial changes in firms' operating environments, of a type that may well be associated with improvements in profitability.

[5] In a sample of eighteen emerging and developing countries, including both transition and nontransition economies, Davis et al. (2000) found that privatized state enterprises had tended to represent a substantial drain on the budget before they were transferred to private hands.

part of the leadership's support or when there is an economic crisis that increases the cost of continuing to subsidize SOEs.

The World Bank identified five components of successful SOE reform, referring to measures that actually tended to improve the operation of such enterprises. These components consisted of divestiture, the enhancement of competition, the imposition of hard budget constraints, the implementation of financial reform, and the introduction of changes in the institutional relationship between state-owned enterprises and governments.

Among the countries that undertook reform of the state-owned enterprise sector, the successful reformers divested more, especially when the state-owned sector was large to begin with. Also, they tended to introduce more competition for these enterprises. In noncompetitive markets, divestiture accompanied by appropriate regulation usually improved efficiency, service, and welfare in the former SOEs. For enterprises that remained in the public sector, the more successful reforming countries hardened the budget constraints facing these firms by reducing direct subsidies, putting their access to credit on a commercial footing, improving the regulation of state-owned firm monopoly prices, and reducing hidden subsidies. The measures were abetted by reform of the domestic financial sector, which tended to help impose market-related access to credit.

Where divestiture did not take place, both successful and unsuccessful reforming countries tried to improve the incentive structure of state-owned firms by changing the relationship between these firms and the government. Noting that all contracts between the government and state-owned enterprises have to deal with common problems (regarding the transmission of information between the government and enterprise managers, the system of rewards and punishments for managers, and mechanisms for ensuring commitment), and that contracts can be crafted as performance contracts (in which government employees run the firms) or management contracts (in which private firms are hired to run them), the World Bank found that the greater the extent of private participation such contracts involved, the better they tended to work.

For example, in the sample of 20 state-owned enterprises on which this part of the World Bank's study was based, in only 3 of 12 cases using performance contracts in an attempt to improve SOE performance did enterprise total factor productivity (TFP) register improvement after the contracts were implemented; 6 firms continued past TFP trends, and 3 performed worse. The World Bank judged that such contracts did not improve a poor incentive structure and did not reduce the enterprise managers' substantial information advantage. Managers were able to use this advantage to generate soft targets in performance contracts. Moreover, from the government side, rewards and penalties did not tend to be applied consistently, so they had little effectiveness, and governments often reneged on key promises in the context of such contracts.

By comparison, management contracts were not widely used in attempting to improve SOE performance. Yet they improved profitability and productivity in two of the three cases in which they were used in the sample. In these cases, governments used competition to reduce management's information advantage: either the SOEs operated in competitive markets or competitive bidding was used to award contracts for monopoly enterprises. Fees were linked to firms' performance to create appropriate incentives. Commitment was enhanced through long contract length, the possibility of renewal, and the provision for arbitration in case of disputes.

IV. SUMMARY

The public sector can enhance its fiscal credibility by taking up-front steps that produce permanent adjustments in the fiscal accounts, either on the expenditure or revenue side. Such measures have the advantage, unlike measures affecting the public sector's access to seignorage, that they work directly to increase the resources available for debt service rather than indirectly by signaling the government's commitment to undertake whatever measures may be necessary in the future to maintain the fiscal accounts in order. Aside from institutional reforms to impose constraints on the overall budgetary process, the key measures on the expenditure side of the budget involve the divestiture or reform of state-owned public enterprises, whereas on the revenue side, they involve the reform of tax laws and/or tax administration.

In this chapter, we have examined how divestiture and reform of SOEs can affect the public sector's budget. The key findings were that the impacts of privatization on the public sector's budget are not well captured by the privatization revenues received by the government during any given year and that the reform of SOEs to improve their economic performance has proven to be a challenging task, in which countries have achieved mixed success.

In a present value sense, the true fiscal impact of privatization depends not just on the revenues received but also on what would have happened had the enterprises involved remained in public hands. The impact on the permanent value of the public sector's primary surplus may be either greater or smaller than the privatization revenues received by the government. With respect to reform, we have reviewed best practice in restructuring such enterprises to improve their economic performance and have found that the international experience suggests that mechanisms that tend to enhance the participation of the private sector in running such enterprises have tended to be most successful in improving their economic performance, with benefits for the economy at large as well as for the pressures on the public-sector budget.

REVIEW QUESTIONS

1. Should the decision on whether to keep an enterprise in the public sector be determined by the fiscal implications of that decision? Why or why not?

2. Explain why the fiscal implications of public enterprise privatization are *not* given by the effects of sale receipts on fiscal deficits.
3. In what sense does privatizing a loss-making public enterprise that would actually be profitable in private-sector hands lock in a future fiscal adjustment?
4. Explain the conditions under which the government's intertemporal budget constraint would and would not be eased by privatizing state-owned enterprises.

EXERCISES

1. Should public enterprises be privatized when they are making cash losses for the government and kept in the public sector if they are making cash profits? Explain.
2. In post-Soviet Russia, corruption caused many productive public enterprises to be sold to oligarchs at substantially less than fair market value. The period immediately following the mass privatizations in Russia was characterized by very high inflation. Using diagrams, explain how these two phenomena may be linked.
3. Assume that for some state-owned enterprise, $r_G - n < 0$, while $r_P - n > 0$. If the government privatizes the enterprise and uses all its privatization revenues to pay back government debt, use our debt-dynamics diagram to show how this transaction would affect the government's rate of debt accumulation, assuming that before the sale, $\Delta d > 0$.
4. When we derived the conditions under which the privatization of state-owned enterprises would or would not improve the government's fiscal prospects, we did not take into account that the government would receive corporate profit tax revenues from the newly privatized firms. Assume that the government imposes a proportional corporate profit tax at the rate τ on the portion of private profits that is not devoted to new investment so that it collects taxes equal to $\tau(r_p - n)$ from each unit of newly privatized capital; show how the comparison between the primary surplus required for solvency if the government keeps the firm (p_G) and that required if it sells it (p_P) would be affected by this change.
5. Consider a public enterprise that is losing money when run by the public sector but that could be profitable if sold to the private sector. Assume that the government sells the enterprise and uses the revenue from the sale to pay off some of its debt. Use the debt-dynamics diagram of Chapter 9 to show how the sale of the enterprise affects the fiscal choices available to the government in the future to retain its solvency.

REFERENCES AND FURTHER READING

Davis, Jeffrey, Rolando Ossowski, Thomas Richardson, and Steven Barnett (2000), "Fiscal and Macroeconomic Impact of Privatization," occasional paper 194 International Monetary Fund.

Megginson, William L., and Jeffry M. Netter (2001), "From State to Market: A Survey of Empirical Studies on Privatization," *Journal of Economic Literature*, vol. 39, pp. 321–389.
Shirley, Mary M., and Patrick Walsh (2000), "Public versus Private Ownership: The Current State of the Debate," unpublished manuscript, World Bank.
World Bank (1995), *Bureaucrats in Business*, Oxford: Oxford University Press.

13

High Inflation and Inflation Stabilization

The last three chapters have examined a variety of ways that a government can avoid fiscal insolvency by making a credible fiscal adjustment, that is, by achieving a credible increase in the prospective future path of its primary fiscal surpluses. Governments that already have the political and economic means at their disposal to make such an adjustment face only the need to convince their creditors that they will indeed do so (Chapter 10), whereas those that do not presently have such means face the need to implement institutional reforms (Chapter 11) or take other up-front actions (such as the privatization of loss-making state enterprises discussed in the preceding chapter) to lock in more responsible future fiscal performance. But what happens in the case of prospectively insolvent governments that fail to do any of these things, that is, that fail to make a credible fiscal adjustment? Such governments face two options: they can either rely on central-bank financing of their continued deficits, or they can cease to service their debt, effectively defaulting on that debt.

We will examine the macroeconomic consequences of debt defaults, known as *debt crises*, in Chapter 25. In this chapter, we will consider the other alternative. Specifically, we will analyze the implications for the economy of the government's handling its prospective insolvency by relying on central-bank financing – in other words, by printing money. We have already seen in Chapter 8 that such a financing policy would be likely to be associated with a high rate of inflation in the medium term. It is therefore no accident that, as shown in Box 13.1, instances of debt default and of high inflation have tended to emerge together because both are symptoms of underlying fiscal weakness. Indeed, you may recall from Chapter 10 that Reinhart et al. (2003) found that a country's past debt-servicing and high-inflation history together were the best predictors of its government's current credit rating.

As suggested by Figure 13.1, high inflation is not an uncommon phenomenon. Table 13.1 lists the number of high-inflation episodes (defined as 12-month periods with annualized inflation in excess of 25%) identified by Fischer et al. (2002) in a

Table 13.1. *Incidence of High-inflation Episodes in a 133-Country Sample, 1960–1996*

Range of Annualized Inflation	Number of Episodes	Number of Countries
25 and above	212	92
50 and above	87	49
100 and above	45	25
200 and above	17	13
400 and above	13	11

Source: Fischer et al. (2002)

Box 13.1. Debt Default and High Inflation

As mentioned in the text, episodes of debt default and high inflation tend to occur together because both are symptoms of unmet fiscal challenges. An interesting historical illustration of this point is provided by Reinhart and Rogoff (2008). Using a new historical database that covers 66 countries in Africa, Asia, Europe, Latin America, North America, and Oceania over the period 1900–2004, Reinhart and Rogoff compute the share of countries with inflation rates above 20 percent and the share of countries going through external debt default episodes in each year. As shown in Figure 13.1, they find a strong correlation between the share of countries in default on debt and the share of countries experiencing high inflation, with a much stronger co-movement in the post–World War II period.

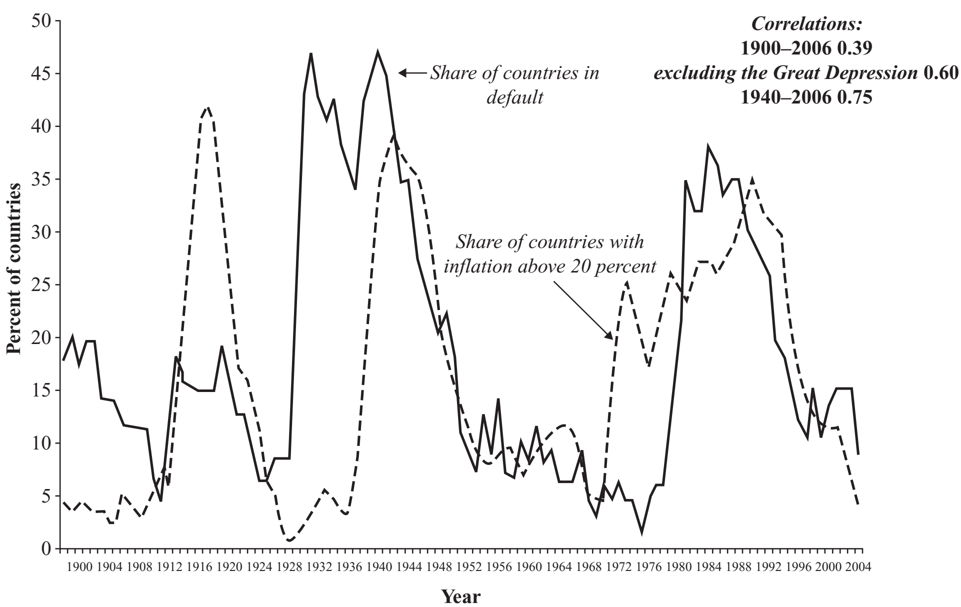


Figure 13.1. Inflation and external default

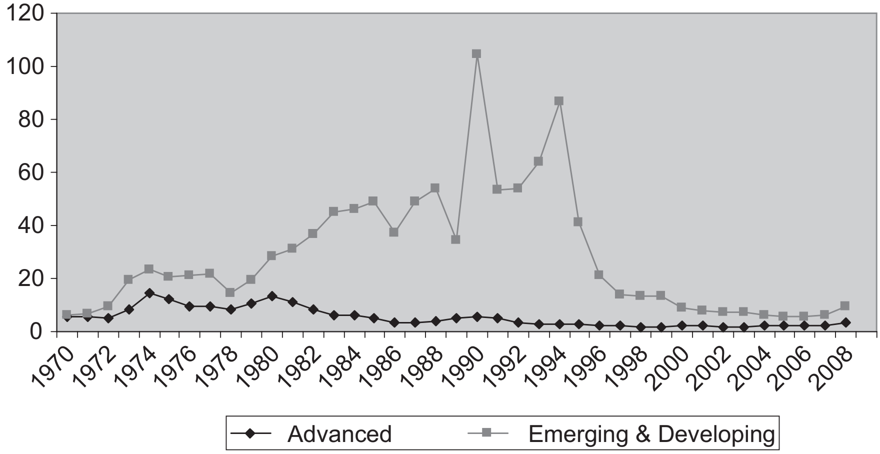


Figure 13.2. Average inflation: Advanced economies versus emerging and developing economies, 1970–2008

sample of 133 countries during the period 1960–1996. Fully one-fifth of the country months in their sample featured an annualized inflation rate in excess of 25 percent during this period, and Fischer et al. indeed found a strong empirical relationship among the high-inflation countries in their sample between inflation and fiscal deficits. Moreover, as shown in Figure 13.2, high inflation has historically been a much more serious problem among emerging and developing countries, where fiscal institutions have traditionally been weak, than among advanced economies.

The issues that we will examine in this chapter are how high inflation affects other aspects of macroeconomic performance and how a government that finds itself in a high-inflation macroeconomic equilibrium can make the transition to one with a lower rate of inflation, a process that is referred to as *inflation stabilization.* Regarding the second of these issues, we saw in Chapter 9 that if the government seeks to achieve a lower rate of inflation in the near future, it will have to accept a lower rate of money growth, which will have to be accompanied by a fiscal adjustment either in the present or in the future if the reduction in inflation is to be sustained. We may wonder, however, how the process of inflation stabilization may itself affect macroeconomic performance and whether other policies can affect the implications of inflation stabilization for macroeconomic performance.

To be specific, in this chapter, we will consider the implications of high inflation for long-run growth and will also analyze alternative approaches to inflation stabilization. These alternative inflation stabilization strategies complement fiscal adjustment with different policies regarding the nominal variables that are actually or potentially controlled by the government. They include *money-based* versus *exchange rate–based* stabilization as well as *orthodox* versus *heterodox* stabilization. The former refer to the *nominal anchor* (the exogenous nominal policy variable that pins down the domestic price level) used by the government during the process of inflation stabilization and the latter to whether the government intervenes directly in the wage-price process during the inflation stabilization effort. The evidence on the effects of disinflation on growth will be reviewed, with the point being made that stabilizing from high inflation may have *positive* effects on growth even in the short run, in contrast with the "sacrifice ratio" analysis typically applied to the disinflation issue at the much lower inflation rates characteristic of industrial countries.

I. INFLATION AND LONG-RUN GROWTH: THEORY

As shown in Figure 13.2, though inflation rates have tended to come down on average during recent years all over the world, high inflation is a problem that emerging and developing economies have traditionally suffered from to a much greater extent than have high-income countries. In analyzing our benchmark model, we saw how inflation was generated in an open economy with a managed exchange rate. Subsequently, we showed that the policies that generate sustained high inflation have their origins in fiscal problems and that fiscal adjustment, either now or later, is necessary to stabilize the domestic price level. Now we need to ask why anyone who is interested in long-run economic growth and development should care about inflation; that is, what is the relationship between economic performance and inflation?

This question can actually be posed in three separate ways:

1. What is the relationship between the *level* of the inflation rate and the *level* of output?
2. What is the relationship between a *change* in the rate of inflation and the *level* of output?
3. What is the relationship between the *level* of the inflation rate and the *rate of growth* of output?

The first two of these are the domain of the traditional Phillips curve.[1] The mainstream view in macroeconomics is that there is no relationship between the *level* of the inflation rate and the *level* of output (the long-run Phillips curve is vertical) but that a *positive* relationship may exist between *changes* in the rate of inflation and the level of output (the short-run Phillips curve has a positive slope). The second of these is controversial. Some have argued that the relationship depends on the rate of inflation, with a positive relationship for low inflation rates and zero or even a negative relationship for high inflation rates.

[1] The Phillips curve is traditionally expressed as a relationship between the unemployment rate and the rate of wage inflation, but it is easily reinterpreted as a relationship between the level of output and the rate of price inflation.

But what about point 3? Why should there be any relationship at all between the level of inflation and the rate of growth of economic capacity? We will consider the analytical links between these phenomena first, then turn to some empirical evidence.

Recall from Chapter 4 (equation (4.5a)) that the growth accounting decomposition allows us to express the rate of economic growth as follows:

$$\Delta Y_P / Y_P = \Delta A / A + \theta \Delta K / K + (1 - \theta)(\Delta L_P / L_P)$$

Thus, if inflation affects the growth rate, it must do so by altering the rate of growth of productivity or the rate of accumulation of productive factors. Could it do so? Several potential channels of influence have been identified by economists. We can classify them into channels that operate through the rate of capital accumulation and those that do so through the rate of growth of total factor productivity. Channels that operate through the rate of capital accumulation include the following.

1. The Mundell-Tobin Effect

An early analysis of the effect of inflation on growth argued that inflation could be conducive to growth by increasing the incentive to accumulate physical capital. To see how such an effect could arise, consider an economy in which non-interest-bearing money is the only financial asset (i.e., there are no bonds) so that the private sector can accumulate savings only in the form of money or physical capital, which is of a single homogeneous type. Suppose for simplicity that the economy's saving rate is constant. In this case, the allocation of saving between money and capital would depend on the relative rates of return paid by the two assets. Because it affects these relative rates of return, ongoing inflation will affect the incentive to allocate savings to the accumulation of physical capital in such an economy. An increase in the steady-state rate of inflation would lower the real rate of return on money, which pays no interest, while leaving the real return on capital, in the form of the marginal product of capital, unchanged. Higher inflation would therefore increase the share of saving devoted to capital accumulation rather than to hoarding (the accumulation of real money balances), thereby increasing the economy's growth rate. This substitution from money to capital is called the *Mundell-Tobin effect*, named after the economists Robert Mundell and James Tobin, who called attention to it.

2. High Real Interest Rates

However, this positive effect of inflation on capital accumulation does not hold up when we introduce other financial assets. In a more general setting in which we allow for the existence of bonds, for example, money and capital may not be close substitutes. Instead, bonds and capital may be close substitutes for each other, and money and capital may even be *complements*, so that decreasing the rate of return on money would actually *decrease* the demand for capital.

To see how expanding the range of assets available to the private sector could influence the effects of inflation on investment, suppose that instead of choosing just between money and physical capital, households can also choose to hold domestic and foreign bonds. To keep things simple, assume that these are perfect substitutes for each other (i.e., the domestic economy is characterized by perfect capital mobility). As we will now see, higher inflation may actually act to *raise* the real return on bonds, and because these compete with capital in private portfolios, this effect would *reduce* domestic investment and growth.

To see how this could happen, suppose that the central bank maintains an officially determined exchange rate. Under officially determined exchange rates, high inflation can create a transitory real exchange rate appreciation when the nominal exchange rate is not immediately adjusted to offset the excess of domestic over foreign inflation. In this case, the resulting expectation of a future real exchange rate depreciation to bring the real exchange rate back to its equilibrium value will increase the expected real rate of return that domestic residents can earn by holding bonds. The reason is the following: recall the uncovered interest parity condition that holds when domestic and foreign bonds are perfect substitutes:

$$R = R^* + \Delta S/S \tag{13.1}$$

Now let π and π^*, respectively, denote the expected domestic and foreign inflation rates. Because the domestic real interest rate is defined as $r = R - \pi$, we can write

$$r = R - \pi = (R^* - \pi^*) + (\Delta S/S + \pi^* - \pi)$$

or

$$r = r^* + \Delta Q/Q$$

where r^* is the foreign real interest rate and $\Delta Q/Q = \Delta S/S + \pi^* - \pi$ is the expected rate of *real* depreciation of the domestic currency. When the domestic currency is *overvalued* (i.e., is more appreciated than its medium-run equilibrium value) and is therefore expected to depreciate in real terms, the real rate of return that domestic residents can earn by holding bonds (whether domestic or foreign ones) will tend to be high. If bonds and physical capital are close substitutes for each other, this will discourage real domestic investment.

3. Working Capital

High inflation could also reduce the demand for physical capital if money and capital are complements, instead of substitutes. One way in which this could happen is if real money balances complement physical capital in domestic production.

This could happen if an expansion of the stock of physical capital carries with it an increase in firms' need for working capital, requiring firms to maintain larger real money balances. Because an increase in domestic inflation would increase the costs of maintaining working capital in the form of real money balances, higher inflation would make the accumulation of physical capital more expensive in real terms, tending to depress the rate of capital accumulation.

4. Capital Flight

Money and capital may also be complements when we allow for the existence of banks. Suppose that households hold capital only indirectly, by holding deposits (which act as money) in banks, which in turn invest in physical capital. Suppose that households can choose between holding deposits in domestic banks or holding foreign bonds. Then an increase in household demand for money (deposits) rather than foreign bonds implies an increase in banks' demand for physical capital. Thus, if high inflation were to reduce household demand for the financial liabilities of banks, it would reduce the flow of funds devoted to the accumulation of physical capital.

Under a set of financial-sector policies called *financial repression*, for example (discussed in Chapter 20), deposit and lending rates of the domestic banking system are artificially kept below their market-clearing levels. Among other things, this means that banking system credit must be rationed and is constrained by the funds flowing into the domestic banking system from domestic savers. When inflation is high and the exchange rate is either actually depreciating or expected to depreciate, financial repression prevents bank deposit rates from adjusting to satisfy arbitrage conditions such as equation (13.1). This causes domestic residents to attempt to place their funds abroad, a phenomenon known as *capital flight*. The reduced flow of resources into the domestic banking system reduces the flow of credit that these institutions can make available for domestic investment, with consequent negative effects on economic growth.

5. The Taxation of Capital

All the mechanisms considered so far have taken the real return on capital as given and unchanged by inflation. However, high inflation may tend to reduce the *after-tax* rate of return on physical capital, which is presumably what provides the incentive for capital accumulation. The reason is that because tax laws tend to be written in nominal terms, high inflation may tend to increase the effective rate of taxation on the return to physical capital. For example, in an inflationary environment, depreciation allowances that are written in nominal terms (i.e., in terms of the original nominal cost of capital) tend to understate the cost in current inflated prices of replacing capital as it wears out. This overstates firms' true profits and consequently generates higher tax obligations than would be due with a proper accounting of the costs of depreciation. Other effects may also be important and operate in a similar direction, for example, the exaggeration of capital gains on real assets.

6. Uncertainty Associated with Unstable Inflation

High inflation has empirically tended to mean unstable inflation. When inflation is unstable, future prices become difficult to predict. Because projections of future prices are an important input into investment decisions, this instability creates uncertainty for firms considering investment in physical capital. When this uncertainty is persistent, risk-averse investors will demand very high returns before undertaking capital accumulation. When it is viewed as transitory and investment is irreversible, capital accumulation is likely to be deferred because the option to wait before committing to an irreversible investment becomes potentially very valuable. In either case, capital accumulation is discouraged, and growth is unlikely to be very high in such a setting.

7. Inflation as an Indicator of Government Competence

Finally, some economists have argued that because economists have produced no good arguments for very high inflation (even the fiscal argument does not support *very* high inflation, as indicated by the seignorage Laffer curve of Chapter 9), a government that promises to prevent it or stop it (they all do) but nevertheless permits it to persist is one that has lost control of the economy (Fischer 1993). This creates an environment of generalized uncertainty. The same negative effects on investment and growth would be expected, as recently discussed.

In addition to these potential effects on the rate of accumulation of physical capital, inflation may also affect the rate of growth of total factor productivity. The channels through which this could happen include the following.

a. Distortions in the Information Content of Relative Prices

High inflation typically tends to increase the variability of relative prices. This tends to reduce the information content of relative prices and interfere with the efficient allocation of resources, reducing the productivity of new investment and thus the rate of growth of total factor productivity in the economy.

b. Shoe Leather Costs

Because high inflation increases domestic nominal interest rates, it creates incentives for economic agents to economize on their holding of real cash balances, as we have seen. Doing so absorbs resources that could otherwise be devoted to producing goods and services, thus reducing the total output of goods and services that the economy can produce from a given level of employment of productive factors. Consequently, this absorption of resources, referred to as *shoe leather costs*, reduces the economy's total factor productivity.[2]

This effect would tend to reduce the *level* of total factor productivity rather than its growth rate. As such, it could have a large temporary negative effect on the economy's growth rate; that is, an economy that suddenly experiences a substantial increase in its inflation rate would divert resources from productive activities to economizing on real cash balances. This would cause it to undergo a reduction in its level of total factor productivity, and thus its productive capacity would decline relative to what it would have been without the inflation acceleration. In the short run, this would show up as a decline in the economy's measured growth rate.

II. Inflation and Long-Run Growth: Evidence

Of all the analytical links between inflation and long-run growth reviewed earlier, only one – the Mundell-Tobin effect – suggested that the association between inflation and long-run growth may be positive. Indeed, there is a fairly strong consensus within the economics profession that high inflation is harmful for the growth of economic capacity.[3] The consensus is based not just on the analytical arguments just discussed but also on a substantial amount of empirical evidence. This section reviews some of that evidence.

An early study testing propositions about the effect of macroeconomic factors on long-run growth was the study by Fischer (1993) discussed in Chapter 3. Fischer included the rate of inflation as one of the determinants of long-run growth performance. He found, based on cross-country growth regressions, that high inflation had a negative effect on long-run growth, even after controlling for other variables – both macroeconomic and others – that may affect long-run growth performance. One problem that Fischer did not resolve, however, was ruling out the possibility of reverse causation between inflation and growth. Though he detected a negative partial correlation between inflation and growth, this may have arisen because low growth causes high inflation rather than the reverse.

Robert Barro of Harvard University has conducted a number of studies on the relationship between inflation and growth that address this question explicitly. In Barro (1994), for example, he dealt with this problem by using *instrumental variables* estimation (see the next paragraph). His study was based on a sample of 100 countries, with data from 1960 to 1990. Barro's procedure, however, differed

2 The (tongue-in-cheek) term *shoe leather costs* comes from the notion that to economize on real money balances involves making many trips to the bank to withdraw currency when it is needed rather than maintaining large stocks of currency outside the bank. All these trips would tend to wear out one's shoes.

3 Though the consensus is not so strong on what "high" means. For recent empirical work on this issue, see Khan and Senhadji (2001).

from that of Fischer in that his regression was based on *panel data* rather than cross-sectional estimates; that is, rather than average his country-specific data across the entire span of time covered in his sample, he used as data point decade averages for the periods 1965–1975, 1975–1985, and 1985–1990.[4]

The advantage of doing this is that, by first running a regression of each of the potentially endogenous variables (such as the inflation rate) on a set of variables correlated with these variables, but not directly with growth itself (called *instruments*), he was able to decompose each of these endogenous variables into a portion that could be predicted on the basis of the instruments (consisting of the fitted values of the regressions of the endogenous variables on the instruments, called the *first-stage* regression) and a portion that could not (the residuals of the first-stage regression). The portion that could be predicted on the basis of the instruments would presumably not be influenced by contemporaneous growth, so any correlation between this portion and the contemporaneous growth rate would presumably have to arise as the result of causation running from current inflation to current growth rather than the other way around. Thus, conditional on the absence of a correlation between the instruments and growth (other than through current inflation), Barro was able to test whether any of the potentially endogenous variables had a causal effect on growth by including in his growth regression only the predicted component, rather than the actual value, of the variable in question. This instrumental variable procedure is a conventional way of attempting to eliminate reverse causation in tests of causal links among economic variables.

Barro included a large number of control variables in his regression. He found that growth within each decade was negatively affected by the initial value of the country's real gross domestic product (*GDP*) per capita and positively affected by the initial value of its secondary enrollment rate and the health of its citizens (as measured by life expectancy). Growth was positively affected by within-decade increases in a measure of the rule of law as well as by improvements in the terms of trade. On the other hand, within-decade increases in fertility, in the ratio of government spending to *GDP*, and in the premium in the parallel foreign exchange market had adverse effects on growth. He found that even after controlling for reverse causation, within-decade inflation had a negative effect on growth, but this effect arose only from the high-inflation observations in the sample, meaning from observations with annual inflation rates in excess of 40 percent.

One problem with this procedure is that the validity of the variables that are used as instruments is hard to establish. Barro used lagged variables of the endogenous variables as instruments on the assumption that, even if growth could potentially affect the *current* values of variables such as the rate of inflation (thereby making

[4] It is worth mentioning that Barro computed the standard deviation of inflation for each country within each decade and found that high inflation was associated with more variable inflation in his sample, as claimed earlier.

them endogenous), current growth is unlikely to have affected the *past* values of those variables. However, lagged endogenous variables may not always be reliable in that role. For example, if there is indeed reverse causation from growth to contemporaneous inflation, and growth itself tends to be persistent across decades (so current growth is correlated with past growth), then a negative association between current growth and *lagged* inflation could arise simply because a positive shock to growth in the last decade would have been associated *both* with lower inflation in that decade and with higher growth in the current decade. In other words, even though current growth may not affect lagged inflation directly, the two variables could nonetheless be correlated through the influence of lagged growth on both of them. If so, because lagged inflation would be correlated with current growth other than through its effect on current inflation, lagged inflation would not be a valid instrument.

Accordingly, in a subsequent study, Barro (1996) revisited the relationship between inflation and growth using a different instrumental-variable procedure. As in the earlier study, he relied on a data set consisting of 100 countries, with annual observations from 1960 to 1990. He again used a panel data set consisting of the three periods 1965–1975, 1975–1985, and 1985–1990. His panel regressions controlled for a large number of additional explanatory variables, including, as in the previous study, initial conditions, contemporaneous exogenous factors, and a diverse set of macroeconomic policy and outcome variables.

An interesting aspect of the more recent study that lends credence to the interpretation of a negative partial correlation between inflation and growth as reflecting causation from inflation to growth was the introduction of some novel instruments for inflation. Rather than relying only on lagged values of endogenous variables as instruments, in this study, Barro used prior colonial history (in the form of separate dummy variables for Spanish or Portuguese colonial background or other colonial background). These turned out to be correlated with long-term inflation performance. Using these as alternative instruments for the inflation rate left Barro's previous results in place.

Barro's central finding regarding inflation and growth was that, other things equal, a 10 percent increase in the rate of inflation reduced long-run growth by about 0.025 percent per year. Barro found that it was the *level* of inflation, rather than its variability, that affected growth adversely. Though this negative relationship between growth and inflation was again driven primarily by high-inflation countries (in the sense that when such observations were excluded, the effect of inflation disappeared), he could not reject the hypothesis that the partial effect of inflation on growth was the same at all levels of inflation. Moreover, the results were robust with respect to the exclusion of a few high-inflation outliers.

The strength of this evidence depends, among other things, on the validity of the new instruments. Unfortunately, these may also be questionable because prior colonial experience may affect growth directly as well as through its effects on

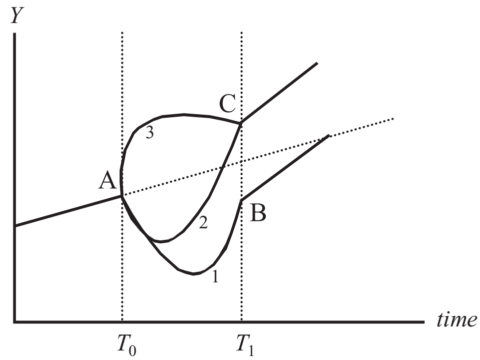


Figure 13.3. Real output during an inflation-stabilization episode

the inflation rate. Therefore, though the jury is still out, the evidence is at worst suggestive of a negative relationship between inflation and capacity growth.

III. INFLATION STABILIZATION: THEORY

From the previous two sections of this chapter, we have learned that there are both theoretical and empirical reasons to believe that high inflation is associated with lower long-run growth. In Chapters 8 and 9, we learned that in a typical emerging or developing economy, high inflation is the product of continuous exchange rate depreciation with monetary accommodation, driven by high fiscal deficits. This told us that reducing inflation requires a fiscal adjustment, either now or later, essentially because it implies a reduction in seignorage revenue. So we have strong reason to believe that high inflation is harmful, and we know what is required to bring it under control.

1. Costs of Inflation Stabilization

But does this mean that stopping high inflation is necessarily a good idea? The answer depends on what happens to an economy over time as we make the fiscal adjustment required to bring inflation down. The reason this matters is that even if inflation is bad for growth, *stopping* inflation may also be bad for growth. If the costs of stopping inflation are sufficiently high relative to the costs of continuing inflation, it may not be worth doing, even if inflation is bad for growth. To see this, consider how the path of real *GDP* might evolve over time in the course of an inflation-stabilization episode, as shown in Figure 13.3.

This figure illustrates three possible paths for the economy's real output during an inflation-stabilization episode. Suppose that the economy's real output follows the shallow low-growth path given earlier up to the point A in a high-inflation equilibrium (the growth rate determines the slope of the path), and that a stabilization is contemplated at time T_0. Suppose further that the transition to a low-inflation high-growth path is completed by time T_1. Then what we have learned about inflation and growth so far tells us only that the path followed by the economy after T_1 should be steeper than the path it followed before T_0 because the economy's long-run growth rate would be expected to be higher in the post-stabilization low-inflation equilibrium than in the initial high-inflation one.

But we do not know, for example, whether the economy will start on the new path at the point B or the point C. This depends on what happens during the transition. Suppose we know it starts from B. Would we advise the country's authorities to undertake a stabilization? The answer depends on whether the domestic political system judges that the eventual gains (beginning where the dashed line consistent with the original growth path intersects the new growth path starting from B) are worth the transitional costs. Would we be able to give a different answer if we knew instead that the post-stabilization path starts at C? The answer is no. It depends on how we get there. If the path from A to C is the one labeled 2, then there is still a trade-off to face, though a more favorable one than in the previous case. However, if it happens to be the one labeled 3, then we can say that the transition is unambiguously good.

2. Modeling Inflation Stabilization

Thus we need to worry about what happens to real output when we stabilize. A natural place to start is to see what the model that we developed in Chapters 4–8 tells us about this issue. Recall that in the short-run version of that model, there were three key endogenous variables: the level of real *GDP* Y, the domestic price level P, and the domestic interest rate R. These three variables were determined by an aggregate supply relationship developed in Chapter 4, a goods-market equilibrium condition developed in Chapter 5, and a bond-market equilibrium condition developed in Chapter 6. Because our focus in Part 2 was on the determination of the domestic interest rate and the real exchange rate, we eliminated Y from this three-equation system by substituting the aggregate supply relationship into the goods- and financial-market equilibrium conditions. This enabled us to analyze the economy's equilibrium graphically using a goods-market equilibrium curve (GM) and a bond-market equilibrium curve (BB) drawn in (P, R) space. Because our interest is now in seeing how real output Y can be expected to behave during an inflation stabilization, we will derive an alternative graphical representation of the model that, instead of eliminating Y from the system, eliminates the domestic interest rate instead and displays the model's solution in (Y, P) space.

The alternative procedure can be described graphically as follows. First, because we will not be substituting the aggregate supply relationship into the goods- and financial-market equilibrium conditions, as we did in Chapters 5 and 6, we begin by redrawing the original GM and BB curves derived in those chapters for *given* values of Y. Beginning with the GM curve from equation (5.7), holding Y constant,

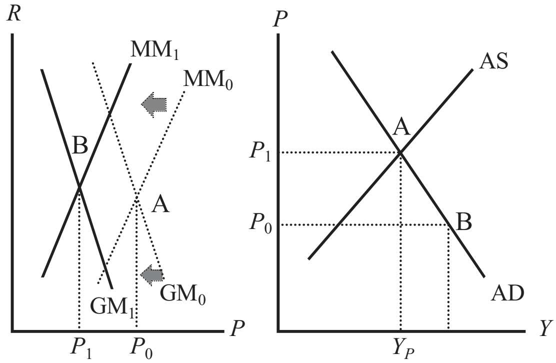


Figure 13.4. Goods- and money-market equilibrium and the aggregate demand curve

the goods-market equilibrium condition becomes

$$Y = \phi(\underset{+}{SP^*/P}, \ldots)A_P[\underset{+}{Y - T}, \underset{-}{R}, \ldots] + G + X(\underset{+}{SP^*/P}, \ldots) \qquad (13.2)$$

In (P, R) space, this can be shown as the negatively sloped dashed curve labeled GM_0 in the left-hand panel of Figure 13.4 (recall that the negative slope arises because a lower domestic interest rate increases aggregate demand for domestic goods and thus requires an increase in the price of such goods to maintain equilibrium in the goods market). Next, consider financial-market equilibrium. As we saw in Chapter 6, the specification of the financial-market equilibrium condition depends on the economy's degree of financial integration with the rest of the world as well as on the monetary policy regime followed by its central bank. The analysis that follows can be reproduced for each of the cases that we considered in Chapter 7, but suppose that for concreteness and simplicity, we consider here the case of imperfect capital mobility and monetary targeting. In that case, in a short-run equilibrium, the BB curve adjusts passively to the intersection of the GM and MM curves. The equation of the MM curve is given by

$$M = PL(R, \ Y) \qquad (13.3a)$$

This equation is depicted in the left-hand panel of Figure 13.4 as the dashed curve MM_0. It has a positive slope (recall that an increase in the domestic interest rate reduces the demand for money; thus, given a constant money supply, a higher domestic price level is required to maintain equilibrium in the domestic money market).

Next, we want to derive the set of all combinations of P and Y that are simultaneously consistent with equilibrium in the goods and financial markets. We will refer to the locus of all such combinations as the economy's aggregate demand curve (AD).[5] To see what this locus must look like, we ask what happens to the equilibrium value of P determined by our GM-MM diagram as we consider changing the level of Y. Suppose that we consider a value of Y that is *higher* than that used to draw the dashed GM and MM curves that intersect at the point A. Then, as long as the private marginal propensity to absorb is less than 1, the increase in Y will create an excess supply of goods. To restore equilibrium in the goods market, a lower value of P would be required because this would depreciate the real exchange rate and increase net exports, thereby increasing demand for domestic goods. The effect is to shift the GM curve to the left. Similarly, an increase in Y creates an excess demand for money, which implies that a reduction in P would be required to restore equilibrium in the money market; that is, the MM curve must shift to the left as well. The equilibrium in the left-hand panel thus moves from A to B, with a lower level of P and an indeterminate change in the domestic interest rate. In the right-hand panel, this corresponds to a movement from A to a point with higher Y and lower P, such as B. If, instead, we consider the effects of a *lower* value of Y, this analysis would be reversed: GM and MM would shift to the right, and the equilibrium price level would rise. Thus the set of all combinations of Y and P that clear the goods and money markets simultaneously, which we previously called the AD curve, must have a negative slope. For any given value of Y, any shock to the economy that would tend to increase P in the GM-MM diagram would shift the AD curve vertically upward by the same amount, whereas any shock that would tend to decrease P would correspondingly shift AD downward.

It is useful to derive the properties of the AD curve a little more formally. To do so, we first solve the money-market equilibrium condition for the domestic interest rate R. From the money-market equilibrium condition (13.3a), the solution can be written as

$$R = R(\underset{-}{M / P}, \underset{+}{Y}) \tag{13.3b}$$

That is, the interest rate consistent with money-market equilibrium is lower the higher the real money supply and higher the higher the level of real *GDP*. Next, substitute this expression for R into the goods-market equilibrium condition. Recall that the GM curve was derived as the set of all combinations of R and P that satisfied the goods-market equilibrium condition (13.2). By examination of equation (13.2), it is easy to see that the properties of this equilibrium must satisfy

$$Y = Y(\underset{+}{G}, \underset{+}{T}, \underset{-}{R}, \underset{+}{SP^*/P}, \ldots) \tag{13.4}$$

[5] What is the relationship between this aggregate demand curve and the one that we derived in Chapter 5? The difference is essentially that the AD curve of Chapter 5 was derived under the assumption of *interest rate targeting*, whereas the AD curve of this chapter is derived under the assumption of *monetary targeting*. Thus the difference concerns only the monetary policy regime that is assumed to be in place.

Substituting for the interest rate, we have

$$Y = Y[\underset{+}{G}, \underset{-}{T}, R(\underset{-}{M}/P, Y), \underset{+}{SP^*/P}, \ldots] \tag{13.5}$$

which represents the AD curve.

The properties of this aggregate demand curve are the following:

- Its slope is determined by the effects of P on the real money stock and the real exchange rate. An increase in P reduces the aggregate demand for domestic goods through two channels. First, the real money supply falls, which raises the domestic interest rate and reduces aggregate expenditure, including on domestic goods (a mechanism called the *Keynes effect*). Second, as P increases, the real exchange rate appreciates. This reduces net exports and thus reduces the demand for domestic goods. This curve is relatively flat when the interest elasticity of the demand for money is small (so changes in M/P have large effects on R), when aggregate spending is highly sensitive to the interest rate, and when changes in the relative price of domestic goods have a large effect on the demand for those goods. That will be the case when the elasticities of export and import demand are large.
- What determines the position of the curve? The curve is shifted vertically up or down by changes in real variables such as fiscal policy and exogenous shifts in domestic absorption or the demand for the country's exports. Increases in government spending, tax cuts, and exogenous increases in domestic absorption or exports all create an excess demand for domestic goods at a given level of output. Consequently, to restore equilibrium in the market for domestic goods, their price has to rise. This means that the AD curve is displaced vertically upward (a higher price level is required at any Y to clear the market). The curve would shift downward if any of these shocks were reversed.
- The AD curve is also shifted by changes in nominal variables – specifically, the exchange rate S and the money supply M. A devaluation of the nominal exchange rate or an increase in the money supply can shift the curve vertically. More important, equal proportionate increases in S and M shift the AD upward in the same proportion because an equiproportionate change in S, M, and P would leave both the real exchange rate and the real money supply unchanged and thus would be consistent with equilibrium in the goods and money markets at a given level of production. This means that the process of ongoing depreciation and money growth can be illustrated in the form of continuous *upward* movements of the AD curve in proportional amounts equal to the ongoing rate of nominal exchange rate depreciation and money growth.

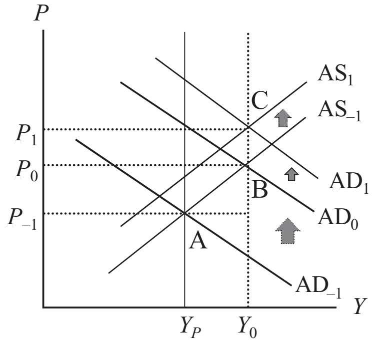


Figure 13.5. The process of ongoing inflation

The final component of our model is the aggregate supply curve given by equation (4.11), reproduced here as equation (13.6):

$$Y = AF[K, L(Pw_P/W_{-1})L_P] = Y(P; A, K, W_{-1}, \ldots) \qquad (13.6)$$

As we saw in Chapter 4, when drawn for a *given* value of W_{-1}, the aggregate supply curve has a positive slope in (Y, P) space, and it must pass above the point on the horizontal axis corresponding to Y_P at a value of the domestic price level P that equals W_{-1}/w_P. The latter means that when W_{-1} changes, the upward-sloping AS curve must shift vertically in proportion to the change in W_{-1}.

Plotting this aggregate supply relationship in the right-hand panel of Figure 13.4, together with the aggregate demand relationship we derived previously, completes the recasting of our model in (Y, P) space and gives us the means to pin down the equilibrium values of P and Y. As drawn in Figure 13.4, the economy is initially in equilibrium at price level P_1, which is assumed equal to W_{-1}/w_P, and real output is therefore initially at its full-employment level Y_P.

To see how ongoing inflation can come about in this context, consider what happens in response to an increase in aggregate demand that takes place, say, in period 0. The AD curve shifts to the right, to a position such as AD_0 in Figure 13.5, causing the economy to move from A to B, with an increase in the equilibrium level of real output from Y_P to Y_0 as well as an increase in the equilibrium price level from P_{-1} to P_0. Recall from Chapter 4 that such an increase in the price level would shift the labor demand curve vertically upward more than the labor supply curve, so the level of employment would rise and the nominal wage would increase, but less than in proportion to the increase in the price level.

How large is the increase in the nominal wage in period 0? Recall from Chapter 4 that the wage adjustment process is given by

$$\Delta W = \sigma(Pw(L) - W_{-1})$$

Dividing both sides of this equation by the lagged nominal wage, we can express the *proportional* change in the nominal wage in period 0 as

$$\Delta W/W_{-1} = (W_0 - W_{-1})/W_{-1} = \sigma[(P_0/W_{-1})w(L_0) - 1] \qquad (13.7)$$

where $w(L)$, the *supply price* of labor, is an increasing function of L. Thus the proportional change in the nominal wage as a result of the aggregate demand shock in period 0 must be larger the larger the increase in the price level P and the larger the increase in employment L – that is, the larger the size of the aggregate demand shock.

But recall that the vertical position of the aggregate supply curve is determined by W_{-1}/w_P. Thus, in the period after the aggregate demand shock (period 1), the economy will inherit a nominal wage that is higher by the percentage amount given by equation (13.7). This will cause period 1's aggregate supply curve to shift vertically upward by exactly that proportionate amount, say, to a position such as AS_1 in Figure 13.5. What would it take to keep the economy at the level of real output Y_0 in period 1 under these circumstances? The answer is that the economy would stay at Y_0 only if the AD curve were to shift up again in period 1, by exactly the same amount as the aggregate supply curve did – that is, by $\Delta W/W_{-1}$ percent. As we have seen, this would happen if the exchange rate were devalued by $\Delta W/W_{-1}$ percent, and the money supply were increased by $\Delta W/W_{-1}$ percent as well, in period 1. In that case, the price level in period 1 would increase by $\pi_1 = (P_1 - P_0)/P_0 = (W_0 - W_{-1})/W_{-1}$. But notice that this would leave P_1/W_0 unchanged at $P_1/W_0 = P_0/W_{-1}$ and would also leave L unchanged at its period-0 value of L_0. From equation (13.7), this means that in period 1, the nominal wage would rise once again by $(W_1 - W_0)/W_0 = (W_0 - W_{-1})/W_{-1} = \pi_1$ percent, which would once again shift the AS curve vertically upward by π_1 percent, requiring an additional π_1 percent vertical shift in the AD curve to keep the level of output unchanged. If the economy is to stay at Y_0, this process would need to be repeated period after period, resulting in ongoing inflation of π_1 percent per period. Because this ongoing inflation rate is higher the higher the level of output produced by the initial shock, the relationship between the ongoing inflation rate and the equilibrium level of output is given by

$$\pi = \pi(\underset{+}{Y}) \qquad (13.8)$$

This relationship is referred to as the economy's *short-run Phillips curve.*

Does this mean that the economy can permanently move to a higher level of output if only it is willing to live with a permanently higher inflation rate? The answer is no. The reason it cannot do so is that our measure of nominal wage flexibility is unlikely to remain unchanged in the face of ongoing inflation. As the rate of inflation increases, workers are likely to demand more frequent revisions of their nominal wage contracts to avoid inflationary erosion of their real wages.

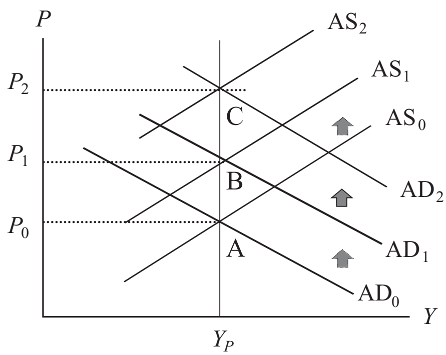


Figure 13.6. Ongoing steady inflation

As they do so, the wage adjustment parameter σ would tend to increase. As we saw in Chapter 4, this would tend to make the aggregate supply curve steeper. If the AS curve were to become steeper as it shifted upward, the rate of inflation associated with any given value of $Y > Y_P$ would increase, which would in turn lead to larger nominal wage increases and therefore to larger vertical shifts in the AS curve. The result would be *accelerating* inflation if the economy were kept at a level of Y that was greater than Y_P.

How is an ongoing steady rate of inflation possible in this model, then? The answer is that a steady rate of inflation is possible only at $Y = Y_P$. To see how, suppose that the economy's equilibrium in period 0 at the point A in Figure 13.6 is arrived at after an increase of x percent in the nominal wage relative to the previous period. Because the nominal wage in period 0 is x percent higher than the lagged nominal wage, in period 1, the lagged nominal wage will be x percent higher than it was in period 0. That means that in period 1, the aggregate supply curve, labeled AS_1 in the figure, must shift upward by x percent so as to pass, say, through the point B. If the central bank depreciates the currency and increases the money supply by x percent in period 1, period 1's aggregate demand curve will shift upward by x percent, say, from AD_0 to AD_1, therefore intersecting AS_1 at the point B. In that case, the price level must rise by x percent in period 1 relative to period 0. But because both the lagged nominal wage and the price level have risen by x percent in period 1, the labor demand and supply curves must both shift up by x percent in period 1, so the labor market would remain in equilibrium with an employment level of L_P but with a new increase in the nominal wage in period 1 relative to period 0 of x percent. Period 2 thus inherits a lagged nominal wage that is x percent higher than it was in period 1, and the whole process starts over again. Notice that this process is not affected by the slope of the aggregate supply curve.

Suppose, then, that an economy is in an equilibrium such as that depicted in Figure 13.6, with a high ongoing rate of inflation of x percent, and that the government decides to bring the rate of inflation down by reducing the rates of money growth and currency depreciation, and that it supports this plan by implementing a tighter fiscal policy. What happens to real output during the process of stabilization?

Note first that in the period that the tighter fiscal policy is implemented, the rate of vertical displacement of the AD curve will slow down, both because the rate of depreciation and money supply expansion will decrease permanently and because of the one-time change in the level of government spending or taxation associated with the adjustment of the public sector's primary surplus. In subsequent periods, the curve will shift upward at the new lower rate of depreciation and monetary expansion.

What will happen to Y during this process? The answer given by our model is that it depends on the degree of wage inertia in the economy. Consider two options. With complete wage flexibility ($\sigma = 1$), if the stabilization is announced before it is implemented, and if this announcement is credible to the workers whose behavior is reflected in the economy's aggregate supply curve, workers will anticipate the slowdown in the rate of upward displacement of the AD curve and will correctly infer, using the model, what this implies for the behavior of the aggregate price level. Thus the rate of upward displacement of the AS curve will slow down to match that of AD, and the economy's rate of price increase will immediately slow, with Y remaining equal to Y_P all the while. In terms of the stabilization paths illustrated in Figure 13.3, the point B (or C) moves to coincide with point A. Capacity growth accelerates after the rate of inflation falls, and because Y never deviates from Y_P, the rate of growth of actual real GDP accelerates immediately as well. These are no transition problems, and stabilization is unambiguously a good thing.

Suppose, on the other hand, that the nominal wage is sticky ($\sigma < 1$). As we have seen, ongoing inflation at a rate of, say, x percent would imply a continuous upward adjustment in the aggregate demand curve by x percent each period, due to money growth and exchange rate devaluation of x percent. In addition, the aggregate supply curve would shift up by x percent each period as well, as lagged nominal wages rise and workers continuously adjust their price-level expectations. Now suppose that the aggregate demand curve, which, in the absence of a policy change, would have shifted, say, from AD_0 to AD_1, is suddenly slowed down through a fiscal adjustment coupled with a reduced rate of monetary growth and exchange rate devaluation. What would happen to the economy in this case?

The answer depends in part on how sudden the slowdown in aggregate demand is. One extreme option is to suddenly slow the pace of monetary growth all the way to a new value that is compatible with low inflation. This is referred to as a *cold turkey* approach. Alternatively, the slowdown could be more gradual, with the rate of monetary expansion only eventually reaching its desired target, a technique called *gradualism*. Taking an intermediate case, suppose that as a result of the change in policies, the vertical displacement of AD is only to AD_1' in Figure 13.7, rather than to AD_1, which would have sustained the ongoing inflation rate. Notice that an upward shift in the aggregate supply curve in period 1 is unavoidable because

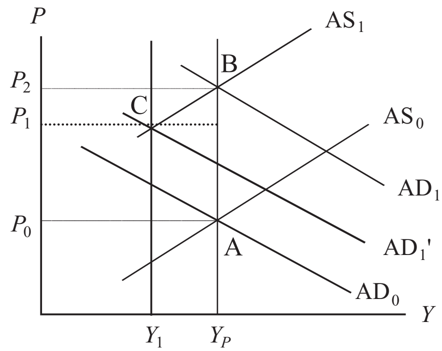


Figure 13.7. Stabilization dynamics with sticky nominal wages

it is determined by what happened to the nominal wage in the period *before* the stabilization was implemented, that is, in period 0. This means that the AS curve will continue to shift upward at the same rate as it would have if the stabilization had not taken place, say, from AS_0 to AS_1. In that case, the economy will move to a new equilibrium at C, with a substantial gap between Y and Y_P, rather than to B, where it would have been without the stabilization. The conclusion is that because of the recession created by the stabilization, the growth rate of *capacity* output may eventually accelerate, but that of *actual* output will fall short of it for some time, perhaps following a trajectory such as 1 or 2 in Figure 13.3. Notice that the contraction in real output is smaller – and the slowdown in the rate of price increase is more pronounced – the steeper the aggregate supply curve. Recalling that the aggregate supply curve is steeper when nominal wages are more flexible, this implies that stabilization is likely to be less costly when nominal wages are more flexible. If nominal wages are likely to become more flexible when the inflation rate is high, as suggested earlier, the implication would be that stabilization from high inflation may be less costly than from low inflation.

All this assumes that the stabilization was anticipated by workers and that the stabilization is *credible*, that is, workers assumed that it would actually be implemented as announced. What happens if the stabilization is *not* credible so that workers expect the price level to continue to behave as it has been, that is, to continue to increase by x percent per period? To answer this question, recall from Chapter 4 (equation (4.8a)) that the supply wage is given by

$$W = (1 - \sigma)W_{-1} + \sigma P^e w(L^S, \ldots)$$

and that one step in the derivation of the aggregate supply curve was the assumption of rational expectations, so $P^e = P$. To capture the lack of credibility of the stabilization, suppose instead that P^e is exogenous, so it is not affected by the announcement of an impending stabilization. In that case, the labor supply curve would not be a function of the current price level. Under those conditions, a change in the current price level would shift the labor demand curve in the same direction and by the same proportion as the change in the price level, but in contrast with the analysis of Chapter 4, it would leave the labor supply curve unchanged rather than shift it vertically less than in proportion to the change in the price level. The implication of that situation is that a change in the current price level would have a *larger* effect on employment than it does under rational expectations, making the AS curve *flatter*. Because labor-market equilibrium would in this case be given by

$$P\,MP_L(L) = (1-\sigma)W_{-1} + \sigma P^e\,w(L,\ldots)$$

you can see that changes in W_{-1}, P^e, and P in the same proportion would leave the equilibrium level of employment unchanged. Therefore, if the lagged nominal wage increased by the ongoing x percent rate of inflation, and if workers continue to expect the price level to increase by x percent despite the announced stabilization, an increase in P by x percent would leave employment unchanged. This means that the aggregate supply curve would shift upward by x percent in this case, just as it did under rational expectations. But because the curve is flatter, the new post-stabilization equilibrium would feature a *larger* drop in Y and a *smaller* slowdown in the rate of price increase. The upshot is that in our model, the cost in terms of foregone output of an inflation stabilization depends on two factors: the degree of wage inertia and the credibility of the stabilization effort. The greater the degree of wage inertia and the lower the credibility of the stabilization policies, the greater the *sacrifice ratio*: the ratio of cumulative lost percentage points of output to the percentage-point reduction in inflation.

3. Stabilization Strategies

We have seen that stabilization entails two necessary components: a reduction in the rate of monetary expansion and exchange rate devaluation and a fiscal adjustment to adapt the public sector's budget to a lower level of seignorage revenues. These components affect the behavior of the economy's aggregate demand curve. A *stabilization strategy*, however, may supplement these components with others intended to reduce the cost of making the transition to low inflation.

Suppose, for example, that workers form their expectations rationally but are unsure about the government's commitment to stabilization, so they cannot perfectly anticipate the behavior of the AD curve. How can a government that truly intends to stabilize – that is, that intends to carry out the necessary monetary and fiscal components of stabilization – influence the behavior of the AS curve to try to avoid recession? By analogy with mechanisms for establishing fiscal credibility discussed in Chapter 9, one approach would be for the government to try to enhance the credibility of its policy intentions by taking supplementary actions that would only make sense if it truly intended to stabilize.

For example, the government could announce a fixed exchange rate or a much slower rate of devaluation for the domestic currency. We have seen that unless domestic credit expansion is complemented with exchange rate devaluation, the central bank's reserves will tend to be depleted, which will eventually make the exchange-rate policy unsustainable. This strategy is referred to as *exchange rate–based* stabilization. When no such commitment is made, the strategy is said to be *money based* because full reliance is placed by the government on the reduction in money growth and fiscal adjustment.

In addition to making special announcements about exchange rate policy, the government may also intervene in the wage-price process directly by supplementing its monetary, fiscal, and exchange rate measures with wage and price controls intended to directly influence the position of the AS curve. Stabilization programs that feature direct intervention with the wage-price process are called *heterodox*, whereas those that do not are called *orthodox*.

To summarize, our model suggests that a reduction in the rate of monetary growth and exchange rate devaluation is a necessary component of inflation stabilization. We have also seen that this will need to be accompanied by a fiscal adjustment that can happen either contemporaneously with the slowdown in money growth or not, though the latter will be feasible only if the government can credibly promise its creditors to make a larger fiscal adjustment in the future. These adjustments can be either sudden (cold turkey) or gradual, but the important thing is that the rate of displacement of the aggregate demand curve must converge to the desired lower rate of inflation. The real effects of this strategy depend on the degree of wage inertia and on how workers' expectations respond, and the government can try to influence these factors – or the behavior of aggregate supply more directly – through supplementary policies such as exchange rate announcements or the implementation of wage and price controls.

When all is said and done, then, whether it is worthwhile to stop inflation depends on the following:

1. how costly the inflation was originally (i.e., how flat and low the initial path of Y was relative to its post-stabilization path)
2. how successful the policies mentioned are in avoiding a long and deep recession during the transition to low inflation – this determines the sacrifice ratio associated with the stabilization
3. the society's *rate of time preference*, which determines the relative value it places on current income compared to future income – this is relevant in the event that stabilization implies a recession during the transition because the social costs of the recession come now, while the benefits of stabilization come later, implying that trade-offs have to be made between current income losses and future income gains

But has stabilization indeed been costly? We now review some international experience with stabilization from high inflation to examine how successful countries have been in avoiding costs of transition.

IV. INFLATION STABILIZATION: EVIDENCE

Theory thus tells us that it may or may not be a good idea to stabilize, depending on the economy's short-run supply response. What does the evidence say about short-run costs of stabilization? The research reviewed in Section II provides at least suggestive evidence that sustained low inflation is conducive to higher economic growth. However, this evidence pertains to the relationship between *sustained* low inflation and *long-run* growth rates, or growth of productive capacity. As we have just seen, this evidence is compatible with a situation in which the *transition* from high to low inflation is associated with a contemporaneous *deceleration* in economic growth. The reason is that short-run adjustment mechanisms may cause the favorable growth effects from disinflation to materialize only with a lag so that growth may actually slow during the transition, and perhaps for some time thereafter, before achieving the higher levels suggested by the cross-country evidence described earlier. What, then, does the evidence say about the growth effects of the transition from high to low inflation in emerging economies?

In this section, we will look at both cross-country and case-study evidence. The first of these will give us an idea what the international experience has been regarding the short-run costs of inflation stabilization. The second will enable us to interpret this experience by looking at *stabilization strategies*, that is, alternative ways to bring about stabilization.

1. Cross-Country Evidence

The effects of inflation stabilization on growth were investigated in a well-known paper by Bruno and Easterly (1995). Bruno and Easterly were interested in looking at the growth effects of temporary inflation crises and their stabilization aftermaths. Their procedure was to collect a sample of macro data on a large group of industrial and developing countries from the 1950s to the 1990s. They defined an *inflation crisis* as occurring in their sample whenever a country had two or more consecutive years of annual inflation greater than 40 percent, ending when inflation fell below 40 percent and stayed there for at least two consecutive years. They then compared the behavior of a number of macro indicators relative to the contemporaneous world average before, during, and after such inflation crises.

What did they find? First, they found that there was a strong correlation between inflation crises and fiscal deficits. Comparing crisis countries to the contemporaneous world average, they found that fiscal deficits were higher in the crisis countries before the crisis, were even larger during the crisis, and were lower after the crisis. This pattern was matched by seignorage revenue, which skyrocketed in the crisis countries during the inflation crisis episode. Their conclusion from this part of the study was that the international evidence is consistent with the view that high inflation is largely a *fiscal* phenomenon, as argued in this chapter.

Second, with regard to growth, they found that it fell by an average of 2.8 percentage points *during* the crisis (i.e., up to time T_0 in Figure 13.3) but *rose* by an average of 3.8 percent after the crisis. This pattern was repeated for measured growth in total factor productivity because the ratio of investment to GDP tended to be below average during the crisis and did not rise above the world average subsequently, suggesting a temporary acceleration of growth of actual production above that of productive capacity. Their conclusion is that the sacrifice ratio may actually be *positive* when inflation is high (greater than 40% on an annual basis).[6] In terms of Figure 13.3, therefore, their evidence is consistent with a typical path of real output during stabilization such as that labeled with the number 3. It should be stressed, however, that the Bruno-Easterly evidence applies to high inflation rates (in excess of 40% annually) and cannot be extended to inflation rates below this value. Notice that this is consistent with the evidence in Barro (1996) concerning the effects of high inflation on long-run growth. Thus the evidence suggests that high inflation is bad for long-run capacity growth and, consistent with our model, that there may be no short-run output costs from stabilizing when inflation is very high.

2. Case Study Evidence: Stabilization Programs

We now turn to case studies of stabilization. These are useful because they permit us to consider stabilization *strategies*, which we defined earlier as the full policy package that gets implemented as part of a complete *stabilization program*. Fiscal adjustment is only part of such a program, the role of which is to stop (or at least slow down) the continuous upward shift of the aggregate demand curve. There are at least two other components, the first of which may affect both the aggregate demand and supply curves and the second of which is directed specifically at the aggregate supply curve:

1. **The choice of a *nominal anchor* during the process of stabilization (money or exchange rate).** As we saw in Chapter 8, all economies must have a nominal anchor to pin down the price level. In general, the money supply, the exchange rate, or some other nominal variable must play this role. Thus, in implementing a stabilization program, the authorities must also make some decision concerning which nominal anchor to adopt during and after the stabilization process.
2. **The *orthodox-heterodox* component of stabilization.** This refers to the issue of whether wage and price controls are used temporarily to support the nominal anchor. The roles of wage inertia, expectations, and credibility in determining the short-run output costs of disinflation analyzed in the previous section suggest why it might be desirable to do so – essentially direct interference with

[6] This result is supported by the results of Fischer et al. (2002).

the wage-price process could be used to influence the position and the shape of the short-run aggregate supply curve during the stabilization period.

Beyond these two key characteristics, the outcomes of individual stabilization programs may also depend on other accompanying policies that may be undertaken at the same time. The most important of these include the state of capital account liberalization and the state of financial reform, to be discussed in more detail in Part 6.

A very useful overview of individual country experiences in stabilizing from high inflation was provided by Vegh (1992). Vegh focused specifically on the distinction between *exchange rate–based* stabilization (in which a predetermined exchange rate is part of the stabilization package, implying the use of the exchange rate as a nominal anchor) and *money-based* stabilization (in which the exchange rate is allowed to float and the domestic money supply is used as the economy's nominal anchor). As in Bruno and Easterly (1995), Vegh also considered threshold effects, in his case distinguishing between countries with *hyperinflation*, defined in the classic Cagan sense (beginning when monthly inflation exceeds 50% for the first time and ending when it falls below 50% and stays below for at least a year), and those with only *chronic* inflation, defined in the Pazos sense (as inflation that is intermediate in intensity and lasts for a long time). To summarize Vegh's findings, we will consider these separately.

a. Hyperinflation

The cases of hyperinflation considered by Vegh included those of central and eastern Europe in the post–World War I period (Austria from October 1921 to August 1922, Germany from August 1922 to November 1923, Hungary from March 1923 to February 1924, Poland from January 1923 to January 1924, and Russia from December 1921 to January 1924), Europe in the post–World War II period (Hungary from August 1945 to July 1946, Greece from November 1943 to December 1945), Taiwan (1945–1949), and more recently, Bolivia (from April 1984 to September 1985). The stylized facts from these stabilization experiences included the following:

- Stabilization in these episodes tended to share common policy elements. The key elements were drastic fiscal reform, the implementation of central-bank independence (which we will discuss in Chapter 14), and the restoration of *currency convertibility* (the removal of foreign exchange controls).
- In each of these cases, inflation stopped immediately. In other words, there was no long, drawn-out period of disinflation. The explanation for this offered by Vegh is that in hyperinflation, immediate price stability can be achieved by using the exchange rate as the nominal anchor because all domestic prices tend to get quoted in foreign exchange. In terms of our model, this can be interpreted as

meaning that W_{-1} in effect gets indexed to SP^* so that the aggregate supply curve shifts up each period by the rate of devaluation.
- The short-run output costs associated with stabilization were small in each of these cases. Output actually *rose* during stabilization in three episodes (Russia, Greece, and Germany) and was unchanged in two others (Taiwan and Bolivia). There was conflicting evidence regarding the short-run behavior of real output in Austria and Hungary in 1946. Real output apparently fell somewhat only in both Poland and Hungary in 1924.

The lessons drawn by Vegh based on the case-study experience in stabilizing from hyperinflation were that credible fiscal reform is needed for success and that in the case of hyperinflation, inflation can be stopped rapidly and without substantial output costs. The credibility of fiscal reform can be enhanced by implementing central-bank independence and currency convertibility. The former eliminates recourse to the inflation tax, and the latter signals the government's intention to pursue prudent aggregate demand policies in the future. Relatively costless stabilization is facilitated in the case of hyperinflation through the role of the exchange rate as a nominal anchor when all prices become indexed to the exchange rate. To make the fixed exchange rate credible during stabilization, the authorities in each case had recourse to large foreign exchange reserves and/or an external line of credit.

b. Chronic Inflation

Vegh classified the chronic inflation cases in his study into three categories. These included the Latin American heterodox programs of the 1960s (Argentina in March 1967, Brazil in March 1964, and Uruguay in June 1968), the Southern Cone episodes of the late 1970s (Argentina in December 1978, Chile in February 1978, and Uruguay in October 1978), and the heterodox programs of the mid-1980s (Argentina's "Austral" program of June 1985, Brazil's "Cruzado" program of February 1986, the Israeli program of July 1985, and the Mexican "Pacto" program of December 1987).

The early Latin American heterodox programs were populist in nature in that they did not feature strong and sustained fiscal adjustments. These programs were characterized by fixed exchange rates and the use of income policies. All achieved an initial decline in inflation, but only Brazil sustained it into the 1970s.

The Southern Cone programs of the late 1970s, by contrast, were orthodox in the sense that no wage or price controls were used, though a centerpiece of each of these programs was the use of a preannounced nominal exchange-rate path, known as a *tablita.* Fiscal balance was achieved in Chile and Uruguay but not in Argentina. All of the Southern Cone programs ended in crises featuring an appreciated real exchange rate triggering large capital outflows as well as a collapse of the financial sector.

These experiences will turn out to be quite important for other issues that we will be discussing later in this book, and we will therefore have cause to revisit the Southern Cone episodes. For now, however, we can just summarize Vegh's assessment of where they went wrong. In his view, there were two key mistakes made in the Southern Cone stabilizations. The first consisted of leaving backward wage indexation in place and the second of combining stabilization with opening the capital account and liberalizing the domestic financial sector. The first of these created inertia in the wage-price process, which caused the real exchange rate to appreciate, whereas the second permitted capital first to flow into an under-regulated financial sector and later to rapidly flow out, resulting in a financial crisis.

Finally, the heterodox programs of the mid-1980s, like those of the 1960s, featured both fixed exchange rates and wage-price controls. The key to success here was the role of fiscal policy. Little fiscal adjustment was achieved in Brazil, and fiscal adjustment proved to be temporary in Argentina. Only in Israel and Mexico, where large and sustained fiscal adjustments were undertaken, did stabilization prove to be successful.

Vegh summarizes experience on all three of the "chronic inflation" stabilization periods in the following way. First, in contrast to the hyperinflation cases, inflation converged slowly. Given the role of the exchange rate as a nominal anchor, a large and sustained real appreciation emerged in all these cases, which contributed to a deterioration of the current account of the balance of payments. Economic growth accelerated initially in all of them and then contracted.

The key lesson drawn by Vegh is that fixing the exchange rate under chronic inflation may indeed reduce the short-run output cost of stabilization and speed the rate of convergence of inflation to its desired value, but successful stabilization under these conditions requires low capital mobility, a stable domestic financial system, and early "flexibilization" of the exchange rate.

V. SUMMARY

This chapter has covered a substantial amount of ground. We saw that there are both analytical and empirical reasons to believe that high inflation is inimical to sustained rapid growth of productive capacity in emerging economies. This does not, however, necessarily imply that the reduction of high inflation is always and everywhere justified. The reason is that the transition from high to low inflation – that is, the process of inflation stabilization – may itself involve high costs in the form of a period during which the growth rate of actual output falls below that of productive capacity. If the growth benefits of lower inflation are small, if the transition involves a prolonged period of slow growth, and if the society weighs current income very heavily relative to future income, then stabilization may not be worthwhile. However, both the cross-country and case-study evidence suggests that, at least when inflation is very high (say, in excess of 40% per year), there may be few – if any – short-run costs associated with the process of bringing inflation down.

A credible and sustained fiscal adjustment is a necessary requirement for doing so. The role of other policies is less clear-cut. Enhanced central-bank independence may increase the credibility of the fiscal adjustment, for reasons that we will explore in more detail in the next chapter.

On the other hand, the effectiveness of the use of the exchange rate as a nominal anchor may depend on circumstances. A fixed exchange rate coupled with currency convertibility (which requires the availability of resources to defend the rate, either in the form of reserves or access to external credit) may enhance the effectiveness of stabilization when inflation is extremely high (as in hyperinflation) because domestic prices tend to be indexed to the exchange rate, and there is little inertia in the domestic wage-price process. Under chronic moderately high inflation, fixing the exchange rate may need to be supplemented with more heterodox elements such as wage and price controls if the stabilization program lacks credibility or if there is institutional inertia in the wage-price process. When the exchange rate is used as a nominal anchor and such inertia is allowed to persist, combining stabilization with domestic and external financial liberalization may provide the recipe for the failure of stabilization and a financial crash, as in the Southern Cone countries.

Thus, while stabilization from high inflation may be desirable in both the short run and the long run, it is certainly possible to do it counterproductively. A durable fiscal adjustment is a necessary condition, but exchange rate and financial-sector policies also matter. As we will see in Chapter 28 these lessons about the roles of exchange-rate and financial-sector management are of broader applicability. They apply to the attainment and preservation of macroeconomic stability in emerging economies more generally rather than just to the issue of inflation stabilization. We will turn to them in Part 7.

REVIEW QUESTIONS

1. Describe five theoretical channels through which high inflation may affect the growth rate of an economy's productive capacity.
2. What are "instrumental variables," and why are they relevant to the study of the effects of high inflation on growth?
3. Why is it not possible for an economy to sustain a level of real output above full employment as long as it is willing to live with a high, but steady, rate of inflation?
4. Using the AD-AS diagram, explain how steady inflation can be sustained in our model.
5. What is the "sacrifice ratio"? What are the roles of nominal wage inertia and stabilization policy credibility in determining the sacrifice ratio?

EXERCISES

1. Suppose that there is no nominal wage inertia ($\sigma = 1$) and that workers' price-level expectations are given by $P^e = P_{-1}(1 + \pi_{-1})$, where P_{-1} and π_{-1} are

the most recently observed price level and inflation rates, respectively. Show diagrammatically how a cold turkey inflation stabilization program (one that freezes the AD curve in place) would affect the output cost of inflation stabilization.

2. How would your answer to exercise 1 change if the economy's labor supply curve were given by

$$W = P\omega$$

where ω is a constant targeted real wage?

3. Once again, suppose that workers' price-level expectations are given by $P^e = P_{-1}(1 + \pi_{-1})$ and that a cold turkey inflation stabilization program is implemented by holding both the nominal exchange rate and the money supply constant at the value they attained in the most recent period. Show what effects the stabilization would have on the real exchange rate, net exports, and capital flows.

4. Suppose that the rate of adjustment of nominal wages σ depends positively on the most recently observed inflation rate, so $\sigma = \sigma(\pi_{-1})$. Consider an economy that departs from an initial equilibrium at full employment and with zero inflation, and suppose that the central bank decides to use monetary and exchange rate policies to increase the level of real output to $Y_1 > Y_P$. Using the AD-AS diagram of this chapter, show what would happen to the inflation rate in this economy over time.

REFERENCES AND FURTHER READING

Barro, Robert J. (1994), "Economic Growth and Convergence," occasional paper 46, International Center for Economic Growth.

Barro, Robert J. (1996), "Inflation and Growth," *Federal Reserve Bank of St. Louis Review*, 78, pp. 153–169.

Bruno, Michael, and William Easterly (1995), "Inflation Crises and Long-Run and Growth," working paper 5209, National Bureau of Economic Research.

Fischer, Stanley (1993), "Does Macroeconomic Policy Matter? Evidence from Developing Countries," occasional paper 27, International Center for Economic Growth.

Fischer, Stanley, Ratna Sahay, and Carlos A. Vegh (2002), "Modern Hyper- and High Inflations," *Journal of Economic Literature*, vol. 40, pp. 837–880.

Khan, Mohsin S., and Abdelhak S. Senhadji (2001), "Threshold Effects in the Relationship between Inflation and Growth," *IMF Staff Papers*, 41, pp. 1–21.

Reinhart, Carmen, and Kenneth Rogoff (2008), "This Time Is Different: A Panoramic View of Eight Centuries of Financial Crises," working paper 13882, National Bureau of Economic Research.

Reinhart, Carmen, Kenneth Rogoff, and Miguel Savastano (2003), "Debt Intolerance," *Brookings Papers on Economic Activity*, vol. 1, pp. 1–62.

Vegh, Carlos (1992), "Stopping High Inflation: An Analytical Overview," *International Monetary Fund Staff Papers*, vol. 39, pp. 626–695.

PART 4

Monetary Institutions and Monetary Policy

14

Monetary Institutions

Up to this point in the book, the central bank has played a very simple role. In the short run, it has conducted exchange rate policy by intervening in the foreign exchange market in a very specific way, that is, by standing ready to buy or sell foreign exchange at the officially announced exchange rate, thereby making the market for foreign exchange. It has also conducted monetary policy by intervening in the domestic bond market with one of three objectives: to target the domestic interest rate, the stock of domestic credit (the quantity of domestic government bonds that it holds), or the money supply, thereby making one of these a policy-determined variable. Over the medium run, we described the central bank as expanding the stock of domestic credit to meet the government's financing needs and adjusting the officially determined nominal exchange rate proportionately to safeguard its stock of foreign exchange reserves.

In all this, the central bank has been treated as acting rather mechanically and passively. It is now time to examine the role of the central bank more closely. In Part 4 of the book, consisting of this chapter and the next, we will consider the institutional and policy frameworks for the conduct of monetary policy. The current chapter will consider the case for central-bank independence from the finance ministry, and in the chapter that follows, we will examine how an independent central bank should conduct monetary policy, that is, how the *monetary policy regime* should be chosen. The focus there will be on *inflation targeting*, a monetary policy regime that has recently been widely adopted among emerging and developing economies.

To set the stage for these issues, this chapter begins with a section considering the role of monetary policy in development, paralleling the similar discussion of the development role of fiscal policy in Chapter 11. We then consider in Section II one of the key challenges in designing an appropriate institutional framework for the central bank, the problem of *time inconsistency* in monetary policy. Section III examines the potential for central-bank independence to mitigate the problems caused by time inconsistency, while Section IV considers the potential role that reputational considerations could play in affecting the behavior of an independent central bank. Section V examines some evidence on the relationship between central-bank independence and inflation. Alternative institutional arrangements for the central bank are considered in Section VI, and Section VII concludes.

I. MONETARY POLICY AND DEVELOPMENT

As was the case with fiscal policy, monetary policy can affect development through both direct and indirect channels.

1. Direct (Microeconomic) Effects

The direct role of monetary policy in development is a microeconomic one: as we saw in Part 3, the creation of base money provides revenue to the government in the form of seignorage. Like other taxes, the inflation tax component of seignorage provides a source of revenue to finance potentially productive government expenditures. Also like other taxes, however, it creates distortions that interfere with the efficient allocation of resources.[1] The principle that an optimal tax structure should feature low tax rates and a wide tax base suggests that the inflation tax should be a component of an optimal tax structure but that excessive reliance on the inflation tax should be avoided. The analysis of the adverse growth consequences of high inflation in the preceding chapter supports this view.

2. Indirect (Macroeconomic) Effects

Monetary policy can play two potentially positive indirect roles in development, both of which operate by fostering macroeconomic stability:

- By avoiding high and unstable inflation, monetary policy can avoid becoming a source of macroeconomic instability.
- By being deployed in a countercyclical fashion, monetary policy can help stabilize the economy in the face of exogenous shocks arising from other sources.

These considerations suggest that a pro-development stance of monetary policy is one that provides the government a relatively minor, but dependable, source of revenue through seignorage; that maintains a low and stable long-run inflation rate; and that helps to stabilize the domestic economy in response to exogenous shocks.

[1] The inflation tax creates a distortion by discouraging the holding of money. Holding money is productive because its use as a medium of exchange enables economic agents to reduce the costs associated with making market transactions. Because the inflation tax induces economic agents to economize on their holding of money, it forces them to incur larger transactions costs than they otherwise would, and such costs absorb resources that could otherwise have been devoted to other types of productive activity.

Unfortunately, monetary policy as practiced in many developing countries and emerging economies often falls short of these principles. Specifically, it tends to display a *pro-inflation bias*, manifesting itself in the form of excessive reliance on seignorage to finance fiscal deficits as well as high and unstable inflation that interferes with the information content of relative prices and thus undermines economic growth. In addition, it also tends to be implemented in pro-cyclical fashion, in the sense that adverse shocks to the economy are often accompanied by a tightening, rather than a loosening, of monetary policy. As in the case of fiscal policy, this pro-cyclical behavior tends to emerge because of credibility problems: specifically, central banks are reluctant to implement expansionary monetary policy in recessions for fear of fueling inflationary expectations. An appropriate institutional framework for monetary policy, therefore, is one that provides short-run flexibility for the central bank to use monetary policy as a stabilization policy instrument without undermining its medium-term anti-inflationary credibility.

II. SOURCES OF PRO-INFLATION BIAS: TIME INCONSISTENCY

To explore the type of institutional framework that could help monetary policy operate in the fashion just described, our first task is to ask why it does not already do so; that is, we investigate the sources of pro-inflation bias in monetary policy. In contrast to fiscal policy, which is typically formulated through the interactions of various independent political actors in the cabinet and legislature, monetary policy decisions are typically made by a single agent: the central bank. Thus, though it is therefore natural to try to understand the pro-deficit bias in fiscal policy as the outcome of a political-economy equilibrium, in the case of monetary policy, the issue is what causes the decision of the central bank as a single decision-making agent to display a pro-inflation bias. In this section, we will build a simple model that explains how the rate of inflation chosen by the central bank is determined as well as why that decision may display a pro-inflation bias.[2]

Of course, a simple answer to this question is that the central bank's pro-inflation bias is just a consequence of the finance ministry's pro-deficit bias: if the central bank is subordinate to the finance ministry, and the finance ministry tends to run large fiscal deficits, then the ministry is likely to exert pressure on the central bank to finance its deficits by printing money to extend credit to the government, as we have seen before. In this case, the pro-inflation bias of the central bank is ultimately attributable to the political economy equilibrium that creates the pro-deficit bias in fiscal policy. As we will see, a central bank that is under the control of the finance ministry would indeed be likely to exhibit a pro-inflation bias. But it turns out that such a bias could arise even if the central bank is fully independent of the finance ministry and if it tries to behave to maximize social welfare.

[2] In general terms, the analysis follows Barro and Gordon (1983).

To see how, consider a central bank that is free to act independently to determine the position of the economy's shifting aggregate demand curve over time, as in Figure 13.4, and that thus determines the economy's ongoing rate of inflation π. What level of π would it choose? Suppose that the bank is indeed motivated by a desire to maximize social welfare, and that social welfare is affected by inflation through two distinct channels. First, recall from Chapter 4 that along the economy's short-run aggregate supply curve, unanticipated changes in the price level are associated with increases in the level of output Y relative to its full-employment level Y_P, even if wages are fully flexible (i.e., even if $\sigma = 1$). If there are distortions in the economy that make the full-employment level of output suboptimally low (e.g., taxes on labor income that would cause the supply price of labor to increase, shifting the labor supply curve upward), then an increase in the *actual* level of output above its full-employment level ($Y > Y_P$) would be socially beneficial. In that case, *unanticipated* inflation would have a positive effect on social welfare, so unanticipated inflation would have a positive coefficient in the social welfare function that the central bank is trying to maximize. Second, we saw in the preceding chapter that sustained high inflation can be expected to have a negative effect on the rate of growth of the economy's productive capacity, and that this negative effect is likely to be more pronounced the higher the rate of inflation. Because countries typically seek to increase the rate of growth of their productive capacity, inflation has a *negative* and nonlinear effect on social welfare through this channel. This negative effect may be partially offset by the seignorage revenue that high inflation provides to finance the fiscal deficit. Putting these two effects together, we can write the social welfare function that the central bank tries to maximize as

$$B(\pi) = \beta(\pi - \pi^e) - (\alpha/2)\pi^2 \tag{14.1}$$

where π is the actual rate of inflation, π^e is the expected rate of inflation, and $\pi - \pi^e$ is therefore the unanticipated rate of inflation. The positive parameters β and α reflect, respectively, the weights given to the excess of real output over its full-employment level and to the rate of inflation in determining social welfare, as perceived by the central bank.[3]

Suppose that the central bank makes an announcement about the level of inflation that it intends to achieve, based on which the public forms its inflationary expectations π^e. After the public has formed its expectations, the central bank sets the rate of inflation, taking the public's expectations as given. What rate of inflation should the central bank announce ex ante (i.e., before the public's expectations are formed)? It is easy to see that if the social welfare function is as in equation (14.1),

[3] Because α is just a constant, there is no harm done in writing it as $\alpha/2$ instead, and you will see later that it is convenient to do so.

the announced inflation rate should be zero because society cannot benefit (but can lose) by having the public expect a higher rate of inflation.[4]

If the bank announces the intention to achieve zero inflation and this is believed by the public ($\pi^e = 0$), then it faces two choices. First, it can treat its announcement as a *commitment* that binds its actions, and follow through on that commitment by indeed producing zero inflation, in which case, $\pi = 0$, $Y = Y_P$, and $B = 0$. Alternatively, the central bank can act in a *discretionary* fashion, reneging on this commitment and attempting to achieve the socially optimal rate of inflation (the one that maximizes B conditional on $\pi^e = 0$). To see what the socially optimal rate of inflation must be, notice that the marginal benefit of inflation is constant at the value β, whereas the nonlinearity of the social costs of inflation means that the marginal cost of inflation is higher the higher the rate of inflation. With the formulation (14.1), the marginal cost of inflation is actually given by $\alpha\pi$.[5] This means that when inflation is very low, the marginal benefit of a little more inflation will exceed its marginal cost, but as the rate of inflation rises, its marginal cost will also rise, while its marginal benefit remains the same. At a sufficiently high rate of inflation, the marginal benefit of inflation will be exactly outweighed by its marginal cost ($\beta = \alpha\pi$), and beyond this point, further increases in inflation would be socially harmful. Thus the rate of inflation that maximizes B, subject to $\pi^e = 0$, is given by $\pi = \beta/\alpha$.

Notice that this inflation rate is higher the more the bank cares about real output (the larger β) and the less it cares about the negative growth effects of inflation (the smaller α). The resulting benefit to society from this choice of the inflation rate is given by

$$B(\beta/\alpha) = \beta(\beta/\alpha) - (\alpha/2)(\beta/\alpha)^2 = \beta^2/\alpha \; - \; \beta^2/2\alpha = \beta^2/\alpha$$

which, being positive, is *larger* than that achieved by following the announced policy of zero inflation. The gain comes from the benefit of "surprise" inflation in raising output above its potential level. This benefit must always outweigh the social cost of inflation when inflation is zero because the quadratic form of the social cost of inflation implies that these costs are zero at the margin when inflation is zero. Thus the central bank always has an incentive to announce zero inflation but then to renege on its announcement and generate a rate of inflation that is larger the more it cares about real output and smaller the more it cares about the social costs of inflation. In other words, the central bank has an incentive to be *time inconsistent*: to say one thing at the beginning of the period and do another later on.

[4] Technically, subject to $\pi = \pi^e$, the social welfare function (14.1) is maximized when $\pi = \pi^e = 0$.

[5] To see this, write π as $\pi_0^2 + \Delta\pi$, where π_0 is the initial inflation rate and $\Delta\pi$ denotes deviations from that rate. Expanding the polynomial and ignoring second-order terms yields $\pi^2 = \pi_0^2 + 2\Delta\pi$, so the marginal cost of inflation is $(\alpha/2)2\pi_0 = \alpha\pi_0$.

The problem, of course, is that the public can be expected to be aware of this incentive and thus to disbelieve the bank's policy announcement. Indeed, if the public is aware of the principles governing the central bank's behavior, it will expect the central bank to produce not zero inflation but the inflation rate $\pi = \beta/\alpha$. How does this affect the bank's subsequent behavior?

The answer is that it does not affect it at all. The public's beliefs π^e affect neither the marginal benefit of inflation β nor its marginal cost $\alpha\pi$, so the central bank will be led to generate the same level of inflation regardless of what the public believes. This means that when the public sets $\pi^e = \beta/\alpha$, the central bank will still set $\pi = \beta/\alpha$, validating the public's beliefs. Notice that the social welfare outcome is very different in this case, however. Substituting $\pi^e = \pi = \beta/\alpha$ into the social welfare function, we get $B = -\beta^2/2\alpha$, which, being negative, is not only a worse outcome than when the central bank's announcement was believed by the public but is even worse than what would happen if the central bank had no discretion whatsoever and was forced to set the inflation rate equal to zero! The reason for this is that there is no inflation surprise in either case, but there is a cost arising from positive inflation when the central bank acts with discretion.

Why does the central bank not just choose a zero inflation rate, then? The answer is that if it has no means to convince the public that it will actually do so, the public will continue to set $\pi^e = \beta/\alpha$, which implies a social welfare outcome of $B = -\beta^2/\alpha$ if the central bank actually sets the inflation rate equal to zero, and this is an even *worse* outcome than with the inflation rate $\pi = \beta/\alpha$. Thus, in the absence of a *precommitment mechanism* that forces the bank to do what it says, and therefore makes the bank's policy announcement credible, the inflationary outcome will exhibit pro-inflation bias – that is, it will be $\pi = \beta/\alpha$ rather than $\pi = 0$.

III. AN INSTITUTIONAL FIX: CENTRAL-BANK INDEPENDENCE

What do we learn from this analysis? First, in a rational-expectations equilibrium, the optimal inflation rate if the central bank could precommit its future actions would be $\pi = 0$. When the central bank can exercise discretion, however, the actual inflation rate turns out to be $\pi = \beta/\alpha$. Thus the extent of pro-inflation bias under central-bank discretion is given by β/α. What determines the size of this bias?

Notice that the bias is larger the larger is β and the smaller is α. The magnitude of β depends on two things: the impact of 'surprise' inflation on real output and the social benefit assigned by the monetary policy decision maker to a given increment in real output. Regarding the first of these, the impact of surprise inflation on real output depends on the slope of the economy's short-run aggregate supply curve: the flatter the short-run aggregate supply curve the larger the effect of a given inflation surprise on real output. In turn, as we saw in Chapter 4, the slope of the economy's short-run aggregate supply curve depends on structural characteristics of the economy that determine the shapes of its labor demand and supply curves. The second determinant of β – the value placed by the monetary policy decision maker on an increment in real output – is more novel. This is likely to depend on who the monetary policy decision maker is. An elected official whose constituents demand increases in real output and employment is likely to place more value on higher output than a nonelected technocrat who does not benefit directly from that outcome. This suggests that the degree of pro-inflation bias is likely to be larger when monetary policy decisions are made by elected officials than when they are made by nonelected ones.

Turning to α, recall that high inflation reduces the rate of growth of productive capacity, an effect that will have more pronounced effects on the actual level of output the longer the time horizon over which it is felt. Thus monetary decision makers with longer time horizons are likely to assign a larger cost to high inflation and therefore will have a larger value of α. It follows that decision makers who are insulated from the political process and have longer time horizons are likely to exhibit less equilibrium pro-inflation bias because they would be expected to assign both a smaller value to β and a larger one to α.

Finally, to the extent that the monetary policy decision maker reflects the finance ministry's preferences – a situation that we previously defined as *fiscal dominance* – he or she may give much more weight to the positive role that inflation can play in financing the government's budget through the inflation tax. Indeed, we saw in Chapter 13 that high inflation usually results precisely because a situation of prospective fiscal insolvency induces the government to attempt to increase seignorage revenue. As discussed earlier, the fiscal benefit of seignorage revenue would tend to reduce the parameter α because the revenue gains would tend to offset the social costs from lower growth. As we have seen, lower α means a higher pro-inflation bias.

All this suggests that the degree of pro-inflation bias can be reduced if monetary policy decisions are put into the hands of an unelected agent with a long time horizon who does not reflect the finance ministry's preferences. This, in short, is the argument for monetary policy to be conducted by an independent central bank – a bank that is governed by individuals who serve for long terms and are insulated from political pressures; in other words, they are not accountable to the elected government. In the presence of the time-inconsistency problem, the theory we have just reviewed suggests that central-bank independence may indeed lower the rate of inflation by reducing the value of β/α.

But notice that the time-inconsistency problem and a pro-inflation bias would tend to arise whether or not the central bank is independent, as long as the central bank reflects the preferences of society, which truly *does* value both higher real output and low inflation. Central bank independence can only be expected to ameliorate the problem. To further reduce pro-inflation bias, central-bank independence can be buttressed with several additional mechanisms[6]:

- The appointment of a "conservative" central-bank governor. This refers to a central-bank governor who places a lower value on higher real output, and assigns a higher cost to inflation, than does society as a whole. In other words, this would be an individual with a social welfare function that has a lower value of β and a higher value of α than does society as a whole. Because the central bank's actions would be guided by the governor's preferences, rather than those of society as a whole, this would result in a lower level of inflation (by reducing the relevant value of β/α) than would be true if the governor's preferences accorded with those of society.
- Enact a legal mandate for the central bank to focus on inflation only. Whatever the governor's own preferences, society can create a legal mandate that determines the form of the social welfare function that the central bank is supposed to maximize. This mandate can be enshrined in the country's constitution or in the central bank's charter. To reduce pro-inflation bias, the central bank would be mandated to assign a small value to β and a large one to α. In the limit, the central bank can be mandated to focus on inflation *only*, that is, to set $\beta = 0$. In principle, such a mandate would completely eliminate pro-inflation bias. As shown in Box 14.1, such mandates are quite common among emerging and developing economies.
- Place legal limits on central-bank lending to the government.

Finally, even with legal independence, a conservative central-bank governor, and a mandate limited to seeking a stable price level, the central bank may face informal pressure from the government to provide it with seignorage revenue. The central bank's ability to resist such pressures can be reinforced if it is legally unable to lend more than a prescribed amount to the government. Such legal restrictions can again be specified in the constitution or in the central bank's charter. Because such restrictions are intended to allow the central bank's preferences to be divorced from those of the finance minister, they would operate formally by reducing the value of α.

IV. REPUTATION

The analysis of Section II should have left you a little bit uneasy. After all, even though a central bank exercising its policy discretion in a myopic (period-by-period) fashion may indeed *choose* to inflate each period at the rate β/α, there is nothing *compelling* it to do so. Knowing that behaving in this myopic way would

[6] For accessible discussions of institutional "fixes" to the pro-inflation bias problem, see Chang (1998) and Mishkin (2000).

Box 14.1. Price Stability Mandates for Central Banks

Table 14.1 contains excerpts from publicly available documents that describe the mandate and functions of six central banks in emerging and developing countries. A common characteristic across the different statements is that they both charge the central banks with the primary responsibility of protecting price stability and highlight the importance of price stability in achieving sustainable growth and economic progress.

Table 14.1. *Price Stability in Emerging and Developing Central-Bank Mandates*

	Institutional Mandate/Mission Statement
Chile	"[To] ensure the stability of the value of the national currency and the normal functioning of domestic and foreign payments. . . . The purpose of keeping inflation low and stable, which is how this price stability concept is interpreted in practical terms, is no mere whim of the law, but rather serves the broader objective of moving the national economy along the path of sustained growth, full employment and, in general, progress and wellbeing for the population."
Colombia	"According to constitutional and legal mandate, the essential responsibility of the Banco de la República is to safeguard the purchasing power of the national currency, therefore implying that inflation control is the primary objective of monetary policy. Accordingly, price stability is recognized as being necessary for economic progress."
Israel	"The main objective of monetary policy is to achieve price stability and financial stability. In achieving its goals, monetary policy helps to create a business environment that supports sustainable economic growth."
Peru	"The purpose of the Bank is to preserve monetary stability. The Central Bank of Peru acknowledges that achieving monetary stability by controlling inflation is the most important contribution it can make to sustained economic growth."
Philippines	"Commitment to promote and maintain price stability and provide proactive leadership in bringing about a strong financial system conducive to balanced and sustainable growth of the economy."
South Africa	"The primary objective of the Bank shall be to protect the value of the currency of the Republic in the interest of balanced and sustainable economic growth in the Republic."

Source: Chile, "About the Central Bank, Central Bank Policies," http://www.bcentral.cl/; Colombia, "The Banco de la República and the Constitution of 1991," http://www.banrep.gov.co/; Israel, "Regulating and Directing Monetary Policy," http://www.bankisrael.gov.il/; Peru, "What Is the Mission of the BCRP?," http://www.bcrp.gob.pe/; Philippines, Central Bank Charter Law, Article 1, Section 3; South Africa, South African Reserve Bank Act 90 of 1989.

cause the economy to incur a social loss equal to $-\beta^2/2\alpha$ every period, why does the central bank not just choose to set $\pi = 0$?

We previously saw that if the central bank did so, it would generate an even worse social outcome than if it acted with discretion. Table 14.2 summarizes the possibilities. It spells out the social welfare outcome associated with each of the central bank's

Table 14.2. *Social Welfare Payoff Matrix for the Central Bank*

		Public's Expectations (π^e)	
		0	β/α
Central Bank Behavior (π)	0	0	$-\beta^2/\alpha$
	β/α	$\beta^2/2\alpha$	$-\beta^2/2\alpha$

actions for given values of the public's inflation expectations. The top row of the table specifies alternative values for the public's inflation expectations, and the left-hand column lists the alternative rates of inflation chosen by the central bank. Each cell of the table describes the social welfare outcome given that the public expects the rate of inflation at the top of the column and the central bank actually generates the rate of inflation corresponding to the row. Notice that whatever the public's inflation expectations, it is always optimal for the central bank to produce inflation equal to β/α (i.e., the entries along the second row of the table are always larger than those in the first row corresponding to the same column). Thus producing a rate of inflation equal to β/α is always a *dominant strategy* for the central bank. Knowing this, the public will always expect inflation equal to β/α. Consequently, an inflation rate of β/α, associated with the social welfare loss of $-\beta^2/2\alpha$, is the Nash equilibrium for this policy game between the public and the central bank. The answer to the question of why the central bank would not just generate $\pi = 0$, no matter what the public expects, is easy to see from the table. By doing so, it would generate the social loss of $-\beta^2/\alpha$, which is larger (in absolute value) than the social loss of $-\beta^2/2\alpha$ associated with $\pi = \beta/\alpha$. The loss associated with lower-than-expected inflation more than offsets the gain from lower *actual* inflation.

But this analysis assumes that the public is well informed about the function that the central bank actually maximizes. Suppose that is not the case. In particular, suppose there are two possible types of central bank. Type I, an inflation hawk, seeks to maximize the function B given by equation (14.1), but with $\beta = 0$, whereas type II, an inflation "dove," maximizes the same function with $\beta > 0$. The central bank cannot change its type, but the public does not know what type of central bank it has. It must infer the central bank's type from the bank's behavior. How does it do so?

Suppose that the public and the central bank interact repeatedly over an infinite number of periods and that the central bank therefore maximizes a function that is the discounted sum of B over all future periods; that is, the central bank maximizes

$$V_0 = \sum_{t=0}^{\infty} (1+\rho)^{-t} B(\pi_t)$$

by choosing π each period, where ρ is the central bank's discount rate.

Notice that a type I central bank maximizing the function B given by equation (14.1), but with $\beta = 0$, would always set $\pi = 0$ each period, no matter what the public expects, because from a hawkish central bank's perspective, all that matters is inflation, and any inflation is worse than zero inflation. Thus, if the public ever were to observe $\pi = \beta/\alpha > 0$ when it was expecting $\pi = 0$, it would immediately know that the central bank is of type II and would henceforth always expect $\pi = \beta/\alpha$. Similarly, if the public expects $\pi = \beta/\alpha > 0$ and indeed observes it, it would also know that the central bank is of type II because a type I central bank would have continued to set $\pi = 0$. Again, the public would henceforth always expect $\pi = \beta/\alpha > 0$. What this means is that if the central bank ever actually produces $\pi = \beta/\alpha > 0$, from then on, it will always have to live with a public expectation of $\pi^e = \beta/\alpha > 0$.

Consider what this means about the incentives facing a type II central bank. If the public expects $\pi = 0$, a dovish central bank can engineer a surprise inflation of β/α in the current period, which provides a positive social benefit of $\beta/2\alpha$. But since the central bank thereby reveals its type to the public, from the next period on the public expects $\pi = \beta/\alpha$, and the central bank can do no better than to set $\pi = \beta/\alpha$ and incur a social welfare loss of $-\beta^2/2\alpha$ in all future periods. Unless the discount rate is very large, these permanent future welfare losses are likely to outweigh the transitory current welfare gain. This creates a strong incentive for the dovish central bank *not* to engineer an inflation surprise in the first period but instead to ratify the public's expectations by setting $\pi = 0$.

Alternatively, suppose that the public expects $\pi = \beta/\alpha > 0$. If the dovish central bank were to set $\pi = 0$ in this case, each period that it does so, it would incur a loss of $-\beta^2/\alpha$, a larger loss than would be implied by ratifying the public's expectations and setting $\pi = \beta/\alpha$. Would it ever be worthwhile to do this anyway? Suppose that after N periods of observing $\pi = 0$, the public revises its expectation to $\pi^e = 0$. In that case, the social welfare loss from the $N + 1$st period on becomes zero, compared to a permanent loss of $-\beta^2/2\alpha$ per period if the central bank had instead set $\pi = \beta/\alpha > 0$. Thus setting $\pi = 0$ initially involves incurring a larger loss through period N but a *smaller* one after that. Again, if the discount rate is not too large and the public does not wait too long to revise its expectations (i.e., if N is also not too large), the reduced permanent future losses could outweigh the larger transitory losses associated with a strategy of setting $\pi = 0$ even by a dovish central bank.

The upshot is that the desire to establish a reputation as an inflation hawk could cause even a type II central bank to set $\pi = 0$ optimally. But notice that unless the public initially believes (falsely) that it has a type I central bank, this strategy will incur short-run costs (up through period N) that are increasing in β/α. The point is that the institutional measures described earlier would still be beneficial even after reputational mechanisms for establishing central-bank credibility are taken into account. In this case, those measures improve social welfare by reducing the social cost of gaining anti-inflationary credibility.

We have seen, therefore, that theory can describe ways in which a specific institutional innovation – the creation of an independent central bank mandated to have a high degree of aversion to inflation – may indeed result in reducing pro-inflation bias or reducing the costs of gaining anti-inflationary credibility. But what has happened in practice under central-bank independence? To answer this question, we would want to determine whether central banks that have been more independent have indeed tended to produce lower inflation, other things equal. The next section will review some of the evidence that is available on this issue.

V. CENTRAL-BANK INDEPENDENCE AND INFLATION: EVIDENCE

As we have seen, in principle, we might expect a relationship to exist between central-bank independence and inflation performance because independent central banks can be endowed with a built-in bias against inflation. Legal independence strengthens their hand in pursuing that objective and can be used to strengthen the objective through legal means. Is there empirical evidence that independence matters?

The first requirement in testing the proposition that central-bank independence tends to lower the rate of inflation is to devise a measure of central-bank independence. How do we do this? Many countries have granted the central bank legal independence over recent years. One possibility would be to simply record when the legal status of the central bank changed in some sample of countries and then test whether the behavior of the inflation rate changed in these countries after that date, controlling for other observable determinants of inflation.

But unfortunately, simply giving the central bank legal independence (i.e., ensuring that the governor of the central bank does not report to the finance minister) is not enough to make it effectively independent because the bank may remain influenced by the government. In other words, what is required is a measure of de facto, rather than de jure, independence.

A classic early study of the effects of central-bank independence on inflation by Cukierman et al. (1992; hereinafter referred to as CWN) provides a useful illustration of how these complications can be addressed in empirical research. CWN proposed some alternative measures of de facto central-bank independence in a large group of countries and tested whether these have been associated with lower inflation. They used both formal and informal indicators of independence, noting that laws tend to be incomplete and that actual practice often deviates from what is specified in them (e.g., they noted that in Argentina, the legal term of the governor is four years, but over their sample period, the average tenure was one year). They used these indicators to construct measures of central-bank independence for each of four decades – 1950–1959, 1960–1969, 1970–1979, and 1980–1989 – for 72 countries, of which 51 were emerging or developing countries. They then examined whether these indicators were correlated with inflation performance in panel regressions, both for the full sample and for industrial and developing countries separately.

The evidence in CWN was based on regressions of transformed decade inflation on an intercept, several variables meant to capture de jure central-bank independence, such as the length of tenure of the governor, whether the government had a voice in monetary policy formulation, and five variables capturing objective limitations on central-bank lending to the government, a variable measuring the frequency of turnover of the central-bank governor to capture de facto independence, and three decade dummies. For the whole sample, none of the separate legal variables was statistically significant. Thus the legal variables were aggregated with subjective weights. The more parsimonious regression thus contained variables measuring legal independence and turnover of the governor as well as decade dummies. The results were as follows:

- For industrial countries, the aggregated legal variable had a significantly negative correlation with average decade inflation, but turnover, which was low for these countries, did not. The results were driven by the lending-limit component of the index because the others did not enter the regression significantly when entered simultaneously.
- For emerging and developing countries, the turnover variable entered with a statistically significant positive sign, whereas the legal variable did not. The interpretation offered by CWN was that legal norms have been adhered to less strictly in developing countries.

How reliable are these results? Notice that CWN did not control for variables, other than central-bank independence, that may affect the rate of inflation such as openness, the incidence of various types of economic shocks, the nature of wage-bargaining institutions, or the efficiency of the tax system.[7] They did, however, address the issue of reverse causality.[8] To address this problem, they used the previous decade's inflation and turnover ratios, legal independence, the legal term of office, and the decade dummies as instruments for the current turnover ratio (legal independence was taken as exogenous because it changed infrequently). The

[7] As noted in a subsequent broader survey of the evidence by Kißner and Wagner (1998), studies that have included other potential inflation determinants have sometimes come up with conflicting evidence about the effects of central-bank independence on inflation (see, e.g., Fuhrer, 1997). A recent survey by Cukierman (2006), however, concludes that the bulk of the evidence supports the CWN conclusions. It is fair to say that though a consensus has emerged on the beneficial effects of central-bank independence for price stability, the empirical evidence remains more suggestive than conclusive.

[8] Granger causality tests on inflation and turnover (with a one-decade lag) suggested bidirectional causality between the two variables.

turnover variable remained positive and statistically significant under this approach, both for the sample as a whole and for the subsample of emerging and developing countries.

VI. OTHER INSTITUTIONAL OPTIONS FOR MONETARY POLICY

Central-bank independence, and the various mechanisms for buttressing it discussed in Sections III and IV, are intended to reduce pro-inflation bias by changing the nature of the social welfare function that the bank seeks to maximize. However, under central-bank independence, the central bank retains discretion over monetary policy, and its monetary policy actions are capable of affecting the real economy and the domestic price level. In principle, therefore, independence of the central bank is an institutional mechanism for strengthening the bank's anti-inflationary credibility without sacrificing its flexibility. However, as we have just seen, de facto independence may be quite different from de jure independence, and de facto independence may be quite difficult to achieve. Consequently, other mechanisms have been proposed to address the pro-inflation bias arising from time inconsistency. Some of these mechanisms operate by limiting the central bank's discretion, and others by reducing the payoff to its exercise of that discretion.

1. Money Growth Rules

One of the most venerable of these mechanisms is Milton Friedman's constant money growth rule. This rule essentially removes the central bank's discretion to engineer inflationary surprises by forcing it to keep the rate of growth of the money supply constant. A constant low rate of money supply growth would deliver low long-run inflation but would do so at the cost of removing the central bank's flexibility to use monetary policy as a stabilization tool.

2. A Fixed Exchange Rate with Perfect Capital Mobility

Under perfect capital mobility, a fixed exchange rate would remove the central bank's incentive to engineer surprise inflation through expansionary monetary policy because the impossible trinity implies that the central bank would have no control over the domestic money supply and the domestic interest rate, so monetary policy would be unable to affect aggregate demand. While this policy regime would therefore remove the pro-inflation bias associated with monetary policy, like a constant money supply rule, it does so by making monetary policy useless for stabilization purposes.

3. A Currency Board

However, because a soft exchange rate peg leaves the central bank with discretion both over domestic credit and over the exchange rate, it does not remove pressure for the central bank to finance fiscal deficits, and it does not remove the incentives for the central bank to seek to engineer inflationary surprises through exchange rate policy (by devaluing the exchange rate) rather than through monetary policy. Thus it is not at all clear that fixing the exchange rate in and of itself – that is, without stronger institutional underpinnings – is capable of addressing the problem of pro-inflation bias.

These institutional underpinnings can be provided by enshrining the fixed exchange rate in a currency board – an arrangement whereby the fixed exchange rate is made very difficult to change and the central bank is required to back the domestic currency fully with foreign exchange reserves. The exchange rate can be made difficult to change by specifying the official exchange rate in the constitution or making it legislatively mandated, subject to change only by parliamentary supermajorities rather than simply at the discretion of the central bank. The requirement that the domestic currency be fully backed by foreign exchange reserves essentially rules out domestic monetary policy because the central bank has no authority to issue domestic credit.

Because a currency board eliminates both exchange rate and monetary policy, it eliminates pro-inflation bias. It does so, however, by completely eliminating the central bank's discretion over seignorage revenue and also eliminating the use of both monetary and exchange rate policies as stabilization instruments. Relative to soft exchange rate pegs, therefore, currency boards purchase additional anti-inflationary credibility at the cost of monetary and exchange rate policy flexibility.

4. Dollarization

Because a country retains its own currency under a currency board, it remains, in principle, possible to change the exchange rate (by amending the constitution or achieving the required supermajority in the legislature) or to change the required backing for the domestic currency. Thus the difference between a currency board and a soft exchange rate peg can be blurred, potentially reducing the anti-inflationary credibility of the currency board and reintroducing pro-inflation bias. This blurring can be avoided by abandoning the domestic currency altogether and adopting the currency of a stable-price trading partner as the domestic legal tender. This act, referred to as *dollarization* (whether the foreign currency involved is the dollar or some other currency), therefore represents an extreme sacrifice of flexibility for the sake of anti-inflationary credibility.

VII. SUMMARY

In this chapter, we considered how monetary policy can contribute both directly and indirectly to the process of economic development and concluded that a pro-development monetary policy is one that maintains a credible low rate of inflation while safeguarding the use of monetary policy as a stabilization instrument. In actual practice, unfortunately, monetary policy in emerging and developing economies has often tended to display a pro-inflation bias and behave pro-cyclically.

We saw that theory – in the form of the analysis of the time-inconsistency problem in the formulation of monetary policy – can explain the source of a pro-inflation bias as well as suggest institutional remedies to address it, specifically by making the central bank independent and buttressing its independence through various legal means such as specifically directing it to concentrate on the objective of price stability. Testing whether central-bank independence has empirically resulted in lower inflation encounters the serious obstacle of devising an empirical measure of central-bank independence. We reviewed some prominent research that devised such a measure and found that indeed, central-bank independence was associated with superior inflation performance.

Though central-bank independence and associated institutional reforms may thus be capable of reducing the pro-inflation bias in monetary policy arising from time inconsistency, this leaves open the question of how an independent central bank should conduct its day-to-day operations to strike a balance between anti-inflation credibility and the flexible use of monetary policy for stabilization purposes. We next turn to an examination of a specific monetary policy regime intended to allow an independent central bank to strike this balance.

REVIEW QUESTIONS

1. What is time inconsistency? Under what conditions would it arise in the formulation of monetary policy?
2. Does a pro-inflation bias imply that the central bank cares less about inflation than does the society at large? Explain why or why not.
3. In the absence of reputational considerations, why does a socially responsible central bank not just set the inflation rate equal to zero?
4. Under what conditions would a concern for its reputation induce a central bank to set an inflation rate equal to zero?
5. Describe five alternative institutional "fixes" to the problem of pro-inflation bias in monetary policy formulation.

EXERCISES

1. In what ways can appropriate fiscal and monetary policies contribute to economic growth and development? For each type of policy, describe a direct (microeconomic) and two indirect (macroeconomic) channels through which appropriate policies can be conducive to growth and development.
2. Should a central bank be a democratic institution (i.e., one whose priorities reflect the will of the people)? Explain why or why not.

3. Alan Blinder, a former vice chairman of the U.S. Federal Reserve System, once got into trouble with the financial press for suggesting that central banks should care not just about reducing inflation but also about reducing unemployment. His critics argued that central banks should worry only about keeping inflation low. Using the description of the inflation-setting behavior of an independent central bank that we developed in this chapter, make the best case you can for the critics' point of view.
4. Comment on the validity of the following statement: "Fiscal dominance is at the root of inflationary episodes, henceforth giving full discretion to the Central Bank implies no more inflation."
5. In our analysis of the role of commitment mechanisms in affecting the time-inconsistency problem, we assumed that the costs of deviating from an announced inflation target were linear. How would our prediction about the equilibrium rate of inflation change if those costs were instead quadratic (i.e., if they depended on the *square* of the deviation of the actual inflation rate from the announced target)?

REFERENCES AND FURTHER READING

Barro, Robert J., and David B. Gordon (1983), "Rules, Discretion, and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics*, vol. 12, pp. 101–121.

Carstens, Agustin, and Luis Jacome (2005), "The 1990s Institutional Reform of Monetary Policy in Latin America," working paper 343, Central Bank of Chile.

Chang, Roberto (1998), "Policy Credibility and the Design of Central Banks," *Federal Reserve Bank of Atlanta Economic Review* (First Quarter), pp. 4–15.

Cukierman, Alex (2006), "Central Bank Independence and Monetary Policymaking Institutions: Past, Present, and Future," working paper 360, Central Bank of Chile.

Cukierman, Alex, Steven B. Webb, and Bilin Neyapti (1992), "Measuring Central Bank Independence and Its Effect on Policy Outcomes," *World Bank Economic Review*, vol. 6, pp. 353–398.

Cukierman, Alex, Steven B. Webb, and Bilin Neyapti (1994), "Measuring Central Bank Independence and Its Effect on Policy Outcomes," occasional paper 58, International Center for Economic Growth.

Friedman, Benjamin (2000), "Monetary Policy," working paper 8057, National Bureau of Economic Research.

Fuhrer, Jeffrey C. (1997), "Central Bank Independence and Inflation Targeting: Monetary Policy Paradigms for the Next Millennium?" *Federal Reserve Bank of Boston New England Economic Review* (January–February), pp. 19–36.

Kißner, Friedrich, and Helmut Wagner (1998), "Central Bank Independence and Macroeconomic Performance: A Survey of the Evidence," discussion paper 255, University of Fern-Hagen.

Mishkin, Frederic S. (2000), "What Should Central Banks Do?" *Federal Reserve Bank of St. Louis Review* (November–December), pp. 1–13.

15

Inflation Targeting

We saw in the preceding chapter that time-inconsistency problems are likely to induce a pro-inflation bias to monetary policy. We also saw that one way to combat this bias was by making the central bank independent of the government, thereby giving it the power to refuse the finance minister's request for money financing. An independent central bank that does not weigh budget financing as heavily in its objectives as does the finance minister would indeed tend to produce a lower equilibrium rate of inflation, even if acting with complete discretion, than if the optimal inflation decision were made by the finance ministry itself.

However, the preceding chapter also showed that giving the central bank independence is not a panacea: though central-bank independence may reduce pro-inflation bias, an independent central bank may still produce too high a rate of inflation. Specifically, when the central bank is unable to precommit to following through on its policy announcements, and thus acts with discretion every period, the equilibrium rate of inflation will tend to be higher than is desirable, as long as the central bank cares about both output and inflation. In the absence of an independent mechanism to make the central bank's policy announcements credible, a preoccupation with its reputation does not solve this problem. A central bank that can only establish anti-inflationary credibility by consistently delivering low inflation in the face of the public's expectation of higher inflation purchases its anti-inflationary credibility at the cost of what may be a prolonged recession.

All this suggests that social welfare could be increased if the central bank were somehow able to bind itself (precommit) to following through on its policy announcements.[1] This raises the question of whether it is possible to design

[1] Notice that this problem arises only when the central bank is independent and is potentially able to exercise policy discretion. In other words, it does not arise under the more extreme institutional shortcuts of monetary unification, dollarization, or the adoption of a currency board because in these cases, there really is no central bank worthy of the name, and if a central bank does exist, it has

an institutional mechanism that would induce an independent central bank to be bound by its policy announcements. This chapter considers the role of such *commitment mechanisms* and, in particular, examines the monetary policy regime known as *inflation targeting* (IT). This regime represents an attempt to create a commitment mechanism that can establish the central bank's anti-inflationary credibility while at the same time preserving enough flexibility for the central bank to use monetary policy in a stabilizing role.

The structure of the chapter is as follows. In the next section, we consider the role of a generic (unspecified) commitment mechanism in the context of the time-inconsistency problem, examine what options are available in practice for implementing a commitment mechanism for monetary policy, and explain why IT has become the mechanism of choice in many countries. Section II examines how IT is intended to work in theory, whereas Section III examines how it can be designed to attempt to reconcile anti-inflationary credibility with the flexibility for monetary policy to be used as a short-run stabilization instrument. The empirical evidence on the effectiveness of IT in reducing pro-inflation bias is briefly reviewed in Section IV, and Section V discusses how IT can be implemented with a specific monetary policy rule. Section VI summarizes.

I. COMMITMENT MECHANISMS

1. Role of Commitment Mechanisms

How would an effective commitment mechanism work in the context of the time-inconsistency problem? To answer that question, go back to the objective function that the independent central bank seeks to maximize:

$$B(\pi) = \beta(\pi - \pi^e) - (\alpha/2)\pi^2 \tag{15.1}$$

Recall how the time-inconsistency analysis works: the central bank makes an announcement about its intentions concerning the rate of inflation, the public then forms its expectations of inflation, and the central bank finally decides what rate of inflation to actually produce. The time-inconsistency problem arises because the central bank has an incentive to announce zero inflation but, once the public has formed its expectations, to renege on its announced intentions and actually deliver the positive rate of inflation β/α.

A commitment mechanism essentially adds an additional term to this function, consisting of the cost to the central bank of deviating from its policy announcement. For simplicity, suppose this loss is asymmetric, so the central bank is penalized for achieving inflation that is higher than the announced value, but rewarded for

no policy discretion. There is nothing it can do in this case to influence the inflation rate, which, in the medium run, will be determined by the country's trading partners.

producing inflation that is below the announced value. In that case, we can express the loss of reputation linearly as $C(\pi - \pi^*)$, where C is a constant that reflects the penalty that the central bank pays per point of inflation in excess of that implied by its announced policy intention π^*. Under this assumption, the central bank's revised objective function becomes

$$B(\pi) = \beta(\pi - \pi^e) - (\alpha/2)\pi^2 - C(\pi - \pi^*) \tag{15.2}$$

Notice that since the inclusion of this term just reduces the marginal benefit of inflation from β to $\beta - C$, the optimal rate of inflation now becomes

$$\pi = (\beta - C)/\alpha \tag{15.3}$$

The penalty thus encourages the bank to seek to achieve a lower rate of inflation than it otherwise would, and the bigger the penalty, the lower the equilibrium inflation rate. To see how large the loss would have to be for the bank to actually hit its target, set π equal to its target value π^*. This gives

$$C = \beta - \alpha\pi^* \tag{15.4}$$

Thus, for the regime to be successful, the institutional arrangement has to be set up such that the loss to the bank is greater the greater its temptation to engineer surprise inflation (i.e., the greater is β and the smaller is $\alpha\pi^*$).

But how could a penalty be imposed on the central bank for failing to comply with its policy intentions? One way to do so would be for the central bank to make a very public announcement that it indeed intends to achieve the rate of inflation π^* and to subject itself to a cost of some type if it fails to comply with those announced intentions. That cost could come in the form of a loss of reputation and credibility, or more concrete penalties could be imposed on the bank by making the salaries or the very job security of the governor and the central bank's board of directors contingent on the bank's compliance with its policy announcements.

2. Targets and Instruments

So far, we have been analyzing the central bank's actions as if it could control the rate of inflation directly. However, as we saw in Chapter 5, what the central bank controls directly in the real world is its balance sheet, that is, its stock of foreign exchange reserves and domestic credit. The bank can conduct exchange rate policy by changing the *composition* of its balance sheet, and it can conduct monetary policy by changing the *size* of its balance sheet. Because the former affects the exchange rate and the latter the money supply, to keep matters simple, we can describe the central bank as controlling the nominal exchange rate and the money supply. These are the central bank's policy *instruments*, while the inflation rate is a policy *target*. As such, what makes it different from a policy instrument is that it may also be affected by variables other than those controlled by the central bank, at least in the short run.

That the central bank does not control the inflation rate directly complicates the issue of how to implement a commitment device because it raises the question of whether the central bank should make an announcement about its intentions regarding its policy instruments or about the actual outcome for its policy target, the rate of inflation. The answer depends in part on the relationship between the central bank's policy instruments and the economy's inflation outcome.

What exactly is this relationship? We saw in Chapter 7 that in the short run, a variety of shocks can affect the domestic price level and thus exert transitory effects on the inflation rate, even if the central bank is holding its policy instruments constant. In Chapter 8, in turn, we showed that, holding other exogenous variables constant, if the central bank fixes a constant rate of nominal exchange rate depreciation, both the economy's rate of inflation and its rate of money supply growth must be equal to that rate of depreciation in the medium run, whereas if it fixes a constant rate of money supply growth, the rate of inflation and rate of exchange rate depreciation must be equal to that rate of money supply growth in the medium run.[2] These observations have three implications:

1. In the short run, if the central bank keeps its policy instruments on a predetermined path, the inflation rate can be affected by a variety of exogenous shocks to the economy.
2. Whichever nominal variable the central bank chooses to set on a predetermined path – the exchange rate or the money supply – the path of the price level will be closely tied to the path of that variable in the medium run.
3. In the medium run, the central bank can control the exchange rate or the money supply, but not both.

The upshot is that the central bank can indeed determine the medium-run inflation rate by choosing an appropriate path for one of its two nominal policy instruments, but if it keeps that instrument on its announced path, the inflation rate may fluctuate around its medium-run target as the result of nonmonetary policy shocks.

In principle, then, the central bank faces several choices in announcing its policy strategy. One choice is to announce a future path for one of its nominal policy instruments – that is, a path for the exchange rate or the stock of money – and let the public infer from that choice what the bank's targeted path for its ultimate policy objective, the average price level, actually is. If it opts to do this, it needs to make a secondary choice about which nominal instrument to use. Alternatively, it could formulate its policy announcement in terms of an intended path for the

[2] This statement assumes, as in Chapter 8, that there is no growth in domestic productive capacity as well as no inflation in the country's trading partners.

price level itself.[3] In this case, it would make no explicit commitment about the path of its nominal policy instruments but would stand ready to alter the exchange rate and/or the money supply as necessary to achieve a certain price-level outcome, depending on circumstances.

3. Choosing a Commitment Device

What choice should the central bank make from among these options? An important consideration in answering this question is that the three options we have considered may not be equally effective as commitment devices, that is, as mechanisms for inducing the central bank to act in accordance with its preannounced policy intentions. Presumably, to solve the time-inconsistency problem, the central bank would want to make the choice that provides the most effective commitment mechanism. Our next task, therefore, is to investigate the relative merits of each of these possible types of policy announcements as commitment mechanisms.

a. Characteristics of Commitment Mechanisms

To do this, consider first what characteristics a good commitment mechanism should have. Recall the sequence of steps in our analysis of the time-inconsistency problem:

- The central bank announces a policy.
- On the basis of this, the public forms expectations of inflation.
- Given the public's inflation expectations, the central bank chooses the actual policy.

This suggests that a good commitment mechanism should have the following characteristics:

If the central bank announces a policy that is consistent with low inflation and indeed acts in good faith (i.e., acts in accordance with its announced policy), inflation should in fact turn out to be low. In other words, the public must believe that the central bank's plans, if they are actually carried out, would in fact generate a low rate of inflation. If this were not so, the public would have no reason to form the expectation that inflation will indeed turn out to be low. And if the public were not to form an expectation of low inflation, the central bank would face a stronger incentive not to abide by what it announced that it would do. To see why this is so, recall from Chapter 14 that a central bank that cares about its reputation has a stronger incentive to keep the inflation rate low when the public expects low inflation than when it expects high inflation. Thus a public belief that the central

[3] Notice that the announcement that the central bank makes determines the exchange rate regime. The exchange rate is fixed if it announces an exchange rate path and floating if it does not.

bank's announced policy is indeed consistent with low inflation makes it more likely that the central bank will actually implement that policy.

Note that the issue of whether a central bank using a particular commitment mechanism is at least in principle capable of delivering low inflation would not be relevant if the bank actually controlled the inflation rate directly, as we assumed in the preceding chapter. The issue becomes important only when this control is indirect and potentially imperfect. Whether the public believes that inflation will indeed turn out to be low if the central bank acts in good faith in turn depends on three things:

1. whether the central bank is technically able to comply with its policy announcement (i.e., the controllability of the announced target)
2. whether the announced policy is indeed effective in affecting inflation
3. whether the link between the central bank's actions and the rate of inflation is actually transparent to the public

The second characteristic of a good commitment device concerns the final step in the time-inconsistency analysis: how the central bank reacts to the public's expectation. If the public expects low inflation, a good commitment device should reduce the central bank's incentive to deviate from its announcement.

Recall from Section I that in the presence of a commitment mechanism, the central bank will tend to produce the inflation rate $\pi = (\beta - C)/\alpha$. Thus the temptation for a central bank to "cheat" depends on β, α, and C. Specifically, it is increasing in β but decreasing in both α and C. Assuming that β and α are the same for different commitment mechanisms, the best commitment device is therefore the one that provides the highest value of C. The value of C is higher:

1. if the commitment device makes it easier for the public to detect any tendency for the central bank to renege on its commitments
2. if it imposes a higher cost on the central bank when it does so

In principle, then, we want an instrument that the central bank can control directly, that has a strong and transparent impact on the rate of inflation and that the public can observe directly so it can monitor the central bank's compliance with its policy announcements. How well do the alternative nominal anchors satisfy these criteria?

b. Comparing Alternative Policy Announcements as Commitment Mechanisms

i. Exchange Rate Announcements. Announcing a predetermined exchange rate path has certain advantages as a commitment mechanism:

- It is a variable that the central bank can – at least, in principle – control directly, so it is reasonable to expect that it will move in the indicated direction if the central bank acts in good faith.

- The exchange rate is directly tied to the price level through the price of imported goods and the equilibrium real exchange rate (i.e., $P = SP^*/Q$).
- It is a simple announcement that everyone can understand.

These considerations suggest that if the central bank follows its announcement, people would have good reasons to expect low inflation. Also, consider the following:

- The exchange rate can be observed directly in the foreign exchange market, so everyone can tell fairly readily whether the central bank is complying with its announcement.
- Fixing the exchange rate lends itself to international agreements that are costly to break, in the form of common exchange rate regimes adopted by groups of countries.

For these last two reasons, breaking promises with respect to the nominal exchange rate has high political costs for the central bank, and these costs may negate any benefits that the central bank may perceive from inflationary surprises engineered through exchange rate changes. This would leave the bank with reduced incentives to generate such surprises and consequently may help to make policy announcements more credible.

Unfortunately, exchange rate announcements also have important disadvantages as commitment mechanisms. These consist of the following:

- If capital mobility is *imperfect*, abiding by a preannounced exchange rate path does not prevent the bank from engineering an inflationary surprise anyway, by implementing an expansionary monetary policy. Another way to say this is that the relationship of the nominal exchange rate to the price level would not be stable in the short run if there are large changes in the short-run equilibrium real exchange rate, and when capital mobility is imperfect, the central bank retains the capacity to produce such changes through monetary policy.
- When capital mobility is *high*, however, this multiplicity of policy instruments at the central bank's disposal is not a problem because the central bank loses monetary autonomy when it announces a predetermined path for the exchange rate. But a different problem arises in this case: that the central bank may not be *able* to control the exchange rate. If this is so, the effectiveness of the exchange rate as a commitment device would be undermined. We will discuss the reason that the central bank may not be able to control the exchange rate in this case in Chapter 27. In brief, the argument is that when capital mobility is very high, an anticipated exchange rate devaluation may cause a large increase in the domestic interest rate. This would create a recession that the central bank would not be able to combat with monetary policy. Under these circumstances, the central bank may be induced to try to stimulate the economy using the only instrument at its disposal: the exchange rate. Because the expectation of a devaluation may therefore itself *cause* a devaluation, the private sector may have

good reason to doubt that any preannounced exchange rate path would actually be implemented.

ii. Money Supply Announcements. Consider next the characteristics of money supply announcements as commitment devices. Just as policy announcements about the future path of the nominal exchange rate have advantages and disadvantages as commitment devices, so do announcements about the future path of the money supply. The advantages include the following:

- In principle, like the exchange rate, the money supply is a variable that the central bank can control directly.
- When the demand for money is stable, the money supply is directly tied to the domestic price level.
- Because information about the money stock is available with high frequency, the bank's adherence to its announced targets can be known fairly quickly, allowing the public to monitor the central bank's compliance with its policy announcements.

On the other hand,

- It may be that the monetary aggregate about which the central bank makes a policy announcement is *not* indeed under the complete control of the central bank (e.g., the bank may make an announcement about the path of a broad monetary aggregate, but it only controls the path of the monetary base directly). If that is so, the signal sent by the path of the targeted aggregate about the future behavior of the inflation rate would be imperfect.
- The relationship of the money supply to the price level may not be as obvious as that of the exchange rate because the link is not direct and immediate to any component of the average price level.
- When capital mobility is imperfect, the exchange rate and money supply can be controlled independently. Thus committing itself to a preannounced path for the money supply does not prevent the central bank from engineering an inflationary surprise by using exchange rate policy.
- Most important, when the demand for money is unstable, adhering to money supply targets may actually *destabilize* the price level.
- Finally, the money supply cannot be observed directly but only through announcements made by the central bank itself. This makes it more difficult for the public to monitor the central bank.

iii. Inflation Announcements. Instead of announcing a path for one of its policy instruments, the central bank could opt to make an announcement about the ultimate objective of policy: the price level. Like the alternatives examined earlier, this has advantages and disadvantages.

Among the advantages are the following:

- This is a simple and transparent announcement.
- The information on the central bank's compliance is available quickly both in the marketplace and through official inflation figures, which are published with high frequency.
- Most important, this announcement allows the central bank to retain substantial flexibility in its choice of instruments. It permits the central bank to respond to changes both in money demand and in the equilibrium real exchange rate without destabilizing the domestic economy.

The main disadvantage, on the other hand, is that the inflation rate is an endogenous variable not controlled directly by the central bank and not always easily predictable (note that because of policy lags, the central bank would actually have to be able to *predict* inflation to take countervailing measures to permit it to hit its target) or easily linked to the instruments controlled by the central bank (e.g., the central bank may be able to control nominal short-term money-market rates, but what affects real aggregate demand is long-run real interest rates). Thus the central bank would find it easier to evade responsibility for changes in this variable. This suggests that an announcement about the future path of the price level is not obviously superior as a commitment mechanism to announcements about the central bank's policy instruments.

iv. Recent Experience with Commitment Mechanisms. Until relatively recently, countries that opted to use commitment devices for monetary policy tended to rely on announcements about the path of the exchange rate or the money supply. However, despite the uncertain superiority of a target-oriented announcement, both of these instrument-oriented policy announcements have tended to fall out of favor in recent years.

The use of the money supply as a nominal anchor has been widely (though not completely) abandoned by central banks around the world – in both industrial and developing countries – because a continuous process of financial reform and innovation has rendered the demand for money highly unstable, loosening the connection between the money supply and the price level. As mentioned earlier, if the central bank persists in pegging the money supply under such circumstances, the effect would be to *destabilize* the price level.

In the case of the exchange rate, high capital mobility among emerging-market economies has made officially determined exchange rates more difficult to defend, calling into question the controllability of this policy instrument and thus undermining the belief than an exchange rate announcement would actually succeed in pinning down the path of the price level. As a consequence, several central banks in both industrial countries and emerging economies have opted, almost by default, to target the rate of inflation itself. This monetary policy regime, dubbed *inflation targeting*, is examined in the rest of this chapter.

II. HOW INFLATION TARGETING WORKS

IT is a monetary policy framework designed to have the public expect, and the central bank actually to deliver, a low rate of inflation, say, π^*. As a commitment mechanism, it is designed to enable the central bank to precommit to following through on its policy announcements, essentially by penalizing it if it does not do so. Knowing that the central bank faces such a penalty, the public is more likely to believe the bank's policy announcements.

IT was first introduced as a monetary policy framework by New Zealand in 1990, but several other small industrial countries (Sweden, the United Kingdom, Canada, Australia, Finland, Israel, and Spain, until it adopted the euro) have also adopted it. Among emerging-market economies, IT has been widely adopted in Latin America. Chile announced an inflation target in 1990, after a new central-bank law was passed in 1989. After its 1994 crisis, Mexico adopted a dual monetary regime focusing on money growth while setting a goal for inflation, but it implemented formal IT in 1999. Brazil also adopted it in 1999, after the collapse of its crawling-peg exchange rate arrangement. Colombia adopted IT in 1999. Peru has been publishing inflation targets since 1994 and adopted formal IT in 2002. Argentina also has an unofficial target range for inflation.[4] Other emerging-market economies outside Latin America that have also adopted IT include the Czech Republic, Hungary, Korea, Poland, South Africa, and Thailand.[5]

1. Defining Inflation Targeting

The key problem involved in designing an IT monetary policy regime is how to cope with the fact that the central bank does not control the price level directly. This creates the challenge of *accountability*, that is, how to tell whether the central bank has done everything in its power to deliver on its commitments. Much of the design of the institutional framework of IT is intended to cope with this challenge.

IT has four components:

1. The central bank makes a public announcement of an inflation rate π^* that it will seek to achieve *within a certain period of time*, *within a certain range*, and possibly *subject to certain conditions*. As described in the last section, this is what distinguishes IT from other possible monetary policy regimes: the choice of announcing – and committing to – the ultimate target rather than the instrument through which it will be attained.

4 See Bernanke (2005).
5 The list of inflation targeters is from Fraga et al. (2003).

2. The bank makes no commitment to target any other nominal variable, such as the exchange rate, the money supply, or domestic credit, or to any specific target for the domestic interest rate. The reason is that under IT, the inflation rate is the overriding policy objective, and the behavior of all these instruments must be subjugated to that objective. To do so, the settings of all these other variables may have to change as circumstances change. This means, among other things, that the economy has no unique nominal anchor. Instead, the monetary policy variable that provides the nominal anchor for the economy changes as circumstances change.
3. In addition to a public commitment to a specific rate of inflation, and non-commitment to specific values of any other variables, the third ingredient of IT is *central-bank transparency*. This means that the central bank periodically explains to the public the relationship between the state of the economy and the setting that it is choosing for its policy instruments, as well as the links between these two things and the inflation target.
4. Finally, under IT, a variety of mechanisms are put in place to make the central bank accountable for attaining its inflation objectives.

2. Inflation Targeting and Time Inconsistency

How do the characteristics of IT described earlier address the challenges posed by the time-inconsistency problem? First, central bank independence and a publicly announced target help make the central bank accountable.[6] A very public commitment to price stability by an institution that has both the means and the discretion to achieve that goal reduces the temptation for the central bank to cheat by producing surprise inflation because it would suffer a significant loss of reputation if it did so (we can interpret this as a high value of C). The idea, therefore, is that the central bank's putting its reputation on the line (and possibly the job or salary of the governor as well) provides the commitment mechanism required to make the inflation target credible.

Second, an operating procedure in which the central bank is free to choose among alternative intermediate variables in *ad hoc* fashion confirms the primacy of the price-level target over any particular monetary policy instrument. Allowing the bank flexibility to respond to any instability that may emerge in the relationship between specific instruments (e.g., the exchange rate or the money supply) and the ultimate price-level target by giving it the discretion to choose among alternative intermediate instruments at different moments in time makes it more likely that it will hit its ultimate inflation target, thus making that target more credible. Note that this means that a central bank that is pursuing an IT regime should

[6] This is not unique to IT because it would also be true if the central bank chose to target the money supply or the price level.

not simultaneously commit itself to any prespecified path for any of its policy instruments because that would deprive it of the flexibility to use that instrument to hit its inflation target.

Finally, the objective of transparency is to facilitate monitoring by the public of what the central bank is doing, thereby enabling the public to confirm that the central bank is indeed committed to achieving its announced inflation target. This is particularly important in the case of IT because it helps to compensate for the fact that the inflation rate is not directly controlled by the central bank, unlike the exchange rate and the monetary base. Enhanced transparency facilitates monitoring, replacing the direct observability of the exchange rate or monetary base.

III. IMPLEMENTING INFLATION TARGETING

1. Prerequisites

The description of IT in the preceding section makes it clear that there are two important prerequisites for IT. First, as already indicated, for the central bank to be accountable, it must be independent. But central-bank independence can mean different things: the bank can have *goal independence* or *instrument independence*. Goal independence refers to a situation in which the central bank can choose the inflation target, whereas instrument independence allows the central bank discretion over how to hit that target. A prerequisite for IT is that the central bank must have *instrument independence*; that is, it must be able to set its policy instruments (intervention in domestic securities markets and in the foreign exchange market) in an unconstrained way. Otherwise, the objective of hitting an inflation target could be compromised by the need to set policy instruments at values that are incompatible with the inflation target to satisfy a side constraint on the instrument.

Second, the central bank must have a model for producing inflation forecasts that link the inflation rate to monetary policy instruments as well as to the state of the economy. In other words, it must understand how monetary transmission works in the economy, both qualitatively and quantitatively. If this were not the case, the public would have no reason to believe that the central bank would actually be able to fulfill its commitment.

If either of these prerequisites is not in place, the central bank's ability to hit its inflation target is compromised, and there would thus be no reason for its promises to do so to be credible.

2. Operational Issues

If the prerequisites listed earlier are in place, implementing an IT regime next requires resolving several operational issues. These involve both the institutional arrangements underpinning the regime and several decisions about the specification of the inflation target.[7]

a. Institutional Arrangements

The implementation of an IT regime needs to address several broad institutional issues:

1. Who specifies the inflation target? If the central bank has instrument independence, but not goal independence, the target is presumably chosen by the government. If it has goal independence as well, the target is chosen by the central bank. Alternatively, it could be chosen in joint consultation between the government and central bank.
2. What is the formal standing of the inflation target? Once the target is announced, is it to be considered as a formally mandated requirement with which the central bank must comply under threat of formal sanctions of some type, or just an operational objective for monetary policy?
3. How much instrument independence does the central bank actually enjoy? In other words, who determines how matters related to the conduct of monetary policy are decided?
4. What specific means will the central bank use to enhance its transparency and accountability? For example, will the central bank issue inflation forecasts, provide regular inflation reports, and release the minutes of the deliberations of its monetary policy committee? Will the central-bank governor provide testimony on the conduct of monetary policy to the legislature? If so, with what regularity?

b. Specification of the Inflation Target

Up to now, we have been focusing on the role of IT in allowing the central bank to achieve anti-inflationary credibility. But can a central bank simultaneously target low inflation and real output stability? As we have seen, this is not easy, because actions taken by the central bank to safeguard output stability may be perceived as inconsistent with its long-run inflation objective. But the two can be reconciled within an IT regime if the regime constrains the central bank's behavior sufficiently in the medium term as to create the expectation than any short-term departures of the inflation rate from its medium-run target are only temporary. Indeed, Bernanke and Mishkin (1997) describe IT as "constrained discretion" – the idea is precisely to strike an optimal trade-off between credibility and flexibility by committing the central bank to a medium-run value for the rate of inflation while allowing it to engage in output stabilization in the short run. The objective is for the temporary deviations from the announced path of the inflation rate that may be required

[7] For a description of the implementation of inflation targeting among some early adopters, see Kahn and Parrish (1998).

to stabilize real output in the short run not to impair the central bank's anti-inflationary credibility in the medium run. The central bank can seek to achieve this short-run flexibility through several choices that it has to make regarding the specific characteristics of the inflation target.

The relevant choices are the following:

1. The central bank has to determine which price index to choose in defining the target. The choice typically involves whether to target the "headline," or conventional, consumer price index (CPI) or a component of the CPI that excludes items with particularly volatile prices such as food and energy (this is often referred to as the "core" component of the CPI). Though the overall CPI is the concept that is of most relevance to the public in assessing its cost of living, pinning down the path of the overall CPI in the face of large and temporary fluctuations in food and energy prices would involve making large compensating changes in broad categories of other prices that are "stickier" than those of food and energy, thereby possibly destabilizing real output. Defining the target in terms of core inflation therefore enables the bank to better perform a short-run output-stabilizing function by preventing monetary policy from responding to a variety of price-level shocks that are likely to be of a highly transitory nature.
2. The bank also has to determine whether to target the price level or the rate of inflation. Targeting the rate of inflation makes the behavior of the price level more difficult to forecast at long horizons, which impedes private-sector planning. However, targeting the price level may increase the variability of real output because it would require creating *deflationary* shocks to offset the effects of transitory inflationary shocks on the price level (see Bernanke and Mishkin 1997). Thus opting for an inflation target rather than a price-level target involves sacrificing some anti-inflationary credibility for the sake of output stability.
3. The next issue concerns the behavior of the target itself. A numerical value has to be chosen for the target, as well the horizon over which the target is to be achieved, and if the inflation rate is currently higher than the target, the time path that the inflation rate will follow in converging to the target must also be chosen. Again, these choices can be made to avoid inducing excessive variability to real output as well as to retain flexibility for the central bank to respond to shocks countercyclically. If inflation is initially higher than eventually desired, setting too ambitious a target over a short time horizon may imply a severe output contraction in the short run. To safeguard output stability, therefore, the bank may set a less ambitious short-run inflation target or specify an ambitious target that would only be achieved over a much longer term. Lengthening the horizon also gives the bank flexibility to respond to short-run shocks to the economy during the transition to the inflation target without being constrained by its ultimate inflation commitment.

4. Finally, the bank must determine whether to specify the target as a point estimate or a band; if it specifies the target as a band, it must choose the width of the band and determine the use of escape clauses or exemptions. A wider band allows for uncertainty regarding the links between the central bank's instruments and the inflation outcome and also allows the bank more scope to accommodate supply shocks without destabilizing real output. However, the width of the band is clearly closely related to the credibility-flexibility trade-off: the wider the band, the more flexibility the bank retains to deviate from its announced target without losing credibility, but the less effective the target is in pinning down the public's inflation expectations. Similarly, more extensive use of escape clauses, or the use of escape clauses that are vague, also safeguards flexibility at the expense of credibility. For a given degree of flexibility, one would expect that the more restricted is the use of escape clauses, the wider the band would have to be.

3. Inflation Targeting in Practice

As indicated previously, IT has been adopted as a monetary policy regime by various countries since 1990. In practice, the regime has been implemented with a concerted effort to gain anti-inflationary credibility, while retaining countercyclical flexibility. For example, to avoid reacting to supply shocks, the targeted price index has typically been one consisting only of core prices; targets have tended to be set in the form of a band, not only allowing for uncertainty about the link between instruments and targets but also allowing the central bank room to address output objectives; targets have been adjusted to accommodate price shocks outside the central bank's control; and the horizon for achieving the target has typically been set several years ahead, with inflation intended to converge from the current rate to something approaching a desirable medium-run rate of inflation, usually a small positive number. In projecting future inflation, a variety of predictors of inflation are used, rather than just one or two intermediate instruments, precisely because the relationship of those intermediate variables to the price level has often tended to break down. Central banks practicing IT have issued regular, detailed assessments of the inflation experience and outlook, with the policy response explained. Finally, practices with respect to accountability have varied, with New Zealand at one extreme making the tenure of a governor depend on hitting the targets, while other countries have relied on penalties based on lost reputation and prestige.[8]

IV. EVIDENCE

Overall, the results from the adoption of IT in emerging and developing countries appear to have been relatively positive, despite the handicap that the IT framework

[8] For detailed descriptions of how inflation targeting has been in four early inflation targeters, see Kahn and Parrish (1998). See also Bernanke et al. (1999) as well as Bernanke and Woodford (2005).

faces in that context (including weak macroeconomic institutions, relatively high initial inflation, low initial credibility, more severe macroeconomic consequences associated with exchange rate changes, and more severe external shocks). Fraga et al. (2003), for example, found that inflation dropped after the adoption of IT in the emerging-market economies listed in the introduction to this chapter, though not immediately to the levels recorded by industrial-country inflation targeters. However, perhaps because of the factors listed earlier, emerging-market inflation targeters have on the whole been less successful at hitting their targets than have industrial countries, even though their target ranges have tended to be broader (Ho and McCauley 2003).

Though the experience to date among countries that have adopted inflation targets suggests that the use of inflation targets has been associated with lower inflation, the move to IT has taken pace in a context in which inflation rates have tended to fall worldwide, including in countries that have not adopted this policy regime. Careful studies that have controlled for other determinants of inflation in evaluating the success of this strategy have derived mixed results. Fuhrer (1997), for example, found that adopting IT made very little difference in central-bank behavior among several industrial-country inflation targeters. On the other hand, Newmann and von Hagen (2002) found not only that adopting IT did change the behavior of the central banks in their sample but also that the regime was associated with lower and more stable inflation as well as improved anti-inflationary credibility for those central banks. On the other hand, there is little evidence to date that adopting inflation targets has lowered sacrifice ratios (the amount of lost output associated with a given reduction in the rate of inflation) or directly lowered inflation expectations – at least until central banks have actually delivered low inflation.

V. TAYLOR RULES

Whereas IT is a monetary policy *framework*, it is not a monetary policy *rule* in that it does not provide simple and mechanical links between the setting of the central bank's monetary policy instruments and the state of the economy. It essentially just directs the central bank to use all available information to set whatever instruments the bank has at its disposal, however it sees fit, to achieve the announced outcome for inflation. As we have seen, the framework can be set up in such a way as to achieve this announced inflation outcome in the medium run while retaining a substantial amount of short-run flexibility for monetary policy to stabilize the real economy in response to exogenous shocks. The next question we have to ask is, what specific policy rules can the central bank follow to meet its inflation target while performing its short-run stabilization function?

An equation called the *Taylor rule* actually seems to describe very well how many central banks behave, and it illustrates how an inflation-targeting central bank could simultaneously pursue a medium-term inflation target while engaging in short-run output stabilization.[9] The rule describes how a central bank that operates a monetary policy regime that uses the interest rate as the key monetary policy instrument manages that instrument in response to the state of the economy. It can be specified as follows:

$$R = r_M + \pi^e + \beta_1(\pi^e - \pi^*) + \beta_2(Y/Y_P - 1)$$

where r_M is the medium-run equilibrium real interest rate, π^e is the expected inflation rate, π^* is the inflation target, Y is the actual level of real output, and Y_P is its full-employment level; β_1 and β_2 are both positive parameters. This rule sets the deviation of the nominal interest rate R from its medium-run equilibrium value of $r_M + \pi^e = r_M + \pi^*$ as an increasing function *both* of expected deviations of the inflation rate from its long-run target value $(\pi^e - \pi^*)$ and of proportional deviations of real gross domestic product (*GDP*) from its capacity level $(Y/Y_P - 1)$.

Notice how this rule works: if the economy is operating at full employment, (so $Y = Y_P$), and if the (actual and expected future) inflation rate is at its targeted level (so $\pi^e = \pi^*$), the bank sets the nominal interest rate at a value equal to the sum of the medium-run equilibrium real interest rate and its inflation target, that is, $R = r_M + \pi^*$. We can call this R_M and refer to it as the *medium-run equilibrium nominal interest rate.* However, if the economy is at full employment, but the inflation rate is expected to exceed its target, the bank sets the interest rate at a value higher than R_M (i.e., it adopts a more restrictive monetary policy), and that value is higher the larger the expected deviation of the inflation rate from its target. Finally, even if the inflation rate is equal to its targeted level, the central bank would still adopt an activist monetary policy (i.e., would set the interest rate at a level different from R_M) if output is not equal to its full-employment level. The interest rate exceeds R_M if output is *above* capacity (so $Y/Y_P > 1$) and falls short of R_M if output is below capacity $(Y/Y_P < 1)$.

You can see that this rule is compatible with IT because the central bank will always adopt an activist monetary policy to move the inflation rate toward its target, as long as it is not already there. It is also compatible with short-run output stabilization because the central bank always adjusts the interest rate in a countercyclical fashion when output deviates from its capacity level. In response to demand shocks, which cause the inflation rate and the actual level of output to move in the same direction, the direction of change in the policy interest rate is unambiguous because both the IT and output stabilization objective would drive the interest rate in the same direction. In response to supply shocks, which would cause the inflation rate and the rate of capacity utilization to move in opposite directions in the absence of a monetary response (e.g., an adverse supply shock would cause inflation to rise above the target and output to fall below its capacity level), the reaction of monetary

[9] The rule is named after Stanford economist John Taylor, who first formulated it to describe the behavior of the U.S. Fed.

policy depends both on the relative magnitudes of $(\pi^e - \pi^*)$ and $(Y/Y_P - 1)$ and on the relative sizes of β_1 and β_2. A central bank that cares relatively more about hitting its inflation target and less about stabilizing real output would have a large value of β_1 relative to β_2. In response to an adverse supply shock, for example, such a central bank would be likely to raise the interest rate, whereas one with the opposite preferences would be more likely to lower it.

Finally, notice that the width of the band around the inflation target is related to the parameters β_1 and β_2. A central bank that gives more weight to stabilizing real output rather than the inflation rate, because it would have a high value of β_2 relative to β_1, would want to accommodate this preference by choosing a relatively wide band for its inflation target. The reason is that even though it would eventually reach its target whether the band is wide or narrow, the priority it gives to stabilizing real output would tend to cause the inflation rate to deviate from its target by larger amounts than would be the case for a central bank facing the same shocks but with a smaller value of β_2 relative to β_1.

VI. SUMMARY

In the preceding chapter, we analyzed institutional mechanisms for enhancing the effectiveness of monetary policy in achieving the development objectives for which it is well suited. An independent central bank is one such mechanism. We saw that an independent central bank would be less likely to suffer from time-inconsistency problems that would result in high equilibrium inflation rates than would a central bank that answers directly to the finance ministry, while avoiding the rigidity that is associated with mechanisms such as currency boards or dollarization. But an independent central bank still faces the challenge of how to conduct monetary policy to strike an appropriate balance between credibility and flexibility.

In this chapter, we addressed how it could do so through the adoption of IT as a monetary policy framework. Under IT, the central bank publicly commits itself to a medium-run inflation outcome rather than to a specific path for one of the policy instruments that it can potentially control. But because the bank does not control inflation directly, it adopts specific procedures to enhance its transparency and thus enable the public to monitor its performance. Committing itself to a medium-run inflation target gives the central bank anti-inflationary credibility while preserving its flexibility to perform a short-run stabilizing function in the economy, as illustrated by Taylor rules for monetary policy. The evidence to date is encouraging regarding the effectiveness of IT in achieving its anti-inflationary objective, but the jury is still out on this issue, especially in the context of emerging and developing economies. Moreover, there is little direct evidence to date on the extent to which IT has preserved the flexibility of central banks in emerging and developing economies to conduct countercyclical monetary policies without sacrificing anti-inflationary credibility.

Central banks are charged with responsibility for the conduct not just of monetary policy but also of exchange rate policy. In the next part of the book, we turn our focus to exchange rate policy and consider how it interacts with the fiscal and monetary policy issues that we have discussed up to this point.

REVIEW QUESTIONS

1. What is a commitment mechanism, and why might one be needed in the context of monetary policy formulation?
2. What are the desirable attributes of a commitment device?
3. Compare the pros and cons of exchange rate and money supply announcements as commitment mechanisms for monetary policy.
4. Why has IT tended to replace fixed exchange rates and monetary rules as commitment device?
5. What are the components of IT, and how do they address time-inconsistency problems?

EXERCISES

1. Compare the advantages and disadvantages of IT and money growth rules as alternative frameworks for monetary policy. How have central banks attempted to overcome the disadvantages of IT in practice?
2. In some countries that maintain an inflation target, the salary of the governor of the central bank is made dependent on how close the economy's inflation rate comes to the target announced by the central bank. Using the tools we developed in this chapter, explain how such an arrangement is likely to affect an economy's inflation outcome.
3. The U.S. Federal Reserve System does not pursue an inflation target. In the last six months of 2009, responding to the credit crisis that emerged in the United States in the summer of that year, it more than doubled the monetary base, yet there were no signs at that time that expectations of future inflation increased dramatically in that country. How can this be?
4. Critics of IT often claim that IT causes the central bank to give too much priority to inflation, in preference to stabilizing output around its full-employment level. Do you agree or disagree? Explain.
5. What do you think would be the relevance of the complexity of the monetary transmission mechanism to a central bank's decision regarding whether to adopt IT?

REFERENCES AND FURTHER READING

Bernanke, Ben S. (2005), "Inflation in Latin America: A New Era?" remarks at the Stanford Policy Institute for Economic Research Economic Summit, February 11, 2005.

Bernanke, Ben S., and Frederic S. Mishkin (1997), "Inflation Targeting: A New Framework for Monetary Policy?" *Journal of Economic Perspectives*, vol. 11, pp. 97–116.

Bernanke, Ben S., and Michael Woodford (2005), *The Inflation Targeting Debate*, Chicago: University of Chicago Press.

Bernanke, Ben S., Thomas Laubach, Frederic S. Mishkin, and Adam S. Posen, eds. (1999), *Inflation Targeting: Lessons from the International Experience*, Princeton, NJ: Princeton University Press.

Carare, Alina, Andrea Schaechter, Mark Stone, and Mark Zelmer (2002), "Establishing Initial Conditions in Support of Inflation Targeting," working paper WP/02/102, International Monetary Fund.

Cecchetti, Stephen G., and Stefan Krause (2002), "Central Bank Structure, Policy Efficiency, and Macroeconomic Performance: Exploring Empirical Relationships," *Federal Reserve Bank of St. Louis Review* (July–August), pp. 47–59.

Fraga, Arminio, Ilan Goldfajn, and Andre Minella (2003), "Inflation Targeting in Emerging Market Economies," working paper 10019, National Bureau of Economic Research.

Fuhrer, Jeffrey C. (1997), "Central Bank Independence and Inflation Targeting: Monetary Policy Paradigms for the Next Millennium?" *Federal Reserve Bank of Boston New England Economic Review* (January–February), pp. 19–36.

Ho, Corinne, and Robert McCauley (2003), "Living with Flexible Exchange Rates: Issues and Recent Experience in Inflation Targeting Emerging Market Economies," working paper 130, Bank for International Settlements.

Kahn, George A., and Klara Parrish (1998), "Conducting Monetary Policy with Inflation Targets," *Federal Reserve Bank of Kansas City Economic Review* (Third Quarter), pp. 5–30.

Newmann, Manfred J. M., and Jurgen von Hagen (2002), "Does Inflation Targeting Matter?" *Federal Reserve Bank of St. Louis Review* (July–August), pp. 127–148.

Truman, Edwin M. (2003), *Inflation Targeting in the World Economy*, Washington, DC: Institute for International Economics.

PART 5

Exchange Rate Management

16

Equilibrium Real Exchange Rates

At various places throughout this book, we have discussed the role of the real exchange rate as a key macroeconomic relative price. As the relative price of foreign goods (goods produced abroad) in terms of domestic goods (goods produced at home), the real exchange rate plays an important role in guiding the broad allocation of production and spending in the domestic economy between these two types of goods. Because of this important allocative role of the real exchange rate, emerging economies are often encouraged to conduct their affairs to get this particular macroeconomic relative price "right" – that is, to make sure that the economy's actual real exchange rate does not stray too far from its equilibrium value, a situation that is known as exchange rate *misalignment.*

Why is the avoidance of misalignment so important? As we will discuss more fully in this chapter, the key reasons are two: when the exchange rate is misaligned, it will not provide the appropriate signal to guide the allocation of resources between domestic and foreign goods. In addition, as mentioned in the previous chapter, when the real exchange rate is perceived to have become severely misaligned, the expectation will be created that it will adjust toward its equilibrium value in the future. To the extent that this adjustment is expected to take place through movements in nominal exchange rates, this will discourage domestic agents from holding assets denominated in domestic currency, which is a potential source of capital-flow reversals and exchange rate crises. Thus large real exchange rate misalignments can have important microeconomic and macroeconomic costs.

But how do we know whether the real exchange rate is misaligned, and how do we avoid misalignment? Part 5 takes up these issues, in the context of a more general discussion of *exchange rate management.* Notice that because the actual real exchange rate can be observed, the detection and measurement of real exchange rate misalignment depends on the ability to define and estimate an *equilibrium* real exchange rate. Thus a logical place to begin a discussion of exchange rate management is to consider what we mean by the notion of an equilibrium real exchange rate, and how we might go about measuring it. This chapter will explore the theoretical determinants of the medium-run equilibrium real exchange rate and consider different approaches to its empirical measurement.

It is worth noting at the outset that this is a controversial area of research in macroeconomics. Objections to the enterprise of defining and measuring an equilibrium real exchange rate tend to take one of three forms:

1. Some economists argue that it is meaningless to distinguish between the actual real exchange rate and the equilibrium real exchange rate because any observed real exchange rate must be an equilibrium one.
2. Others would say that even if it is meaningful to make such a distinction, it is useless to do so, because gaps between actual and equilibrium real exchange rates have no policy implications.
3. Finally, a third view holds that even if it is possible to make the distinction and is useful to do so, measuring the equilibrium real exchange rate is impossible.

We will address each of these doubts during the course of this chapter.

I. THE REAL EXCHANGE RATE: DEFINITION AND MEASUREMENT

The problem of defining and measuring the equilibrium real exchange rate is complicated by a variety of factors. The most basic one is that no single definition of the real exchange rate is widely accepted among economists. Recall that in broad terms, the real exchange rate is simply the relative price of foreign goods in terms of domestic ones. But what constitutes domestic and foreign goods depends on the particular analytical framework – the specific macroeconomic model – being used. Because different types of macro models are used by economists for different purposes, a variety of analytical real exchange rate definitions tend to be used, as we will see. A second problem is that though the notions of what are domestic and foreign goods may be relatively easy to pin down within specific theoretical models (most of which contain at most three goods), in the real world, things tend to become more complicated. In particular, in a context in which there are many goods, it is not always obvious how to come up with empirical counterparts to the prices of domestic and foreign goods that we are trying to measure. Different approximations result in different empirical definitions, even when the theoretical concept is unambiguous.

1. Analytical Issues

The key factor that affects the definition of the real exchange rate in analytical models is the assumed production structure of the model. Some of the most widely used modeling frameworks are the following.

a. One (Tradable) Good Model

Some models contain a single good that is assumed to be internationally traded, and arbitrage is assumed to equalize its price everywhere. Such models are useful for the analysis of purely monetary phenomena (such as inflation as well as certain approaches to the explanation of the determinants of the balance of payments). Obviously, there can be no real exchange rate in such models because with only a single good, there is no distinction to be drawn between domestic and foreign goods.

b. Complete Specialization (Mundell-Fleming) Models

An alternative framework assumes that the domestic economy and the rest of the world are each specialized in the production of a single good and that these goods, which are traded internationally, are imperfect substitutes for each other. This is the framework that we used in the model of Part 2 of this book. It is applicable to countries whose trade consists largely of manufactured goods because these tend to be imperfect substitutes for what the rest of the world produces.

In this context, it is easy to define the real exchange rate. As already noted, it is the number of units of the domestically produced good that have to be given up for each unit of the foreign good. As we saw in Part 2, in the context of our analytical model, the role of the real exchange rate was to determine the composition of absorption (both domestic and foreign) between goods produced at home and those produced abroad. As such, the real exchange rate determined the aggregate demand for the domestic good and was also an important determinant of the country's net exports.

In Mundell-Fleming models, the real exchange rate happens to coincide with the (inverse of the) country's *terms of trade* (the price of exports relative to the price of imports). This is an artifact of the assumption of complete specialization in production, however, and these two concepts are in general quite different from each other in analytical frameworks that do not make this assumption, as we will see later.

c. Dependent-Economy (Swan-Salter) Models

An alternative framework also has a production structure that contains two goods, but in this case, one is produced at home and consumed only at home (the *nontraded* good), whereas the other is produced and consumed both at home and abroad. Because only the latter can be bought and sold across international boundaries, it is the *traded*, or foreign, good. In this case, the definition of the real exchange rate is straightforward – it is the number of units of the nontraded good required to purchase one unit of the traded good:

$$Q = P_T/P_N$$

where P_T and P_N are, respectively, the domestic-currency prices of the traded and nontraded goods. This is sometimes called the *internal* real exchange rate. Notice that because there is only one type of foreign good in this framework, there are no terms of trade in this model.[1] This model is useful to analyze issues for which the role of exogenous changes in the terms of trade are not important, in the context of economies whose terms of trade are exogenous. For example, the model is widely used to analyze the effects of domestic macroeconomic policies in small countries.

d. Three-Good (Exportable-Importable-Nontraded) Model

If changes in the terms of trade do matter, however, then a three-good model is required, consisting of *exportable* and *importable* goods (both of which may be produced and consumed at home, but one of which is exported and the other imported) as well as nontraded goods. But in this case, there are two foreign goods and therefore *two* real exchange rates as well as a separate and distinct definition of the terms of trade. Letting P_X denote the domestic-currency price of the exportable good and P_{IM} the domestic-currency price of the importable good, we have the *exportables real exchange rate* $Q_X = P_X/P_N$ and the *importables real exchange rate* $Q_{IM} = P_{IM}/P_N$. In turn, the terms of trade (TOT) are defined as $TOT = P_X/P_{IM}$.[2] This model is useful for analyzing the macroeconomic effects of terms of trade changes as well as of changes in commercial policies that affect the domestic relative prices of exportables and importables.

2. Measurement Issues

Given the analytical framework that is suitable for the problem at hand, the next issue in empirical applications is how to translate the relevant concept into an empirical measure of the real exchange rate. The most common way to measure the real exchange rate empirically is to take a foreign price index, express it in domestic-currency terms (by multiplying it by a nominal exchange rate index), and then divide it by a domestic price index. But this still leaves some problems, for example:

- What do we mean by a foreign price index when there are many foreign countries?
- If there is more than one exchange rate against some foreign currency (i.e., in the case of parallel exchange rates), which one do we use?

[1] Another, more general way of thinking about this model is that there is more than one type of foreign good, but because the home country is small, it cannot affect the relative prices among them. In particular, it cannot affect the relative price between the foreign goods that it tends to export and the foreign goods that it tends to import, i.e., its terms of trade.

[2] In this case, the terms of trade are expressed as the number of importable goods that can be bought with one exportable good, which is the inverse (foreign goods per domestic good) of the way the real exchange rate was defined (domestic goods per foreign good).

- Do we use the same price index for the foreign and the domestic countries? If not, why not?
- Which price index do we use (e.g., consumer prices, producer prices, gross domestic product (*GDP*) deflator, etc.)?

The most common choice is to use the trade-weighted consumer price index in partner countries (measured in domestic currency) for the foreign price index and the consumer price index for the domestic price index. This is called the *real effective exchange rate* (*REER*), or sometimes the *multilateral real exchange rate.* This has several problems, however:

1. Because the denominator is a weighted average of domestic traded and nontraded goods prices, this indicator will only show changes in the internal real exchange rate in muted form (e.g., if the price of nontraded goods changes by x percent, the denominator will change in the same direction, but by less than x percent).
2. Because the numerator contains foreign traded and nontraded goods prices, the information it contains about the incentives facing domestic agents to consume and produce traded goods or nontraded goods (captured by the internal real exchange rate) will be contaminated by changes in the foreign internal real exchange rate, which is not relevant for domestic agents. For this reason, some economists prefer to use the trade-weighted producer price index for partner countries, on the grounds that it contains a larger share of traded goods.
3. Finally, because the numerator cannot pick up the effects of changes in commercial policies on the domestic relative prices of traded and nontraded goods, when such policies change, this measure will fail to indicate the appropriate changes in the incentives facing domestic agents.

II. THE MEDIUM-RUN EQUILIBRIUM REAL EXCHANGE RATE: CONCEPTUAL ISSUES

Notice that, whether a country operates with a fixed or floating *nominal* exchange rate, the real exchange rate is an endogenous variable.[3] If we adopt the dependent-economy framework, for example, we can define the real exchange rate as follows:

$$Q = P_T/P_N = SP_T^*/P_N$$

Under the floating exchange rate regime, the nominal exchange rate S is endogenous. But even under fixed rates, P_N is an endogenous variable. In either case, then, Q is

[3] We have already seen this to be so in the Mundell-Fleming fixed-exchange rate model of Part 2, in which the real exchange rate was endogenous both in the short run and in the medium term.

endogenous, and as such, it must be determined as the outcome of the economy's macroeconomic equilibrium.

This observation has caused some economists to question the very notion of distinguishing between the actual real exchange rate and its notional equilibrium value, but this view, which was the first objection to defining and measuring the equilibrium real exchange rate mentioned in the introduction to this chapter, is fundamentally misguided. The distinction between the actual real exchange rate and its equilibrium value is not one between disequilibrium and equilibrium but rather between different types of equilibria (i.e., equilibria conditioned on different values of macroeconomic variables).

1. Defining Equilibrium

The traditional definition of the equilibrium real exchange rate is that it is the value of the real exchange rate that is simultaneously consistent with internal and external balance, conditioned on *sustainable* values of exogenous and policy variables.[4] The term *internal balance* in this definition refers to a situation in which the markets for domestic goods and labor are both in equilibrium. Thus it corresponds to what we referred to in Chapter 7 as a short-run macroeconomic equilibrium with full employment. External balance, on the other hand, refers to a situation in which the economy's current account deficit is equal to the value of the *sustainable capital inflows* that it can expect to receive. So when we refer to the equilibrium real exchange rate, we are not referring to the real exchange rate that falls out of just any arbitrary macroeconomic equilibrium but rather to an equilibrium that is *sustainable*.

What does it take for a macroeconomic equilibrium to be sustainable? To answer this question, we need to formalize the dynamic structure of an economy. At any moment in time, we can think of the economy's endogenous variables as being determined by three types of variables: *predetermined variables*, *exogenous policy variables*, and *other exogenous variables*. Predetermined variables, which we discussed previously in Part 2, are endogenous variables that change slowly over time. In the model of Part 2, this referred to the economy's capital stock and technology as well as its international investment position (*IIP*). From a Keynesian perspective, in which the nominal wage adjusts only gradually over time, it would also refer to the nominal wage. Exogenous policy variables include fiscal and monetary policies, trade policies, and other variables under the control of the domestic authorities. Other exogenous variables can usefully be classified into three types: observable variables such as the weather, the terms of trade, world interest rates, and so on; unobservable variables that can be treated as random shocks; and so-called *bubble variables*, that is, variables that affect the economy only through their influence on expectations.

[4] This definition is from Nurkse (1945).

Because the real exchange rate is an endogenous variable, we can express it as determined by the reduced-form relationship

$$RER(t) = F[X_1(t), \boldsymbol{X}_2(\boldsymbol{t}), \boldsymbol{X}_3(\boldsymbol{t}), B(t)] \tag{16.1}$$

where X_1 represents the current values of a set of predetermined variables, $\boldsymbol{X}_2$ represents the current *and expected future* values of a set of real policy variables, $\boldsymbol{X}_3'$ corresponds to the current and expected future values of a set of exogenous (observable and unobservable) variables, and B includes any bubble variables that may affect the economy.[5]

At the same time that the economy determines the values of endogenous variables such as the real exchange rate, it also determines the values of all other endogenous variables. These include the *rates of change* of the predetermined variables:

$$\dot{X}_1(t) = G[X_1(t), \boldsymbol{X}_2(\boldsymbol{t}), \boldsymbol{X}_3(\boldsymbol{t}), B\,(t)] \tag{16.2}$$

Notice that the actual real exchange rate observed at any moment may thus be influenced by speculative bubble factors, by the *actual* values of predetermined variables, and by transitory values of policy and (observable and unobservable) exogenous variables. When at least some of the variables on which the actual equilibrium real exchange rate depends are unsustainable, the actual real exchange rate will tend to change over time. It is possible, then, to think of alternative equilibrium real exchange rates, where the notion of equilibrium is defined over different time horizons.

a. The Short-Run Fundamental Real Exchange Rate

Because speculative factors are transitory and are generally likely to be short-lived, we can derive a concept that we can call the short-run equilibrium real exchange rate (*SRER*) by setting such factors to zero:

$$SRER(t) = F(X_1(t), \boldsymbol{X}_2(\boldsymbol{t}), \boldsymbol{X}_3(\boldsymbol{t}), 0) \tag{16.3}$$

This equilibrium concept is short run in the sense that it is conditioned on the *current* values of X_1, $\boldsymbol{X}_2$, and $\boldsymbol{X}_3$, which can be referred to as the *short-run fundamentals.*

b. The Long-Run Equilibrium Real Exchange Rate

The short-run equilibrium real exchange rate will itself not be sustainable, for two reasons:

1. The policy and exogenous variables that affect it can deviate from their sustainable, or permanent, values. Call these permanent values $\boldsymbol{X}_2^*$ and $\boldsymbol{X}_3^*$. Thus

[5] X_1 does not necessarily contain *all* policy variables. E.g., in an economy operating with a fixed exchange rate, it would not contain the nominal exchange rate. Instead, the value of the nominal exchange rate affects the value of the predetermined variables.

SRER can be expected to change when the policy and exogenous variables are themselves expected to change.

2. Even if the policy and exogenous variables are at sustainable levels, the predetermined variables may not have completed their adjustments to sustainable positions. Endogenous changes in the predetermined variables would then cause the short-run real exchange rate to keep moving, even with no further changes in the policy and predetermined variables. The predetermined variables will stop moving when they reach a *steady state*, that is, when they reach a value that satisfies

$$0 = G(X_1, \boldsymbol{X}_2^*, \boldsymbol{X}_3^*)$$

We can solve this equation for the long-run (steady state) values of X_1 as follows:

$$X_1^* = X_1(\boldsymbol{X}_2^*, \boldsymbol{X}_3^*) \tag{16.4}$$

Substituting these values of X_1 back into equation (16.1), we have the following:

$$LRER = F[X_1(\boldsymbol{X}_2^*, \boldsymbol{X}_3^*), \boldsymbol{X}_2^*, X^*] = H(\boldsymbol{X}_2^*, \boldsymbol{X}_3^*) \tag{16.5}$$

The *long-run equilibrium real exchange rate* (*LRER*) depends only on the *sustainable* values of the exogenous and policy variables that affect the real exchange rate directly or indirectly (through X_1). These are called the *long-run fundamentals*.

c. The Desired Equilibrium Real Exchange Rate

But the policy variables in $\boldsymbol{X}_2^*$, though sustainable, may not be desirable. Thus $\boldsymbol{X}_2^*$ itself may be changing over time, as policies are adjusted in the desired direction. Suppose we call the optimal levels of the policy variables $\boldsymbol{X}_2^{**}$. In general, $\boldsymbol{X}_2^{**}$ will depend on the values of the exogenous variables:

$$\boldsymbol{X}_2^{**} = J(\boldsymbol{X}_3^*)$$

Substituting this into equation (16.5), we have

$$DRER = H[J(\boldsymbol{X}_3^*), \boldsymbol{X}_3^*] \tag{16.6}$$

We can call this the *desired equilibrium real exchange rate.*

2. Misalignment

Corresponding to each definition of the equilibrium real exchange rate offered earlier (i.e., *SRER*, *LRER*, and *DRER*) is a corresponding measure of misalignment – defined as the difference between the actual real exchange rate and its equilibrium value. Thus we can define the following alternative concepts of misalignment:

1. *Short-run misalignment:*

$$RER - SRER$$

2. *Long-run misalignment:*

$$RER - LRER = (RER - SRER) + (SRER - LRER)$$

3. *Ultra-long-run misalignment:*

$$RER - DRER = (RER - SRER) + (SRER - LRER) + (LRER - DRER)$$

In floating exchange rate industrial country applications, the term misalignment usually refers to short-run misalignment. Long-run misalignment, on the other hand, is more often what is meant in fixed exchange rate applications, when reserves and capital flows pick up speculative factors. Ultra-long-run misalignment is what is often at issue in the context of structural reform.

Thus, depending on the definition employed, misalignment can arise from any of the following factors:

1. bubble factors (accounting for the difference between *RER* and *SRER*)
2. transitory policy shocks and slow adjustment of predetermined variables (accounting for the difference between *SRER* and *LRER*)
3. inappropriate policies (accounting for the gap between *LRER* and *DRER*)

In the rest of this chapter, we will focus on long-run misalignment. There are two reasons for not emphasizing instead the more comprehensive concept of ultra-long-run misalignment. A practical reason is that the detection of ultra-long-run misalignment requires using an optimal policy framework as a benchmark – this is certainly difficult to do in general – but a more fundamental reason is that misalignment may be costly even if policies are optimal. Thus long-run misalignment involves two types of costs: the costs of misalignment itself and the costs of suboptimal policies.

The claim that misalignment itself may be costly, even if policies are set at their optimal levels, is a controversial one. Some economists believe that real exchange rates are always appropriate, conditional on the fundamentals, and consequently, that so-called inappropriate exchange rates reflect unsustainable or otherwise undesirable macroeconomic policies. This was the second objection to the enterprise of defining and measuring equilibrium real exchange rates mentioned in the introduction to this chapter. This conclusion is implied by the view that deviations of the short-run real exchange rate from the long-run equilibrium real exchange rate are the (short-run) equilibrium outcomes of the optimizing behavior of agents operating in an undistorted environment. If so, then such deviations are optimal and cannot be improved on by exchange rate policy.

The statement thus concerns the role of nominal rigidities in generating temporary deviations of the real exchange rate from its long-run equilibrium value and reflects the view that nominal rigidities do not help to account for such deviations. This amounts to an application to the exchange rate arena of the broader question of the role of nominal rigidities in explaining macroeconomic fluctuations.

The maintained view among those who consider the estimation of equilibrium real exchange rates a worthwhile endeavor is that nominal rigidities have an important role to play in explaining such fluctuations, and consequently, that temporary deviations of the real exchange rate from its long-run equilibrium value do indeed have implications for nominal exchange rate policies. Specifically, the presence of nominal rigidities implies that nominal exchange rate adjustments may often be useful in helping to guide the real exchange rate back to its long-run equilibrium value from an initial position of disequilibrium. The mainstream perspective, therefore, is that long-run misalignment has costs, and the costs are greater the greater the extent of financial integration.

What are the costs of misalignment in the presence of nominal rigidities? First, for any given episode of misalignment, because the elimination of misalignment in the absence of nominal exchange rate adjustments would require changes in the domestic price level, it would tend to be associated with macroeconomic instability, in the form of a period of below-average growth (i.e., a recession) or above-average inflation, even in the absence of capital mobility. With an open capital account and high capital mobility, on the other hand, these macroeconomically harmful events may be magnified through capital outflows and inflows. Thus, with substantial capital mobility, the emergence of substantial misalignment may be associated with very severe macroeconomic instability. The evidence indeed suggests both that real exchange rate misalignment helps to explain the incidence of exchange rate crises and that such crises can have substantial real output costs (see Chapter 27).

Thus a large real exchange rate misalignment may prove very costly to reverse. But repeated smaller episodes of misalignment may also prove to be costly. As argued in Chapter 3, for example, repeated minor episodes of misalignment would tend to increase the noise component in real exchange rate movements, thereby undermining the role of the real exchange rate as a relative price signal. This would have harmful effects on growth, as we have seen in previous chapters, by impairing the efficiency of resource allocation as well as by discouraging capital accumulation. The evidence from the cross-country growth literature (e.g., Corbo and Rojas 1995) indeed suggests that real exchange rate variability is negatively related to growth. Though this does not necessarily mean that it is real exchange rate misalignment itself that has this effect (because such variability could reflect variability in the *equilibrium* real exchange rate), there is more direct evidence that this is so (Razin and Collins 1997).

Do we have evidence on the role of nominal rigidities in generating misalignment? The answer is not much, and what we have is indirect. For example, real exchange rate variability tends to increase sharply when countries switch from fixed to floating exchange rates and to decrease when they revert from floating exchange rates back to fixed (see, e.g., Mussa 1986). In addition, large episodes of overvaluation tend to end in devaluation rather than price-level adjustments, suggesting that the latter are perceived as more costly by the countries involved (Dornbusch et al. 1995). Finally, the evidence that large estimated overvaluation of the real exchange rate tends to be associated with slower growth (as in the paper by Razin and Collins cited previously) is consistent with the effects of nominal rigidities.

III. HOW LONG IS THE LONG RUN?

A tricky issue in the definition of the long-run equilibrium real exchange rate is what we mean by the *long run*. In terms of the formal definition, the answer seems obvious – it is a situation when all the variables in X_1 have come to rest. But on closer inspection, this definition does not seem to be very useful. The problem is that, as we saw in Part 2, variables such as the nominal wage, the country's net international creditor position, and sectoral capital stocks may approach their long-run values at very different speeds, and some of these may be very slow. Indeed, that is how we distinguished between the short run, the medium run, and the long run in Part 2. In the short run, the nominal wage had adjusted fully, but the country's international investment position and its capital stock had not; in the medium run, the nominal wage and international investment position had fully adjusted, but the capital stock had not; and in the long run, all three variables had completed their adjustments.

Should we require all these to reach a stationary position in our definition of the long run, as the preceding exposition would seem to suggest? The problem is that if we did, the *LRER* concept derived may be of little operational usefulness because it may be approached too slowly.

For guidance on this problem, we can consult the common definition of the equilibrium real exchange rate as the value of the real exchange rate that is simultaneously consistent with internal and external balance, conditioned on sustainable values of exogenous and policy variables. If this is the definition that has proven most useful for policy guidance, perhaps we can use it to help us choose where to draw the line on what meaning of *long run* is most relevant for policy purposes.

What do these have to do with the long-run equilibrium perspective adopted earlier, and how do we use them to determine the relevant meaning of the term *long run*? Recall that internal balance refers to a situation in which the markets for domestic goods and labor are both in equilibrium. Thus it corresponds to short-run macroeconomic equilibrium (full employment). External balance, on the other hand, refers to a situation in which the current account deficit is equal to the value of the sustainable capital inflow. Thus, because it supposes a situation of full employment, internal balance requires that cyclical adjustment mechanisms operating through the labor market cease to be operative. At the same time, this definition appears to reflect the view that allowing for full capital stock adjustment is too much; that is, the span of time required would be too long to be of much policy relevance.[6] What this means is that the operationally-relevant definition of "long run" in this context is not the steady-state equilibrium of growth theory, but rather what we referred to as a *medium-run* equilibrium in Chapter 8.

On the other hand, the traditional definition unfortunately does not help us know what to do with the country's international investment position. The reason is that the term "sustainable capital inflows" is somewhat ambiguous. It can be interpreted as consistent with taking the country's international investment position as given and thus treating it as one of the fundamentals, or with requiring it to have completed its adjustment. For example, if sustainable capital inflows are interpreted as those required to keep the country's international investment position at its medium-run equilibrium level, then the country's international investment position should be treated symmetrically with the nominal wage. On the other hand, if sustainable capital flows are those not driven by speculative or cyclical factors, then external balance does not necessarily impose the requirement that the economy's international investment position be unchanging. In that case, the economy's international investment position would be taken as given in the definition of the medium-run equilibrium real exchange rate. It would be treated as a fundamental, symmetrically with the capital stock, and net capital inflows would just be another endogenous variable.

The key point is that the asymmetric adjustment speeds among the economy's several predetermined variables imply that there is *no unique equilibrium real exchange rate*. The relevant conception of what we mean by *long run*, and therefore the relevant equilibrium concept, depends on the policy issue that motivates our concern. If the policy concern is with the effect of the real exchange rate on the level of economic activity in a sticky-wage context, for example, then the operationally relevant definition of the equilibrium real exchange rate is the one that would be consistent with internal balance only, because deviations of the actual real exchange rate from that equilibrium value will give rise to expansionary or contractionary pressures in the goods market. In this case, the equilibrium real exchange rate is conditioned on the economy's international investment position, which is treated as a fundamental. We can refer to this as the "classical" notion of equilibrium because it pertains to a situation of classical full employment. On the other hand, if the policy concern is with competitiveness – that is, with the evolution of the economy's international investment position – then the relevant concept of equilibrium is one for which *IIP* has also reached a sustainable value. This is the medium-run equilibrium of Chapter 8, which we can refer to as the "Nurksian" notion of equilibrium. This distinction will be illustrated in the section that follows.

[6] Note that in such a long run, there would be ongoing investment (a flow), and thus ongoing changes in the size of the capital stock. Because changes in the capital stock would in principle alter the long-run equilibrium real exchange rate under our definition, the implicit assumption is that ongoing investment flows have a negligible effect on the size of the capital stock, as in the short-run model of Part 2.

IV. A MODEL OF THE MEDIUM-RUN EQUILIBRIUM REAL EXCHANGE RATE

To illustrate these issues, as well as to investigate the determinants of the equilibrium real exchange rate, we can make use of our medium-run model in Chapter 8. That model consisted of the following components:

1. goods-market equilibrium condition (equation (8.4a)):

$$Y_P = \phi(Q, \ldots)A_P[Y_P - T, r, \ldots] + G + X(Q, \ldots) \tag{16.7}$$

2. bond-market equilibrium condition (equation (8.5b)):

$$\begin{aligned}(B - B_C)/S = {} & [b(r - r^*) + b^*(r - r^*)]IIP^* \\ & + b(r - r^*)[B/S - Q^{-1}L(r + \pi, Y_P)] \\ & + b^*(r - r^*)(F^* - L^*)\end{aligned} \tag{16.8}$$

3. external balance condition (equation (8.10a)):

$$\begin{aligned}0 = {} & [Q^{-1}NX(Q, r, \ldots) + r^* IIP^* \\ & -(r + \pi - r^*)b^*(r + \pi - r^*)(F^* - IIP^* - L^*)\end{aligned} \tag{16.9}$$

where $\pi = \hat{S}$ has been set as the common rate of currency depreciation and credit expansion (you can verify that setting $\pi = \hat{S}$ in equation 8.5b yields equation 16.8). Recall also from Chapter 8 that by solving equation (16.7) for the domestic real interest r and substituting into equation (16.8), the two equations could be combined into the short-run equilibrium relationship (see equation (8.11)):

$$Q = S(IIP^*; Y_P, G, T, \pi, r^*, F^* - L^*) \tag{16.10}$$

We showed in Chapter 8 that this relationship can be plotted as a curve with a negative slope in (IIP^*, Q) space, which we referred to there as the IB (for internal balance) curve. Notice that the equilibrium value of the real exchange rate Q required for internal balance to hold (the "classical" definition of equilibrium) depends not just on the economy's international investment position IIP^* and policy-determined rate of currency depreciation π but also on the exogenous variables Y_P, G, T, r^*, and $F^* - L^*$. The variables to the right of the semicolon in equation (16.10) therefore represent the set of *real exchange rate fundamentals* in the model of Chapter 8. The equilibrium real exchange rate conditioned on all these fundamentals, including the arbitrary initial value of IIP^*, denoted IIP_0^* in the figure, is found at the point on the internal balance curve above IIP_0^*, labeled B in the figure. The classical equilibrium real exchange rate is labeled Q_C.

The external balance condition (16.3) is depicted in Figure 16.1 as a curve that also has a negative slope but is flatter than the internal balance curve (see Chapter 8). The "Nurksian" equilibrium holds at the point of intersection of the IB and EB curves, labeled A in the figure, and the corresponding equilibrium real exchange

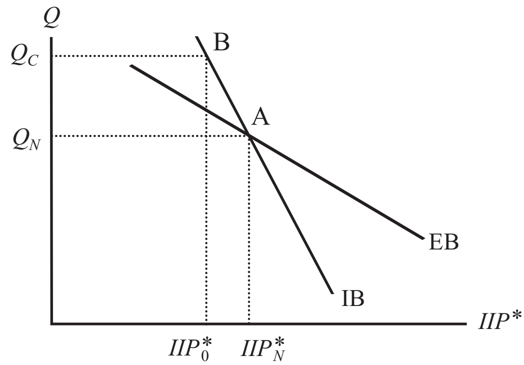


Figure 16.1. **Classical and Nurksian equilibrium real exchange rates**

rate is Q_N. Notice that under this definition, the medium-run equilibrium real exchange rate and international investment position are determined jointly, and IIP^* is therefore not in the set of fundamentals for the equilibrium real exchange rate.

Another important difference between the two equilibrium concepts is that the effects of the remaining fundamentals on the equilibrium real exchange rate will in general be different. Consider, for example, the effect on the equilibrium real exchange rate of a permanent increase in government spending. Suppose the economy starts at the medium-run equilibrium point A in Figure 16.2. Holding the real exchange rate Q constant in equation (16.7), an increase in G requires an increase in the domestic real interest rate r to clear the goods market. Because an increase in r causes the demand for domestic bonds to increase at any given value of Q in equation (16.8), it requires a *reduction* in IIP^* to sustain equilibrium in the domestic bond market. Thus the internal balance curve shifts to the *left* in Figure 16.2, from a position such as IB_0 to one such as IB_1. Under the classical definition of equilibrium, the equilibrium real exchange rate moves from Q_0 to Q_c, the point on the new IB

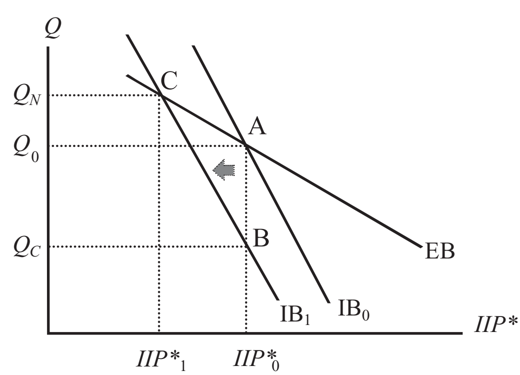


Figure 16.2. **Effects of a permanent increase in government spending**

curve that lies directly above the original international investment position IIP_0^*. Because G does not affect the EB curve, it remains stationary in response to this shock. Consequently, when adjustment in IIP^* is allowed for, as under the Nurksian definition of equilibrium, the new equilibrium is at C. Notice that the equilibrium real exchange rate *appreciates* in response to this shock under the classical definition but *depreciates* under the Nurksian one. The key point is that both the specific combination of variables that comprise the set of fundamentals and the direction and magnitude of the effects of those fundamentals on the equilibrium real exchange rate depend on the definition of the real exchange rate that is relevant for the issue at hand.

V. THE EQUILIBRIUM REAL EXCHANGE RATE: ESTIMATION

How has the equilibrium real exchange rate been estimated in practice? In this section, we will consider four alternative approaches to estimation: the purchasing-power parity (PPP) approach, the trade elasticities approach, the use of general-equilibrium model simulations, and estimation of single-equation reduced-form models for the real exchange rate. All these methods have been widely used by economists to estimate the equilibrium real exchange rate.[7]

1. The Purchasing-Power Parity Approach

The most venerable approach to the estimation of the equilibrium real exchange rate is based on the PPP hypothesis. In brief, this hypothesis states that the real exchange rate is constant in the long run. Thus, using the notation of the previous section, the long-run equilibrium real exchange rate can be written as

$$Q^* = Q_0$$

where Q_0 is a constant, and the actual real exchange rate Q_t that prevails at any moment t is given by

$$\log(Q_t) = \log(Q_0) + \varepsilon_t \tag{16.11}$$

where ε_t is a mean-zero *stationary* random variable.[8] This would suggest estimating the equilibrium real exchange rate in one of two ways: either as the value of the real exchange rate in some past period when ε_t was known to be zero, or as the average value of the real exchange rate over some number of past periods, if it is not known when ε_t may have been equal to zero. The assumption here is that because the equilibrium real exchange rate Q^* is taken to be unchanged and the *actual* real exchange rate is distributed randomly around it, we can use the *past* values of the

[7] For an overview, see Hinkle and Montiel (1999).

[8] A random variable is said to be stationary if the parameters that determine its statistical distribution are not a function of time.

real exchange rate to estimate the *current* value of the equilibrium real exchange rate.

Is this a valid procedure? The answer is that it is valid only if the real exchange rate is *mean reverting*, that is, only if ε_t is indeed stationary. Otherwise, we have no reason to expect the real exchange rate to return to its mean value. How do we know whether this is so? The appropriate procedure is to apply statistical tests for stationarity to the observed real exchange series before using this method. Notice that, in terms of the theory of the previous section, for the PPP hypothesis to be valid, the relevant set of long-run fundamentals that we identified previously must on average have remained unchanged over the sample period; that is, the fundamentals must not themselves have undergone permanent changes. Whether this is so for any given country over any given period of time is a purely empirical matter. Many stationarity tests have been applied to real exchange rates of emerging economies, and the vast majority of them tend to reject PPP.[9]

2. The Trade Elasticities Approach

The most common approach to the estimation of the equilibrium real exchange rate that is capable of taking account of permanent changes in the fundamentals is the trade elasticities approach. We can describe it as follows. Going back to the model of Part 2, recall that the country's net exports could be expressed as $N(Q, Y, \dots)$. Assuming that external effects on net exports depend on partner-country incomes Y^*, we can write the external balance condition in that model in the following form:

$$-RB = N(Q,\, Y_P,\, Y_P^*) \tag{16.12}$$

where RB is the sustainable level of the resource balance (net capital inflows minus net interest payments on external debt), and Y_P and Y_p^* are the full-employment levels of domestic and foreign real *GDP*. We can use this equation to estimate the equilibrium real exchange rate by following the following sequence of steps:

1. First, we need to estimate the level of domestic and foreign real *GDP* at full employment and the sustainable level of the resource balance.
2. Next, we require estimates of the parameters of the function $N(\ )$, including those that determine how net exports respond to the real exchange rate, that is, the trade elasticities involved in the functions $\varphi(\ )$ and $X(\ )$ of Chapter 5.
3. With an empirical estimate of the function $N(\ )$ in hand, estimating the equilibrium real exchange rate amounts to substituting into equation (16.12) the estimated full-employment values of Y and Y^* as well as the estimate of the sustainable resource balance. The equilibrium real exchange rate is the value of Q that solves this equation.

[9] For references, see Hinkle and Montiel (1999).

This method has several advantages. Notice that, unlike the PPP approach, the equilibrium real exchange rate estimated using this method is not necessarily constant. It is affected by the fundamentals Y_P, Y_P^*, and RB. Moreover, this approach is structural, so it has the advantage that the determination of the long-run equilibrium real exchange rate can be easily understood. Moreover, it makes use of a small set of behavioral parameters that are widely estimated and thus readily available, making it relatively simple to apply. On the other hand, the estimated trade elasticities tend to be small for most countries and to be imprecisely estimated. This creates a large range of uncertainty for the estimated long-run equilibrium real exchange rate. A very difficult problem is how to estimate RB. For small, low-income countries that are not well integrated with international capital markets and rely heavily on external transfers and concessional financing from official sources, RB may reasonably be taken to be exogenous. But for countries that are financially open, most procedures for estimating RB will have to rely on extensions of this simple structure.

Because these shortcomings render the method of questionable reliability, it is perhaps best suited for first-pass, back-of-the-envelope estimation of the equilibrium real exchange rate. However, in country applications where data limitations do not permit the implementation of the more sophisticated approaches described subsequently, or in circumstances in which time and resource constraints on the analyst are severe, this may be the best that the analyst can do. Under these circumstances, the trade elasticities approach, despite its limitations, may be the method of choice.

3. General-Equilibrium Model Simulation

One way to extend this simple procedure is to embed net exports in a full empirical general-equilibrium model of the domestic economy that simultaneously explains the determinants of the resource balance as well as the full-employment levels of real *GDP* at home and abroad. Such a model could then be simulated for sustainable values of the fundamentals and solved for the relevant definition of the equilibrium real exchange rate (classical or Nurksian) to produce an estimate of the equilibrium real exchange rate. Indeed, in principle, simulation of empirical general-equilibrium models should dominate other estimation methods because the approach has several important advantages. First, it is structural, so the underlying mechanisms determining the equilibrium real exchange rate can be understood. Second, this approach allows the estimated equilibrium real exchange rate to reflect the full range of known macroeconomic interactions in the economy. Third, because the entire dynamic path of the real exchange rate from its current value to the steady state (or semi–steady state, if conditioned on some slow-adjusting variable) equilibrium real exchange rate can be simulated, it provides maximum flexibility concerning dynamics.

This model-based general-equilibrium method, however, places relatively strong demands on economic theory, on the power of statistical techniques, and on the availability and quality of data. Made-to-suit models for individual developing countries with limited data and possibly unstable economic structures are vulnerable to doubts about model specification and parameter stability. Estimates derived from such models may thus fail to command much credibility, particularly when the models on which they are based have no existing track record. For the near future, it is likely that estimates of the equilibrium real exchange rate derived from simulations of such models would be treated as indicative and used to supplement and inform other approaches to estimation. Model simulations may be most attractive in applications in which an existing model has demonstrated its usefulness through an established record of tracking the macroeconomic performance of a particular economy.

4. Single-Equation Reduced-Form Estimation

In the absence of a believable (multiequation) model, single-equation estimation based on the reduced-form equation for the equilibrium real exchange rate becomes the alternative. One obvious possibility would be to estimate the reduced-form equation (16.5) directly. Why not do so? The answer is, of course, that neither its dependent nor any of its independent variables are observable.

However, if the real exchange rate turns out to be *nonstationary*, then the theory described in the preceding section suggests that at least some subset of the fundamentals must be so as well. This suggests the use of *cointegrating equations* (stationary linear combinations of nonstationary variables) linking the real exchange rate with the relevant set of fundamentals, based on the observable current values of the real exchange rate and the fundamentals. Econometric theory suggests that when neither the real exchange rate nor some subset of its fundamental determinants is stationary, the parameters of the function $H(\ )$ in equation (16.5) linking the equilibrium real exchange rate to the permanent values of its nonstationary fundamentals can be estimated from these observable values. These estimated parameters would reveal the long-run relationship between the real exchange rate and its fundamental determinants, that is, the parameters of the function $H(\ )$. Coupled with a technique for estimating sustainable values of the fundamentals, this would permit us to estimate the equilibrium real exchange rate.

This approach obviously follows naturally from the time series tests for stationarity required to assess the applicability of the simple PPP-based approach. Relative to the simulation of empirical general-equilibrium models, it places fewer demands on economic theory because the theory required is about long-run relationships, not short-run macroeconomic dynamics. Moreover, fewer data (time series) are required to implement it because the researcher needs time series only for the variables that can be expected to appear in the reduced-form equation for the real exchange rate in short-run macroeconomic equilibrium.

To illustrate the method, consider the results of the cointegrating equation reported subsequently for the real exchange rate in Thailand, estimated over the period 1970–1996. In this application, statistical tests for stationarity suggested that the real exchange rate, the ratio of government consumption to *GDP* (a proxy for government spending on nontraded goods), Thailand's terms of trade, and the world rate of inflation were all nonstationary variables, and that these variables were jointly cointegrated. This means that it is possible to find a linear combination of these four variables that is stationary. This linear combination is the "cointegrating equation" referred to earlier. If this equation is normalized so that the coefficient of the real exchange rate is unity, the coefficients of the remaining variables – the fundamentals – correspond to the parameters of the function $H(\ )$ in equation (16.5). The resulting equation – the cointegrating equation for the Thai real exchange rate – is given by

$$\log(RER) = 6.113 - \underset{(0.028)}{0.746} \log(CGR) - \underset{(0.016)}{0.581} \log(TOT) + \underset{(0.002)}{0.012}\, INFL$$

where *RER* is the real exchange rate, *CGR* is the ratio of government consumption to *GDP*, *TOT* is the terms of trade, *INFL* is the world rate of inflation, and the numbers in parentheses are standard errors of the parameter estimates. This equation suggests that the equilibrium real exchange rate in Thailand appreciates when the ratio of government consumption to *GDP* rises and the terms of trade improve and depreciates when external inflation increases.

When estimated permanent values of the fundamentals are substituted into the right-hand side of this equation, the fitted values provide an estimated time series for the equilibrium real exchange rate. The resulting estimates for the sample period and the actual real exchange rate are both plotted in Figure 16.3. Notice that a comparison of the actual real exchange rate with the estimated equilibrium real exchange rate suggests that the Thai real exchange rate had become overvalued by 1996. We will come back to this observation when discussing the Asian financial crisis of 1997 in Chapter 27.

This technique is quite attractive, for the reasons mentioned earlier, and there has been a substantial amount of research in recent years on the determination of equilibrium real exchange rates utilizing this method. Unfortunately, however, the technique has some important disadvantages. The first is that the statistical tests involved have *low power* in small samples, which means that they are unable to discriminate very closely between variables that are stationary and those that are not, which is a key requirement in the implementation of this approach. Moreover, in emerging economies, there are typically small numbers of annual observations with which to work, and the prevalence of structural change reduces the number

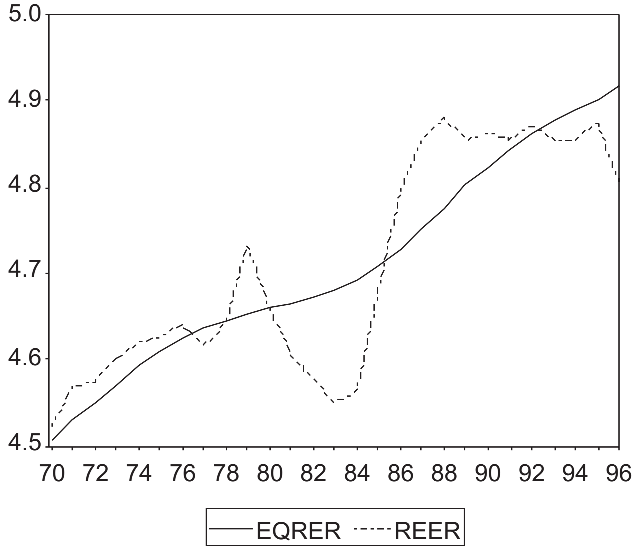


Figure 16.3. Thailand: Actual (RER) and estimated equilibrium real exchange rate (EQRER), 1970–1996

of useful observations even further. This is complicated by the fact that, as we have seen, theory can identify many potential fundamentals. Taken together, these observations imply that there are typically few degrees of freedom available in estimation, so it is hard to extract the parameters of the cointegrating relationship from the data. Moreover, the dependent and independent variables are typically measured with error because the available information often provides poor proxies for the theoretically correct variables, as we saw in Section I with respect to the real exchange rate itself. The former means that coefficients are measured imprecisely, the latter that they are biased toward zero. Finally, it is difficult to estimate the permanent component of the fundamentals, and the procedures adopted for this purpose are often arbitrary.

5. An Evaluation

How can we tell how well we are doing in estimating the equilibrium real exchange rate? We cannot judge this either by looking at the in-sample fit of estimated regressions or by our ability to forecast the long-run equilibrium real exchange rate because the equilibrium real exchange rate is unobservable. There are three ways that we might go about answering this question:

1. We can check to see how well estimates of misalignment based on theory and estimation of the equilibrium real exchange rate fit the macroeconomic history of the country in question (e.g., do episodes of extreme estimated overvaluation correspond to poor performance in the country's external accounts?).

2. We can test the extent to which estimated misalignment can help predict *future* movements in real exchange rates, holding constant changes in fundamentals and in strictly short-run influences on the real exchange rate. The relevant question in this case is, does past misalignment help predict future real exchange rate movements?
3. We can examine the goodness of fit of equations designed to predict movements in *actual* real exchange rates based on estimates of misalignment. Equations that express changes in current exchange rates as functions of past misalignment, controlling for current and lagged changes in fundamentals as well as in nonfundamental influences on the real exchange rate, are called *error-correction* models. This approach would ask, how much of the variation in real exchange rates can we explain on the basis of our theory using such equations?

The evidence from applying methods such as these is mixed (see Hinkle and Montiel 1999). The bottom line seems to be that, though we cannot estimate the equilibrium real exchange rate with any great precision, when there are large movements in the real exchange rate, we probably know enough to say whether they can plausibly be explained by the behavior of the underlying fundamentals.

VI. SUMMARY

This chapter has emphasized that real exchange rate misalignment is costly, and it may become more so as enhanced financial integration adds to the microeconomic (resource misallocation) costs of misalignment, the potential macroeconomic costs of capital-flow reversals, and currency crises. The best way to avoid these costs, of course, is to avoid the emergence of a large misalignment in the first place. In principle, there are three ways that misalignment can be avoided:

1. through movements in the nominal exchange rate that adjust the actual real exchange to its equilibrium value
2. through movements in prices of domestic goods that achieve the same purpose
3. through changes in the policy components of the fundamentals to move the equilibrium real exchange rate closer to the prevailing value of the real exchange rate

The problem with the third option, of course, is that such policies may be required to meet other objectives (e.g., should commercial policies be tightened to cope with a depreciation of the equilibrium real exchange rate caused by an adverse movement in the terms of trade?). The choice between the first two alternatives raises the issue of nominal exchange rate management.

There are essentially three options in this regard: allow the nominal exchange rate to achieve the adjustment endogenously by adopting a floating exchange rate system, place the burden of adjustment on the price of domestic goods by adopting an exchange rate system in which the nominal exchange rate cannot be changed, or manage the nominal exchange rate actively through adjustments in an officially determined exchange rate.

The considerations involved in choosing among these alternatives are the subject of the next two chapters. What we have done in this chapter, however, serves as an input into that discussion because, as we shall see in the next chapter, being able to estimate the value of the equilibrium real exchange rate empirically only matters for the third option, in which the exchange rate is to be managed actively. Given what has been said here about the costs of misalignment, an empirical assessment of the value of the equilibrium real exchange rate would presumably be an important input into the management of the nominal exchange rate. The question is, do we know enough about the empirical measurement of the equilibrium real exchange rate to be able to guide nominal exchange rate management?

This was the third doubt about equilibrium real exchange rate definition and estimation expressed in the introduction to this chapter. Unfortunately, this one is not as easy to dismiss as the first two. The answer to the question posed at the end of the last paragraph is that unfortunately, we do not yet have a reliable, workhorse approach to estimation of the equilibrium real exchange rate in which we can place a lot of confidence. Theory is ahead of estimation in this area because we do know, at least in principle, the types of economic variables that represent the fundamental determinants of the equilibrium real exchange rate. At present, we cannot estimate the value of the equilibrium real exchange rate with any great precision. However, that empirical estimates of real exchange rate misalignment are often consistent with the episodic experience of countries, that such estimates help predict future real exchange rate movements, and that explanations of real exchange rate changes based on such estimates explain a substantial part of the variation in real exchange rates together suggest that we know enough to detect severe episodes of misalignment. We will come to the implications of this state of knowledge in Chapter 18.

REVIEW QUESTIONS

1. Why is the estimation of equilibrium real exchange rates a problematic proposition?
2. In what sense can we distinguish between the real exchange rate that prevails in an economy at any moment in time and an equilibrium real exchange rate?
3. List four possible causes of long-run real exchange rate misalignment.
4. When is it appropriate to define the equilibrium real exchange rate in classical terms? In Nurksian terms?
5. How can we tell how successful our techniques for estimating the equilibrium real exchange rate are?

EXERCISES

1. Explain in intuitive terms why an increase in government spending causes the classical definition of the real exchange to appreciate in Section IV but the Nurksian definition to depreciate.
2. What effect would a permanent increase in tax collection (the variable T in our model) have on the Nurksian definition of the real exchange rate?
3. Explain in intuitive terms the effect that an increase in the policy-determined rate of devaluation of the domestic currency (π in Section IV) would have on the classical definition of the real exchange rate.
4. Assume that domestic and foreign bonds are perfect substitutes. What would Figure 16.2 look like in this case? Explain why.
5. On the assumption that domestic and foreign bonds are perfect substitutes, explain what effect a permanent increase in the domestic economy's total factor productivity (the parameter A in the aggregate production function) would have on the economy's equilibrium real exchange rate in the classical sense.

REFERENCES AND FURTHER READING

Christiansen, Lone, Alessandro Prati, Luca Antonio Ricci, and Thierry Tressel (2009), "External Balance in Low-Income Countries," working paper WP/09/221, International Monetary Fund.

Cline, William R., and John Williamson (2008), "New Estimates of Fundamental Equilibrium Real Exchange Rates," working paper PB 08-7, Peterson Institute for International Economics.

Corbo, Vittorio, and Patricio Rojas (1995), "Exchange Rate Volatility, Investment, and Growth: Some New Evidence," unpublished manuscript.

Dornbusch, Rudiger, Ilan Goldfajn, and Rodrigo O. Valdes (1995), "Currency Crises and Collapses," *Brookings Papers on Economic Activity*, vol. 2, pp. 219–294.

Dunaway, Steven, Lamin Leigh, and Xinagmin Li (2006), "How Robust Are Estimates of Equilibrium Real Exchange Rates: The Case of China," working paper WP/06/220, International Monetary Fund.

Egert, Balazs (2004), "Assessing Equilibrium Exchange Rates in CEE Acceding Countries: Can We Have DEER with BEER without FEER? A Critical Survey of the Literature," discussion paper 1, Bank of Finland Institute for Economies in Transition.

Hinkle, Lawrence, and Peter J. Montiel (1999), *Exchange Rate Misalignment: Concepts and Measurement for Developing Countries*, Oxford: Oxford University Press.

Isard, Peter, and Hamid Faruqee, eds. (1998), *Exchange Rate Assessment: Extensions of the Macroeconomic Balance Approach*, occasional paper 167, International Monetary Fund.

Isard, Peter, Hamid Faruqee, G. Russell Kincaid, and Martin Fetherston (2001), "Methodology for Current Account and Exchange Rate Assessments," occasional paper 209, International Monetary Fund.

Lee, Jaewoo, Gian Maria Milesi-Ferretti, Jonathan Ostry, Alessandro Prati, and Luca Antonio Ricci (2008), "Exchange Rate Assessments: CGER Methodologies," occasional paper 261, International Monetary Fund.

MacDonald, Ronald (2000), "Concepts to Calculate Equilibrium Exchange Rates: An Overview," discussion paper 3/00, Deutsche Bundesbank.

Mussa, Michael (1986), "Nominal Exchange Rate Regimes and the Behavior of Real Exchange Rates: Evidence and Implications," in K. Brunner and A. Meltzer, eds., *Real Business Cycles, Real Exchange Rates, and Actual Policies*, Amsterdam: North-Holland, pp. 117–214.

Nurkse, Ragnar (1945), *Conditions of International Monetary Equilibrium*, Essays in International Finance 4 (Princeton: Princeton University Press).

Razin, Ofair, and Susan M. Collins (1997), "Real Exchange Rate Misalignments and Growth," working paper 6174, National Bureau of Economic Research.

Zhang, Zhichao (2001), "Real Exchange Rate Misalignment in China: An Empirical Investigation," *Journal of Comparative Economics*, vol. 29, pp. 80–94.

17

The Benchmark Model with Floating Exchange Rates

As shown in Chapter 1, most emerging-market and developing economies maintain some type of officially determined exchange rate regime. Accordingly, the benchmark model of Part 2 was built on the assumption of a fixed exchange rate. However, as also shown in Chapter 1, a substantial number of emerging-market economies – and a few developing economies as well – operate exchange rate regimes that are best described as managed or free floats. To examine how the economy behaves in the short run under this alternative exchange rate regime, in this chapter we will revisit the benchmark model of Part 2 and investigate how it would need to be modified to describe an economy with floating exchange rates. This extension will not only allow us to study the macroeconomics of some important emerging economies but will also give us a frame of reference for analyzing optimal exchange rate regimes in Chapter 18.[1]

The structure of this chapter is as follows: the first section spells out the analytical framework that we will use to study a floating exchange rate economy. Section II takes the first step in analyzing such an economy by examining how the equilibrium nominal exchange rate is determined when the central bank does not intervene in the foreign exchange market. Section III assembles the complete short-run model under floating rates, building on the groundwork laid out in the two preceding sections. To see how the model works, it analyzes how an economy with floating exchange rates responds to the familiar set of shocks that we have studied under fixed exchange rates. As we will see, under floating exchange rates, the expected

[1] There is an extensive literature on the experience with floating exchange rates among industrial countries. See Obstfeld (1985, 1995), MacDonald and Taylor (1992), Taylor (1995), and Meese (1990). Among emerging and developing economies, an important issue has been whether such countries really float at all, even when they say they do. The classic reference is Calvo and Reinhart (2002). See also Reinhart (2000), Borensztein et al. (2001), Goldfajn and Olivares (2001), Frankel et al. (2002), Edwards and Levy Yeyati (2003), and Shambaugh (2003).

duration of shocks to the economy turns out to be very important in determining the macroeconomic impacts of such shocks. Our initial examination of the effects of shocks in Section III assumes that the shocks last for a single period and then go away. Such one-period shocks are referred to as *transitory* shocks. Section IV extends the analysis to longer-lasting shocks, including shocks that are expected to be permanent. Section V concludes.

I. ANALYTICAL FRAMEWORK

To extend the benchmark model to the case of floating exchange rates, we will make several simplifying assumptions. First, recall that in the sticky wage framework of Chapter 4, the supply wage was given by equation (4.8a):

$$W = (1-\sigma)W_{-1} + \sigma P^e w(L^S, \ldots)$$

where W is the current nominal wage, W_{-1} is the lagged nominal wage, P^e is the price level that workers expect to prevail this period based on the information that was available to them last period, L^S is the supply of labor, and σ is the share of nominal wage contracts that are renegotiated this period. The parameter σ is our index of nominal wage stickiness, with nominal wages ranging from fixed when $\sigma = 0$ to fully flexible when $\sigma = 1$. To simplify our model by removing the predetermined variable W_{-1}, we will focus on the case of flexible nominal wages; that is, we will assume $\sigma = 1$. In this case, the supply wage takes the following simpler form:

$$W = P^e w(L^S, \ldots)$$

By setting the demand wage equal to the supply wage, we can write labor market equilibrium as

$$PMP_L(L, \ldots) = P^e w(L, \ldots)$$

where MP_L is the marginal product of labor, a decreasing function of the level of employment. Dividing both sides of this equation by P^e and solving for L, the equilibrium level of employment can be written as

$$L = L(\underset{+}{P/P^e})$$

with $L(1) = L_P$. In words, the economy is at full employment when the actual price level is equal to the price level that workers had expected to prevail during the current period; it will be at *more* than full employment when the actual price level exceeds the expected price level and at less than full employment when the actual price level falls short of the expected price level. Substituting this expression into the aggregate production function, the flexible-wage aggregate supply curve becomes

$$Y = AF(K, L\ \underset{+}{P/P^e} = Y(\underset{+}{P/P^e}; A, K)$$

Thus, in the flexible-wage case, output is equal to potential output when the actual price level equals the expected price level, exceeds potential output when the actual price level exceeds the expected price level, and falls short of potential when the actual price level is less than the expected price level.

Next, we have to decide how to describe the economy's links with international capital markets. Because among the group of emerging and developing economies, the majority of countries that maintain some type of floating exchange rate regime tend to be emerging economies rather than developing ones, we will develop the floating exchange rate version of our model under the assumption that the country's capital account is open and its economy is highly integrated with international capital markets, as is likely to be true for most emerging economies. This being so, it is simplest just to assume perfect capital mobility.

We will retain the assumptions of flexible wages and perfect capital mobility throughout this chapter. Our final assumption concerns the *type* of floating exchange rate regime that we will be analyzing in this chapter: specifically, whether we are modeling a free float or a managed one. We will initially develop the model under the assumption that the country operates a free, or clean, float but will describe how a managed float would differ later in the chapter.

With these assumptions in place, we can begin to construct our model. To do so, recall from Chapter 7 that in a short-run macroeconomic equilibrium, three markets have to clear at the same time: the (flow) market for domestic real output and the (stock) markets for domestic money and domestic bonds. The equilibrium condition in the market for domestic real output is given by the flexible-wage version of equation (5.7):

$$\begin{aligned} Y(\underset{+}{P/P^e}, \ldots) = \varphi(\underset{+}{SP^*/P}, \ldots)A_P(Y(P/P^e, \ldots \underset{+}{)} - T, \underset{-}{R}, \ldots) \\ + G + X(\underset{+}{SP^*/P}, \ldots), \qquad (17.1) \end{aligned}$$

with the symbols being defined as in Chapter 5; that is, φ is the share of private absorption A_P spent on domestic goods; S is the nominal exchange rate (price of foreign currency in terms of domestic currency); P^* is the foreign price level expressed in foreign-currency terms; T is real taxes paid by the private sector; R is the domestic nominal interest rate (the interest rate paid on domestic government bonds); G is real domestic government spending, which is assumed to be devoted entirely to domestic goods; and X is the real foreign demand for the domestic economy's goods (exports).

Turning to the financial markets, if the economy operates a clean float, as assumed earlier, the central bank does not intervene in the foreign exchange market. Consequently, it has no reason to hold foreign exchange reserves. In this case, the stock of base money consists only of credit extended to the government, B_C, and we can write the money-market equilibrium condition as

$$M = B_C = PL(\underset{-}{R},\ \underset{+}{Y(P/P^e)}) \tag{17.2}$$

where the function $L(\ )$ represents the real demand for money. Finally, under the assumption of perfect capital mobility, domestic and foreign bonds are perfect substitutes, and the equilibrium condition in the domestic bond market requires that the expected rate of return, *measured in the same currency*, be the same for domestic and foreign bonds. Here we encounter a first important difference between fixed and floating exchange rates: under floating exchange rates, we cannot generally assume that the exchange rate tomorrow will be the same as that today, as we typically did in Part 2, because there is no central-bank commitment to maintain the nominal exchange rate at any specific officially determined value. Equalization of expected same-currency rates of return on domestic and foreign bonds therefore requires the version of the uncovered interest parity (UIP) condition that explicitly allows for expected future exchange rate changes:

$$R = (1 + R^*)S^e_{+1}/S - 1 \tag{17.3a}$$

where R^* is the interest rate on foreign bonds and S^e_{+1} is the exchange rate expected to prevail one period ahead.[2]

The three equations just introduced describe, respectively, goods-market, money-market, and bond-market equilibrium in our floating exchange rate economy. The next step in specifying our model is to note that these equations contain two expectations variables: the expectations that workers formed last period about the price level that would prevail *this* period, P^e, and the expectations that domestic and foreign portfolio managers form this period about what the exchange rate will be *next* period, S^e_{+1}. How are these expectations formed?

In this chapter, we will confine ourselves to the analysis of the macroeconomic effects of *unanticipated* shocks. That means, as you will recall from Part 2, that shocks arriving this period come as a surprise to workers and thus do not affect their previously formed expectations of what the current price level would be. As long as shocks are unanticipated, therefore, we can hold P^e constant in the face of shocks to the economy, and we will indeed do so throughout this chapter.

2 On the conditions under which UIP is likely to hold, see Ito and Chinn (2007). Dooley and Isard (1983) describe a floating exchange rate model with imperfect capital mobility.

However, this assumption does not entitle us to similarly hold S^e_{+1} constant because this variable refers to the expectation of the exchange rate that will prevail *next* period and, even if shocks come as a surprise when they arrive this period, once they are known to have arrived, their future effects are unlikely to be ignored. These effects may or may not be expected to persist into the next period, depending on the expected total duration of the shock, and it is also possible that entirely new shocks may be expected to arrive then, whether or not they actually do.

How, then, is the expectation S^e_{+1} formed? That expectation has two components: agents have to form expectations about the shocks that will affect the economy in the future, and they have to interpret how such shocks will affect the future exchange rate. Regarding the first component, we will keep things simple by assuming that no *new* shocks are expected to affect the economy in the future, so the only question concerning the effects of shocks in the future is how long *current* shocks are expected to last. We will consider single-period shocks, shocks of intermediate duration, and permanent shocks. Regarding the second issue (how shocks affect the future exchange rate), we will assume that the economic agents in our model form their expectations about the future exchange rate using rational expectations; that is, given their expectations about the future behavior of the model's exogenous variables (the shocks), they use the model itself to interpret what those anticipated future shocks would imply about the value of the future exchange rate.

It may seem that this assumption will make things complicated, but it actually makes our lives relatively simple. To see how, suppose that we identify the last, current, and next periods successively with the labels -1, 0, and $+1$. Let us assume that no new shocks are expected to hit the economy next period (in period $+1$) and that any shocks that affect the economy in the current period (period 0) are expected to be purely *transitory*; that is, they last only a single period. That means that any shocks that affect the economy in period 0 are expected to be gone by period $+1$. If that is so, then under rational expectations, given the structure of our model, the state of the economy in period $+1$ should be expected to revert to what it was in period -1, that is, before the shock arrived.[3] The implication is that $S^e_{+1} = S_{-1} = \bar{S}$, where $\bar{S}$ is the equilibrium value of the exchange rate in the absence of shocks. That is, like the rest of the economy, the exchange rate in period $+1$ would also revert to its undisturbed (no shock) value. Thus, under transitory shocks and rational expectations, we can simply set $S^e_{+1} = \bar{S}$, an exogenous constant. We can therefore rewrite equation (17.3a) as

$$R = (1 + R^*)\bar{S}/S - 1 \tag{17.3b}$$

[3] The phrase "given the structure of our model" refers to the fact that the version of our model that we are studying in this chapter does not contain any predetermined variables. If it did, then even one-period shocks would affect the future state of the economy because of the gradual adjustment of the predetermined variables.

With this information about P^e and S^e_{+1} in hand, we are ready to solve our floating-rate short-run model.

II. THE ASSET-MARKET APPROACH TO THE EXCHANGE RATE

As a first step, notice from equation (17.3b) that the domestic interest rate determines the unique current-period nominal exchange rate that is consistent with equilibrium in the bond market. To see how, notice that equation (17.3b) can be solved for the nominal exchange rate S to yield

$$S = (1 + R^*)\bar{S}/(1 + R) \tag{17.4}$$

Because both R^* and $\bar{S}$ are exogenous, $(1 + R^*)\bar{S}$ is exogenous too. This means that, holding other (exogenous) things constant, the equilibrium value of the exchange rate is determined by the domestic interest rate R.

Why should this be so? It is easiest to see why by rewriting equation (17.3a) in the equivalent form

$$R = (1 + R^*)[1 + (\bar{S} - S)/S] - 1$$

Notice that $(\bar{S} - S)/S$ is just the expected rate of depreciation of the domestic currency (or the expected rate of *appreciation* of the foreign currency) between the current period and the next, and that with the expected future exchange rate S^e_{+1} given by the constant $\bar{S}$, the current level of the exchange rate S relative to S^e_{+1} determines this expected rate of depreciation. Thus, given the domestic interest rate, the role of the current exchange rate S is to adjust so as to equalize the nominal returns on the domestic and foreign bonds, thereby satisfying the *bond-market equilibrium condition (UIP)*. This mechanism for determining the equilibrium value of the exchange rate is referred to as the *asset-market approach to the exchange rate*. We can explore how it works by plotting the relationship between R and S given by equation (17.4) in Figure 17.1.

The curve that plots S as a function of R, labeled UIP in Figure 17.1, has two important properties: it has a negative slope, and it must pass through the point $(R^*, \bar{S})$, labeled A in Figure 17.1. The slope must be negative because as the domestic interest rate rises, UIP requires that foreign bonds also yield a higher return when measured in domestic currency. But for that to be the case, given that the foreign interest rate is exogenous, the domestic currency must be expected to depreciate more – and the foreign currency therefore to *appreciate* more – with a higher domestic interest rate. Because the expected future exchange rate $\bar{S}$ is given, this can only happen if the domestic currency *appreciates* more in the *current* period.

To see why the curve must pass through the point $(R^*, \bar{S})$, note that if domestic and foreign interest rates are equal $(R = R^*)$, there can be no expected exchange rate change between this period and the next if UIP is to be satisfied, meaning that

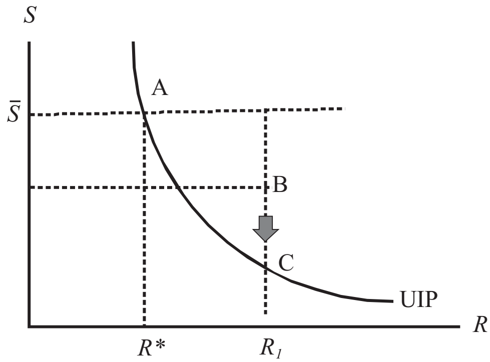


Figure 17.1. The asset-market approach to the exchange rate

if the expected future exchange rate is $\bar{S}$, the current exchange rate must be $\bar{S}$ as well. Because the curve must always pass through the point $(R^*, \bar{S})$, it must be true that changes in R^* must shift the curve *horizontally* by an amount that is exactly equal to the change in R^*, while changes in $\bar{S}$ must shift it *vertically* by the same amount as the change in $\bar{S}$.

To gain some intuition about how the exchange rate is actually determined, suppose that, at a given value of the domestic interest rate, such as R_1, the exchange rate happens to be above the UIP curve, say, at the point B. Notice that at points like B that are above the UIP curve, the current exchange rate is closer to the expected future exchange rate $\bar{S}$ than it would be at a point like C that actually lies on the curve. This means that the expected depreciation of the currency between this period and next (which depends on the vertical distance between S and $\bar{S}$) is *smaller* than would be required for the domestic-currency return on foreign bonds to match the interest rate paid by domestic bonds. The implication is that at a point such as B, domestic bonds yield a *higher* return than foreign ones. Under our assumption of perfect capital mobility, this triggers massive capital flows into the domestic economy, which create a strong demand for the domestic currency in the foreign exchange market. Because under floating exchange rates, the central bank does not intervene to meet this demand at the original exchange rate, the domestic currency must appreciate; that is, S must fall. Because the expected depreciation of the domestic currency remains too small to close the gap between domestic and foreign interest rates as long as the economy remains above the UIP curve at the interest rate R_1, this process must continue until the current exchange rate falls all the way down to its UIP value at the point C. A similar mechanism operates to depreciate the domestic currency if it is initially *below* the UIP curve at any given value of the domestic interest rate. In this case, the expected future depreciation of the domestic currency is too *large* to satisfy UIP, and domestic bonds thus yield a *lower* return than do foreign ones. Capital now flows *out* of the domestic economy, and because the central bank does not intervene to supply the foreign exchange market with the additional foreign currency that it demands at the prevailing exchange rate, the domestic currency *depreciates* until it reaches a new equilibrium value on the UIP curve.

Notice a rather surprising aspect of this story: given the value of the domestic interest rate, the equilibrium value of the exchange rate does not at all depend on what is happening to the current account of the balance of payments. The reason this can be is that once the exchange rate has adjusted to the equilibrium value implied by UIP, domestic and foreign bonds yield exactly the same rate of return, so both domestic and foreign agents are indifferent about which type of bond they hold. This makes capital flows into the domestic economy indeterminate in magnitude; that is, they can be whatever they need to be to satisfy the excess demand or supply for foreign exchange created by current account transactions. The bond-market equilibrium condition thus determines the exchange rate, but the current account balance actually determines the volume of capital flows. Because those flows must be exactly equal, but opposite in sign, to the current account surplus or deficit, the foreign exchange market is always in equilibrium.

For notational convenience, we can summarize what has just been said about the determination of the equilibrium value of the exchange rate by rewriting equation (17.4) in the form

$$S = [(1 + R^*)/(1 + R)]\bar{S} = h(\underset{-}{R}, \underset{+}{R^*})\bar{S} \tag{17.5}$$

where the function $h(\ )$ has the property that if initially $R = R^*$, equal changes in R and R^* leave h unchanged.

III. SHORT-RUN EQUILIBRIUM WITH FLOATING EXCHANGE RATES

Now that we have seen how the exchange rate is determined, we can go back to our short-run model. Notice that equation (17.5) is simply a rewritten version of the bond-market equilibrium condition (17.3b). Thus the short-run model under floating rates consists of the goods-market equilibrium condition (17.1), the money-market equilibrium condition (17.2), and the new bond-market equilibrium condition (17.5). These three equations determine three endogenous variables: the short-run equilibrium values of the domestic price level P, the domestic interest rate R, and the exchange rate S.

As in Part 2, we will solve the model graphically. To compare our results most easily with those of Part 2, we will solve the model for R and P first and then use equation (17.5) to see what happens to S. To do this, we first need to rewrite the model in the form of two equations in the two endogenous variables R and S. We can do that by using the bond-market equilibrium condition (17.5) to substitute for S in the goods-market equilibrium condition (17.1). This allows us to write the

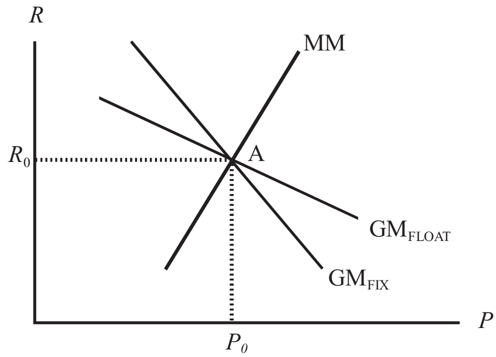


Figure 17.2. Short-run equilibrium with floating rates

goods-market equilibrium condition in the following revised form:

$$\begin{aligned} Y(P/P^e, \ldots) = {}& \phi(h(R, R^*)\bar{S}P^*/P, \ldots)A_P(Y(P/P^e, \ldots) - T, R, \ldots) \\ & + G + X((h(R, R^*)\bar{S}P^*/P, \ldots) \end{aligned} \tag{17.6}$$

Notice that this equation now only contains two endogenous variables: R and P. We can therefore plot it in the form of a GM curve in (P, R) space, as in the curve labeled GM_{FLOAT} in Figure 17.2. Notice that this new GM curve must have a negative slope, just like the GM curve of Part 2 (which is drawn as the curve GM_{FIX} in the figure). The reason is that an unanticipated increase in the domestic price level creates an excess supply of domestic goods, which requires that the domestic interest rate adjust to increase demand for domestic goods by just enough to absorb this excess supply. Because the adjustment required to increase aggregate demand is in the form of a *reduction* in the domestic interest rate, the curve has a negative slope.

What is different between this GM curve and that of Part 2 is the *size* of the required adjustment in the domestic interest rate. Because changes in the domestic interest rate have much more powerful effects on aggregate demand under floating exchange rates than under fixed rates, the required reduction in the domestic interest rate is much smaller in this case than under fixed exchange rates, making the new GM curve *flatter* than that of Part 2. Why is this? The reason is that a reduction in the domestic interest rate now affects demand for domestic goods through two channels: an *expenditure increasing-decreasing* channel and an *expenditure switching channel*. The former is the direct effect of changes in the interest rate on private absorption A_P, which was also present under fixed exchange rates. The latter is what is new under floating exchange rates. Because a reduction in the domestic interest rate causes a capital outflow that induces a depreciation in the exchange rate, domestic goods become cheaper, and both domestic and foreign residents are induced to switch spending from foreign to domestic goods. The effect on the composition of domestic spending is captured through the role of R in the function $\varphi(\ )$ in equation (17.6), while the effect on the composition of foreign spending is captured through the role of R in the export demand function $X(\ )$. It is this additional expenditure-switching effect that makes the floating-rate GM curve flatter than the fixed-rate one. Note that this effect is stronger the higher the degree of substitutability between domestic and foreign goods, so the new GM curve is flatter the greater the degree of substitutability between the two types of goods.

The new GM curve also differs from that of Part 2 in another important aspect. Recall that in Part 2, foreign financial shocks, in the form of exogenous changes in the interest rate on foreign bonds or in the expected future exchange rate, only affected the economy by shifting the BB curve vertically. However, now that we have substituted the bond-market equilibrium condition into the goods-market equilibrium condition, anything that disturbs the domestic bond market will also shift the GM curve. In particular, foreign financial shocks will shift the floating-rate GM curve vertically. To see how, recall that the UIP condition was written as

$$S = h(R, R^*)\bar{S} = [(1 + R^*)/(1 + R)]\bar{S} = (1 + R^*)\bar{S}/(1 + R)$$

Consider what happens to the vertical position of the GM curve if there is an increase in $(1 + R^*)\bar{S}$. Because this shock would trigger capital outflows that would cause the domestic currency to depreciate, it is expansionary and thus requires an increase in the domestic interest rate (at a given value of the domestic price level) to maintain equilibrium in the goods market; that is, the GM curve must shift vertically upward. The key point, however, is that the increase in $(1 + R)$ cannot be as large as that in $(1 + R^*)\bar{S}$ because if it were, $h(R, R^*)\bar{S}$ would remain unchanged, but the value of R would be higher than it would have been in the absence of the shock, leaving the goods market in a state of excess *supply* through the negative effect of the higher interest rate on private absorption. In other words, an increase in the domestic interest rate sufficient to leave $h(R, R^*)\bar{S}$ unchanged would be too large to restore equilibrium in the goods market: it would *more* than compensate for the expansionary effect of the shock. The upshot is that the increase in $(1 + R)$ at a given level of P must be smaller than that in $(1 + R^*)\bar{S}$, resulting in a higher value of $h(R, R^*)\bar{S}$. In other words, at an unchanged value of the domestic price level, to restore domestic goods-market equilibrium after an increase in $(1 + R^*)\bar{S}$ requires a combination of a higher domestic interest rate and a more depreciated exchange rate.

Other than these two differences, the other properties of the GM curve are exactly as they were in Part 2. Specifically, goods-market shocks such as changes in government spending cause the floating-rate GM curve to shift *horizontally* by exactly the same amount as the fixed-rate GM curve. You can see that this must be so because horizontal shifts are measured at a fixed value of the domestic interest rate, and with an unchanged interest rate, the exchange rate would be unchanged as well, making the situation exactly analogous to that with a fixed exchange rate.

The money-market equilibrium condition (17.2) is depicted as the upward-sloping curve MM in Figure 17.2. This curve is exactly the same as that in Part

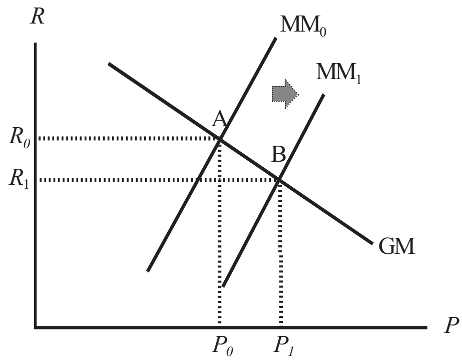


Figure 17.3. Monetary expansion

2, and its properties are unchanged, because the exchange rate does not enter the money-market equilibrium condition. Because the GM curve incorporates goods- and bond-market equilibrium, while the MM curve describes money-market equilibrium, the economy's short-run general equilibrium is at the point of intersection of the floating-rate GM curve and the MM curve, such as at the point A in Figure 17.2. Once the domestic interest rate and price level are determined in this way, the exchange rate can be determined directly from equation (17.3a). To understand how the economy works under floating rates, we can now consider the effects of the three familiar types of shocks that we analyzed in Part 2: money-market shocks, goods-market shocks, and external financial shocks.

1. Money-Market Shocks

To analyze the effects of money-market shocks, consider, for example, how the economy would respond to an unanticipated expansion in the stock of domestic credit, that is, to an increase in B_C. Recall that under perfect capital mobility, the BB curve of Chapter 6 was flat, and we saw in Chapter 7 that a monetary expansion would have no effect on the domestic economy in that case because central-bank intervention in the foreign exchange market would undo the effect of the credit expansion on the money supply, leaving the MM curve unchanged. What happens under floating exchange rates?

The answer is that because the central bank does not intervene in the foreign exchange market, there is nothing to undo the expansionary effect of the increase in domestic credit on the money supply, so the MM curve shifts to the right, as in Figure 17.3. The GM curve is undisturbed. As you can see in the figure, the monetary expansion results in a reduction in the domestic interest rate and an increase in the domestic price level as the economy moves from A to B.

How does monetary policy affect the real economy? At the original value of the domestic price level P_0, the monetary expansion requires a reduction in the domestic interest rate to clear the money market. This reduction in the domestic

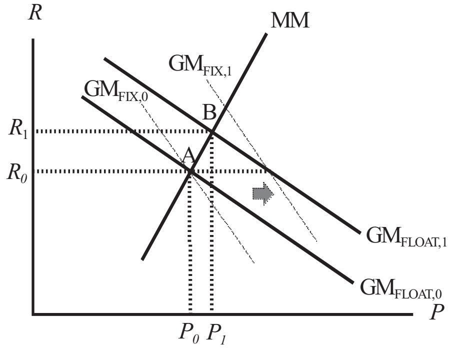


Figure 17.4. Fiscal expansion

interest rate induces capital outflows, which cause the exchange rate to depreciate. The combination of a lower domestic interest rate and a more depreciated exchange rate creates an increase in demand for domestic goods, which in turn causes the domestic price level to rise. The higher domestic price level helps to increase the demand for money, allowing the decrease in the domestic interest rate and therefore the depreciation of the exchange rate to be smaller than they would have been at an unchanged value of the price level. The important point is that monetary policy works through both expenditure-increasing and expenditure-switching channels; that is, it stimulates aggregate demand both by reducing the domestic interest rate and by depreciating the domestic currency. You can see that the domestic currency must have depreciated by inspecting equation (17.3b): the only determinant of S that has changed in that equation is the domestic interest rate R, and because R has fallen, S must increase.

Finally, notice that the monetary expansion must have a larger effect on the domestic price level – and therefore a larger effect on real output – the flatter the GM curve is. This is because a flatter GM curve implies closer substitutability between domestic and foreign goods and therefore a more powerful expenditure-switching effect triggered by the exchange rate depreciation.

2. Goods-Market Shocks

Next, consider the effects on the economy of a goods-market shock under floating exchange rates. For concreteness, let us take the case of a fiscal expansion in the form of an increase in G.

As previously mentioned, a goods-market shock has the effect of shifting the GM curve under floating rates by a horizontal amount that is exactly equal to the amount by which it would have shifted under fixed rates. This is shown in Figure 17.4 in the form of equal rightward shifts in the GM_{FLOAT} curve and the GM_{FIX} curves. The increase in government spending results in an increase both in the domestic price level and in the interest rate. However, notice that the increase in

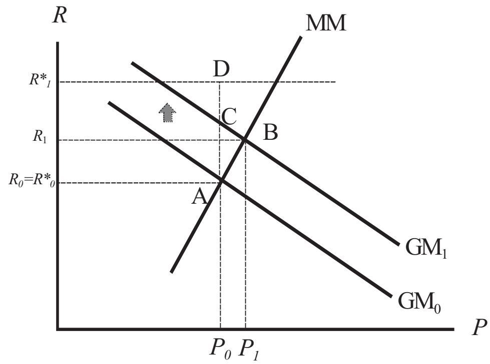


Figure 17.5. Effects of an increase in the world interest rate

the domestic price level is smaller than it would have been under fixed rates, where the domestic interest rate would have remained at R_0. Why is this?

The answer is that the higher price level caused by the fiscal expansion increases the demand for money, which in turn increases the domestic interest rate. The higher domestic interest rate attracts capital inflows that, in the absence of central-bank intervention in the foreign exchange market, cause the nominal exchange rate to appreciate. Because today's nominal exchange rate appreciation creates an expected future exchange rate *depreciation*, the higher domestic interest rate can persist in equilibrium, in contrast to what happens under fixed rates. In addition, the nominal exchange rate appreciation causes the real exchange rate to appreciate at any given value of the domestic price level, and this real exchange rate appreciation, by making domestic goods more expensive, induces expenditure switching away from domestic and into foreign goods. Thus the expansion in aggregate demand caused by increased government spending is partly offset under floating exchange rates by the expenditure-reducing effect of a higher domestic interest rate and the adverse expenditure-switching effect of an appreciated nominal exchange rate.

3. External Financial Shocks

Finally, consider the effects on the domestic economy of an external financial shock, which, for concreteness, we can initially take to be an increase in the foreign interest rate R^*. As we saw earlier, an increase in the foreign interest rate shifts the GM curve upward, but by an amount that is less than the increase in the foreign interest rate, as in Figure 17.5. Notice that this shock is *expansionary* under floating exchange rates: the domestic interest rate and price level both rise. As you can see from the figure, the equilibrium increase in the domestic interest rate must be smaller than that in the foreign interest rate, so the exchange rate must depreciate on impact to satisfy UIP.

Recall that under fixed exchange rates, an increase in the foreign interest rate had a *contractionary* effect on the domestic economy. Why the difference here? There are two reasons. The first is that, though the domestic interest rate increases, the increase (to R_1) is smaller under floating exchange rates than under fixed (where the domestic interest rate would increase by the same amount as the foreign rate, i.e., to R_1^*). Though even this smaller increase in the domestic interest rate under floating rates continues to have a contractionary effect on the economy, this effect is more than offset by an effect that is missing under fixed exchange rates: the depreciation of the exchange rate caused by the capital outflows that are triggered by the increase in the foreign interest rate.

How do we know that the negative effect of the higher domestic interest rate must be more than offset by the positive one arising from the depreciation of the exchange rate? The answer is that if this were not so, the money market could not clear. The higher domestic interest rate reduces the demand for money, so money-market equilibrium can only be sustained if a higher domestic price level causes the demand for money to increase. A higher domestic price level, in turn, can only emerge if the shock has a net expansionary effect on the domestic economy.

For the purposes of the next section, this is a good place to consider what would happen if the external financial shock instead were to take the form of an increase in the exchange rate that is expected to prevail next period. In the case of a transitory (one-period) shock, this would not happen, of course, but in the next section, we will be considering longer-lived shocks, so it is worth investigating the effect here, building on the analysis we have just completed.

Notice that $\bar{S}$ enters the model only through the term $(1 + R^*)\bar{S}$, so the effects on the economy of a change in $\bar{S}$ must be very similar to the effects of a change in the same direction in R^*. Given the current exchange rate, an increase in $\bar{S}$, for example, would create the expectation of an exchange rate depreciation between this period and the next, which would induce a capital outflow from the domestic economy and therefore a depreciation of the exchange rate. Because exchange rate depreciation is expansionary, at a given level of P, a higher domestic interest rate would be required to clear the domestic goods market, so the GM curve must shift upward. But the increase in the domestic interest rate cannot be as large as would be required to completely offset the expected depreciation of the exchange rate, thus preventing capital outflows and keeping the exchange rate unchanged, say, at a point like D in Figure 17.4. The reason is that if the domestic interest rate were to rise by that much, while the exchange rate remained unchanged, thereby eliminating the expansionary effect on the domestic economy, the higher domestic interest rate would exert a *contractionary* effect on the economy, meaning that an excess demand for domestic goods at the original point A would be replaced by an excess supply at D. What is required to restore goods-market equilibrium is that the domestic interest rate rise by less than this amount, so the contractionary effect of a higher domestic interest rate is just offset by the expansionary effect of a depreciated exchange rate. The GM curve must therefore shift upward by an amount that is smaller than what would be required to keep S unchanged. In Figure 17.4, the GM curve would shift up to a point like C rather than D. Because the new equilibrium interest rate is found at the point B, rather than C, the final increase in the domestic interest rate must be even smaller, so the equilibrium value of the exchange rate unequivocally depreciates as a result of this shock. In other words, a depreciation of the expected *future* exchange rate results in a depreciation of the equilibrium value of the *current* exchange rate.

However – and this is the important point – it is easy to see that this depreciation must be *less than proportionate* to the depreciation in the expected future exchange rate. The reason is that because the domestic interest rate increases, uncovered interest parity can continue to hold only if the domestic currency is expected to depreciate between the current period and the next. For that to be the case, the depreciation of today's exchange rate must be *smaller* than the depreciation in the exchange rate that is expected to prevail tomorrow. We will make use of this result in the next section.

4. Managed Floats

Up to now, we have been discussing floating exchange rates under the assumption that the central bank does not intervene in the foreign exchange market, that is, that the economy operates a "clean," or "free" float. How would our analysis change if it were to maintain a "dirty," or "managed" float instead?

Because this situation involves the central bank intervening in the foreign exchange market, the bank must have the means to conduct such intervention; that is, it must hold some foreign exchange reserves. In that case, we need to modify the money-market equilibrium condition to the more familiar form that we used under fixed exchange rates:

$$M = B_C + SF_C^* = PL(R, Y(P/P^e)) \tag{17.7}$$

where F_C^* is the foreign-currency value of the central bank's stock of foreign exchange reserves. The question, then, is what happens to the economy if the central bank changes F_C^*, that is, if it intervenes in the foreign exchange market either to buy or sell reserves?

The answer depends crucially on whether the central bank's foreign market intervention is sterilized. If it *is* sterilized, then its intervention in the foreign exchange market, by definition, will have no effect on the domestic money supply. That means that the MM curve will not be displaced, and the economy's equilibrium will not be disturbed. In other words, sterilized intervention has no effect on the economy, except to change the composition of assets in the central bank's balance sheet, and that composition does not matter because domestic and foreign bonds are perfect substitutes anyway. However, if the intervention is *unsterilized*, then

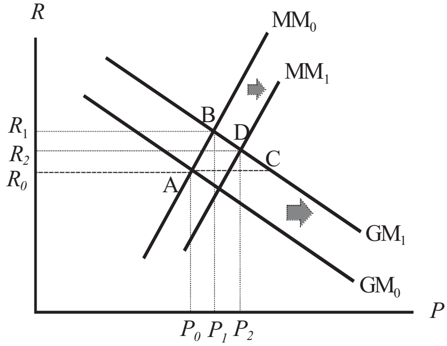


Figure 17.6. Managed floating

central-bank purchases of foreign exchange increase the domestic money supply, and central-bank sales of foreign exchange reduce it. In the former case, the MM curve shifts to the right, lowering the domestic interest rate and causing the exchange rate to depreciate as the result of incipient capital outflows; in the latter case, MM shifts to the left, and the exchange rate appreciates. In other words, central-bank intervention in the foreign exchange market is just another way that it can conduct monetary policy.

We can now provide an interpretation of managed floating. Suppose there is a shock to the economy, like an exogenous increase in the world demand for domestic goods. As shown in Figure 17.6, in the absence of a response by the central bank, the economy's equilibrium would move from the point A to B, where the higher domestic interest rate would induce a capital inflow that would cause the exchange rate to appreciate. Under a managed float, the central bank may want to resist that appreciation. To do so, it can intervene in the foreign exchange market by buying up the foreign exchange that is being offered for sale by those that are bringing capital into the country, in exchange for newly printed domestic money. This increase in the domestic money supply would cause the MM curve to shift to the right. To avoid the appreciation of the exchange rate completely, it would have to intervene sufficiently to cause the MM curve to pass through the point C, which would be equivalent to maintaining a fixed exchange rate. To resist it partially, thereby "managing" the rate without keeping it fixed, it would have to intervene less aggressively, making the shift in the MM curve smaller and causing it to intersect the new GM curve at a point like D.

As this example illustrates, managed floating just means using monetary policy actively to keep the domestic interest rate closer to the foreign rate than it would have been in the absence of central-bank action. Doing so prevents the exchange rate from fluctuating as much as it otherwise would and, in that sense, is a "managed" float.

IV. NONTRANSITORY SHOCKS

Up to now, we have been restricting our attention to transitory shocks – shocks that last for a single period and then go away. How would our analysis change in the case of shocks that are expected to last for more than a single period, that is, in the case of nontransitory shocks?

To analyze this case, we can proceed just as before; that is, for each shock, we begin by determining the value of the expected next-period exchange rate S^e_{+1} and then, given S^e_{+1}, solve for the current-period values of R, P, and S. The key difference between this case and that of transitory shocks, however, is that when shocks are expected to last more than one period, we can no longer determine S^e_{+1} simply as $S^e_{+1} = S_{-1} = \bar{S}$. The reason is that, because the shock is still expected to be in place in the next period, the next period can no longer be expected to look exactly like the pre-shock period, so there is no reason to believe that the exchange rate that will prevail next period is the same that would have been observed in the absence of the shock. Our first task, therefore, is to determine exactly what exchange rate will be expected to prevail next period under these new circumstances.

1. Permanent Shocks

It is simplest to begin by considering the case of permanent shocks, that is, shocks that arrive unexpectedly but, once in place, are not expected ever to go away. This situation differs from that described in the preceding section in two ways:

1. Though the shock may have been unanticipated in the period in which it arrived, from that period on, the change in the economic environment is recognized by everyone, so from the next period forward, the presence of the shock becomes anticipated. This means that because from the next period onward nothing unexpected is happening, we must have $P = P^e$ in all those periods.
2. Similarly, once the presence of the shock is known, if the shock is permanent and no other shocks are expected, the economy's equilibrium cannot be expected to change from one period to the next after period $+1$. Therefore we must also have $S_{+1} = S_{+2} = S_{+N}$ for all future periods N.

But what will that unchanging equilibrium value of the exchange rate be in all those future periods? To answer this question, notice that because the exchange rate is constant from the next period on, in each of those periods, UIP implies $R = R^*$. Imposing both $P = P^e$ and $R = R^*$ in our previous equations (17.6) and (17.2), the goods- and money-market equilibrium conditions in all those future periods must be

$$Y_P = \phi(SP^*/P, \ldots)A_P(Y_P - T, R^*, \ldots) + G + X(SP^*/P, \ldots) \qquad (17.8)$$

$$M = PL(R^*, Y_P) \qquad (17.9)$$

where we have used the fact that $P = P^e$ implies $Y = Y_P$. Notice that the money-market condition (17.9) contains only one endogenous variable: the domestic price level P. The goods-market equilibrium condition, in turn, can also be interpreted as containing only one endogenous variable, the real exchange rate $Q = SP^*/P$. Thus, from the next period on, the money market must determine the domestic price level, and given this equilibrium value of P and the value of the real exchange rate SP^*/P that clears the goods market, the goods market determines the nominal exchange rate S as $S = QP/P^*$.

Notice what this means:

- Because the domestic price level is determined in the money market, permanent fiscal policy changes can have no effect on the domestic price level or real output after the first period in which they arrive. Their only effect is to alter the nominal exchange rate. This means that the effects of a permanent increase in government spending on aggregate demand, for example, must be completely offset by a decrease in net exports caused by real exchange rate appreciation. What is the economic intuition for this? The answer is that as long as fiscal policy has any effect on the price level at all, it will cause an increase in money demand, which will put upward pressure on the domestic interest rate and pull capital into the country, causing the exchange rate to appreciate. This pressure on the exchange rate will continue until the reduction in net exports exactly offsets the positive effects of the fiscal expansion on aggregate demand.
- Permanent monetary policy changes have proportional effects on the domestic price level after the first period in which they are implemented. They also have proportional effects on the nominal exchange rate, leaving the real exchange rate unchanged. The reason is that because monetary policy cannot affect the domestic interest rate when changes in the money supply are expected to be permanent, they cannot affect the value of the ratio SP^*/P required to clear the goods market. Thus, if P changes at all, S and P must change in the same proportion – and P must indeed change because a change in P is required to cause the demand for money to change by as much as the money supply and thereby clear the money market.
- A permanent increase in the world interest rate increases the domestic price level permanently after the first period in which it takes effect because it reduces the real demand for money with an unchanged nominal money supply. Because it also reduces aggregate demand for domestic output, it must cause the *real* exchange rate to depreciate after the first period. Accordingly, it results in a depreciation of the nominal exchange rate that is *more* than proportional to the increase in the domestic price level.

All this describes what happens after the first period in which a shock arrives, that is, from period $+1$ onward. But what happens in the first period (period 0), in which each of the considered shocks arrives unexpectedly?

In the first period, because the shock is unanticipated, the economy's equilibrium is once again described by the equilibrium conditions (17.6) and (17.2) analyzed in the preceding section, with one exception: $S^e_{+1} = \bar{S}$ is no longer equal to the pre-shock exchange rate S_{-1} but rather to the post-shock *permanent* exchange rate that prevails from the next period forward – that is, the one we just derived from equations (17.8) and (17.9). How does this modification in the expected *future* exchange rate affect the equilibrium outcome for the economy in the *first* period?

We can interpret that outcome as the result of two separate shocks: the specific exogenous shock under consideration, and the separate "shock" implied by the change in the expected future exchange rate. It is this change in the expected future exchange rate that makes the analysis of this section different from that of Section III. To see what is implied about the economy's first-period response, consider each of the three shocks analyzed earlier in turn.

a. Fiscal Expansion

In the case of the fiscal expansion, the expected permanent future exchange rate appreciates after the first period, as we have just seen. In the graphical analysis of the preceding section, therefore, two things must happen: first, the fiscal shock itself causes the GM curve to shift to the right, but second, the appreciation of the expected next-period exchange rate now causes the GM curve to shift back to the *left*, or down. This second shift of the GM curve is what is new here. The issue in deriving the impact (first-period) effect of the permanent fiscal expansion on the economy, then, is determining the extent to which the leftward shift caused by the appreciation of the expected next-period exchange rate offsets or does not offset the rightward shift caused by the fiscal expansion itself.

It turns out that the leftward shift of the GM curve caused by the expected future exchange rate appreciation must *exactly* offset the rightward shift caused by the fiscal expansion itself, leaving the domestic interest rate, price level, and level of output unchanged in the face of a permanent fiscal shock. To see why, consider the goods-market equilibrium condition written in the form of equation (17.1), reproduced here for convenience:

$$\begin{aligned} Y(P/P^e, \ldots) = \phi(SP^*/P, \ldots)A_P(Y(P/P^e, \ldots) - T, R, \ldots) \\ + G + X(SP^*/P, \ldots) \end{aligned} \tag{17.1}$$

As we have just seen, *after* the first period, because the shock is both anticipated and permanent, this equation must have $P = P^e$ and $R = R^*$. We have shown that there is some equilibrium permanent exchange rate $S = S_P$ for which this equation is satisfied with $P = P_{-1}$, so a change in fiscal policy does not cause the domestic interest rate or price level to be different from what either of them would have been without the fiscal policy change after the first period. In the *first* period, because the shock is unanticipated, all we know is that $P^e = P_{-1}$, and we have to solve for the endogenous variables R, P, and S. But because we already know that the equation holds when $R = R^*$, $P = P_{-1}$, and $S = S_P$, those must be the first-period equilibrium values as well. In other words, a permanent fiscal expansion causes an exchange rate appreciation and a reduction in net exports that exactly offsets the effect of the fiscal expansion on aggregate demand, both on impact and in all future periods.

b. Monetary Expansion

In the case of a monetary expansion, we saw that after the first period, the only effects on the economy were an increase in the price level and an exchange rate depreciation that were both proportional to the increase in the money supply. What happens on impact?

Again, on impact, $P^e = P_{-1}$, and the impact effect of the permanent monetary expansion can be decomposed into two parts: the effects of the monetary expansion itself and those of the change in the expected future exchange rate. The monetary expansion shifts the MM curve to the right, as in the preceding section. However, the depreciation of the expected future exchange rate now shifts the GM curve to the right as well. The domestic price level must therefore unambiguously increase on impact, and because this increase in the price level was unanticipated, real output must unambiguously rise. The effect on the domestic interest rate, however, is ambiguous. If the rightward shift in the MM curve exceeds that in the GM curve, the interest rate falls, and if the opposite is true, it must rise. Notice the surprising result that a permanent monetary expansion could actually *increase* the domestic interest rate on impact. For this to happen, the rightward shift in the GM curve must be large; that is, goods-market equilibrium must be very sensitive to the expected future exchange rate, which requires a high degree of substitutability between domestic and foreign goods. As long as this substitutability is not too high, the domestic interest rate falls on impact. In this case, UIP implies that the exchange rate must appreciate between the first period and the second so that on impact, it must actually *overshoot* its eventual equilibrium value.

c. Increase in the World Interest Rate

As shown earlier, an increase in the world interest rate causes an increase in the domestic price level and a more-than-proportional nominal exchange rate depreciation after the first period. On impact, it affects the domestic economy both directly, through the effect of the higher world interest rate on the GM curve, and indirectly, through the effects of the expected future depreciation, again on the GM curve. Both effects cause the GM curve to shift to the right in the first period, while leaving the MM curve undisturbed. The implication is that the domestic interest rate and price level must both rise on impact in response to an increase in the world interest rate that is expected to be permanent. Because the increase in the domestic price level is unexpected, real output must rise on impact. Finally, because a higher domestic interest rate and higher domestic price level both induce an excess supply of domestic goods, the goods market can clear in the first period only if the exchange rate depreciates.

2. Two-Period Shocks

Next, let us assume that the shock is expected to remain in place for exactly two periods – this period and the next – but to go away after that (i.e., it will be gone by period +2). This means that in the next period (period +1), the exchange rate that will be expected to prevail in the following period must once again be the pre-shock exchange rate: $S^e{}_{+2} = S_{-1} = \bar{S}$. This is useful because once we know what exchange rate is expected to prevail in period +2, we can use our model, in accordance with rational expectations, to determine what the equilibrium exchange rate must be in period +1.

Will it simply be the same as the *current* period exchange rate that we solved for in Section III? One might think so, because after all, that analysis investigated what would happen to the equilibrium value of the exchange rate in a context in which the economy was affected by a shock that was expected to be gone by the next period. With a one-period shock, that was true of the current period. With a two-period shock, it would be true for the *next* period instead. But there is one difference. The one-period shock of the preceding section was unanticipated in the period in which it materialized, but a two-period shock cannot be unanticipated in its *second* period. Because nothing happens in that period that is unanticipated, to solve for the exchange rate in period +1, therefore, we have to impose $P = P^e$ in equations (17.6) and (17.2). Because $Y(P/P^e) = Y_P$ when $P = P^e$, the result is as follows:

$$\begin{aligned} Y_P &= \varphi(h(R, R^*)S_{-1}P^*/P, \ldots)A_P(Y_P - T, R, \ldots) \\ &\qquad + G + X((h(R, R^*)S_{-1}P^*/P, \ldots) \\ M &= PL(R, Y_P) \end{aligned}$$

These equations can be used to solve for the equilibrium values of R and P in period +1, following a procedure that is identical to what we did in the preceding section for period 0 in the case of a one-period shock. Equation (17.4) can then be used to determine the value of the equilibrium exchange rate in period +1.

The GM and MM curves that can be derived from equations (17.8) and (17.9) are not the same as those of Section III. In particular, because changes in the domestic price level have weaker effects on the goods and money markets when they are anticipated than when they are not, the GM and MM curves derived from equations (17.8) and (17.9) must be *flatter* than those of Section III. Because both equilibrium conditions are less sensitive to changes in the domestic price level, these equations must also experience *larger* horizontal shifts in response to shocks. However, because the slopes of the curves and the directions in which they are shifted by shocks are qualitatively the same as in Section III, the *qualitative* effects of shocks to the economy must be exactly the same as those derived in that section. Specifically, the following results must hold

- Expansionary goods-market shocks and increases in world interest rates both shift the GM curve to the right, causing increases in the domestic interest rate and price level. Because expansionary goods-market shocks cause the domestic interest rate to rise above the foreign interest rate, they result in an *appreciation* of the exchange rate in period +1. On the other hand, because increases in world interest rates leave the domestic interest rate lower than the world rate, they cause it to *depreciate*.
- An expansionary money-market shock would cause a *depreciation* of the exchange rate in period +1.

We now have all the pieces in place to describe how multiperiod shocks affect our floating exchange rate economy. We begin by considering a two-period shock that arrives in period 0 and lasts only through period +1, and then we describe how the analysis can be extended to longer-lasting shocks.

a. Fiscal Expansion

Suppose that the shock in question is an expansionary fiscal shock. As we have just seen, such a shock would cause the exchange rate to appreciate in period +1, the last period in which the shock is in place. In the current period, therefore, the shock once again affects the economy through *two* channels: the expansionary fiscal shock itself and the expected appreciation of the exchange rate in the next period.

The magnitude of the leftward shift in the GM curve depends on the size of the change in the expected period +1 exchange rate. We saw earlier that when the shock was permanent, the appreciation in the period +1 exchange rate was exactly enough to shift the GM curve all the way back to where it started, leaving R and P unchanged. How big will it be now?

The key is period +2. Recall that when the shock was permanent, the exchange rate appreciated to S_p in every period, including period +2. With a shock that is gone by period +2, however, the exchange rate must revert to the pre-shock value S_{-1} in that period. But because we have already seen that $S_P < S_{-1}$, this means that from the perspective of period +1, the one-period-ahead exchange rate is more appreciated in the case of a permanent shock than in the case of a two-period shock. The implication is that the appreciation of the period +1 exchange rate must be *smaller* when the shock lasts two periods than when it is permanent. In turn, this means that the leftward shift of the GM curve in the current period must be smaller in the case of the two-period shock than in the case of a permanent shock, and therefore it must not be large enough to completely offset the rightward shift of the GM curve caused by the fiscal expansion itself.

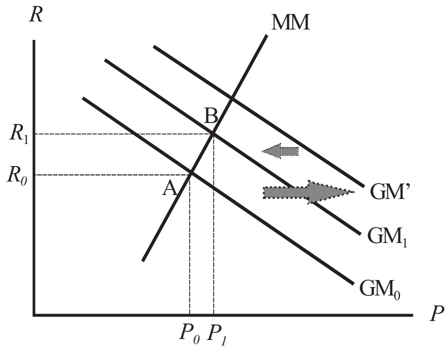


Figure 17.7. Impact effects of a two-period fiscal expansion

In short, the situation is as in Figure 17.7. The effect of the fiscal expansion alone on the GM curve is shown as the curve GM′, shifted rightward from its initial position GM_0. The appreciation of the expected period +1 exchange rate shifts the GM curve back to the left, but not by enough to restore its original position. The fiscal expansion thus increases the domestic price level as well as the domestic interest rate, as in the case of a transitory shock, but the effects are weaker because the expected future exchange rate is more appreciated than in the case of a purely transitory (one period) shock. Notice that because the domestic interest rate is higher than the world rate, the exchange rate must be expected to *depreciate* from the current period to the next. This means that the two-period fiscal expansion must cause a larger appreciation of the exchange rate in its first period than in its second.

This last observation makes it straightforward to extend this analysis to fiscal shocks of any length. Consider a fiscal expansion that lasts T periods and goes away in period $T + 1$. Then the analysis just conducted can be used to describe what happens in the shock's last period (which would correspond to period +1 above) and the one before that (corresponding to the current period in the analysis above). You can see that as we work back in time toward the present, the exchange rate becomes progressively more appreciated relative to its no-shock value as we move closer to the current period. This means that the longer the duration of the shock, the more appreciated the exchange rate must become on impact and therefore the closer the GM curve must move back to its initial location. In the limit, as we saw earlier, when the shock is permanent, the GM curve moves all the way back to its initial location. Thus the impact effects of an unanticipated fiscal expansion on the domestic price level and real output are smaller the longer the shock is expected to last.

b. Monetary Expansion

The analysis of a two-period monetary expansion is similar to that of a two-period fiscal expansion. There are two effects on impact: the one that operates directly

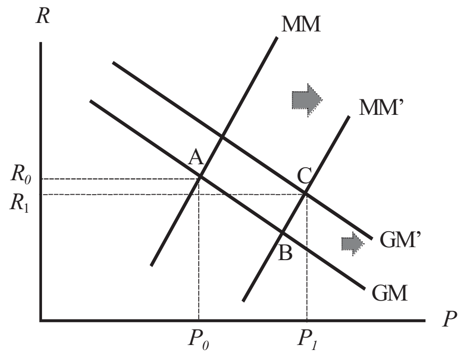


Figure 17.8. Impact effects of a two-period monetary expansion

through the increase in the money supply and one that operates indirectly through the expected depreciation of the next period's exchange rate. Again, working back from the second period of the shock, an analysis similar to that conducted in Section III indicates that the exchange rate will be more depreciated in the next period (period + 1, when it can be treated as an anticipated shock that will be gone by the period after) than it was in the pre-shock equilibrium.

In period 0, therefore, the increase in the money supply shifts the MM curve to the right, say, from MM to MM′ in Figure 17.8, while the depreciation of the expected period +1 exchange rate shifts the GM curve to the right, say, from GM to GM′. It is easy to see from the figure that the rightward shift of the GM curve magnifies the monetary expansion's impact on the price level (and therefore on real output) on impact compared to the one-period shock but weakens its effect on the domestic interest rate. It must also make the exchange rate more depreciated at the new equilibrium at point C than it would have been made to be by the increase in the money supply alone at point B (recall that a depreciation in the expected future exchange rate causes the current exchange rate to depreciate, though by less than the future rate). Thus the exchange rate in period 0 is more depreciated than it would otherwise have been as the result of the expected depreciation of the exchange rate in period +1, at the same time that the gap between the current and expected future exchange rates becomes smaller, so there is less appreciation for the exchange rate to do between period 0 and period +1.

If the monetary expansion lasts for T periods rather than just two, then the analysis in the next-to-last preceding paragraph would apply to period T. Because of the influence of the depreciated exchange rate in period T, the exchange rate in period $T - 1$ would therefore be more depreciated than it would have been as the result of the monetary shock alone, and thus more depreciated than the period T exchange rate itself. Because the period $T - 1$ exchange rate is more depreciated than that in period T, the period $T - 2$ exchange rate must be more depreciated than that in period $T - 1$, and so on. It follows that the monetary shock has more

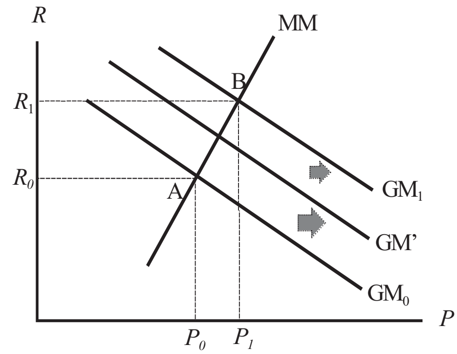


Figure 17.9. Impact effects of a two-period increase in the world interest rate

powerful effects on the price level and the exchange rate and weaker effects on the domestic interest rate, the closer we get to the present, and similarly, that monetary shocks will have stronger effects on the price level, real output, and the exchange rate, but weaker effects on the domestic interest rate on impact, the longer such shocks are expected to last.

c. An Increase in the Foreign Interest Rate

Finally, consider a two-period increase in the foreign interest rate. By now, the analysis should be familiar. In the current period, the shock is unanticipated and affects the economy through two channels: the direct effect of the higher world interest rate and the indirect effect of the higher interest rate expected to persist into the next period on today's expectation of next period's exchange rate. The higher interest rate today causes the GM curve to shift to the right, as in Figure 17.9. The higher world interest rate in period +1 is an *anticipated* shock that causes the exchange rate to be more depreciated in period *+1* than it would have been in the absence of the shock. This increase in S^e_{+1} causes GM to shift even further to the right. The new current-period equilibrium is therefore at the point B in Figure 17.9.

Notice that the current exchange rate is more depreciated relative to the non-shock exchange rate than the exchange will be next period. Again, for an interest rate shock that lasts T periods, the same relationship will hold between the exchange rates in period $T - 1$ and period T, and for reasons similar to those discussed in the analyses of the previous shocks, the exchange rate becomes progressively more depreciated as we move closer from period T to the present. The upshot is that the indirect effect on the economy today that operates through the secondary shift in the GM curve caused by a change in the expectation of next period's exchange rate is larger the longer the shock is expected to last and largest when the shock is permanent.

V. SUMMARY

In this chapter, we have modified the macroeconomic model that we developed in Part 2 to investigate how an economy that is highly integrated with world financial markets operates under floating exchange rates. The key change is that the exchange rate becomes an endogenous variable under floating exchange rates, while the central bank's stock of foreign exchange reserves is a policy variable. The analysis of the model is complicated by the fact that the expected future exchange rate also becomes an endogenous variable under floating rates. We handled that problem by making three assumptions:

1. Shocks to the economy arrive unexpectedly in the current period and are of known duration.
2. No new shocks are expected to arrive in future periods.
3. Expectations about the future value of the exchange rate are formed rationally (i.e., using the model itself).

Using these assumptions, we derived the asset-market approach to the exchange rate and embedded it in a complete floating exchange rate model of an emerging economy.

We found that the economy behaves very differently under floating exchange rates than under fixed exchange rates. Most important, monetary policy remains an effective tool to affect aggregate demand even when capital mobility is perfect, fiscal policy becomes weaker because it "crowds out" net exports, and the effects of changes in the world interest rate on the domestic economy are the opposite of what we found them to be under fixed exchange rates. All these effects become more pronounced the longer each of these respective shocks is expected to last.

REVIEW QUESTIONS

1. Under what conditions is it possible to hold the expected future exchange rate constant when exploring the effects of shocks on the current equilibrium exchange rate?
2. What is the asset-market approach to the exchange rate? What is the role of current account transactions in exchange rate determination under this approach?
3. What is the transmission mechanism for monetary policy under floating exchange rates? How do the effects of monetary policy under floating rates differ from those under fixed rates?
4. Compare the effects of foreign financial shocks on real gross domestic product (*GDP*) under fixed and floating exchange rates, assuming perfect capital mobility in both cases.

5. Compare the impact effects on equilibrium real *GDP* of an increase in government spending that is expected to be transitory to one that is expected to be permanent. Why is there a difference?

EXERCISES

1. What is the asset-market approach to the exchange rate? Explain in intuitive terms how changes in the domestic interest rate affect the equilibrium value of the exchange rate under a clean float, assuming that expectations about the next period's exchange rate are unchanged.
2. Consider a small open economy characterized by perfect capital mobility. Explain why the effects on the domestic price level of an unanticipated monetary expansion differ when the central bank maintains a floating exchange rate from what they would be under a fixed rate.
3. A target zone is an exchange rate regime under which the exchange rate is free to move within specified bands. Use the model developed in this chapter to explain how a central bank would maintain such a regime in the face of goods-market shocks, shocks to money demand, and external financial shocks.
4. In December 1994, Mexico experienced a currency crisis that caused it to abandon its fixed exchange rate regime in favor of a floating exchange rate. Assume that Mexico is perfectly integrated financially with the United States. Explain how the effects on the Mexican economy of an increase in U.S. interest rates would differ before and after December 1994.
5. A small open economy that is perfectly integrated with international financial markets is accused by its trading partners of manipulating its exchange rate to make its products more competitive in the international market. On the basis of what you have learned in this chapter, is it possible to assess the truth of this accusation by examining the amount of foreign exchange reserves accumulated by the economy's central bank? Explain why or why not.
6. In the early 1990s, Germany maintained a floating exchange rate against the rest of the world, while Italy maintained a fixed exchange rate against Germany. German reunification resulted in a fiscal expansion and monetary contraction in Germany at that time. Assuming that that mix of policies was perceived as transitory, and that real GDP was left unchanged in Germany, describe the implications of those policies for Italy.

REFERENCES AND FURTHER READING

Borensztein, Eduardo, Jeromin Zettelmeyer, and Thomas Philippon (2001), "Monetary Independence in Emerging Markets: Does the Exchange Rate Regime Make a Difference?" working paper WP/01/1, International Monetary Fund.

Calvo, Guillermo, and Carmen Reinhart (2002), "Fear of Floating," *Quarterly Journal of Economics*, vol. 117, pp. 379–408.

Dooley, Michael P., and Peter Isard (1983), "The Portfolio-Balance Model of Exchange Rates and Some Structural Estimates of the Risk Premium," *IMF Staff Papers*, vol. 30, pp. 683–702.

Edwards, Sebastian, and Eduardo Levy Yeyati (2003), "Flexible Exchange Rates as Shock Absorbers," working paper 9867, National Bureau of Economic Research.

Frankel, Jeffrey, Sergio Schmukler, and Luis Serven (2002), "Global Transmission of Interest Rates: Monetary Independence and Currency Regime," working paper 8828, National Bureau of Economic Research.

Goldfajn, Ilan, and Gino Olivares (2001), "Can Flexible Exchange Rates Still 'Work' in Financially Open Economies?" *discussion paper* 8, Group of 24.

Ito, Hiro, and Menzie Chinn (2007), "Price-Based Measurement of Financial Globalization: A Cross-Country Study of Interest Rate Parity," unpublished manuscript.

MacDonald, Ronald, and Mark P. Taylor (1992), "Exchange Rate Economics: A Survey," *IMF Staff Papers*, vol. 39, pp. 1–57.

Meese, Richard (1990), "Currency Fluctuations in the Post-Bretton Woods Era," *Journal of Economic Perspectives*, vol. 4, pp. 117–134.

Obstfeld, Maurice (1985), "Floating Exchange Rates: Experience and Prospects," *Brookings Papers on Economic Activity*, vol. 2, pp. 369–450.

Obstfeld, Maurice (1995), "International Currency Experience: New Lessons and Lessons Relearned," *Brookings Papers on Economic Activity*, vol. 1, pp. 119–191.

Reinhart, Carmen (2000), "The Mirage of Floating Exchange Rates," *American Economic Review Papers and Proceedings*, vol. 90, pp. 65–70.

Shambaugh, Jay (2003), "The Effect of Fixed Exchange Rates on Monetary Policy," *Quarterly Journal of Economics*, vol. 119, pp. 301–352.

Taylor, Mark (1995), "The Economics of Exchange Rates," *Journal of Economic Literature*, vol. 33, pp. 13–47.

18

Exchange Rate Regimes

We saw in Chapter 3 that the avoidance of persistent real exchange rate misalignment should be an important macroeconomic objective for emerging economies. The most important policy decision that a country can make in pursuit of this objective concerns the management of the nominal exchange rate. And in managing the nominal exchange rate, the most fundamental decision that has to be made is the choice of exchange rate regime.

The considerations that govern the choice of exchange rate regime are subject to influence by the surrounding domestic and international macroeconomic environment. In the context of emerging economies, an important aspect of that environment is capital account liberalization. As we will see in Chapter 22, technological and institutional developments in the international economy, as well as capital account liberalization in many emerging economies, has caused many of these economies to become strongly integrated with world financial markets. This development has important implications for macroeconomic management in such countries, not least because, as we have seen, the well-known impossible trinity of open-economy macroeconomics implies that perfect international capital mobility, monetary autonomy, and an officially determined exchange rate cannot coexist. The implication is that capital account liberalization may place restrictions on the types of exchange rate arrangements that are feasible for emerging economies.

This chapter will explore these issues and review arguments for alternative exchange rate arrangements in emerging economies under conditions of high capital mobility. It will focus specifically on two types of arrangements that have recently been adopted among some emerging economies and advocated for others: currency boards and floating exchange rates. These represent different choices among the options offered by the impossible trinity and have very different implications for domestic macroeconomic management.

It is worth noting at the outset that though the subject of exchange rate management is an old one (going back at least to David Hume), it is a difficult topic on which there has always been substantial disagreement and on which there has been a very active debate in recent years. There are probably two main reasons why this subject has proven to be so controversial. An old reason is that though the exchange rate is a nominal variable (a price expressed in terms of units of domestic currency), what we ultimately care about are its *real* effects. The latter inevitably depend on the behavior of other nominal prices in the economy. Unfortunately, the behavior of nominal prices is the issue on which there has been the most extensive disagreement among economists since the time of Keynes.

The second reason is more recent. As we saw in Chapter 14, it turns out that well-meaning central banks may have an incentive to announce one policy and do another (the time-inconsistency problem) and that the public's reaction to this situation may result in a worse macroeconomic equilibrium than what would emerge if the central bank could commit itself to follow through on its policy announcements. As we also saw in Chapter 14, whether central banks follow through on their policy intentions may actually depend on the kind of exchange rate policy they adopt. In other words, when analyzing exchange rate policy, we often have to consider the behavior of the monetary authorities as an *endogenous* variable. This complicates things because the analysis of the behavior of the monetary authorities as an endogenous phenomenon is a relatively new topic, one that is still not well understood. In short, the issue of exchange rate policy is tied up with some important and very problematic areas in macroeconomics.

This chapter begins with an overview of the alternative exchange rate regime choices that are available to emerging and developing economies. Section I also examines the relationship between exchange rate regimes and monetary policy. Section II sets out the framework within which the choice of exchange rate regime has to be made by identifying several macroeconomic principles that operate as constraints on exchange rate policy under every regime. Given this framework, the next step in exploring the determination of the optimal exchange rate regime is to spell out the criteria that have been used to evaluate the desirability of alternative regimes. These criteria are described in Section III. The evaluation of how alternative exchange rate regimes fare under these optimality criteria is divided into two parts. Section IV compares the performance of alternative regimes according to criteria that involve a direct effect on development, whereas Section V turns to criteria that involve indirect effects on development through macro stability. On the basis of the results of those sections, Section VI considers the case for and against extreme exchange rate regimes. These regimes have the common feature that all decisions about exchange rate policy are made at the time the regime is chosen, with no decisions about the day-to-day management of the exchange rate being required. The day-to-day management of the exchange rate in intermediate exchange rate regimes is the subject of Chapter 19.

Box 18.1. Varieties of Exchange Rate Regimes

There is a broad range of exchange rate regimes from which countries can choose, and there are many ways to compile a list of possible regimes. One such list is presented here, with regimes listed according to the extent of central-bank intervention in the foreign exchange market, ranging from least intervention at the top to most at the bottom:

1. A clean float, with the exchange rate determined in the market alone and no central-bank intervention.
2. Floating with a "feedback rule" for intervention – in this case, the central bank may be active in the foreign exchange market, but it intervenes in the market only to smooth exchange rate movements, with no systematic accumulation or depletion of its foreign exchange reserves.
3. A managed float. In this case, intervention is more systematic and results in changes in reserves, but the central bank does not announce an exchange rate target.
4. A "band, basket, and crawl (BBC)" system. This is a regime with an announced exchange rate peg that can crawl according to either a backward- or forward-looking rule, that can be adjusted in response to changes in fundamentals, that can have hard or soft margins, and that can accommodate a variety of intervention patterns within the band.
5. A target zone. This is a special case of a BBC regime with a fixed central parity, hard bands, and no intervention within the bands.
6. A fixed, but adjustable peg. This is essentially a very narrow band, with fluctuations of the exchange rate limited to ± 2.25 percent on either side of the central parity.
7. Currency board.
8. Dollarization or monetary unification.

I. FOREIGN EXCHANGE REGIMES, EXCHANGE RATE MANAGEMENT, AND MONETARY POLICY

The most direct and obvious way that the central bank can manage the exchange rate is through its behavior in the foreign exchange market (i.e., by buying or selling foreign exchange, as in the model of Part 2). When we speak of the exchange rate *regime*, we are typically referring to the rules that govern direct central-bank intervention in this market. Box 18.1 provides a menu of exchange rate regimes from which central banks can choose, ranked according to the degree of central-bank activism in the foreign exchange market, from least (clean float with no central-bank activity in the foreign exchange market) to most (dollarization, in which the foreign exchange market is effectively abolished). Regimes 1–3 in the table can loosely be referred to as floating exchange rates and 4–8 as fixed rates. Within the fixed exchange rate regimes, 4–6 can be classified as *soft* (adjustable) exchange rate pegs and 7–8 as *hard* (theoretically immutable) pegs.

However, the central bank can also influence the foreign exchange market indirectly – that is, without buying or selling foreign exchange itself – through what it does in other markets. It can do so because the behavior of the central bank in other markets can influence the behavior of private agents in the foreign exchange market. Specifically, as we saw in the floating exchange rate model of the preceding chapter, the central bank can influence the exchange rate through its conduct of monetary policy. This involves the buying or selling of domestic securities (intervention in the domestic securities market) rather than of foreign exchange.

It is worth noting that the central bank does not always manage the exchange rate, either directly or indirectly. Specifically, there are two extreme cases in which it does not:

1. Under cleanly floating exchange rates, the central bank does not intervene in the foreign exchange market in any way – it neither buys nor sells foreign exchange. Or, if it buys or sells foreign exchange because the public sector needs to acquire it or needs to convert foreign exchange earnings into domestic currency, it does so without trying to influence the exchange rate; that is, it acts as any private bank would do on behalf of the government. The extreme case of a floating exchange rate system occurs when the central bank not only does not intervene in the foreign exchange market on its own account (a clean float) but when, in addition, *it does not allow its monetary policy to be influenced by what happens in the foreign exchange market.* This can happen, for example, when monetary policy is directed at fixing the path of a monetary aggregate, such as the monetary base or M1, or at fixing the path of the price level, without taking the exchange rate into account.
2. But these are not the only circumstances under which the central bank does not actively manage the exchange rate. Ironically, this situation also prevails under the polar opposite fixed exchange rate regime. As we saw in the model of Part 2, a fixed exchange rate is an extreme form of intervention in the foreign exchange market, in which the central bank announces that it stands ready to buy or sell any quantity of foreign exchange that agents wish to trade at some predetermined and unchanging price. The polar case occurs when the central bank not only fixes the exchange rate unalterably but, in addition, allows its monetary policy to be entirely determined by this goal, either not extending credit to domestic agents at all or tying its domestic credit policy rigidly to the behavior of its foreign exchange reserves. This is, in effect, what happens under a currency board, as we discussed in Chapter 14. We will discuss currency boards more extensively later in this chapter. We also saw in Chapter 14 that a currency board is taken to an extreme under a full currency union or dollarization.[1]

[1] The difference between currency boards and currency unions has to do with their differing fiscal consequences (the availability of seignorage revenue) and with credibility issues (the difficulty of altering the exchange rate), not with their effects on day-to-day management of the exchange rate policy or monetary policy.

This discussion, then, suggests that there is an important relationship between exchange rate policy and monetary policy. In the extreme cases in which (1) the central bank does not intervene directly in the foreign exchange market and, in addition, does not adjust monetary policy to influence the behavior of the exchange rate – that is, it directs monetary policy entirely to domestic objectives – or (2) the central bank intervenes to fix the exchange rate indefinitely, and in addition, it directs monetary policy completely to this objective (which is an external objective), there are no day-to-day exchange rate policy decisions to be made. All decisions on exchange rate policy are taken at the time that either of these extreme exchange rate regimes is chosen. But in all other possible cases, the central bank has to decide how to coordinate intervention in the foreign exchange market with monetary policy to influence both domestic and external policy objectives.

The implications are that when it comes to exchange rate policy, the monetary authorities have to make two independent decisions:

1. They must decide whether to adopt an extreme exchange regime, in which there are no further issues of exchange rate management.
2. If they do not, and they opt instead for some form of intermediate exchange rate regime, then they have to decide how to manage the exchange rate on a continuing basis.

The first decision is the subject of the sections that follow, whereas the second is taken up in the next chapter.

II. CONSTRAINTS ON EXCHANGE RATE MANAGEMENT

The decisions just described have to respect certain universal macroeconomic principles, which essentially function as constraints on the choices that the central bank can make. We can classify these principles as those that apply only in the medium run and those that operate in the short run as well.

1. Medium-Run Constraints

a. Nominal Anchors in Open Economies

The first principle is that every economy that uses a monetary unit of account must have a nominal anchor (i.e., a variable that determines the average level of domestic prices in terms of domestic currency). Because establishing a nominal anchor is a public good, it is properly a task for the government to undertake. To do so, the government has to make use of a nominal instrument; that is, it has to fix some nominal price or nominal quantity. The reason is that because the economy's general equilibrium determines a set of *relative* prices and *real* quantities, fixing the price of any single good in terms of domestic currency, or fixing the value of any nominal quantity, would determine the prices of all other goods and services in terms of domestic currency.

This observation has two important implications for exchange rate policy:

1. The decision made with respect to the foreign exchange regime will determine the economy's nominal anchor, that is, the variable that will determine the domestic price level in the medium run. If a fixed, or officially determined, exchange rate is chosen, the nominal anchor will be the nominal exchange rate, as we saw in the context of the model in Part 2 (Chapter 8). If the exchange rate is allowed to float, on the other hand, the nominal anchor must be something else. It could, for example, be the money supply, or some combination of nominal variables that could in principle vary over time, as discussed in our examination of inflation targeting in Chapter 15.
2. The choice of exchange rate regime and nominal anchor management will have important effects on the average price level, so exchange rate management and inflation can be expected to be closely related.

b. Monetary Neutrality

What difference does it make what is chosen as the nominal anchor? In answering this, we need to distinguish between two subquestions:

1. How does the choice of nominal anchor affect the difference between how the government says it is going to manage the nominal anchor ex ante and what it subsequently actually does ex post?
2. How does the actual ex post behavior of the variable chosen as the nominal anchor affect the economy?

As we saw in Chapter 14, there is no uniformity of opinion among economists regarding the answer to the first of these sub-questions. The next macroeconomic principle that we need to discuss here, however, concerns the second one.

In answering this question, we can break it up into two parts, corresponding to how the actual behavior of the nominal anchor affects two different characteristics of the economy: the average price level and the economy's real equilibrium. The answers that economists give to these questions are governed by the principle of *monetary neutrality*: the economy's medium-run real equilibrium is not affected either by the identity of the nominal anchor (the particular variable that is chosen as the anchor) or by the specific value that is chosen for the nominal anchor. This implies that the domestic price level must be proportional to the level of the nominal anchor in the long run. The reason behind this is that economists tend to believe that people do not suffer from *money illusion*; that is, what they care about are *relative* prices and *real* quantities, not prices and quantities measured in nominal terms. If this is so, then the economy's equilibrium should not be affected by a change in nominal magnitudes because neither preferences nor production opportunities are affected by the units in which goods and services are measured.

Box 18.2. Nominal Anchors and Real Exchange Rate Targets

Can the monetary authorities choose the real exchange rate itself as a target? There are indeed countries in which the authorities have fixed a value for the real exchange rate. This was the case, for example, in Indonesia before the 1997 Asian crisis, for Chile during the 1980s, and for Costa Rica until quite recently. Real exchange rate targets are implemented by adopting a *crawling peg* exchange rate regime (a peg that is changed continuously over time), in which the nominal exchange rate is depreciated by the difference between domestic and foreign inflation, a so-called *purchasing-power parity* (*PPP*)-*based* rule. How can we explain this phenomenon in light of the analysis in this section?

The answer is that to choose the real exchange rate as a policy target (e.g., by following a PPP-based exchange rate rule), a country has to dedicate a *real* policy instrument to this goal. This policy instrument has to be one of the fundamental real determinants of the long-run real exchange rate that we discussed in Chapter 15. Note that among the possible candidates for this role is the rate of inflation, which is a real variable because it is a form of tax. If the government does *not* adjust some other real policy and nonetheless tries to fix the real exchange rate, the implication of the analysis in this section is that the rate of inflation will adjust endogenously. That being so, the paths followed by the nominal exchange rate and stock of money must become endogenous (see Montiel and Ostry 1991).

This has an important implication for exchange rate policy: *the real exchange rate is not affected in the long run by the value of the nominal anchor*. As we have seen, the real exchange rate is the price of foreign goods and services in terms of domestic goods and services. In the model of Part 2, it was given by $Q = SP^*/P$. We have seen throughout this book that this is an important relative price for the economy because it affects production and spending decisions that determine resource allocation. What this corollary says is that the real exchange rate is an endogenous macroeconomic variable that in the medium run does not depend on the exchange rate regime, the conduct of nominal exchange rate policy, or the specific value of the nominal exchange rate. What exchange rate policy does affect in the medium run is the absolute level of prices in domestic currency, not relative prices.

Despite this result, some emerging and developing countries have at times chosen to conduct their exchange rate policies to target the real exchange rate. Box 18.2 explains how this policy can be reconciled with the result we have just derived.

c. Fiscal Solvency

We saw earlier that monetary neutrality implies that, depending on the choice of nominal anchor, exchange rate policy or monetary policy determines the domestic price level in the long run. That means that it must also determine the long-run rate of inflation. But we have seen previously (Chapter 9) that for the public sector to be solvent, the domestic rate of inflation (determined by the rate of depreciation of an officially determined exchange rate or by the rate of growth of the money supply) has to be consistent with the fiscal accounts, in the form of the government's intertemporal budget plans. What this means is that the path chosen for the nominal anchor has fiscal implications; that is, nominal anchor policy and fiscal policy are related through the government's intertemporal budget constraint.

Recall from Chapter 9 that for the consolidated public sector to be seen as solvent, actual and potential creditors have to expect that the present value of future primary surpluses plus seignorage revenue, discounted at the safe rate of interest, should be at least equal to the face value of the existing debt. Because the existing debt is predetermined at any given moment, the expected future path of the primary surplus has to be consistent with the expected future path of inflation in the long run.

We can now combine principles 1–3. Together they imply that there has to be consistency between the path of the primary fiscal deficit over time, the domestic rate of inflation, and the rate of change of the economy's nominal anchor. This means that if the government, for example, determines a given path for its primary deficit, it can choose either the exchange rate or the money supply as the economy's nominal anchor. But the rate of growth of whatever variable the government chooses must be consistent with the rate of inflation determined by the fiscal deficit. The other nominal variable must adjust endogenously. On the other hand, the government can choose the rate of change of the nominal anchor. This determines the rate of inflation in the long run. Then the government has to adjust its fiscal deficit to this rate of inflation, and the other nominal variable will adjust endogenously. Notice that in both cases, the real exchange rate will adjust endogenously.

2. A Short-Run Constraint: The "Impossible Trinity"

The three previous observations refer to restrictions that have to be obeyed by fiscal, exchange rate, and monetary policies in the medium run. The last restriction to be discussed in this section is one that must apply even in the short run.

Recall that one consequence of monetary neutrality was that the paths followed by the money supply and the nominal exchange could not be chosen independently of each other in the long run. If one is chosen exogenously, the other is determined endogenously. We have previously encountered, however, an additional constraint that must be respected even in the short run: *perfect capital mobility, a fixed exchange rate, and monetary autonomy cannot coexist, even in the short run.* As we have seen, this is known as the impossible trinity of open-economy macroeconomics.

Why is it true? It holds because, as we saw in Chapter 6, with perfect capital mobility, the elasticity of capital flows with respect to the differential between domestic and foreign interest rates becomes infinite. Under these circumstances, if the government allows the exchange rate to float, as in the last chapter, it can control the money supply because it is not committed to buying or selling foreign exchange. That means that changes in the monetary base can only happen at the government's own initiative. On the other hand, if the government wants to maintain a fixed exchange rate, it cannot sustain a money supply that would produce an interest rate different from the world rate because that would produce an infinite excess demand or supply for foreign exchange reserves.

It may be worth noting that perfect capital mobility requires two conditions: that domestic and foreign interest-bearing assets are perfect substitutes and that there are no policy-imposed barriers to capital movements. Free capital mobility just refers to the latter. Thus the condition does *not* say that free capital mobility, fixed exchange rates, and monetary autonomy cannot coexist because different risk characteristics between domestic and foreign interest-bearing assets may cause them to be considered imperfect substitutes, even in the absence of policy-imposed barriers to capital movements.

The importance of the impossible trinity arises because it means that, if there are no natural reasons that cause assets in different currencies to be treated as imperfect substitutes, the government will only be able to choose among two of the three options: free capital mobility, monetary autonomy (control over the domestic money supply), and an officially determined exchange rate. Thus this constraint tells us the conditions under which a country can have both an exchange rate policy and an independent monetary policy in the short run. The necessary condition is that it be imperfectly integrated with world capital markets. There is another way to put this that is also informative: the exchange rate and the money supply are only independent policy instruments in the short run when capital mobility is imperfect. When this does not hold – that is, when capital mobility is perfect – the choice of the exchange rate dictates the choice of the money supply, and the choice of the money supply determines the exchange rate (at the value that satisfies uncovered interest parity). Because the choice of one is determined by the choice of the other, they cannot be chosen independently and do not represent different policy instruments.

III. CRITERIA FOR CHOOSING AMONG EXCHANGE RATE REGIMES

Given the constraints that exchange rate regimes must respect, we can now turn to the objectives that they have been intended to achieve. These objectives can be used as criteria for evaluating alternative exchange rate regimes. In practice, multiple independent criteria have been widely used in the analysis of optimal exchange rate regimes. As a point of departure in specifying these objectives, we can begin by noting the roles that exchange rate policy can potentially play in promoting growth and development.

Consider first the *direct* roles. There are several such roles. First, as we saw in Chapter 16, the real exchange rate is a key macroeconomic relative price. Thus exchange rate policy can play an important direct role in development by ensuring that the economy receives clear signals about the equilibrium value of the real exchange rate. Second, the existence of different currencies in different countries creates transaction costs in international commerce associated with currency conversion and exchange rate uncertainty. By mitigating such costs, exchange rate policy can promote openness in the form of increased international commerce and capital flows. Finally, as we saw in the last section, the exchange rate can provide an alternative nominal anchor for the economy. When the exchange rate is cast in that role, exchange rate policy can enhance development prospects by promoting long-run price stability and financial development.[2]

As with other macroeconomic policies, the *indirect* role of exchange rate policy in promoting development is through the safeguarding of macroeconomic stability. It can do so in several ways. One way is through the do-no-harm principle, that is, by avoiding the exchange rate itself becoming a source of macroeconomic instability, as happens in the context of currency crises (Chapter 27). Additionally, as we will discuss later, like fiscal and monetary policy, exchange rate policy can help to stabilize aggregate demand in response to exogenous shocks.

1. The Implications of Multiple Objectives

These diverse roles that exchange rate policy can play make for multiple potential objectives that can be sought in choosing an optimal exchange rate regime, and thus for multiple criteria that can be used in evaluating the desirability of alternative regimes. The existence of multiple criteria for evaluating exchange rate regimes has an important implication: it means that even if we can come to the conclusion that a single regime best satisfies a specific criterion, the same regime may not do so for all possible criteria. If that is so, in choosing a specific regime for an individual country, we may face a need to choose among alternative desirable objectives, implying that the optimal choice of regime may depend on which objective is most important for that country at a particular moment in time.

Given the roles of exchange rate policy mentioned earlier, this means that in reaching a judgment about the optimal exchange rate regime for a particular country, we would have to pose questions such as the following:

- How critical is the structure of production (the allocation of resources between goods for domestic consumption and those for export) to the country's development strategy?

[2] Financial development is the subject of Part 6 of this book.

- How critical is it for the country to become more closely integrated with a dominant trading partner both commercially and financially?
- How costly is high inflation to the country, and how large is the gain in lower inflation when one switches from one exchange rate regime to another?[3]
- How large are shocks to the economy, how costly would macroeconomic adjustment be without an explicit policy response, how productive is the exchange rate regime in facilitating adjustment, and what other policies are available to facilitate adjustment to shocks?

As mentioned earlier, the important point is that, in light of considerations like these, countries with different preferences or facing different circumstances may choose different exchange rate regimes, and the same country may choose different regimes at different points in time as its circumstances change.[4]

In the next section, we will consider the implications that each of the potential *direct* effects of exchange rate policy on development may have for optimal regime choice, before turning to the indirect effects in the section that follows.

IV. DIRECT EFFECTS OF EXCHANGE RATE POLICY ON DEVELOPMENT

1. Observability of the Equilibrium Real Exchange Rate

As mentioned in Box 18.2, several emerging and developing countries have in the past opted for real exchange rate targeting. But even for those that have not explicitly done so, the real exchange rate has not been a matter of indifference. In particular, in the past, many developing countries, particularly in East Asia, have pursued *outward-oriented development strategies*, a development strategy focusing on the production of manufactured goods for export. China has been the most prominent example of a country pursuing such a strategy during recent years.[5] For such countries, a relatively depreciated real exchange rate is important in sustaining the competitiveness of manufactured exports, and for a depreciated real exchange rate to play that role effectively, it has to be easily observable in the marketplace so that it can provide an appropriate relative price signal.

Countries in this situation are likely to find BBC exchange rate arrangements particularly attractive because such arrangements limit the variability of the real exchange rate while allowing the actual real exchange rate to be managed – at

[3] We will see later that the gain may arise from differential properties of exchange rate regimes as precommitment mechanisms for the central bank. This gain may not be large, e.g., if the central bank already has an independent anti-inflationary reputation.

[4] E.g., we have already seen that the exchange rate may play a role during the process of stabilizing from high inflation. Once stabilization is completed, exchange rate–based stabilization may give way to "flexibilization."

[5] See Dooley et al. (2003).

least in principle – to closely track the equilibrium rate (see Williamson 2000). A second-best option for such countries is to adopt a managed float, which can be managed to stabilize the real exchange rate around some (unannounced but widely understood) target.

2. Reducing Transactions Costs to Promote International Commerce and Capital Flows

By contrast, countries that attach a high priority to economic and financial integration with a dominant trading partner would be led to make a different exchange rate choice. This criterion acknowledges that the nature of the foreign exchange regime may affect the costs of doing business across international boundaries, through the resources involved in maintaining the currency conversion mechanism as well as through the uncertainty created by possible exchange rate changes, and regards a regime as superior if it minimizes these costs, which act as a tax on international commerce.

This criterion amounts to an argument for a single currency to eliminate the costs of currency conversion and exchange rate risks, through dollarization or currency unification. The criterion has been applied most notably in the arguments for a single currency in the European Union. The real question about it is how large these costs really are and how much of a disincentive they provide to international commerce. Recent research suggests they may be significant (Rose 2000).

3. Medium-Run Inflation Stabilization

The third direct effect of exchange rate policy on development has to do with medium-run price-level performance. It concerns the effects that the exchange rate regime may have on the medium-run equilibrium rate of inflation, with an exchange rate regime judged superior if it results in a lower medium-run rate of inflation.

As we saw in Chapter 13, macroeconomists agree that the medium-run Phillips curve is vertical – that is, the economy's full-employment equilibrium can be associated with *any* sustained rate of inflation – and that high inflation has substantial economic costs. These two observations together imply that an important target for monetary and exchange rate policy should be to achieve a medium-run macroeconomic equilibrium with a relatively low rate of inflation. But we have also seen that an economy's medium-run rate of inflation is an endogenous variable that depends on the path followed by the economy's nominal anchor. Moreover, in principle, the same rate of inflation could be consistent with different choices of nominal anchor (i.e., the exchange rate or the money supply), so why should we believe that the exchange rate regime should affect the long-run rate of inflation?

The reason has to do with the analysis of the time inconsistency of optimal policies that we discussed in Chapter 14. Recall the general setting: a well-meaning central bank chooses the economy's optimal rate of inflation based on a social objective function in which an unanticipated increase in the rate of inflation is socially beneficial, and in which the social costs of *actual* inflation rise as a nonlinear function of the rate of inflation. The private sector forms its expectations about the rate of inflation and acts on them by setting wages and prices, before (less often than) the central bank makes its decision about the nominal policy instrument.

We saw that under these circumstances, the central bank may find it advantageous to announce a low rate of inflation and, after the private sector has acted on these expectations, to produce an inflationary surprise to improve social welfare. The problem was that the private sector knew that the central bank had incentives to do this, and consequently, the private sector expected the inflation rate to be greater than that announced by the central bank. When inflation expectations were high, the central bank would no longer find it appropriate to comply with its original announcement because this could create large social costs. The equilibrium in this situation was one in which the inflation rate had to be sufficiently high to deprive the central bank of any incentive to try to spring an inflation surprise on the economy. But the resulting social outcome was suboptimal – the economy would bear the costs of a high level of actual inflation without any of the benefits of surprise inflation. The problem in this situation was the central bank's lack of credibility; that is, the central bank had no way to commit itself to actually doing what it had promised to do. To solve this problem, it was necessary to find a mechanism that could commit the central bank to fulfilling its promises. The question is whether an official exchange rate announcement – and therefore a fixed exchange rate – could provide such a mechanism.

As we saw in Chapter 14, theory does not provide a clear-cut answer to this question. However, because there is substantial international experience with alternative exchange rate regimes, it would seem straightforward to examine the evidence on this matter to settle the issue empirically, for example, by simply running regressions on long-run average inflation rates on a set of inflation determinants as well as on the nature of the exchange rate regime.

Though such investigations have indeed been undertaken, their results are unfortunately not easy to interpret. Complications arise from three sources. First, it is difficult to classify many regimes as either fixed or flexible. This is both because there are many intermediate regimes, as illustrated in Box 18.1, and because countries often report to international institutions that maintain such data that they are observing one regime while, in practice, behaving in accordance to another (e.g., declaring that they maintain a freely floating regime while intervening heavily to stabilize the exchange rate).[6] Second, the set of third factors that may influence

[6] There is a large body of literature attempting to distinguish de facto from de jure exchange rate regimes; see, e.g., Calvo and Reinhart (2002), Reinhart and Rogoff (2004), Levy-Yeyati and Sturzenegger (2005), and Frankel and Wei (2008). Tavlas et al. (2008) provide a useful review of this literature.

inflation together with the exchange rate regime is extensive, and it is difficult to be confident that all these have been controlled for. Finally, the exchange rate regime may be endogenous to the country's inflation performance, if countries with a preference for low inflation for other reasons choose one exchange rate regime over the other.

A recent study by Ghosh et al. (2003) is particularly careful about all these issues. Using a sample of 140 countries with data over 1960–1990, these authors found that the rate of inflation tended to be lower in countries that maintained a fixed exchange rate. On the other hand, the growth rate was lower in countries with fixed exchange rates, despite these countries having higher average rates of investment. This means that they had lower average productivity growth.[7] The conclusion would seem to be that fixed exchange rate regimes (i.e., regimes in the range 4–8 in Box 18.1) are probably more effective than floating rate regimes (those in the range 1–3) as commitment devices for central banks and thus that fixed regimes tend to be more successful in promoting medium-run price stability. However, because of the empirical problems just mentioned, and the scarcity of studies that have addressed them, at this point the evidence is at best suggestive rather than conclusive.

V. CURRENCY CRISES AND THE MACROECONOMIC ADJUSTMENT PROCESS

As argued, exchange rate policy can affect development indirectly through its effect on macroeconomic stability. These effects can emerge when exchange rate policy itself becomes a source of instability (currency crises) or when exchange rate policy is (or is not) used to mitigate instability arising elsewhere in the economy, thus affecting the process of adjustment to macroeconomic shocks. This section considers the implications of these indirect effects for the choice of exchange rate regime.

1. Currency Crises

Currency crises are the subject of Chapter 27, so the treatment here will be brief and will focus specifically on the implications of vulnerability to currency crises for the choice of exchange rate regime. One consequence of the multiple currency crises that emerging economies experienced during the 1990s was indeed the emergence of a particular view of the implications of such crises for optimal exchange rate regimes, referred to as the *bipolar view.*

This view emphasized the interaction between increased financial integration and the nature of currency crises. The argument was that as countries became financially more integrated with the rest of the world, the expectation of a currency devaluation would have stronger impacts on the domestic economy by causing

[7] The authors explain this through a lower rate of growth of trade in these countries.

larger movements in domestic interest rates. In the language of Part 2, the impact of an expected exchange rate devaluation would have a larger positive impact on the domestic interest rate as increased financial integration made the BB curve flatter and more responsive to changes in the rate of return on foreign assets. This prediction is consistent with our model, as we saw in Part 2. Coupled with the perception that the abandonment of an exchange rate peg through a sudden devaluation or a transition to a floating exchange rate is something that central banks *choose* to do, rather than are *forced* to do, this observation suggests that soft exchange rate pegs would become increasingly vulnerable to devaluation or abandonment (i.e., to a *currency crisis*) as financial integration increased because the mere expectation of a devaluation would induce central banks to abandon the prevailing exchange rate peg to protect the domestic economy from contractionary high interest rates.

Under this view, therefore, maintaining a soft exchange rate peg in a context of high capital mobility is a recipe for currency crisis and the attendant macroeconomic instability. The implication for the choice of optimal exchange rate regime is that when capital mobility is high, countries should avoid soft exchange rate pegs that are easily altered or abandoned. Either of the extreme ends of the exchange rate regime spectrum – floating rates or hard pegs – would be less crisis-prone because they would not give rise to expected discrete exchange rate changes. In the case of the former, this is because, under floating rates, exchange rate adjustments would be continuous, and in the case of the latter, it is because, under hard exchange rate pegs, central banks would not have the option to alter the peg or change the regime, making it less likely that an expected devaluation would materialize in the first place. This preference for the two extreme poles of the exchange rate regime spectrum explains the term *bipolar view*.

2. Macroeconomic Adjustment

As we have discussed throughout this book, economies are typically afflicted by a variety of shocks to which they need to adjust, in the sense that such shocks may call for new configurations of relative prices and real quantities. Because the reallocation of productive factors is costly, and because of the role of short-run stickiness in nominal wages and prices, moving from one equilibrium to another is not always a smooth process. The next criterion to consider concerns the effect that the exchange rate regime can have on the path that the economy follows to regain its full-employment equilibrium after it is affected by a shock; that is, the question is what type of exchange rate regime best promotes the adjustment process, in the sense of minimizing adjustment costs.

Shocks can be classified along two important dimensions: their *duration* (i.e., whether they are expected to be temporary or permanent) and their *origin*. As we have seen, shocks to the domestic economy can originate in the domestic or

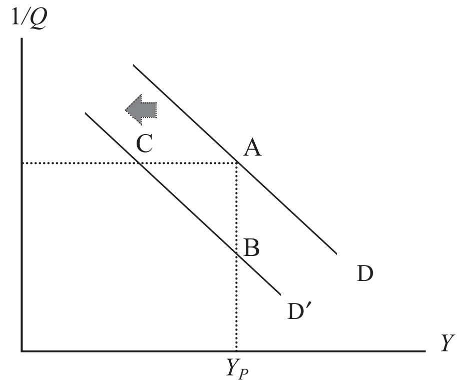


Figure 18.1. Adjustment to a permanent real shock

foreign goods markets (real shocks), in domestic asset markets (nominal shocks), and in foreign asset markets (external financial shocks). To see how the exchange rate regime can play a role in the adjustment to macroeconomic shocks, consider how adjustment would work in response to a permanent real (i.e., goods market) shock that affects the demand for domestic goods adversely (causes it to fall).

In principle, there are several options for adjustment:

1. The relative price of the domestic good could fall (price adjustment).
2. Domestic factors of production could relocate to another country (quantity adjustment).
3. Fiscal policies (undertaken either by the domestic government or by the rest of the world) could be designed to sustain the level of demand for the domestic good (policy response).

These alternative options are illustrated in Figure 18.1. From an initial equilibrium at a point like A, the economy undergoes an anticipated permanent reduction in the demand for its goods.[8] Assuming the vertical aggregate supply curve of Part 2, adjustment through mechanism 1 implies a new equilibrium vertically below the original one (at B), mechanism 2 implies one horizontally to the left of it (at C), and mechanism 3 implies a return to the original equilibrium at A through a policy-induced rightward shift of the demand curve. If neither mechanism 2 nor mechanism 3 can happen (because domestic factors of production are not mobile internationally and there are no extranational institutions that can undertake the required fiscal policies), then adjustment must eventually occur through mechanism 1. This means that the medium-run equilibrium real exchange rate must depreciate. The exchange rate regime determines the extent to which this happens

[8] Note that in Figure 18.1, the real exchange rate is expressed as the price of domestic goods in terms of foreign goods $1/Q$, instead of its inverse, as used in the rest of the book, to permit us to draw the demand curve for the domestic good with a negative slope.

through a change in the nominal exchange rate (an increase in S) or a change in relative price levels (a reduction in P).

What difference does it make how this adjustment in relative prices is brought about? Notice that at A, the economy's real exchange rate is misaligned. The key point about misalignment made in Chapter 16 was that if the price level is sticky, then bringing about a real exchange rate depreciation through adjustment in relative price levels could require a recession, which implies substantial social costs and adds an excess burden to the real income loss caused by the demand shift. Thus adjustment from A to B is less costly if brought about by a change in the nominal exchange rate. This advantage of nominal exchange rate adjustment is not general, however. More generally, as we shall see, the conclusion depends on the source of the shock.

More generally, exchange rate flexibility has two important consequences for the costs of adjustment to macroeconomic shocks:

1. Whether the exchange rate is free to adjust influences the real effects on the economy of macroeconomic shocks in the absence of a discretionary policy response. In general, whether exchange rate adjustment amplifies or reduces these real effects depends on the source of the shock.
2. Moreover, whether the exchange rate adjusts freely also affects the availability of macroeconomic policy instruments that can be used in a discretionary fashion to respond to shocks.

Consider these one at a time.

a. Consequences of Macroeconomic Shocks

To see the implications of the exchange rate regime for the consequences of macroeconomic shocks, suppose initially that the authorities have a single policy objective – consisting of internal balance – and that they are constrained from undertaking a discretionary policy response when shocks arrive. Thus the superior regime is the one that proves to be the best automatic stabilizer, that is, the one that is most successful in stabilizing real output automatically (i.e., without explicit countervailing policy actions) in response to shocks.

The classic analysis of this situation derives the following standard results:

- When nominal shocks dominate (i.e., shocks to the demand or supply of money), fixed exchange rates are better able to stabilize output because the endogenous response of the money supply under fixed rates contains the adjustment in the money market, thereby protecting domestic aggregate demand and production from the effects of such shocks.
- When real shocks dominate (shocks that directly affect the demand or supply of domestic goods), floating exchange rates are superior because they allow

relative prices to adjust without requiring quantity adjustments that would involve changes in domestic output, as in Figure 18.1.

- Foreign financial shocks tend to affect both domestic goods and financial markets. Which regime is better when shocks take this form depends on structural characteristics of the economy such as the interest elasticities of demand for money and goods.

b. Availability of Macroeconomic Instruments

Now suppose that we allow for the possibility of a policy response to shocks and that the authorities care not only about the stability of real output (internal balance) but also about an external balance target such as the outcome for the current account of the balance of payments. Then there may exist an important difference between the two regimes, in the sense that the instruments that are available to achieve these targets – and thus the possibility of achieving them both – may differ between the two regimes.

A basic principle of macroeconomic policy making is that to achieve *n* different policy targets, one needs at least *n* independent policy instruments (*independent* means that they are truly different, in that they have different effects on the targets). As we have already seen, if there is high capital mobility, money and the exchange rate are not independent policy instruments, even in the short run, so we can choose one or the other.

Under these circumstances, there is an important difference between the two instruments: if the central bank chooses the money supply as a policy instrument, it can be used in a discretionary fashion, whereas if it chooses the exchange rate, it typically cannot. Why do these differences exist? The answer is because, as mentioned earlier, under conditions of high capital mobility, fixed but adjustable exchange rate systems tend to be susceptible to currency crises.

The implications for the availability of policy instruments are the following: assuming there is fiscal flexibility, with high capital mobility and fixed exchange rates, the economy would be left with only one policy instrument (fiscal policy) to achieve two targets – internal balance and external balance. This means, for example, that if there is a recession and a current account deficit, the fiscal authorities would have to choose between one objective and the other. With money as the policy instrument, however, different mixes of monetary and fiscal policies can be used to achieve internal and external balance targets. On the other hand, if there is no fiscal flexibility, then fixing the exchange rate means being left with no policy instruments, whereas fixing the money supply means having one instrument to achieve two targets.

To summarize the analysis of exchange rate regimes and adjustment costs, floating exchange rates have the advantage in facilitating adjustment when shocks are real because they permit the government to continue to hit targets for both internal and external balance in the face of such shocks. On the other hand, fixed rates have the advantage when shocks are nominal because they protect both internal and external balance targets from the effects of such shocks by confining these effects to the domestic money market. In the face of external financial shocks, neither regime is unambiguously better in the absence of a discretionary policy response, but floating rates have the advantage that they allow policy more discretion in responding to such shocks by allowing the domestic authorities to bring two policy instruments to bear rather than the single instrument that is available when the nominal exchange rate is fixed.

VI. EXTREME FOREIGN EXCHANGE REGIMES: ARGUMENTS PRO AND CON

In the prior two sections, we considered how the two extreme exchange rate arrangements under consideration fared under each of the two main criteria that have been used to evaluate such regimes. In this section, we take a different cut at the problem by briefly evaluating each of the extreme regimes independently.

1. Currency Boards

As explained in Chapter 14, a currency board is an exchange rate arrangement in which the central bank commits itself to issuing domestic currency only in exchange for foreign exchange reserves at a predetermined and unchanging rate of exchange. This means that the domestic monetary base is backed fully by foreign exchange reserves, and the central bank is unable to issue domestic credit. Currency boards have been advocated for many emerging economies. Among the larger such economies, currency boards have been put in place in Argentina (1991), Bosnia (1998), Bulgaria (1997), Estonia (1992), Hong Kong (1983), and Lithuania (1994). Currency boards have also been advocated at various times for Mexico, Indonesia, and Russia, among major economies that have recently undergone currency crises.

Advocates of this arrangement claim that they assure the convertibility of the currency (because the stock of foreign exchange reserves is more than sufficient to buy back the monetary base); impose macroeconomic discipline, thereby enhancing the credibility of inflation objectives (by denying the government discretionary access to seignorage); and provide a mechanism for automatic balance of payments adjustment (because the monetary implications of current account surpluses or deficits induce automatic offsets through the capital account of the balance of payments). For these reasons, currency boards are perceived as creating confidence in the domestic monetary system, which helps promote trade, investment, and growth.

None of these arguments is fully convincing, however. The convertibility of the domestic currency is not assured if the public perceives that the government would not allow the persistence of the very high interest rates that would be implied by an extreme contraction of the monetary base in the event of wholesale conversions into the foreign currency. The gains from macroeconomic discipline, on the other hand, may not be very important if the central bank has already established an anti-inflationary reputation. Finally, automatic balance of payments adjustment is also available through freely floating exchange rates, so the last advantage cited for currency boards does not distinguish them from fully floating exchange rates.

The arguments against currency boards, on the other hand, emphasize the loss of monetary autonomy that they imply, the loss of control over the inflation tax as a source of revenue, and the loss of the lender-of-last-resort function for the central bank (because, as we have seen, the restriction that the monetary base be fully backed by foreign exchange reserves prevents it from extending domestic credit to commercial banks in times of crisis).[9]

Again, however, the importance of these limitations depends on the circumstances facing the domestic economy. The loss of monetary autonomy is more important when there are no other stabilization policies available and when the central bank is independent and competent, that is, when monetary autonomy would have been used appropriately under an alternative monetary arrangement. The loss of discretion over seignorage revenue is more important when the domestic tax system is very distortionary, and the inability of the central bank to extend credit in fulfilling its lender-of-last-resort function is more important when the domestic financial system is weak. Moreover, it can be mitigated through changes in banks' required reserve ratios and through external lines of credit maintained by the domestic government to finance emergency lending to domestic banks.

Taking a more general perspective, it is not clear that confidence is easy to establish with a currency board because if the economy were to become more vulnerable to shocks without monetary autonomy and a lender of last resort, it would not be easy for the government to convince the market that it will not abandon the system if it is unable to reap the credibility effects that are the system's major advantage. This may create a situation of multiple equilibria, in the following sense: the major advantage of the currency board comes through the benefits of credibility. Thus, if credibility cannot be achieved because the shortcomings of the system are too costly for the country in question, the government will be tempted to abandon it, ratifying the absence of credibility.

The difficulty of achieving credibility has been demonstrated by the two best-known currency boards during recent years: those of Argentina and Hong Kong. Argentina was one of the countries most affected by the Mexican financial crisis of early 1994. Argentina went through a severe crisis in early 1995, as markets saw Mexico-like symptoms (exchange rate appreciation, large current account deficit) in Argentina that raised doubts about the sustainability of the exchange rate peg in that country. The domestic banking sector lost 30 percent of its deposits in the

[9] The lender-of-last-resort function is discussed in more detail in Chapter 22.

first three quarters of 1995, domestic real interest rates rose, unemployment hit 18 percent, and growth slowed. Hong Kong also suffered from financial instability elsewhere. It was buffeted by the Asian crisis in fall 1997, as the sustainability of its peg came into question in view of the exchange rate devaluations undergone by the currencies of countries whose exports competed with those of Hong Kong. The stock market in Hong Kong collapsed, and very high domestic interest rates were required to defend the currency, calling the health of the domestic financial sector into question. Nonetheless, both currency boards withstood these shocks and remained in operation.

To conclude, what do these considerations suggest about the conditions under which the adoption of a currency board would make sense? The answer is that there are indeed circumstances that favor the adoption of currency boards, but these circumstances are far from universal. These circumstances include the following:

1. when fixed exchange rates are very useful because the country is very open and its trade is dominated by a single country with stable prices
2. when gaining credibility is very important
3. when asymmetric real shocks are *not* very important, or if they are, when the monetary authorities are not very competent and the fiscal authorities are (e.g., when fiscal policy is flexible and can be used for countercyclical purposes, perhaps through fiscal transfers from a supranational authority)
4. when domestic wages and prices are likely to become more flexible as a result of the adoption of the currency board, reducing the costs to the economy of adjusting automatically to those real shocks that do affect it
5. when the domestic financial system is strong, or when there are external sources of liquidity (such as lines of credit) in place to handle domestic liquidity crises

2. Floating Exchange Rates

Recall that under a floating exchange rate regime, the central bank does not intervene in the foreign exchange market, which means that the price of foreign exchange is determined completely by the behavior of private agents. In principle, this regime has several advantages. We have seen that it facilitates adjustment when domestic prices are not flexible and when real shocks are important and that it increases the availability of policies by allowing for the preservation of monetary autonomy. Unlike currency boards, floating exchange rates permit the government to retain control over the inflation tax as a source of revenue and permit the central bank to perform the lender-of-last-resort function by using its own resources. Like a currency board, floating exchange rates provide an automatic mechanism for balance of payments adjustment, with the added benefit that they are not vulnerable to speculative attacks.

A disadvantage of floating rates is that long-term price-level credibility depends on internal restraints (such as reputation, as in Japan, or the institution of an independent central bank with a conservative governor, as in pre-euro Germany). But as we have already seen, it is not obvious that in theory, fixed rates are a superior disciplining device, and the evidence on this issue is limited. For these reasons, many economists have been supportive of the adoption of floating exchange rates in principle.

In practice, however, other issues have arisen that call the superiority of this arrangement into question. Under floating rates, the exchange rate is determined by the interaction of supply and demand in the foreign exchange market. Demand arises from the desire of domestic agents to import goods and services or to purchase foreign financial assets, whereas the supply arises from exports or from the sale of domestic financial assets to foreigners.

The important fact is that, as we saw in the previous chapter, the demand for foreign exchange and supply of foreign exchange that arise from the buying and selling of financial assets is much larger than the demand that arises from the buying and selling of goods and services. The implication of this situation is that for the exchange market to be in equilibrium, agents have to be willing to hold the existing composition of their portfolios between assets denominated in foreign exchange and assets denominated in domestic currency. The nominal exchange rate must adjust to make it so. This means that at a given moment in time, the nominal exchange rate has to be consistent with equilibrium in the international market for financial assets. Thus, under floating exchange rates, the nominal exchange rate must behave like what it really is: the price of a financial asset. In the preceding chapter, we referred to this as the asset-market approach to the exchange rate.

Why does this matter? It matters because the prices of financial assets in the present depend in an important way on what those prices are expected to be in the future. That means that if prices are expected to rise in the future, they will rise in the present, because there are profits to be made in buying assets before their prices rise, and that creates demand for those assets in the present. The point is that expectations about the future have an important effect on the prices of financial assets, and because these expectations tend to be volatile, this transmits the volatility to the prices of the financial assets, among them the nominal exchange rate. The question is, is this a good thing or a bad thing?

An implication is that under floating exchange rates, the nominal exchange rate is much more volatile than relative price levels. The effect is that the real exchange rate will be much more variable under a floating-exchange-rate system than under a fixed-rate system. As we saw in the previous chapter, there is substantial evidence that this is the case.

Why should we worry about this? One argument is based on the view that when the real exchange rate is unstable (i.e., when it changes much more than could be justified on the basis of changes in its fundamental determinants), this creates large costs because it is costly to reallocate resources, and if factors of production move from one sector to the other on the basis of the current real exchange rate, too many resources would be wasted in transitory factor reallocations.

But the real problem is probably exactly the opposite. Precisely because changing the allocation of resources is costly, economic agents will not respond if they are faced with real exchange rate changes that may prove to be transitory. This being so, the problem would be not that there would be too much movement of factors from one sector to another but rather that the movement of factors between sectors would not be sufficient due to the noise created by transitory real exchange rate movements. As we saw in Chapter 16, there is empirical evidence that volatility in the real exchange rate has negative effects on investment and growth (Bleaney and Greenaway 2001), and the argument just made is one way to interpret this evidence.

But real exchange rate variability may not necessarily be a bad thing. One argument is based on the presence of distortions in the financial system, which we will study in Part 6. When (explicit or implicit) deposit insurance is in place, prudential regulation is nonexistent or weak, and the net worth of financial institutions is small, we will see that bank owners have an incentive to invest the institution's resources in very risky assets because if those assets turn out to yield high returns, the stockholders will reap large gains, whereas if they fail, the stockholders can only lose the amounts that they have invested in the institution. In this case, the expected gains from the investments from the private (stockholder) viewpoint may be much larger than the expected social gains. This would create a distortion in resource allocation, which may be a very serious matter if financial institutions in such conditions have the opportunity of intermediating a large volume of resources, as they would have if the economy's capital account is open.

This is where the exchange rate system comes in. From the point of view of external creditors, what they care about is the return that deposits pay in foreign exchange. If deposits are guaranteed, and in addition, the government fixes the exchange rate, then investing in the domestic economy through the banking system carries very little risk for those who bring money in from outside. As we will see in Part 6, this has two negative consequences: it can destabilize the economy, through large capital inflows, and it may also immiserize the economy because the existence of these distortions means that the external resources will tend to be misallocated. Under these circumstances, moral hazard–driven external borrowing can be deterred by removing the perceived exchange rate guarantee for external creditors. This can be done by giving greater flexibility to the exchange rate.[10]

[10] E.g., apparently, this was an important element in the formulation of Chile's exchange rate policy during the 1990s.

What, then, are the conditions under which the adoption of floating exchange rates would seem to be indicated? Several such conditions are suggested by the preceding discussion:

1. when long-run price stability is not a problem because the country possesses an independent central bank with a well-established anti-inflationary reputation
2. when real shocks are important, domestic wages and prices are sticky in nominal terms, and monetary and fiscal policies can be deployed flexibly to achieve internal and external balance
3. when short-run real exchange rate variability is not too costly (e.g., because financial instruments are available to trade such risk and allocate it efficiently)

All this suggests that floating exchange rates would be suitable for countries with strong central banks, advanced financial systems, and the ability to use monetary and fiscal policies in a flexible countercyclical manner. Generally, these criteria are more likely to be satisfied by industrial countries or by relatively advanced emerging economies.

VII. SUMMARY

In this chapter, we have taken the view that issues of nominal exchange rate management in emerging economies can be decomposed into two parts: whether to adopt an extreme exchange rate regime, in the form of a fixed exchange rate with monetary policy fully devoted to external objectives (a currency board), or a clean float, with monetary policy fully devoted to domestic objectives, and if not, how to manage the rate in the context of an intermediate regime.

This chapter was devoted to the first question. We saw that exchange rate regimes could be evaluated by criteria such as their effects on the observability of the equilibrium real exchange rate, on the costs of transacting internationally, on long-run domestic price stability, and on macroeconomic stability. Effects on real exchange rate observability favor BBC arrangements, and those on reducing transaction costs favor currency unification. Exchange rate regime prescriptions based on price stability effects are more problematic. The theoretical arguments are ambiguous in this case, but the limited careful empirical evidence on this issue favors fixed exchange rates. With respect to macroeconomic stability, the bipolar view suggests a preference for extreme exchange rate regimes when capital mobility is high. Finally, floating exchange rates seem to have the advantage as automatic stabilizers when macroeconomic shocks are real, fixed rates when they are nominal, and neither when they originate in the external financial environment.

These considerations suggest that extreme exchange rate arrangements may indeed be appropriate for specific economies under well-defined circumstances. When credibility issues are important, domestic macroeconomic management is weak, domestic wage and price flexibility are high, and real shocks are not important, a currency board may be an attractive exchange rate arrangement. On the other hand, when the country possesses a central bank with a strong anti-inflationary reputation, domestic fiscal and monetary policies are well managed, wages and prices exhibit substantial short-run stickiness, and real shocks can be expected to be important, floating exchange rates may be preferable.

The problem is, of course, that many countries do not meet all these conditions. For example, many emerging economies

1. do not have central banks with strong reputations and so would like to use the exchange rate as a nominal anchor
2. want to give some predictability to the real exchange rate but not necessarily to the nominal exchange rate
3. want to retain some monetary autonomy because fiscal policy is not very flexible
4. are concerned about the adequacy of their regulatory and supervisory capabilities in the financial sector and thus want to discourage excessive intermediation of external funds by that sector

For such countries, intermediate exchange rate regimes that try to strike compromises among the various characteristics of the extreme regimes – for example, between the objectives of stabilizing the price level and promoting adjustment – might be appropriate. In the next chapter, we will consider how the exchange rate might be managed under such arrangements.

REVIEW QUESTIONS

1. What are the differences between hard and soft exchange rate pegs? Between managed and clean floats?
2. What criteria have been used to evaluate the desirability of exchange rate regimes? What are the implications of the existence of multiple criteria?
3. What is the bipolar view of exchange rate regimes? Explain the theoretical basis for adopting this view.
4. Make the best arguments you can for and against dollarization as an exchange rate regime.
5. Make the best arguments you can for and against a clean float as an exchange rate regime.

EXERCISES

1. "To stabilize real gross domestic product, countries that are primarily vulnerable to goods market shocks should open their capital accounts when they maintain

a floating exchange rate, and close them when they maintain a fixed exchange rate." Do you agree or disagree? Explain why.

2. Is the bipolar view of exchange rate regimes implied by the impossible trinity? Explain why or why not.
3. Can countries characterized by imperfect capital mobility target the money supply and the nominal exchange rate simultaneously in the short run? In the medium run? Explain.
4. Emerging and developing economies are often criticized for pursuing so-called inconsistent fiscal, monetary, and exchange rate policies. On the basis of the analysis in this chapter, can you explain what this term means? What would it take for these policies to be consistent?
5. What are the differences between currency boards and dollarization as alternative hard pegs? Under what conditions might a country that has opted for a hard peg be led to choose one over the other?

REFERENCES AND FURTHER READING

Bleaney, Michael, and David Greenaway (2001), "The Impact of Terms of Trade and Real Exchange Rate Volatility on Investment and Growth in Sub-Saharan Africa," *Journal of Development Economics*, vol. 65, pp. 491–500.

Calvo, Guillermo, and Carmen Reinhart (2002), "Fear of Floating," *Quarterly Journal of Economics*, vol. 117, pp. 379–408.

Chang, Roberto, and Andres Velasco (2000), "Exchange-Rate Policy for Developing Countries," *American Economic Review Papers and Proceedings*, vol. 90, pp. 71–75.

Corden, W. Max (1994), "Exchange Rate Policy in Developing Countries," in *Approaches to Exchange Rate Policy*, International Monetary Fund Institute (Washington D.C.), pp. 65–89.

Dooley, Michael P., David Folkerts-Landau, and Peter Garber (2003), "An Essay on the Revived Bretton Woods System," working paper 9971, National Bureau of Economic Research.

Frankel, Jeffrey (2003), "Experience of and Lessons from Exchange Rate Regimes in Emerging Economies," working paper 10032, National Bureau of Economic Research.

Frankel, Jeffrey, and Shang-Jin Wei (2008), "Estimation of De Facto Exchange Rate Regimes: Synthesis of the Techniques for Inferring Flexibility and Basket Weights," paper presented at the 8th annual Jacques Polak Research Conference, November.

Ghosh, Atish R., Anne-Marie Gulde, and Holger Wolf (2003), *Exchange Rate Regimes: Choices and Consequences*, Cambridge, MA: MIT Press.

Husain, Aasim, Ashoka Mody, and Kenneth S. Rogoff (2005), "Exchange Rate Regime Durability and Performance in Developing Countries versus Advanced Economies," *Journal of Monetary Economics*, vol. 52, pp. 35–64.

Levy-Yeyati, Eduardo, and Federico Sturzenegger (2005), "Classifying Exchange Rate Regimes: Deeds vs. Words," *European Economic Review*, vol. 49, pp. 1603–1645.

Montiel, Peter J., and Jonathan Ostry (1991), "Macroeconomic Implications of Real Exchange Rate Targeting in Developing Countries," *IMF Staff Papers*, vol. 38, pp. 872–900.

Reinhart, Carmen, and Kenneth S. Rogoff (2004), "The Modern History of Exchange Rate Arrangements: A Reinterpretation," *Quarterly Journal of Economics*, vol. 119, pp. 1–48.

Rogoff, Kenneth, Aasim M. Husain, Ashoka Mody, Robin Brooks, and Nienke Oomes (2004), "Evolution and Performance of Exchange Rate Regimes," occasional paper 229, International Monetary Fund.

Rose, Andrew (2000), "One Money, One Market: The Effects of Common Currencies on Trade," *Economic Policy*, vol. 30, pp. 7–46.

Tavlas, George, Harris Dellas, and Alan C. Stockman (2008), "The Classification and Performance of Alternative Exchange-Rate Systems," *European Economic Review*, vol. 52, pp. 941–963.

Williamson, John (2000), *Exchange Rate Regimes for Emerging Markets: Reviving the Intermediate Option*, Washington, DC: Institute for International Economics.

19

Managing an Officially Determined Rate

In the previous chapter, we reviewed arguments for and against various exchange rate regimes. We saw that in the special cases of clean floats or hard pegs, the authorities have no need to make day-to-day decisions about how the exchange rate should be managed. In these cases, all decisions concerning exchange rate management are effectively made at the time that the regime is adopted. It may indeed be optimal for a country to adopt a regime of one of these polar types, depending on its circumstances, and the previous chapter reviewed some of the considerations involved in determining whether it would be appropriate to do so. But as we saw at the end of the preceding chapter, for many countries, the conditions that would make such an exchange rate regime appropriate are not likely to be met. When such a regime is not adopted, the exchange rate will have to be managed in some fashion, whether through intervention in the foreign exchange market or by using monetary policy.

This chapter examines some basic principles of exchange rate management in such intermediate regimes. We will use as an organizing framework the assumption that a country maintains a band, basket, and crawl (BBC) exchange rate arrangement because, with a slight modification, all the managed exchange regime options available to an emerging economy can be thought of as variations on such an arrangement. A BBC arrangement essentially involves the setting of a central parity for the domestic currency, together with fluctuation margins around that parity that may or may not commit the central bank to buy or sell the domestic currency in unlimited amounts whenever its value reaches previously specified lower or upper bounds. The decisions that have to be made in this context concern the setting of the central parity, the width of the band, the rules for intervening at the edges of the band, and the rules for intervening inside the band. Target zones and fixed but adjustable pegs fit naturally as special cases of such an arrangement. The modification that we would have to make to this definition to encompass all the circumstances under which the exchange rate is managed is to allow for the possibility that the central parity would not be announced. In that case, we can also interpret a managed float as a version of a BBC regime.

The considerations involved in managing an officially determined exchange rate are not of great interest, however, if operating an officially determined exchange rate is not a feasible option for a specific country. As we saw in the last chapter, in recent years, the argument has indeed been made that this is the case for emerging economies that are characterized by high degrees of capital mobility. We will examine this argument in more detail in this chapter and consider how empirical evidence might be brought to bear on it. A key question is whether attacks on currency pegs are primarily driven by fundamentals or self-fulfilling expectations. We will review some of the evidence that is available on this issue in this chapter.

The organization of the chapter is as follows. In the next section, we consider the arguments for adopting an exchange rate regime that requires active management of the exchange rate. The question of whether such intermediate regimes are indeed feasible is taken up in Section II. Sections III–V then consider in succession several decisions that have to be made in the course of managing a BBC arrangement: how to set the central parity, how to set the bands, and how to intervene inside the bands. Section VI summarizes.

I. WHY MANAGE THE EXCHANGE RATE?

The nominal exchange rate can, in principle, play two key macroeconomic roles: in the medium run, a policy-determined nominal exchange rate can be used to pin down an economy's average price level, whereas in the short run, changes in the nominal exchange rate can be used to influence its real exchange rate. Each of these roles is related to one of the criteria that have been adopted for evaluating exchange rate regimes: the former to the promotion of medium-run price stability and the latter to the reduction of the costs of adjustment to macroeconomic shocks. The problem is that stabilizing the price level and adjusting to shocks may be competing objectives for exchange rate policy in the sense that promoting medium-run price stability may call for stabilizing the nominal exchange, but reducing the short-run costs of macroeconomic adjustment may call for allowing the nominal exchange rate to change.

However, these objectives may not always be in conflict. The previous chapter noted that the value placed by a country on the role that the nominal exchange rate can play in attaining these alternative objectives depends in part on whether other means exist to achieve the same objectives at relatively low cost. For example, a country whose central bank has independently attained a substantial amount of low-inflation credibility under a floating exchange rate would place little value on the credibility gains that might be attainable by fixing the exchange rate in the medium run, while a country with very flexible domestic wages and prices, or one not typically afflicted by real shocks, may not place much value on the role that the exchange rate can play in facilitating the short-run adjustment of relative prices in response to such shocks. The *real* problem is that many (most?) emerging and developing countries tend to find themselves in situations in which they value *both* objectives because their central banks have not independently achieved a sufficient amount of anti-inflationary credibility, because they have a domestic wage-price mechanism that generates a relatively sticky wage-price process in the short run, and because they can frequently expect to be hit by macroeconomically important real shocks.

Even in such a situation, if the price stability or adjustment objective turns out to be of overwhelming importance at a specific time, it may still be optimal for a country to adopt an extreme exchange rate regime that emphasizes one of the two macroeconomic roles of the exchange rate at the expense of the other. For example, Argentina is a country with strong labor unions, and its exports are dominated by primary commodities. It is likely that the wage-price process in that country would exhibit substantial stickiness in normal times and that real shocks – for example, in the form of changes in the terms of trade – would tend to be frequent. However, it may still have made sense for Argentina to have adopted a currency board in 1991, a regime that promotes price stability at the expense of adjustment, because the government's anti-inflationary credibility had been shattered by a series of failed anti-inflation plans during the late 1980s, and hyperinflation was looming at the beginning of the 1990s.

However, when both roles that the nominal exchange rate can play have value for a country, and neither the price stability objective nor the adjustment objective is of overriding importance, the country has an incentive to seek to implement an exchange rate regime that strikes a different balance between the competing objectives than do the more extreme arrangements. This is essentially the argument for the fixed-but-adjustable, or pegged, exchange rate arrangements that have typically been maintained by developing countries. Such arrangements can be officially described as a fixed rate against one currency or a basket of currencies, with fluctuation margins of varying widths, with varying rules for adjusting the parity and with varying amounts of intervention within the margins. Alternatively, they can be officially described as a float, but typically with heavy intervention that tends to stabilize the rate in accordance with some (typically unannounced) rule. As mentioned previously, all such arrangements can be subsumed into the general classification of BBC arrangements, and we will refer to them as pegged arrangements throughout the rest of this chapter.

II. Are Pegged Exchange Rates Feasible under Conditions of High Capital Mobility?

While pegged exchange rate arrangements have been common among emerging economies, in the wake of a series of currency crises that afflicted both industrial and developing countries during the 1990s, many economists have begun to argue that, in view of the high degree of integration that many emerging economies have attained with international financial markets over the past decade, the time for such exchange rate arrangements is past. In this view, emerging economies have to choose between the extreme arrangements of dollarization (or full currency union) or floating exchange rates. In other words, pegged arrangements are not feasible in the context of high capital mobility.[1]

1. Self-Fulfilling Speculative Attacks

The argument goes as follows: a central bank's ability to defend an exchange rate peg depends on whether it can generate the resources to buy back with foreign exchange any amounts of its own currency that are presented to it at the officially determined exchange rate. Its ability to do so has often been taken to depend on a comparison of its liquid assets (its foreign exchange reserves) with its liquid liabilities (the monetary base). The reason is that a central bank with a sufficiently large stock of foreign exchange reserves could, in principle, buy back its entire monetary base, and that is all that an officially determined exchange rate commits it to do. Indeed, that is one of the arguments that is sometimes made for why a currency board guarantees the convertibility of the currency (see Chapter 18).

However, a central bank's liquid assets and liabilities are not, in practice, limited to the stock of foreign exchange reserves and the monetary base. On the liability side, if the central bank maintains *deposit guarantees* (a commitment to exchange bank deposits for currency, discussed further in Chapter 22), its liquid liabilities may extend beyond the monetary base to the entire stock of currency plus deposits (broad money). Thus the commitment to buy back the monetary base is extended to the entire stock of broad money by the deposit guarantee. On the asset side, a public sector that is solvent should easily be able to augment its foreign-currency assets through external borrowing. Indeed, because the conversion of domestic-currency liabilities into foreign currency is in effect a portfolio transaction by the public sector that does not affect its net worth, a solvent public sector should, in principle, be able to borrow enough to convert *all* its liquid liabilities into foreign exchange, including the stock of deposits.

The question is not, therefore, the *ability* of the central bank in a solvent public sector to defend the official exchange rate peg but rather its *willingness* to do so. Why is this an issue? The reason is that if a substantial part of the monetary base is indeed presented to the central bank for conversion into foreign exchange, the domestic money supply will contract and domestic interest rates will rise, perhaps dramatically. High domestic real interest rates will have a number of harmful effects on the domestic economy. They can trigger or aggravate a recession, they can imperil

[1] These arguments are developed in Eichengreen (1994) as well as Obstfeld and Rogoff (1995).

the solvency of the domestic banking system and trigger a banking crisis (Chapter 26), or they can imperil the solvency of the domestic public sector itself by sharply escalating its domestic-currency debt-servicing costs. Under these circumstances, the central bank may prefer to abandon the currency peg than to continue to defend it, simply because the country's economic welfare will be higher if it does so. If it indeed abandons the currency peg and the currency depreciates sharply, those who have converted domestic currency into foreign exchange will reap substantial gains.

The problem is that this situation could arise any time that the currency comes under attack, *for whatever reason*. In other words, even if the real exchange rate is not misaligned, and there is no other reason to believe that the nominal exchange rate would otherwise have been changed (i.e., even if the nominal exchange rate would otherwise have been sustainable), the emergence of a speculative attack on the currency might itself cause the central bank to abandon the currency peg, to avoid incurring the economic costs associated with the high interest rates that would be required to sustain it. In this case, a currency crisis could be *self-fulfilling*; that is, a currency peg that might have been sustainable for an indefinite period in the absence of a speculative attack is actually abandoned simply because an attack occurred, thereby validating the expectations of the speculators who launched the attack.

When would this be likely to happen? The answer is when agents become convinced that the central bank is indeed likely to abandon the peg when speculation against the currency drives domestic interest rates up. When it takes a relatively small increase in domestic interest rates to cause the peg to be abandoned, we say that the peg *lacks credibility*, and pegs that lack credibility will be particularly vulnerable to speculative attack. When credibility is lacking, currency pegs will be costly to maintain because central banks will have to bear very high domestic interest rates to demonstrate their resolve. The problem is that credibility is hard to achieve, so any exchange rate peg will, in principle, be vulnerable to a self-fulfilling speculative attack. Countries are thus driven to the extremes of monetary unification, where a change in the peg is impossible, or to a clean float, where speculative pressures show up in the form of continuous exchange rate changes that do not create the possibility of large capital gains.

What does high capital mobility have to do with this story? Notice that what puts pressure on central-bank reserves are its purchases of the domestic currency in foreign exchange markets. When the capital account is closed and the domestic currency is only convertible for current transactions, economic agents who hold the domestic currency cannot collectively sell it to the central bank to acquire foreign assets (because the central bank will not buy it), making it more difficult for speculators to attack the central bank's foreign exchange reserves. On the other hand, when the capital account is open but capital mobility is limited because domestic and foreign assets are imperfect substitutes, the prospect of a speculative gain on foreign assets of a given magnitude would trigger a smaller conversion out of domestic-currency assets – and thus fewer sales of domestic currency on the foreign exchange market and less pressure on the central bank's stock of foreign exchange reserves than would be the case under a higher degree of financial integration (as we saw in the context of the model of Part 2). As the degree of capital mobility increases, however, a given expected speculative gain triggers larger and larger movements out of domestic-currency assets, increasing the pressure on the central bank.

2. The Role of Fundamentals

However, the logic of the preceding argument does not imply that pegged exchange rates are *universally* infeasible because it suggests that pegged rate regimes should not be *uniformly* vulnerable to self-fulfilling speculative attack. What matters in determining the vulnerability of a pegged regime to a speculative attack is the credibility of its exchange rate peg, that is, the likelihood that the central bank would abandon the peg if it were subjected to an attack of a given degree of severity. But because the decision of whether to abandon the peg is made by the central bank by trading off the costs and benefits of doing so, and because these costs and benefits themselves depend on the state of the domestic economy, there is a role for fundamentals to determine the sustainability of an exchange rate peg. In particular, if the characteristics of the domestic economy (the fundamentals) are such that the benefits of defending the peg are relatively high and the costs of doing so are relatively small, then the currency should be in a strong position to resist speculative attack, and the currency peg should be sustainable. This would be the case, for example, when the real exchange rate is not misaligned (so there is no benefit from a nominal depreciation), the financial sector is strong, the public sector does not have a large stock of domestic-currency debt, and the economy is growing rapidly. Alternatively, if the currency appears to be overvalued, the domestic financial sector is weak, the public sector's solvency is precarious, and it has a large stock of short-term domestic debt – and/or the economy is in recession – the prospects for a successful speculative attack would tend to be strong.

In theory, then, high capital mobility does not mean that pegged exchange rate regimes cannot be sustained. It merely means that the probability of a crisis when the domestic fundamentals are poor is likely to be much higher; that is, the financial markets will punish countries more quickly and more severely for so-called "bad" fundamentals. At bottom, then, the issue of the feasibility of pegged rates boils down to two questions:

1. Do fundamentals matter in determining a country's vulnerability to speculative attack? If they do, as the theory just outlined suggests they should, then well-managed countries should find it feasible to sustain pegged exchange rates, even under conditions of high capital mobility.

2. But even so, pegged exchange rates would not be feasible if countries cannot avoid bad fundamentals (i.e., fiscal insolvency, exchange rate misalignment, financial-sector fragility, etc.). So the second question is, can countries reliably avoid bad fundamentals? If they cannot, then if the costs of attacks on their currencies are very high in the event that their fundamentals turn poor, they may be well advised to adopt more robust exchange rate regimes (such as one of the two extremes advocated by many economists), even if pegged rates would be feasible for them most of the time.

3. Empirical Evidence

The first question is an empirical one, and many studies have been conducted in recent years on the issue of the role of fundamentals in determining vulnerability to currency crises. A very simple version, for example, is described in Chapter 24 in the form of a discussion of a World Bank (1997) study on the factors that determined vulnerability to capital flow reversals in the wake of the Mexican crisis in 1995. That study relied on comparing the economic characteristics of countries that were more or less affected by the Mexican crisis and detected some interpretable differences in their characteristics, suggesting that fundamentals helped to determine susceptibility to contagion from the Mexican crisis.

More sophisticated studies have used econometric techniques (called *probit analysis*) to estimate from cross-country data the contributions made by various sets of fundamental factors to the probability that a country would undergo a currency crisis. Such techniques essentially involve expressing the probability of a crisis as a function of a set of potential fundamental variables and then using information on the occurrence or nonoccurrence of crises in a large panel of countries to estimate the parameters of this probability distribution.

For example, a widely cited study by Frankel and Rose (1996) looked at the determinants of what they referred to as currency "crashes."[2] They used two methodologies: a graphical "event study" approach, in which the behavior of individual macroeconomic variables during periods surrounding currency crashes is compared with the behavior of the same variables during periods of tranquility, as well as probit analysis, in which the contributions of individual fundamentals to the probability of a crash could be estimated. Their sample consisted of annual data on 105 developing countries over the period 1971–1992, and it included 117 currency crashes.

Among the potential fundamentals they considered were the rate of growth of domestic credit, the ratio of the fiscal deficit to gross domestic product (*GDP*),

[2] They defined a currency crash as a nominal depreciation of more than 25% that was also more than 10% greater than that in the previous period, with the latter criterion added to rule out high inflation episodes. To be included in their sample, such crashes had to be separated by at least three years from previous episodes.

the ratio of foreign exchange reserves to imports, the ratio of the current account deficit to *GDP*, the rate of growth of real *GDP*, the degree of overvaluation of the real exchange rate (i.e., its deviation from an equilibrium real exchange rate calculated using a period-average PPP methodology), the ratio of debt to *GDP*, world interest rates, the average growth rate of the Organization for Economic Co-operation and Development (OECD) countries, and various debt and capital inflow composition variables (such as the share of debt from commercial banks; the share of debt that is concessional; and variable-rate, public, short-term, loaned by multilaterals, and foreign direct investment (FDI) flows as a percentage of the total debt stock).

The results of the graphical analysis were that, except for the current account and fiscal deficit ratios, all the fundamental variables behaved during the period surrounding currency crashes as theory predicts they should if such variables indeed help to determine the likelihood of such crashes. But of course, such graphical analysis only considers the potential role of each fundamental one at a time and provides no indication of the statistical confidence that can be associated with the results. Frankel and Rose complemented this with probit analysis to deal with both these problems. The results from the probit estimation were as follows:

- Most debt composition variables proved not to be statistically significant predictors of currency crashes, except for the FDI variable. A lower share of FDI in capital inflows tends to increase the probability of a crash (we will come back to this in the next chapter).
- Neither the current account nor fiscal deficit variables proved to be important contributors to the probability of currency crashes.
- Except for the rate of growth of OECD countries, all the external variables (world interest rates, the reserve adequacy measure, the ratio of debt to GDP, and the extent of real exchange rate overvaluation) proved to be very important.
- Finally, high credit growth (associated with financial-sector fragility) and a recession in the domestic economy both increased the probability of a crash.

These results were derived with the fundamentals entered contemporaneously with the qualitative dependent variable measuring the incidence of currency crashes. The results proved to be even stronger, however, when the explanatory variables were lagged. A low share of concessional debt, a low share of FDI in capital inflows, and a larger public share in total debt all increased the likelihood of a crash, as did low foreign exchange reserves, real exchange rate overvaluation, high international interest rates, and high domestic credit growth.

The evidence from this study and others suggests that fundamentals indeed matter in determining susceptibility to speculative attack and currency crises. Thus they are consistent with the theory described earlier. The implication is that well-managed countries should face a reduced susceptibility to speculative attack. Whether bad fundamentals can be consistently avoided, however, is another matter. The crisis that afflicted the previously highly successful miracle economies of East and Southeast Asia in 1997 certainly calls into question whether they can, even under very favorable circumstances. We will return to this when we discuss the Asian crisis in Chapter 27. For now, we conclude that the evidence that fundamentals matter is at least consistent with the possibility that exchange rates can still be managed, and we turn next to how this should be done.

III. HOW TO MANAGE AN OFFICIALLY DETERMINED EXCHANGE RATE: FIXING THE CENTRAL PARITY

As indicated previously, if exchange rates are to be managed, a useful way to organize the discussion of how this should be done is to consider exchange rate management within the framework of an exchange rate band. This breaks down the question of how to manage the rate into the three subsidiary questions of how to fix the central parity, how wide to set the fluctuation margins, and how to intervene within those margins. In this section, we will first take up the question of how to set the central parity, leaving the other two questions for the next two sections.

Notice that the central parity is what governs the medium-term path of the nominal exchange rate under a band. In setting it, then, a central bank has first to confront the question of why it would be preferable in the first place to announce any central parity at all rather than simply floating the exchange rate and intervening to influence it when desired. As mentioned previously, managed exchange rates include regimes that officially describe themselves as pegged and those that do not. The difference between them is that the former do, and the latter do not, announce a central parity.

1. Using the Central Parity as Nominal Anchor

There are two reasons that governments may find it desirable to announce a central parity, and both have to do with making known the medium-run path of the exchange rate. The first reason is the one that has been alluded to previously as a reason why a government might want to choose a managed exchange rate rather than a clean float in the first place: to use the exchange rate as a nominal anchor.

If this is the reason for opting for a managed rate, then the central parity should be set to depreciate at a rate equal to the difference between the *desired* (not actual) domestic inflation rate (which, owing to the requirements of fiscal solvency, would have to be determined simultaneously with the desired path of fiscal variables) and the weighted-average inflation rate of the country's trading partners. This rule, therefore, adopts a *forward-looking* domestic inflation measure (the desired *future* inflation rate) as a guide for the adjustment of the central parity. It would tend to stabilize the real exchange rate as long as the domestic inflation target is met, and it would avoid misalignment as long as the real fundamental determinants of the long-run equilibrium real exchange rate do not change. If the inflation target is not met over some time period, or if the fundamental determinants of the equilibrium real exchange rate were to change, real exchange rate misalignment would emerge. Under this rule, such misalignment would be expected to be eliminated through adjustments in the domestic price level, which is what it means to use the exchange rate as nominal anchor.

2. Choosing the Peg

A separate question that arises in this context is to which foreign currency (or currencies) we should peg the exchange rate. This makes a difference, of course, only when trading partners' inflation rates differ and when their *real* exchange rates fluctuate substantially against each other. If their real exchange rates did not fluctuate against each other, then their inflation differentials would be offset by nominal exchange rate changes, and their price levels would be equalized when expressed in a common currency. The basic choices are whether to peg to a single country or to a basket of currencies weighted to reflect the shares of the country's trade with individual trading partners. Pegging to a low-inflation country would promote long-run domestic price stability, but if that country has a relatively small share of the domestic economy's external trade, doing so would tend to destabilize the domestic real exchange rate because the domestic-currency prices of goods produced by countries with which the home country does a substantial amount of trade would be affected by fluctuations among partner-country currencies.

If a single low-inflation country represents the dominant trading partner for the domestic economy, of course, then the choice is easy: peg to that country's currency. The optimal choice will, in general, depend on the weights placed by the domestic economy on price stability versus real exchange rate stability as well as on the distribution of trade weights across trading partners with different long-run inflation rates and different degrees of real exchange rate variability.

3. Targeting the Real Exchange Rate

However, using the exchange rate as a nominal anchor is not the only reason to opt for managed rates instead of a clean float. Countries whose central banks have sufficient anti-inflationary credibility to dispense with this role of the nominal exchange rate may nonetheless opt for managed rates out of the desire to avoid the excessive volatility in the real exchange rate that might be associated with a clean float. This motivation for managing the exchange rate does not necessarily require the announcement of a central parity. A float in which the central bank intervenes to smooth out fluctuations in the exchange rate without committing itself to a central parity could achieve the desired goal. However, in this case, private agents are given no assurances about the medium-term evolution of the real exchange rate. As the experience of industrial countries under floating exchange rates shows, the real exchange rate can exhibit substantial medium-term fluctuations under this system. If the motivation for adopting a managed rate rather than a clean float was to avoid excessive noise in the real exchange rate, this situation may not be satisfactory.

An alternative is to manage the central parity to stabilize the medium-term path of the *real* exchange rate (see Box 18.2). That means adopting a rule under which the rate of depreciation of the nominal exchange rate is set as the difference between the *past* domestic inflation rate and that of the country's trading partners (say, on a monthly basis). This will ensure that the real exchange rate will not deviate far from its initial value because fluctuations in the domestic rate of inflation are accommodated by changes in the rate of depreciation of the central parity to ensure that this does not happen. Notice that in this case, because the objective is to stabilize the effective real exchange rate, it makes sense to adjust the nominal rate in accordance with the trade-weighted inflation rates in partner countries rather than that of a single country.

A difficulty with this approach is that the exchange rate can no longer play the role of the nominal anchor. Though it may be thought that this simply means that the money supply would have to play the role of nominal anchor, this is not actually an option because we are considering an *officially determined* exchange rate regime (not a floating rate regime), in which the domestic money supply is rendered endogenous by the central bank's commitment to intervene in the foreign exchange market to defend its announced parity.

The problem can be illustrated with the model of Part 2. Recall that in the short run, under nominal exchange rate targeting, the economy's price level was determined by the intersection of a positively sloped BB curve (under financial autarky or imperfect capital mobility) and a negatively sloped GM curve in (P, R) space. When the central bank targets the real exchange rate, by contrast, the price level drops out of the goods-market equilibrium condition (SP^*/P is replaced by the real exchange rate target; call it Q^*). This turns the GM curve into a horizontal straight line parallel to the price axis. As long as the BB curve retains its positive slope (i.e., as long as capital mobility is not perfect), the price level remains determinate under these conditions in the short run. But the endogeneity of the money supply in the medium run (and, under perfect capital mobility, in the short run as well) leaves the domestic price level indeterminate in this model under real exchange rate targeting.

In a model that incorporates wealth effects on private spending, however, this property would not hold. The reason is that wealth effects on spending would reintroduce a negative slope to the GM curve as long as not all of the private sector's wealth is indexed to the price level. It can be shown that under these circumstances, the domestic long-run rate of inflation actually becomes endogenous and is determined in such a way as to eliminate the gap between the real exchange rate targeted by the authorities under this exchange rate regime and the equilibrium real exchange rate implied by the economy's fundamentals.[3]

This has an important implication for the setting of the central parity under real exchange rate targeting. If the central bank simply chooses a real exchange rate target and leaves it unchanged in the face of permanent changes in the underlying fundamental determinants of the economy's long-run equilibrium real exchange rate, the domestic rate of inflation will change whenever these fundamentals do. This means that the real exchange rate would be stabilized at the cost of destabilizing the domestic inflation rate. To avoid the latter, it would be necessary to alter the targeted real exchange rate in response to permanent changes in the real fundamentals. Thus the central parity would have to be managed to avoid the emergence of real exchange rate misalignment. As shown in Chapter 16, our state of knowledge on the empirical measurement of misalignment might make this a challenging proposition.

In short, the choices to be made in setting the central parity depend on the trade-offs that the country is prepared to make in using the exchange rate as a nominal anchor and in using it to stabilize the real exchange rate. If the exchange rate is to be used as a nominal anchor, the central parity implies a path for the nominal exchange rate that depends on the targeted domestic rate of inflation and the expected rate of inflation in a relatively important, low-inflation trading partner. If the nominal exchange rate is to be managed to stabilize the real exchange rate, then a PPP-based rule for the central parity is required, based on past domestic inflation rates as well as trade-weighted partner-country inflation rates. In this case, to avoid destabilizing the domestic price level, the real exchange rate targeted by the central parity will have to be adjusted in accordance with permanent changes in the underlying real fundamental determinants of the country's long-run equilibrium real exchange rate.

IV. HOW TO MANAGE AN OFFICIALLY DETERMINED EXCHANGE RATE: SETTING THE WIDTH OF THE BAND

The next issue that arises in managing a pegged exchange rate is how much fluctuation to allow around the central parity, that is, using the exchange rate band framework, how wide to set the exchange rate band.

1. Announcing the Band Width

As in our discussion of the central parity, the first question that should be posed in this context is whether it makes sense to adopt and announce an explicit band width at all. An announced band width commits the central bank to unlimited intervention at the margins of the band. In the absence of an announced band width,

3 See Montiel and Ostry (1991).

the central bank retains discretion whether to intervene throughout the range of fluctuation of the nominal exchange rate.

One argument for announcing an explicit band width has been made by Krugman (1993). It is that, given the behavior of the fundamental determinants of the nominal exchange rate, the mere announcement of an explicit band width would tend to stabilize the nominal exchange rate automatically within the band, even in the absence of central-bank intervention. Krugman models the link between the exchange rate and its fundamental determinants inside the band as follows:

$$s = f + E(ds/dt)$$

where s is the log of the nominal exchange rate; f is the log of its fundamental determinants, given by $f = m - v$, where m is the log of the domestic money supply; and v (for velocity) is an exogenous shock, taken to be a *random walk* (i.e., a variable with the property that its *change* from period to period is a mean-zero random variable). In this equation, $E(ds/dt)$ is the expected change in the log of the nominal exchange rate from one period to the next.

Now suppose that the band is perfectly credible and that intervention only takes place at the edges of the bands. Then, in the absence of an explicit band, the relationship between the exchange rate s and the fundamental f could be depicted as a 45 degree line (because with the random walk assumption, the expected future change in the exchange rate would be zero, making $E(ds/dt) = 0$, and leaving $s = f$). However, with an explicit band in place, this changes in the following ways:

1. The slope of the relationship between s and f becomes *less* than unity inside the band because the exchange rate would be more likely to appreciate on net (making $E(ds/dt)$ negative) the closer it is to the upper edge and to depreciate the closer it is to the lower edge. This is called the *honeymoon effect*, and it means that some exchange rate stability is achieved for free because the exchange rate becomes more stable than it would be in the absence of the band even without central-bank intervention.
2. The curve depicting the relationship between s and f is asymptotic to the edges (a property known as *smooth pasting*), essentially because the exchange rate must become insensitive to the fundamental at the edges. This is because there is a discontinuity in the expected change in the fundamental at the edge of the band (it can only move in one direction owing to the expected offsetting change in policy). The expected change in the fundamental is zero inside the band (recall that v is a random walk) but suddenly not zero at the band. But because there can be no discontinuity in the expected change in the exchange rate at the edge (the reason is that this would mean a discontinuity in the path of s itself, because the other determinant of s, the fundamental, is continuous in levels), this means that the exchange rate must be insensitive to the fundamental at the edge.

The main argument *against* explicitly announced bands, on the other hand, is the argument against pegged exchange rates: speculation against the currency, for example, would quickly drive its value to the upper bound of the band. At that point, the situation becomes identical to that of a fixed exchange rate that is under speculative attack. Speculators face a one-way option for the exchange rate to move: they stand to make large gains if the parity is adjusted in a discrete fashion at that point and face little likelihood of losses through a movement of the exchange rate back toward the center of the band. The central bank would face the same choices that it does under a fixed rate and, if the fundamentals are wrong, would be tempted to devalue or abandon the band. This situation would not arise if the central bank did not make an explicit commitment to defend upper and lower bounds for the value of its currency.

These arguments suggest that the desirability of announcing explicit bands for the exchange rate depends on the likelihood that such bounds can be made credible. If they can, then the honeymoon effect would tend to add stability to nominal exchange rate movements. If they cannot, then the announcement of such bands is simply an invitation to a speculative attack. As in the previous section, the achievement of such credibility depends on governments' ability to adequately manage their fundamentals.

2. Determining the Width of the Band

Suppose, then, that a country can achieve a sufficiently high degree of credibility to warrant the adoption of an explicitly announced band. How wide should it be? In principle, the width of the band should depend on several factors.

a. The Degree of Fiscal Flexibility

Inside the band, the domestic economy retains monetary autonomy because the exchange rate can move by a maximum amount given by the width of the band, allowing domestic interest rates to differ from foreign ones by this amount at a maximum. One factor that should affect the optimal width of the band, then, is the desired degree of monetary autonomy that the domestic central bank finds it optimal to retain. In turn, this depends on the value of monetary autonomy to the domestic economy, which depends on the availability of other instruments of stabilization policy. The implication is that the less flexible fiscal policy is as a stabilization instrument, the more important is monetary autonomy, so the stronger is the case for a wide band.

b. The Degree of Financial Integration

But the degree of monetary autonomy that the country can achieve does not just depend on the width of the band. As we saw in Part 2, the higher the country's degree of financial integration, the less monetary autonomy it will have. Economies that are highly integrated with world capital markets will have very little monetary autonomy with a narrow band, so for such economies to achieve a given degree of monetary autonomy, they will require a larger band than countries that are less financially integrated with the world economy. Thus, other things equal, increased financial integration strengthens the case for a wider band.

c. The Nature and Amplitude of Macroeconomic Shocks

The width of the band should also depend on whether shocks to the economy tend to be permanent or transitory, real or nominal, and how large they tend to be. As we have seen previously, when the motivation for managing the exchange rate is to stabilize the real exchange rate, the central parity should respond to real shocks that are perceived to be permanent, but the response to transitory real shocks should occur within the band. On the other hand, the exchange rate should not respond at all to shocks that are perceived to be nominal in origin. Thus, the more susceptible the country is perceived to be to transitory real shocks, and the larger the perceived amplitude of such shocks, the wider the band should be.

d. The Cost of Maintaining Foreign Exchange Reserves

Finally, maintaining a stock of foreign exchange reserves will be required to defend the upper bound of the band. Because the costs of such reserves will depend in part on the width of the band, the optimal band width will be influenced by the costs of reserves. The costs of maintaining foreign exchange reserves depend on the total stock of reserves the central bank holds as well as on the cost of maintaining each unit of reserves. The latter depends on the opportunity cost of the resources tied up in reserves (say, the domestic marginal product of capital minus the interest receipts on reserves). The required stock of reserves, in turn, is likely to depend on the credibility of the band as well as on its width. The greater the cost of maintaining reserves for other reasons (i.e., because large reserves are required due to limited confidence or because the cost of maintaining a given stock of reserves is high), the wider should be the amplitude of the band.

V. HOW TO MANAGE AN OFFICIALLY DETERMINED EXCHANGE RATE: INTERVENTION WITHIN THE BAND

Finally, there is relatively less to say about the factors that should govern the rules for intervention within the band. Three such factors are worth mentioning.

First, the optimal extent of intervention should depend on the perceived source of specific macroeconomic shocks. As we have seen, the optimal response to permanent real macroeconomic shocks – which alter the equilibrium real exchange rate – involves a change in the central parity. From the perspective of stabilizing aggregate demand, the optimal extent of intervention in foreign exchange markets in response to transitory shocks depends on the source of such shocks. We argued earlier that less intervention is indicated for real shocks, so that transitory fluctuations in real exchange rates can bear some of the adjustment burden, whereas more intervention would be optimal for shocks that are perceived to be nominal in origin, so that the adjustment burden is picked up by accommodating changes in the money supply.

In the case of foreign financial shocks, the optimal degree of intervention depends on structural characteristics of the domestic economy.

This analysis, however, may need to be modified if the domestic authorities perceive a need to stabilize the real exchange rate for domestic resource allocation reasons. In response to perceived real shocks, the authorities face a choice between shifting the adjustment burden to the real exchange rate through transitory movements in the nominal exchange rate inside the band (produced by minimizing intervention inside the band) and allowing the response to be reflected in changes in aggregate demand (if they intervene to stabilize the nominal exchange rate within the band). The question is how much noise is introduced into the real exchange rate as a relative price signal by fluctuations that occur within the band. The answer, of course, depends on the width of the band as well as on private agents' perceptions that real exchange rate movements inside the band are indeed transitory. This confidence can be enhanced if the authorities indeed manage the central parity successfully to respond to perceived changes in the long-run equilibrium real exchange rate. If they do, then movements inside the band should be regarded as transitory and should not have adverse implications for resource allocation. In that case, less intervention would be required inside the band in response to transitory real shocks.

The final factor to be considered is how desirable variability in the *nominal* exchange rate might be. The more desirable such variability is (e.g., because there are distortions in the financial system that create incentives for capital flows into the country), the less intervention there should be inside the band to stabilize the nominal exchange rate. Indeed, when distortions in the domestic financial system are important, and have the effect of inducing overborrowing by domestic financial intermediaries, the authorities may choose to intervene within the band to *increase* variability in the nominal exchange rate to create uncertainty in the domestic-currency returns faced by foreign agents who bring capital into the country.

VI. SUMMARY

Managed exchange rates have the attractive feature that they can combine characteristics of both of the extreme exchange rate arrangements discussed in the preceding chapter, allowing them to strike a different balance between the alternative objectives to which exchange rate policy can be devoted than is done by the more extreme arrangements. As we have seen, striking a balance between those objectives different from those available through extreme exchange rate arrangements may well be desirable for many emerging and developing economies.

Depending on how the parity is administered, managed exchange rates can be designed to use the nominal exchange rate as a nominal anchor or to stabilize the real exchange rate. In either case, they can, in principle, permit the exchange rate to be used for adjustment if desired (through movements in the rate within a band as well as through changes in the central parity). Moreover, by placing the parity within a band, they can allow the authorities to preserve some degree of monetary autonomy. Thus, with a managed exchange rate, by setting a nominal target for the central parity, the monetary authorities can seek to combine the nominal anchor properties of a fixed exchange rate with the adjustment advantages of a floating rate, or, if they have an independent source of price-level credibility, by setting a real exchange rate target for the central parity, they can attain the adjustment advantages of floating rates without the real exchange rate instability that floating exchange rates can involve.

The obvious advantage of managed rates, then, is the flexibility that they allow to the domestic monetary authorities. But are pegged exchange rate arrangements feasible under conditions of high capital mobility? An important problem is that pegged exchange rate arrangements are potentially vulnerable to self-fulfilling speculative attacks. Theory and evidence give us some hope, in the sense that both indicate that such attacks are more likely to happen when the economy's fundamentals are weak. Under such conditions, the authorities would be unable to resist a speculative attack, and the peg would have limited credibility. Thus, not all currency pegs are equally vulnerable. But can the domestic economy be managed to avoid bad fundamentals that would leave the peg vulnerable to speculative attack?

This is an open question to which we will turn in Chapter 27 by examining some case studies. The implications of the answer to this question are the following:

1. If the fundamentals can be managed appropriately, then managed exchange rates remain an option, even under conditions of high capital mobility. They may not necessarily be optimal for all countries and at all times because some countries will find it preferable to adopt one of the extreme arrangements discussed in the preceding chapter, but they at least remain feasible without restricting the country's degree of integration with world capital markets.
2. If, however, small errors of policy or small random shocks make countries vulnerable to crises with very large costs under pegged rates and high capital mobility, then this option becomes much less attractive. The alternatives in this case are to retain a high degree of financial integration and move to one of the extreme exchange rate regimes or to impose restrictions on capital movements and continue to manage the exchange rate.

REVIEW QUESTIONS

1. Why do fixed-but-adjustable regimes tend to be attractive for developing countries?
2. What is the argument for infeasibility of soft exchange rate pegs under high capital mobility? Under what conditions might such exchange rate arrangements remain feasible?

3. What is the difference between a managed float and a soft peg? Under what conditions might a soft peg be preferable to a managed float?
4. Should the central parity in a BBC arrangement be set according to the difference between the desired future domestic inflation rate and foreign inflation or according to the difference between the lagged domestic inflation rate and foreign inflation? Explain.
5. What factors should determine the width of the band in a BBC arrangement?

EXERCISES

1. Use the short-run model of Chapter 7 to explain the role of capital mobility in determining the sustainability of a fixed-but-adjustable exchange rate regime.
2. Use the analysis of formal commitment mechanisms in Chapter 15 to explain why a hard exchange rate peg might be less likely to succumb to a speculative attack than a soft peg.
3. Consider the case of a small open economy that is characterized by perfect capital mobility and operates a target zone. Using the GM-MM analysis of Chapter 17, explain how the central bank should respond to a goods-market shock that, in the absence of central-bank reaction, would
   a. leave the nominal exchange rate inside the target zone
   b. cause the exchange rate to depreciate out of the target zone

   Should the indicated central-bank reaction take the form of exchange rate policy or monetary policy? Explain why.
4. The empirical evidence reviewed in this chapter suggested that speculative attacks on a currency tend to be influenced by a variety of fundamentals that are consistent with theory rather than simply by what type of exchange rate regime happens to be in place. Does this mean that soft exchange rate pegs remain feasible under conditions of high capital mobility? Explain why or why not.
5. For most of the 1990s, Chile operated an officially determined exchange rate regime that targeted the real exchange rate. Does this mean that during this period, the Chilean economy operated without a nominal anchor? Explain why or why not.

REFERENCES AND FURTHER READING

Edwards, Sebastian (2002), "The Great Exchange Rate Debate after Argentina," working paper 9257, National Bureau of Economic Research.

Eichengreen, Barry (1994), *International Monetary Arrangements for the 21st Century*, Washington, DC: Brookings Institution.

Eichengreen, Barry (2008), "Exchange Rate Regimes and Capital Mobility: How Much of the Swoboda Thesis Survives?" working paper 14100, National Bureau of Economic Research.

Fischer, Stanley (2001), "Exchange Rate Regimes: Is the Bipolar View Correct?" *Journal of Economic Perspectives*, vol. 15, pp. 3–24.

Frankel, Jeffery A. (1999), "No Single Currency Regime Is Right for All Countries or at All Times," Graham Lecture, Princeton University.

Frankel, Jeffrey A., and Andrew K. Rose (1996), "Currency Crises in Emerging Markets: An Empirical Treatment," *Journal of International Economics*, vol. 41, pp. 351–366.

Frankel, Jeffrey, and Shang-Jin Wei (2004), "Managing Macroeconomic Crises: Policy Lessons," in J. Aizenman and B. Pinto, eds., *Managing Volatility and Crises: A Practitioner's Guide* (Cambridge: Cambridge University Press), pp. 1–86.

Glick, Reuven (2000), "Fixed or Floating: Is It Still Possible to Manage in the Middle?" working paper PB00-02, Federal Reserve Bank of San Francisco Pacific Basin.

Krugman, Paul R. (1993), "Target Zones and Exchange Rate Dynamics," in P. Krugman, ed., *Currencies and Crises*, Cambridge, MA: MIT Press, pp. 77–89.

Montiel, Peter J., and Jonathan Ostry (1991), "Macroeconomic Implications of Real Exchange Rate Targeting in Developing Countries," *IMF Staff Papers*, vol. 38, pp. 872–900.

Obstfeld, Maurice, and Kenneth Rogoff (1995), "The Mirage of Fixed Exchange Rates," *Journal of Economic Perspectives*, vol. 9, pp. 73–96.

World Bank (1997), *The Road to Financial Integration: Private Capital Flows to Developing Countries* (New York; Oxford University Press).

PART 6

The Financial Sector and Macroeconomic Performance

20

Finance, Welfare, and Growth

In Parts 2–5 of this book, we have taken the domestic financial system for granted. The model of Part 2, the analysis of public-sector solvency in Part 3, and the discussions of monetary institutions in Part 4 and of exchange rate policy in Part 5 all assumed – explicitly or implicitly – a domestic financial structure in which financial assets consisted of cash and securities that were traded in open markets. There were no specialized financial institutions as such, and in particular, there were no banks.

But management of the domestic financial system is actually one of the key challenges facing policy makers in emerging and developing economies, and macroeconomic performance in such economies depends critically on the state of the financial system. The functioning of the financial system strongly affects the economy's long-run growth performance as well as its short-run macroeconomic stability. Distortions in the domestic financial sector can present serious obstacles to long-run growth by impairing both capital accumulation and growth in total factor productivity. On the other hand, financial-sector weaknesses can themselves be the source of macroeconomic instability or can serve to propagate and magnify macroeconomic shocks arising elsewhere in the economy.

Part 6 of the book takes up the role of the financial system as well as of policies directed to the financial system. We will begin by examining in the present chapter how the financial system can affect economic welfare and long-run growth. The perspective to be adopted is that, in a world of imperfect information, financial intermediation is an economic activity that is productive in a very real sense; that is, it uses scarce productive inputs to produce valuable outputs. We will see that a well-functioning and efficient financial sector can contribute to growth in three ways: by allocating capital to its most productive uses, by providing incentives for the accumulation of physical and human capital, and by minimizing the amount of resources that are absorbed in the process of intermediation. These links between financial-sector performance and economic growth will be illustrated in the context of a simple model, and some of the existing evidence bearing on these issues will be described.

Subsequent chapters in Part 6 will consider policy issues that arise in the context of financial intermediation. The causes and consequences of *financial repression* – a set of policies that severely constrains the functioning of the financial system – are described in Chapter 21, followed by two chapters dealing, respectively, with domestic and external dimensions of financial reform, in the form of financial liberalization and financial openness. In Chapter 24, we will examine how the benchmark model of Part 2 needs to be modified to take into account the role of banks in financial intermediation. The final chapter in Part 6 will consider one of the implications of financial openness: the need to manage capital inflows. We will return to the financial sector in Chapter 26, when we consider how financial-sector vulnerabilities have produced some of the most severe episodes of macroeconomic instability experienced by emerging economies over the past two decades.

The current chapter presents an overview of the role of the financial sector. Section I defines terms and explores the sense in which the financial system performs a productive role in the economy. Section II explores the properties of financial-market equilibrium in an idealized frictionless world, which serves as a frame of reference for the introduction of real-world financial frictions in Section III. Section IV explains how banks arise naturally as a market response to those real-world frictions and describes how the presence of banks helps to reduce the costs of those frictions. The extent to which banks can successfully perform this function depends in part on the support of public policy, and the relevant policies are considered in Section V. Finally, Section VI draws out the implications of the preceding analysis for the links between financial-sector performance and aggregate economic activity.

I. FINANCIAL TRANSACTIONS AND FINANCIAL MARKETS

It is useful to begin our examination of the role of financial intermediation with some definitions that will be used in this and subsequent chapters. A financial *instrument* is a contract that commits an economic agent to make a future payment under some specified conditions. Notice that the role of the *future* in this definition means that the stipulation that payments are made at different points in time is an essential feature of financial instruments. A *financial transaction* is one in which a financial instrument is traded, either in exchange for a current payment or for another financial instrument. *Financial markets* are the markets in which financial transactions are carried out.

To see why financial transactions exist, consider the following economic environment. Suppose that time passes in discrete periods, and in each period, the world can exist in different "states of nature." Nobody knows ahead of time which state of nature will prevail in each period, and no one has any control over those outcomes, so the state of the world is inherently uncertain. There are many individuals, each of whom is endowed with a stochastic stream of income (i.e., a stream of income received each period that depends on the state of nature that prevails in that period). The welfare of each person depends on his or her stream of consumption (i.e., the amount he or she can consume each period).

Individuals in this economy seek to maximize their economic well-being by engaging in trade with each other. How can they improve their well-being by doing so? To answer this question, suppose that each person's preferences regarding consumption are characterized by the following:

- *nonsatiation* (they regard consuming more as better than consuming less)
- a *positive rate of time preference* (that means that they regard consuming now as better than consuming later)
- *risk aversion* (they prefer certainty to uncertainty)

Assume that all the individuals in this economy are identical. Would they have any reason to engage in financial trades with each other? The answer is no, because any trade that would benefit one party would harm the other; that is, there would be no opportunity for *mutually beneficial* trades.

Now let us introduce *heterogeneity* among individuals. In particular, suppose that people can differ from each other in five ways:

1. their distribution of income across the possible states of nature within each period
2. their degree of risk aversion
3. the intertemporal profile of their income streams
4. their rate of time preference
5. their entrepreneurial talents; that is, some (but not all) individuals have the opportunity to buy new stochastic streams of income from nature

In this case, there may be several motivations for financial transactions to take place. These include the desire to *reduce risk*, to *reallocate consumption over time*, and to *finance investment opportunities*. To the extent that financial markets facilitate transactions of these types, we would expect them to contribute to enhancing economic welfare, just as other markets do that facilitate mutually beneficial exchanges. To see how each of these motives could give rise to mutually advantageous financial trades, consider each of them in turn.

1. Risk Reduction

The risk reduction motive can give rise to mutually advantageous financial trade when individuals have similar attitudes toward risk but differ with respect to their income streams, or when individuals with identical income streams differ with respect to their degree of risk aversion. For example, consider two individuals A and B with income streams given by the following table under states of nature 1 and 2, each of which occurs with probability $p = 1/2$:

	State of Nature		
Individual	1 ($p = 1/2$)	2 ($p = 1/2$)	Expected Income
A	0	100	50
B	100	0	50

Suppose that individual A agrees to pay B 50 if the state of nature turns out to be 2, while B pays A 50 if it turns out to be 1. Then the situation becomes as follows:

	State of Nature		
Individual	1 ($p = 1/2$)	2 ($p = 1/2$)	Expected Income
A	50	50	50
B	50	50	50

Note that all risk has been eliminated by this financial transaction, making both parties better off (because they are both risk averse). Thus this transaction would be entered into voluntarily and would improve social welfare.

Alternatively, assume instead that both parties have the same stochastic distribution of income, but individual A is risk averse, while B is *risk neutral* (does not care about risk). The situation is as follows:

	State of Nature		
Individual	1 ($p = 1/2$)	2 ($p = 1/2$)	Expected Income
A	30	70	50
B	30	70	50

Suppose that B promises to pay A 18 if the state of nature is 1, while A promises to pay B 22 if the state of nature is 2. Then the situation would become as follows:

	State of Nature		
Individual	1 ($p = 1/2$)	2 ($p = 1/2$)	Expected Income
A	48	48	48
B	12	92	52

Could this be a voluntary exchange? The answer is yes, if A is sufficiently risk averse. The reason is that A has lost two units of expected income but has succeeded in eliminating all risk from his or her income stream. This financial transaction reallocates risk from A to B, and both parties are better off. Social welfare again improves.

Notice that in the first case considered, risk was reduced for both parties, whereas in the second, it was simply reallocated. The conclusion is the same in both cases, however. Whether risk is reduced or reallocated, social welfare can be improved through voluntary financial transactions.

2. Intertemporal Reallocation of Consumption

Now forget risk. A second motivation for financial transactions is to reallocate consumption over time. This could happen even if individuals have the same rate of time preference, if they have income streams with different patterns over time, or if they have income streams with the same pattern over time but different rates of time preference.

For example, suppose A and B have the same rate of time preference, but A's income stream increases over time, whereas B's decreases over time. Because consumption in each period is subject to diminishing marginal utility, A would prefer to shift consumption from the future to the present, whereas B would prefer to shift it from the present to the future. Both parties could benefit by a trade in which A makes payments in the future to B in exchange for payments made by B to A in the present.

Alternatively, suppose the two have the same (say, flat) time profile of income, but A is very impatient to consume now, whereas B is more willing to defer consumption. Again, mutually advantageous financial trades are possible, with B making payments to A in the present and A paying B in the future.

3. Financing Investment

Finally, suppose that some individuals ("entrepreneurs") can "buy" new streams of income from nature ("projects"). If purchased today, a project entitles its owner to a stream of payments in the future that may vary across states of nature. We can describe these payments by listing some of their characteristics: their expected value in each period, their risk (how much payments vary across possible states of nature in each period), and their maturity (i.e., how far into the future these expected payments extend). What determines how attractive these projects are for the agents that have access to them? If we assume that people prefer more consumption to less, that they prefer present consumption to future consumption, and that they dislike risk, then projects will be more attractive when they have high expected returns, when these returns are close in time to the present, and when they have little risk associated with them.

If some of the projects available to entrepreneurs yield them enough income in the future, with enough certainty, and close enough in time, they may wish to undertake them. But suppose that at least some of these attractive projects are characterized by indivisibility so that they must be undertaken at some minimum scale. In that case, the costs of the projects may exceed the current incomes of the entrepreneurs or the amount of their current consumption that they are willing to forgo. In that case, they could offer to make future payments out of the proceeds of the project that would be more than sufficient to compensate other individuals for making enough payments to them in the present to finance the purchase of the project. If the project leaves enough left over to make the entrepreneur better off, then all parties are better off and would thus voluntarily undertake the transaction. Moreover, the economy as a whole would grow because its stream of future income would be increased by the returns to the project. Furthermore, this growth must be welfare enhancing because by assumption, it is enough to compensate everyone for deferring present consumption.

Our conclusion is that heterogeneity among individuals can give rise to financial trades that are both welfare enhancing and conducive to faster economic growth. But so far, we have been talking only about the motivation for individual trades. What does the market equilibrium look like?

II. FINANCIAL-MARKET EQUILIBRIUM IN AN IDEAL WORLD

We have just seen that, given heterogeneity among economic agents along the various dimensions we have explored, there may be mutually advantageous financial exchanges to be made in the economy. How do these exchanges happen?

As a point of departure, consider first how such exchanges would take place in a world characterized by *perfect competition*, with *public information* and *costless contract negotiation and enforcement*. The assumption that information is public means that each potential lender knows the identity of each potential borrower as well as the potential payoff from the project of each entrepreneur in each possible future state of nature. Costless contract negotiation and enforcement means that there are no costs to reaching agreement on future payments to be made in all possible states of nature and that once the agreement is reached, both sides will abide by the terms of the contract. In other words, under this assumption, commitments once made are automatically binding. Finally, perfect competition means that no agents – neither deficit nor surplus units – have market power. Financial markets would have two important properties under these circumstances.

1. *Financial intermediation would be costless (would absorb no real resources) and could be carried out by individuals*

This must be so because each potential lender knows the identity of each potential borrower as well as all of the characteristics of the income streams available to each borrower in each possible future state of nature. Thus, in such a world, each lender could costlessly contract with any number of borrowers to provide purchasing power today in exchange for payments to be made at previously specified points in the future under previously specified states of nature.

2. *Financial intermediation would be efficient (in the economists' Pareto sense)*

We know that this must be so because if goods are identified by the time and state of nature in which they exist, all the usual Pareto conditions would hold in this case, and the competitive outcome would, by the usual economic reasoning, be Pareto efficient.

In this environment, the price of current resources in terms of future resources (the real interest rate) would be determined by market conditions (the total supply of funds from net lenders and demands from net borrowers). People with access to very attractive projects or with a high degree of impatience would be net borrowers, and people without very attractive projects or with a low degree of impatience would be net lenders. The amount of investment that goes on in this economy would depend on the net cost of borrowing facing entrepreneurs, whereas the amount of saving that goes on would depend on the net return available to savers. In short, the market would costlessly allocate income efficiently between consumption and saving and also allocate saving efficiently among investment projects.

Now consider the pricing and allocation of risk in such a world. The total risk of each project can be decomposed into *systematic risk* (the common risk element in all projects) and *idiosyncratic risk* (the component of risk that is not correlated with systematic risk). Notice that in the idealized world we are describing, no lender would ever have to bear any idiosyncratic risk because such risk could always be diversified away simply by holding claims on a large number of individual projects. The implication is that borrowers would not have to compensate lenders for idiosyncratic risk because lenders would always compete any such compensation down to zero. Thus only systematic risk is left to influence saving and investment. In equilibrium, the price of systematic risk would depend on the degree of risk aversion among the agents in the economy. This price, the amount of compensation that borrowers have to pay risk-averse lenders for assuming risk, is known as the *risk premium*. The risk premium paid by individual borrowers would differ across borrowers only as a function of the amount of systematic risks in their investment projects. Given the market price of risk, agents with a high tolerance for risk (a low degree of risk aversion) would acquire claims on risky projects, while agents with a low tolerance for risk (a high degree of risk aversion) would prefer to acquire safe income streams. Thus risk would be borne by those most willing to do so; that is, it would be *optimally shared*. Notice also that the following would apply in this idealized world:

- There would be no role for specialized financial institutions.
- For a potential borrower, the cost of funds would be the same whether those funds were internally or externally generated.
- A borrower's net worth would play no role in determining the terms on which he or she could borrow in the market.

- The financial structure of enterprises would not affect their value.
- *Liquidity problems* (unexpected needs for cash) would never arise because borrowers and lenders could make fully contingent arrangements to ensure against unanticipated needs for funds.
- There is no role for public policy.

III. THE REAL WORLD: IMPERFECT INFORMATION AND OPPORTUNISTIC BEHAVIOR

The real world, however, is very different from this. In the real world, *information is imperfect* (not everyone knows what a given individual knows), *behavior is opportunistic* (people cannot credibly commit their future actions), and *perfect competition is at best an approximation.* We will focus on the first two problems. As we will see, these two features of economic life create a productive role for specialized financial institutions.

Suppose, then, that instead of public information and costless contract enforcement, the economic environment is characterized by information deficiencies and opportunistic behavior. The kind of information deficiency involved is one in which information is private rather than public; that is, each individual knows his or her own circumstances but not those of other individuals. This leads to two types of information problems:

1. **Searching and matching.** Borrowers and lenders do not know where to find each other.
2. **Asymmetric information.** Borrowers know the properties of their projects, but lenders do not.[1]

On the other hand, opportunistic behavior means that any commitments made by individuals today will not be considered by them to be binding in the future. Thus promises about future behavior are not self-enforcing; instead, they have to be enforced through legal sanctions.

1. Financial Intermediation as a Costly Activity

What are the implications of these phenomena for financial intermediation? The central point is that in this environment, financial intermediation – the transformation of saving into investment and allocation of risk – is not automatic or costless. In a world in which information is not free and agreements cannot be costlessly enforced, intermediation requires the expenditure of resources; that is, it becomes

[1] Notice that borrowers cannot credibly transmit this information costlessly because in this world, everyone knows that he or she may behave opportunistically, i.e., that it is in his or her interest to dissemble.

a form of production. To see why, notice that to strike a loan contract between saver and borrower, several costly (resource-absorbing) activities have to take place:

First, lender and borrower must find each other. This means that *brokerage costs* must be incurred by both parties. For future reference, note that such costs are likely to be independent of the size of the loan that is subsequently negotiated between them.

Second, having found each other, lender and borrower have to strike a bargain. It turns out that this is not trivial in the context of asymmetric information.

To see why, suppose that borrowers are entrepreneurs, all of whom have access to projects with the same expected return but with different distributions across states of nature. Furthermore, suppose that borrowers and lenders are trying to agree on a *debt contract*. Debt contracts stipulate fixed payments (principal plus interest) in all states of nature in which the project yields enough to make such payments and maximum payments in all other states of nature (so creditors get paid before the entrepreneur does). If the lender could observe the characteristics of the individual borrowers, he or she would presumably demand higher returns from riskier projects in good states of nature to compensate him or her for the likelihood of nonpayment in bad states of nature; that is, loan contracts would be tailor-made to the circumstances of the borrowers, and borrowers with riskier projects would have to pay higher interest rates.

But asymmetric information means that lenders cannot costlessly observe the properties of individual projects. This being so, the best the lender can do is to charge everyone an interest rate that would yield the desired expected return for the project with *average* risk. But that means that riskier-than-average borrowers (who would expect to have to pay the interest rate demanded by the lender less often) would in effect face *lower* borrowing costs than safer-than-average borrowers. The implication of this is that the average borrower who actually takes up the lender's offer would tend to be riskier than the population at large. But this being the case, the lender could not expect to receive his or her required expected return under these circumstances (because this particular group of borrowers would on average tend to pay *less* than that). Knowing this, the lender would not be willing to make these loans. This problem, known as *adverse selection*, makes it impossible that a bargain could be struck that both sides would consider advantageous. Neither raising nor lowering the interest rate demanded by the lender would help because the same phenomenon would then repeat itself as at the original terms. Thus markets cannot handle this problem through the price mechanism. Indeed, an important implication of adverse selection is that it causes the price mechanism to break down.

To get around this problem, the lender has to become informed about the likelihood of repayment. Because lenders cannot rely on disclosure by the borrowers, lenders would have to produce an independent evaluation of the loan. This requires obtaining general information about the type of economic activity involved as well as specific information about characteristics of the borrower. All of this implies incurring *evaluation costs*. It is reasonable to think of this as a fixed cost per project (a given fee has to be paid by each lender for each project to which it lends) to acquire the information.

What determines the magnitude of evaluation costs? Evaluation costs are likely to be larger the more *opaque* a borrower is (i.e., the more difficult it is to obtain information about that specific borrower's activities) and smaller the more "transparent" he or she is. In turn, the opaqueness of a borrower depends on factors such as the type of economic activity in which he or she is engaged (new activities, of course, tend to be more opaque), the borrower's previous track record, the amount of publicly available information on the borrower, prevailing accounting and disclosure standards, and so on.

Third, once the loan has been struck and money has changed hands, a resource has been entrusted by one entity (the *principal*) to another (the *agent*) who will actually make the decisions on how to manage that resource. At this point, a new problem arises in the form of the borrower's tendency to behave opportunistically; that is, given a choice, the agent will always have an incentive to make decisions that benefit him or her at the expense of the principal, a situation known as *moral hazard*.

Moral hazard is a general problem in economic interactions, and it can take many different forms. In the context of financial intermediation, the form that moral hazard takes depends on the nature of the contract between borrower and lender. Consider, for example, a debt contract. Suppose that once money changes hands, the borrower can choose from several projects. What project characteristics will the borrower find attractive under such a contract? Assuming no risk aversion, from the standpoint of society (i.e., taking the borrower and lender together), the optimal project choice would be the one with the maximum expected *social* (total) return. But the borrower, behaving opportunistically, would tend to choose the project that maximizes his or her *private* return. Given the nature of the debt contract, this would tend to be a project that pays very high returns (which the borrower gets to keep) in some states of nature, even if its expected return is low because of offsetting losses in other states of nature (because under a loan contract, these losses are borne by the principal). The upshot is that the debt contract creates incentives for the opportunistic borrower, left to his or her own devices, to choose very risky projects (projects with returns that vary greatly across states of nature), even if they have low expected returns. This creates the problem of how, in the absence of precommitment, the principal can ensure that the agent will make resource-management decisions that are in the interests of the principal. This is known as an *agency problem*. It creates the need for measures to safeguard the interests of the lender (*control*), all of which involve costs, known as *agency costs*.

How can borrowers and lenders cope with agency costs? There are several possibilities.

a. The Use of Collateral

Under debt contracts, the key is ensuring that losses in bad states of nature are borne by the borrower. This can happen if the borrower offers sufficient *collateral* (assets that can be seized by the lender in the event of nonpayment) or has sufficient *collateralizable net worth* at stake (defined as net liquid assets plus the collateral value of nonliquid assets).[2] With full collateral, lenders can be completely protected from the possibility of nonpayment. In that case, no agency problems arise because the actions of the agent do not affect the returns received by the principal. The effectiveness of this solution depends on the size of the borrower's collateralizable net worth relative to that of the loan as well as on the efficiency of the legal system that makes it possible to seize collateral.

b. Reputation

Alternatively, if the borrower has an established reputation that would be harmed by nonpayment, then agency problems can be reduced. The reason is that opportunistic behavior that harms the interests of the principal is penalized because the borrower's loss of reputation will make it harder for him or her to borrow in the future.

c. The Imposition of Restrictive Covenants

To the extent that agency problems cannot be fully resolved by ensuring that the borrower has a large stake in the project's outcome (through the offer of collateral and/or reputation effects), lenders could resort to formally restricting the behavior of the borrowers through the imposition of restrictive covenants on loans. These are provisions in loan contracts that reduce moral hazard by ruling out undesirable behavior and encouraging desirable behavior by borrowers. For example, they may stipulate that loans can be used only for specific activities or may encourage behavior like keeping net worth high (i.e., low debt-equity ratios), require the borrower to keep ownership of collateral and maintain it in good condition (e.g., auto loans, home mortgages), impose periodic reporting by firms, or give the lender the right to inspect and audit the firm's books. However, this involves incurring monitoring costs to make sure that the covenants are complied with.

d. Requiring Short Maturity

Restricting the maturity of the debt also facilitates control because it gives the lenders a sanction over the borrowers in the form of withholding funds. Again, this

[2] Note that under *limited liability*, which restricts the borrower's losses to the amount that he or she has invested in the project, not all of the borrower's net worth will be at stake.

involves incurring monitoring costs to determine whether loans will be renewed and enforcement costs if the funds are required to be repaid.

e. Designing Incentive-Compatible Contracts

The form that agency problems take depends on the nature of the contract between the principal and the agent. To reduce such problems, therefore, the two parties can agree to structure the contract between them to cause their interests to be more closely aligned; that is, they can attempt to design *incentive-compatible* contracts. For example, a lender can become an owner by transferring resources in the form of *equity* (a share of the profits), thus seeking to align the incentives of the entrepreneurs with those of the financiers. Unfortunately, however, modifying the nature of the contract may not always eliminate agency problems. In the case of equity contracts, for example, *costly state verification* (that *insiders* involved in managing the firm possess information about the firm that is not costlessly available to some of its owners (*outsiders*)) creates opportunities for managers and majority shareholders to dispossess minority shareholders. This means that the latter would still have incentives to incur monitoring costs. Costly state verification by outsiders makes it possible for managers to exploit owners, for majority owners to exploit minority owners, and so on. The upshot is that, though agency costs may be reduced through these means in some circumstances, they cannot be eliminated.

Fourth and finally, because opportunistic behavior means that contracts are not self-enforcing, lenders may have to be prepared to enforce them through the legal system. *Contract enforcement costs* can be thought of as fixed costs that have to be paid for access to the legal system to enforce agreed future payments.

2. The External Finance Premium

Thus the activity of financial intermediation involves incurring brokerage costs, evaluation costs, agency costs, and contract enforcement costs. All these costs create a wedge between the *gross* return paid by a borrower and the *net* return received by a lender. Because these costs only have to be incurred when funds are acquired externally, the same wedge applies to the difference between the cost of borrowed funds and internally generated funds. This wedge is referred to as the *external finance premium* (Gertler 1988). The costs of financial intermediation are generally not likely to increase in proportion to the size of the loan, so there will be economies of scale in each of these activities. In other words, because these costs have to be incurred per loan (not per dollar loaned), the sum of brokerage, evaluation, agency, and contract enforcement costs would tend to be very high per dollar loaned for small loans. The implication is that because each individual surplus unit makes small loans, the external finance premium would tend to be extremely high if intermediation were to be done by individuals.

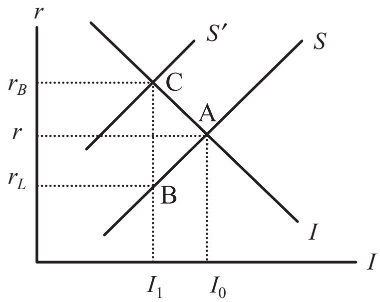


Figure 20.1. Social implications of the external finance premium

To see some of the social implications of this situation, consider the fifth motive for engaging in financial transactions that we considered previously: the financing of investment.

In the absence of the external finance premium, the real interest rate would be determined by the intersection of an upward-sloped saving schedule and a downward-sloped investment schedule, as at the point A in Figure 20.1. When the external finance premium is high, lenders would have to receive a very high gross return for the use of their funds, to compensate them for these costs of intermediation. This shifts the saving schedule to S′. The wedge between the interest rate paid by borrowers, r_B, and that received by lenders, r_L, is determined by the magnitude of the required external finance premium. When the external finance premium is large, the resulting high costs of obtaining external funds would deter entrepreneurs from undertaking all but the most remunerative projects, reducing aggregate investment from I_0 to I_1. The upshot is that potentially highly productive projects that involve a sacrifice of current consumption in return for higher future incomes would not be likely to be undertaken, and the rate of economic growth would thus tend to be low.

IV. BANKS AND CAPITAL MARKETS

We can now address the question of what financial intermediaries do. Financial intermediaries are essentially a market response to the situation just described. They are specialized agents who arise to exploit the high external finance premium that would exist in their absence. Essentially, economies of scale and scope created by the implications of imperfect information and opportunistic behavior make it profitable for firms to arise that specialize in producing financial intermediation.

1. What Do Banks Do?

Banks are firms that announce their willingness to supply liquid assets with attractive features (security, return) and to make information-intensive loans. Their main activity is on the asset side of their balance sheets, consisting of identifying potential borrowers (the brokerage function), gathering information about them (the evaluation function), monitoring their loans, and enforcing the payment of loan contracts. Essentially, banks are profitable because they can carry out these functions more cheaply than individuals can. This means that if they were to charge an external finance premium for their services comparable to that which would exist in their absence, they would make excess profits. It is the allure of these excess profits that triggers the emergence of banks as a market response to the opportunities that would exist in their absence.

How can they do all this? Their key advantage over individuals arises from the benefits of *pooling*. Pooling gives banks advantages over individuals both as lenders and as borrowers. As lenders, pooling enables banks to make a wider range of loans than individuals can, and thus potentially to earn a higher return on their lending. It also permits them to administer those loans at lower cost per dollar lent. As borrowers, on the other hand, pooling permits banks to compete effectively for funds by offering lenders assets that feature an attractive combination of return, safety, and liquidity.

Consider first the advantages that banks have as lenders. Pooling the resources of many lenders permits banks to fund lumpy projects that could not be funded by individuals, thus allowing them to overcome the indivisibility problem that we discussed earlier. To the extent that some of these projects yield high returns, this enlarged menu of options gives banks the option to hold a more attractive portfolio of assets than is available to individual lenders. Moreover, by allowing banks to make many relatively large loans, pooling allows banks to take advantage of economies of scale and scope in lending, thus permitting them to intermediate more cheaply than individuals. Just as individuals do, banks must incur the costs of locating borrowers, evaluating them, monitoring them, and enforcing loan contracts. But because costs of information gathering, monitoring, and contract enforcement are incurred per loan (rather than per dollar lent), and because engaging in some of these activities reduces the costs of engaging in others, banks can take advantage of economies of scale and scope by making a large number of relatively large loans, reducing such costs per dollar lent.

Asset pooling also provides benefits to banks as borrowers. Pooling assets permits diversification, which protects banks from idiosyncratic risk, leaving them exposed only to systematic risk. In other words, by financing a large number of projects, they can achieve the high expected return that these projects offer at relatively low (systematic) risk. This reduces the monitoring costs that have to be borne by bank creditors, allowing the banks to raise funds at lower interest rates. Liability pooling also protects banks against liquidity risk, permitting them to offer assets of short maturity. By pooling the resources of lenders, they can take advantage of the law of large numbers to issue liquid liabilities while keeping only a small share of these resources in liquid assets. They invest the rest in financing the productive projects of entrepreneurs. The short maturity of bank liabilities reduces monitoring costs for those who provide resources to banks and thus makes it possible for banks to attract such resources more easily. Finally, operating on a large scale makes it profitable for banks to make reputation investments that reassure lenders about the dangers of moral hazard. For all these reasons, banks are able to offer savers assets that, while yielding reasonably high returns, are also relatively safe and highly liquid.

In short, banks can attract savings and still be profitable because asset and liability pooling, together with economies of scale and scope, permit them to offer assets with attractive combinations of expected return, risk, and liquidity that dominate cash in the portfolios of prospective lenders.

In performing these functions, banks improve the welfare of both borrowers and lenders, and they promote economic growth. By decreasing the amount of productive resources, per unit of intermediation, that society has to devote to overcoming costs associated with imperfect information and opportunistic behavior, banks enhance the efficiency of intermediation and thus reduce the external finance premium. This permits an increase in the return to saving and a reduction in the cost of borrowing, improving the welfare of both borrowers and lenders. At the same time, banks promote economic growth by improving the composition of the economy's real assets and thus increasing the rate of return of the social portfolio of productive capital. They do so by allowing the economy to undertake projects that may have a high expected return but that would not otherwise be undertaken because their riskiness, lumpiness, and/or illiquidity would not permit them to attract financing from individual savers.

But despite these important contributions, banks are not a complete answer to the problems that the real world poses for the activity of financial intermediation. By definition, banks are institutions that offer safe, liquid assets to savers. They also tend to engage in information-intensive lending. This specialization on the liability and asset sides of their balance sheets means that they do not universally dominate other forms of financial intermediation; that is, there are market niches that banks leave unfilled.

The advantages that banks possess are more important for certain types of lending and for certain types of borrowing than for others. Because banks make information-intensive loans, they will seek out borrowers who are not too opaque (e.g., who are known to bank managers or otherwise have characteristics that make information gathering about them less costly), and they will not have special advantages in competing for borrowers that are very transparent. Banks also will not have special advantages in competing for borrowers whose large stakes in their own projects (because of reputation or collateral) reduce the agency costs of lending to them. Moreover, because they issue liquid liabilities and acquire information-intensive assets, banks are likely to prefer relatively liquid assets or illiquid assets of short maturity. The latter are likely to be less costly to maintain because the shorter the maturity of bank assets, the smaller the stock of liquid reserves that banks will be forced to keep, and the lower monitoring costs will be. Finally, banks are also likely to seek to diversify their portfolios to avoid idiosyncratic risk. This does two things for them: it reduces their need to monitor, and it reduces the need for bank creditors to monitor the banks, thus allowing banks to attract deposits at more favorable rates. The desire for diversification means that banks will be unwilling to make loans that are too large relative to the size of their portfolios.

2. Capital Markets

We have seen that the cost of bank lending is likely to rise with the maturity of the loan and is also likely to increase as the loan becomes large relative to the bank's total portfolio. Thus, in a bank-dominated environment, a relatively high demand price for long-term borrowing in large amounts will exist, set by the cost to banks of making such loans. In this situation, there is a market niche that can be filled by bond financing.

Whether this niche will be filled depends on whether the supply price for this type of lending (i.e., the price at which such lending would be forthcoming) is less than the demand price. The factors that determine the supply price include the following:

- *The magnitude of brokerage and evaluation costs associated with this type of lending*. In turn, this will depend on the visibility and transparency of the potential borrowers as well as the presence and efficiency of specialized firms (underwriters and credit rating agencies) to perform these functions.[3]
- *The magnitude of the collateralizable net worth of the potential borrowers relative to the size of the prospective loans (projects), as well as the value placed by such borrowers on their financial reputations*. Borrowers with large net worth, or who place a high value on their reputations, do not present the types of agency problems that banks can address through loan evaluation and repeated monitoring.
- Because contract enforcement costs also have to be borne under bond financing, the supply price of this type of lending will also depend on the *adequacy of legal provisions for creditor protection (bankruptcy law) as well as the quality of its enforcement*.
- Finally, whether bond markets will arise will also depend on the *existence of a class of lenders that does not place a high value on liquidity and/or on factors that affect the liquidity of bonds such as the potential "thickness" of the bond markets*.

Once again, filling this market opportunity is welfare enhancing because it permits certain investments (i.e., long-term investments by transparent, low-agency-cost

[3] Recall the discussion of Moody's and Standard & Poor's in Chapter 10.

firms) to be carried out that yield returns exceeding those required by lenders to make the funds available; that is, they permit the execution of mutually advantageous trades.

Another reason that banks may not dominate as intermediaries is because firms may prefer equity financing. They may do so because characteristics of the firm may make the agency costs associated with equity contracts lower than those associated with debt contracts, because the size of the project for which financing is sought may be large relative to the firm's equity (thus raising the *relative* agency costs associated with debt financing) and/or because equity financing, by reducing leverage, reduces risk to the entrepreneur by making profits less vulnerable to fluctuations in revenues and interest rates. Thus firms that are risk averse or that are seeking to raise large amounts of funds relative to their owners' equity will tend to generate a high demand price for equity financing.

As with bond markets, whether a supply of equity financing will be forthcoming depends on the attributes of equity claims such as their access to insider information, their legal standing, and their liquidity. All these are enhanced by well-functioning stock markets, the existence of which depends on factors such as the presence of specialized firms to provide brokerage and evaluation services (*stock brokerage* firms tend to do both), the legal protection provided to shareholders (*corporate law*) and the quality of its enforcement, and the potential thickness of equity markets.

Like banks and bond markets, equity markets also create a number of economic benefits that improve the welfare of borrowers and lenders:

- By facilitating trade, stock markets make equity claims on long-gestation projects liquid. This makes equity claims more attractive and thus increases the supply of equity financing.
- By ameliorating agency problems and reducing risk, equity financing may prevent firms from being too small because they forgo large, risky investment projects that would be costly to finance with debt. The result is to foster economic growth.
- Large, liquid stock markets make it easier for an agent to make money on private information, which creates an incentive for the acquisition of such information.
- Through effects on the cost of capital, the near-continuous valuation of firms based on all available information promotes efficiency in capital allocation.
- Liquid stock markets also facilitate takeovers of inefficient firms, which promotes efficiency by imposing discipline on managers.
- Better stock markets can promote corporate control by allowing owners to index the salaries of managers to the value of the firm, thus reducing agency problems by aligning their incentives with the interests of the managers.
- Finally, stock markets facilitate the trading of risky assets, thus allocating risk to relatively less risk-averse individuals who are better able to bear it, thereby improving economic welfare.

V. FINANCIAL MARKETS AND PUBLIC POLICY

However, the desirable welfare and growth outcomes associated with a well-functioning financial system are not automatic. The next three chapters will argue that public policy has an important role to play in ensuring that these benefits indeed materialize. As a framework for that discussion, this section will present a brief overview of the role of public policy in the financial system that follows directly from our analysis so far in this chapter of financial intermediation as a productive activity.

In brief, the efficiency of financial intermediation can be enhanced by public policies of three types: enabling policies, policies directed at the development of financial-market infrastructure, and policies designed to deal with special problems of the financial sector. We now consider each of these in turn.

1. Policies to Promote an Enabling Environment

Enabling policies are policies that improve the environment in which financial intermediaries operate, without necessarily being directed at the financial sector itself. They include policies that facilitate information gathering and contract enforcement as well as policies that reduce risks and financial institutions' ability to monitor it.[4] They also include avoiding the imposition of handicaps on the financial sector through excessive taxation of financial intermediaries or their customers.

Enabling policies have both institutional and macroeconomic dimensions. Institutional policies that would promote the development of the financial system are those that make it less costly for financial intermediaries to cope with the credit-market imperfections that give rise to the external finance premium in the first place. This essentially refers to policies that reduce information costs and costs of contract enforcement through the imposition of an appropriate legal framework. The components of an appropriate legal framework include well-established property rights; adequate accounting and disclosure standards; corporate and bankruptcy laws to protect shareholders and creditors, respectively; and an efficient judicial system to enforce contracts and punish fraud.

Macroeconomic policies can influence the evolution of the financial system in many ways. The key to understanding the influence that macroeconomic policies can exert on financial development, however, is the recognition that the size of the external finance premium can be affected by macroeconomic policies through at least two channels:

1. Macroeconomic policies can influence the financial health of potential borrowers (their net worth) and thus their ratio of collateralizable net worth to the size

[4] These include not just the risks associated with the economic environment but also those created for financial institutions by moral hazard problems (through low borrower net worth).

of prospective loans. This is an indirect influence running from macroeconomic policies to long-term economic performance to financial development.

2. Macroeconomic policies can influence financial development in more direct ways. Taxation of the financial sector, either explicitly or implicitly, will retard its growth and development in obvious ways.[5] Less obviously, the quality of macroeconomic management will affect the degree of uncertainty that characterizes the domestic economic environment and thus the magnitude of loan evaluation and monitoring costs. This channel of influence may be operative at any level of income per capita and any stage of financial development.

With respect to indirect influences mediated through long-term economic performance, the role of macroeconomic stability in promoting long-run economic growth was the subject of Chapter 3, and little additional elaboration is required here. The only additional link to emphasize is that faster growth of income per capita will tend to produce more creditworthy firms as well as to improve the institutional environment for financial intermediation (for evidence, see Easterly 1999). On both accounts, the external finance premium will fall, promoting the growth of financial intermediation, and is likely to fall differentially for arm's-length intermediation through securities markets, increasing the relative role of such markets.

Given the level of income per capita (and thus the average level of net worth of potential borrowers) as well as the quality of the institutional environment, macroeconomic policies can also exert more direct effects on the prospects for financial-sector growth and development. The most obvious way in which this can happen – the explicit or implicit taxation of the financial system – is the subject of the next chapter. But less obviously, the uncertainty associated with an unstable macroeconomic environment increases the costs of financial intermediation and is thus inimical to the process of financial development.

One can imagine a variety of ways in which macroeconomic uncertainty could discourage the development of the domestic financial sector. For example, the emergence of a debt overhang associated with prospective fiscal insolvency, as discussed in Chapter 24, would discourage the acquisition of assets within the taxing authority of the domestic government and thus act as an implicit tax on all domestic financial instruments. Alternatively, instability itself, in the form of the boom-bust cycles that too often have afflicted emerging economies, may make it much more difficult for financial intermediaries to evaluate and monitor the activities of their borrowers, not only increasing the costs of doing business (the external finance premium) but also lowering the average quality of loans and thus

[5] E.g., as we will see in the next chapter, an important reason for the persistence of financial repression in many developing countries was that it met a fiscal need, i.e., it enabled the public sector to capture resources more cheaply than it might otherwise have done. Thus an extensive literature on sequencing of financial reform that emerged in the 1980s typically recommended that fiscal adjustment precede the liberalization of the domestic financial system to wean the government from reliance on the financial repression "tax."

imperiling the health of the financial system itself, as emphasized by Gavin and Hausmann (1996).

In short, financial development requires a conducive environment, and macroeconomic stability is an important component of such an environment. Both indirectly, through the effects of growth on the financial health of potential borrowers and the quality of the institutional environment, and directly, through the degree of uncertainty in the macroeconomic environment, poor macroeconomic policies can undermine the growth and development of financial markets.

2. Proactive Policies Directed at the Financial Sector

Proactive policies directed at the financial sector concern the provision of a regulatory and supervisory framework that promotes competition in the financial sector (preventing collusion) and avoids excessive risk taking due to moral hazard problems. Three separate functions are involved:

1. Implementing and enforcing bank licensing standards and prudential regulations.
2. Developing market infrastructure for capital (bond and stock) markets.
3. Implementing and enforcing antitrust policies directed at the banking and securities industries.

These policies are the subject of Chapter 22, and we can defer further discussion of them until then.

3. Policies to Deal with Financial-Sector Instability

Finally, policies are also required to deal with special problems to which the financial sector is vulnerable. These problems include banking crises and asset-market bubbles.

We will discuss banking crises in more detail in Chapter 26, but for now, we can note that because bank assets are information intensive, they are illiquid, and there tends to be no secondary market for bank assets. At the same time, banks issue highly liquid liabilities. Thus they are inherently illiquid institutions. This makes them vulnerable to liquidity (confidence) crises, which can actually be self-fulfilling. The disruption of bank credit in the context of a crisis affects the real economy through a variety of channels, and governments have consequently sought to avoid banking crises by implementing a variety of financial safety nets such as the creation of a lender of last resort or the implementation of a system of deposit insurance. The lender-of-last-resort function protects individual banks from liquidity risk, while deposit insurance functions provide systemic protection against liquidity crises.

Asset-price bubbles arise when asset prices are driven higher solely by the expectation that such prices will be higher in the future. Bubbles can arise in markets for assets that have no maturity and no defined face value at maturity. Thus they tend to arise in stock and real estate markets, and not in bond markets, because stocks and real estate satisfy these criteria, whereas bonds do not. The value of a stock depends on the expected present value of its dividends plus the present value of its expected price at the time of sale. This last component creates the scope for self-fulfilling expectations that, by causing the value of the stock to deviate from its fundamental determinants, can undermine the efficiency-enhancing effects of equity markets that we discussed previously. When equities loom large in household wealth, stock market bubbles and collapses can have significant real economic effects.[6]

VI. FINANCIAL INTERMEDIATION AND AGGREGATE ECONOMIC ACTIVITY

The theory we have reviewed in this chapter suggests that the premium for external finance depends inversely on borrower net worth. As we will see, this suggests that financial intermediation and aggregate economic activity will tend to be interdependent both in the long run and in the short run. The influence of borrower net worth on the size of the external finance premium will tend to create an interdependence between long-run growth and financial development. Over a shorter time horizon, shocks to the external finance premium may tend to generate business cycles, and a financial accelerator may operate to magnify business cycles originating outside the financial sector.

1. Effects of Growth on Financial Development

The influence of the external finance premium on financial development causes the relationship between financial development and growth to be bidirectional – high income per capita promotes financial development, and financial development promotes long-run economic growth. High income per capita affects financial development for several reasons:

- Countries with high income per capita tend to have large firms. Because large net worth of firms lowers monitoring costs (both by providing collateral and by aligning the incentives of owners more closely with those of creditors), this reduces the costs of intermediation (the premium for external finance) and thus promotes financial intermediation.
- The presence of large numbers of firms with substantial net worth permits the emergence of thick securities (bond and equity) markets that provide liquidity for savers, thus fostering the development of such markets by lowering the costs of issuing securities.

[6] As we saw in Chapter 17, the foreign exchange market may be vulnerable to bubbles as well, for reasons similar to those that apply to stock and real estate markets.

- High income per capita is likely to be associated with greater availability of public goods that facilitate financial intermediation such as an established system of property rights, accounting standards, and an efficient judiciary.

This means that financial development is likely to proceed sequentially. At low levels of income, self-finance and informal finance linked to social relationships are likely to be dominant because the external finance premium will be very large outside the context of social relationships. As incomes rise, firms' net worth will tend to increase, and the availability of public goods required to reduce the external finance premium is likely to improve. Thus the share of total intermediation done by the formal financial system is likely to rise. Though the number of large firms is relatively small and collateral remains limited, however, the dominant intermediary is likely to be the commercial bank because the evaluation and monitoring functions in which banks specialize will continue to be important. Arm's-length securities transactions with little or no monitoring (bond and equity markets) become possible only when net worth is high.

The evidence is consistent with this kind of evolution. For example, among the "stylized facts" about financial development and growth listed by Levine (1992) are that (1) as real income rises, the ratio of financial institutions' assets to gross domestic product (*GDP*) tends to grow, and (2) the importance of intermediaries tends to change in a common pattern as income rises. Central banks become less important, deposit banks grow in importance initially, and then other intermediaries (mutual funds, pension funds, etc.) grow in importance.

Similarly, Gertler and Rose (1991), using annual data for 69 developing and 21 industrial countries over the period 1950–1988, found a positive correlation in panel data between the log of real per capita *GDP* and both the log of the ratio of credit to the private sector to *GDP* and the ratio of quasi-money to *GDP* among developing countries. They found that these results were robust to a variety of perturbations, with little evidence of nonlinearities. They concluded that countries with higher incomes have deeper financial systems.

Finally, Levine (1997), using annual data for 80 countries over the 1960–1989 period, found that the ratio of liquid liabilities of intermediaries to *GDP* increased as income per capita rose. The ratio of bank credit to bank plus central-bank assets also rose with per capita *GDP*. The ratio of credit to private firms to total domestic credit and the ratio of credit to private firms to *GDP* were both positively correlated with real per capita *GDP*.

2. Effects of Financial Development on Growth: Theory

As we saw in Chapter 3, economic growth can occur through improved total factor productivity (TFP) or through the accumulation of productive factors.

From this perspective, financial intermediation can affect growth through three mechanisms:

1. The more efficiently funds are allocated among competing investment projects (including to lumpy, illiquid, high-return projects), the greater is the productivity of the capital stock, and the greater is TFP.
2. The smaller the cost of intermediation, the greater is the amount of investment corresponding to a given amount of saving because savers and investors jointly have to bear the costs of intermediation.
3. The greater the returns to investment, and the lower the cost of intermediation, the greater is the net return to saving, and thus the greater is the incentive to save.

The interactions among these three mechanisms can be illustrated in a simple aggregate growth model that allows a role for the allocative functions of the financial system as well as for the resources absorbed in the process of financial intermediation (see Montiel 1996). Suppose that aggregate output is produced using two kinds of capital, denoted K_1 and K_2, under conditions of constant returns to scale. Thus the aggregate production function can be written as follows:

$$Y = F(K_1,\ K_2) \tag{20.1}$$

The total capital available to the economy is given by

$$K_1 + K_2 = K \tag{20.2}$$

which changes over time according to the amount of investment undertaken each period:

$$\Delta K = I \tag{20.3}$$

Investment, in turn, is equal to effective saving, that is, the portion of aggregate saving not absorbed by the process of financial intermediation. Aggregate saving is proportional to the level of output:

$$I = \sigma s Y \tag{20.4}$$

Here $(1 - \sigma)$ is the cost of financial intermediation per unit of saving, in the form of spreads between borrowing and lending rates, commissions, and so on. In other words, $(1 - \sigma)$ refers to the resources absorbed in producing intermediation services.

To see what this model tells us about the role of financial intermediation in the growth process, let $\theta = K_1/K_2$. Then we can write equation (20.1) as follows:

$$Y = F(\theta,\ 1)\,K_2 = F(\theta,\ 1)K/(1+\theta) = A(\theta)K \tag{20.5}$$

where $A(\theta) = F(\theta, 1)/(1 + \theta)$. This means that the change in Y over time (denoted ΔY) is given by

$$\Delta Y = A(\theta)\Delta K = A(\theta)\sigma sY$$

or

$$\Delta Y/Y = A(\theta)\sigma s \tag{20.6}$$

This model tells us that growth depends on the productivity of the capital stock A, the efficiency of financial intermediation σ, and the rate of saving s. Now notice the following:

- By allocating funds to their most productive uses, the financial sector can increase A. The value of θ that maximizes A satisfies

  $$f'(\theta)/(1 + \theta) - f(\theta)/(1 + \theta)^2 = 0$$

  where f' is the derivative of the function $f(\theta) = F(\theta, 1)$, or

  $$f'(\theta) = f(\theta) - f'(\theta)\theta$$

  which is the requirement that the marginal product of the two types of capital be equalized. This outcome will emerge if financial institutions are able to identify the marginal product of capital in alternative uses and channel funds in such a way as to give priority to high-productivity projects.
- By operating efficiently (at lowest cost per dollar of funds intermediated), the financial sector can increase σ.
- Finally, the combination of high return on investments and low intermediation costs means a potentially large return to savers, which may increase s.

We conclude that a well-functioning financial system can foster economic growth. Moreover, these effects are interactive. In other words, the positive effects of efficient financial intermediation on growth tend to reinforce each other.

3. Effects of Financial Development on Growth: Evidence

There is substantial empirical evidence in support of the proposition that a deep and well-functioning financial system is conducive to faster long-run growth. This evidence has taken a wide variety of forms, including cross-country growth regressions, country case studies, industry studies, and studies of the effects of structural adjustment programs supported by the World Bank, among others.

Several studies have examined the effect of financial development on growth using a cross-country growth framework. The methodology involves adding an indicator of the adequacy of financial intermediation either to the core growth regression mentioned in Chapter 3 or to one that includes other potential policy determinants of growth such as openness, government size, the conduct of monetary policy, and so on. A large number of variables have been used as proxies for financial development:

- indicators of financial depth, such as the ratios of narrow (M1) and broad (M2) money, of quasi-money, and of liquid liabilities (currency outside banks plus demand and interest-bearing liabilities of banks and nonbank financial intermediaries) to GDP
- the share of financial intermediation done by commercial banks (measured by the ratio of the domestic assets of deposit money banks to the domestic assets of deposit money banks and the central bank together, as in King and Levine (1992, 1993))
- the volume of lending to the private sector, measured by ratio of the flow of credit to the private sector to *GDP*, as in De Gregorio and Guidotti (1996), or by the share of total domestic banking-system credit extended to the private (as opposed to the public) sector (King and Levine 1992, 1993)

The most comprehensive study of the cross-country type that addressed the robustness issue raised about such studies in Chapter 3 is that by King and Levine (1993). These authors found, in a sample of 77 countries with data averaged from 1960 to 1989, that their financial indicators were closely correlated with each other as well as with measures of *GDP* growth, capital accumulation, and total factor productivity. In standard cross-country growth regressions applied to each of these growth indicators, they found that when entered separately, the average value of each of the financial indicators was statistically significant at the 5 percent level. These results were confirmed with pooled cross section–time series regressions using decade averages and instrumental variable techniques to deal with the potential problem of reverse causation arising from the fact that financial depth is itself affected by economic growth, as described earlier. They also found that initial financial depth helped to explain subsequent growth performance, again using a variety of indicators both of growth and of financial depth. All these results tended to be robust across samples of countries and time periods as well as with respect to the set of other explanatory variables included in the regressions.

More recently, Levine (1997) used the same sample, but with data averaged over the full period, to conduct a more extensive study of the effects of financial variables on growth. He estimated three cross-country regressions of the form

$$g(j) = \alpha_0 + \alpha_1 F(j) + \alpha_i X(ij) + \varepsilon_j$$

in which the dependent variable $g(j)$ referred successively to growth of income per capita, of the capital stock, and of total factor productivity in country j; F was a vector of four financial indicators (used one at a time); and $X(ij)$ was a matrix of control variables (including initial income per capita, initial education, political stability, and indicators of exchange rate, trade, fiscal, and monetary policy). In each regression, all financial variables were statistically significant and economically important, indicating that improved financial-sector performance enhances growth both by improving total factor productivity growth and by inducing more factor accumulation, as suggested by our model.[7]

4. The Financial Accelerator

Finally, the analysis of this chapter points to a mechanism through which financial variables can aggravate short-run macroeconomic instability, referred to as the *financial accelerator*. Because macroeconomic cycles tended to be accompanied by swings in the net worth of firms, and because these swings in net worth cause the external finance premium to move in the opposite direction (countercyclically), balance-sheet effects on firms will cause borrowing to move pro-cyclically. This tends to aggravate the severity of macroeconomic shocks and thus aggravate macroeconomic instability.

Also, phenomena such as deflation or exchange rate devaluation can, when debt is unindexed, result in a contraction of net worth that sharply increases the premium on external finance. The Great Depression in the United States may have been an example of the former. As we shall see in Chapter 27, the crises that many emerging-market economies experienced during the 1990s provided several examples of the latter.

VII. SUMMARY

This chapter has provided an overview of the economic role of the financial system, laying the basis for the discussion of financial-sector policies in the chapters that follow. We have seen that financial transactions take place for the same reason other economic transactions do: to exploit mutually advantageous trades. Though in an ideal world, these transactions are costless and result in efficient market allocations, imperfect information and opportunistic behavior make financial intermediation a costly activity (i.e., one that absorbs resources) in the real world. Costs of financial intermediation include brokerage, evaluation, monitoring, and contract enforcement costs. These give rise to external finance premia. Banks and other financial intermediaries arise to exploit profit opportunities that exist in the form of external financial premia.

Though financial intermediaries are in business to make profits, a well-functioning financial sector enhances economic welfare by decreasing the real resources, per unit of intermediation, that society has to devote to coping with the problems posed by imperfect information and opportunistic behavior. A well-functioning financial system can also enhance economic growth by allocating capital

[7] Wachtel (2003) provides a critical review of the empirical literature on finance and growth, which nevertheless concludes that the causal effect of financial development on growth is well established empirically.

more efficiently, including to risky, high-return projects; by increasing the resources available for investment out of a given amount of saving; and by improving the incentives to save in the form of higher risk-adjusted rates of return for savers.

But achieving these benefits depends on the stance of public policy. Public policy can help through the adoption of institutional and macroeconomic policies that create an enabling environment, through the specific development of market infrastructure for financial intermediation, and through specific policies to cope with special problems of the financial sector.

In the next two chapters, we will analyze policies toward the financial sector more explicitly, before turning in Chapter 23 to the reformulation of our benchmark model of Part 2 to account for the presence of banks. Chapter 24 will conclude this part of the book with an important application: the interaction between financial-sector and macroeconomic policies that arises in the management of capital inflows.

REVIEW QUESTIONS

1. How do financial transactions affect economic welfare?
2. In what sense is financial intermediation a productive activity?
3. What is the social (economic) role of banks? Why are banks able to intermediate funds at less cost than private individuals?
4. Why do capital (stock and bond) markets exist; that is, why is all financial intermediation not performed by banks?
5. Can public policy affect the external finance premium? If not, explain why not. If so, explain how.

EXERCISES

1. Use the standard introductory-economics analysis of consumer and producer surplus in Figure 20.1 to show how a high external finance premium can reduce social welfare.
2. Interpret the emergence of banks as a social innovation that reduces the external finance premium. Use the analysis of exercise 1 to show how the emergence of banks can increase *both* welfare and growth.
3. Using the model of Section VI, show how each of the following would be expected to affect an economy's growth rate:
   a. a tax on financial intermediation
   b. a differential tax on capital invested in one of the two sectors of the model
4. What effects would you expect each of the following to have on an economy's external finance premium, and why?
   a. poorly defined property rights so that title to real assets cannot be clearly established
   b. lack of harmonization across firms in accounting standards
   c. a cumbersome and expensive legal system
   d. an uncompetitive banking system

5. Using the analysis of financial frictions in this chapter, explain why banks, which engage in information-intensive lending, typically finance themselves through deposits that can be withdrawn on demand.

REFERENCES AND FURTHER READING

De Gregorio, Jose, and Pablo Guidotti (1996), "Financial Development and Economic Growth," in A. Solimano, ed., *Road Maps to Prosperity: Essays on Growth and Development*, Ann Arbor: University of Michigan Press, pp. 237–266.

Easterly, William (1999), "Life during Growth," *Journal of Economic Growth*, vol. 4, pp. 239–276.

Gavin, Michael, and Ricardo Hausmann (1996), "The Roots of Banking Crises: The Macroeconomic Context," in R. Hausmann and L. Rojas-Suarez, eds., *Banking Crises in Latin America*, Baltimore, MD: Johns Hopkins University Press, pp. 27–63.

Gertler, Mark (1988), "Financial Structure and Aggregate Economic Activity: An Overview," *Journal of Money, Credit, and Banking*, vol. 20, pp. 559–588.

Gertler, Mark, and Andrew Rose (1991), "Finance, Growth, and Public Policy," Policy Research Working Paper WPS 814, World Bank.

King, Robert G., and Ross Levine (1992), "Financial Intermediation and Economic Development," unpublished manuscript.

King, Robert G., and Ross Levine (1993), "Finance, Entrepreneurship, and Growth: Theory and Evidence," *Journal of Monetary Economics*, vol. 32, pp. 513–542.

Levine, Ross (1992), "Financial Structure and Economic Development," Policy Research Working Paper WPS 849, World Bank.

Levine, Ross (1997), "Financial Development and Economic Growth: Views and Agenda," *Journal of Economic Literature*, vol. 35, pp. 688–726.

Montiel, Peter J. (1996), "Financial Policies and Economic Growth: Theory, Evidence, and Country-Specific Experience from Sub-Saharan Africa," *Journal of African Economies*, *Part I*, vol. 5, pp. 65–98.

Wachtel, Paul (2003), "How Much Do We Really Know about Growth and Finance?" *Federal Reserve Bank of Atlanta Economic Review* (First Quarter), pp. 33–47.

21

Financial Repression

Despite the contributions that an efficient financial system can make to economic welfare and growth, many developing countries have traditionally applied a combination of policies toward the financial sector that has impeded the functioning of their domestic financial systems. This policy package has become known as *financial repression*. This chapter will describe this set of policies and consider how they affect the functioning of the financial system as well as how they may impinge on each of the channels through which the financial system can promote growth. We will also review some of the evidence on the growth effects of financial repression in developing countries. As we will see, a substantial amount of evidence is available suggesting that financial repression has had harmful effects on economic growth.[1]

In part because of this evidence, the view that financial repression is harmful for growth has carried the day among policy makers in emerging markets, and most emerging economies are well along in the process of revamping their policies toward the financial sector, a process known as *financial liberalization*. But this process has not proven to be an easy one. We will examine it in detail in the next chapter. As we will see there, a key question is what conditions have to be in place in the domestic economy for a liberalization of the domestic financial system to have a chance of being successful.

To answer this question, we have to consider why financial repression may have existed in the first place, that is, what economic role financial repression plays in the domestic economy. This chapter will argue that though financial repression can be interpreted as just an extreme form of regulation, the central motivation for financial repression has often been a fiscal one. In effect, governments have relied on implicit taxation of the financial sector in part because of their difficulties in raising the resources they required through more conventional means.

[1] The classic references on financial repression are McKinnon (1973) and Shaw (1973). See also Fry (1988).

This fiscal role of financial repression links this part of the book with Part 3. Just as the excessive reliance on seignorage or the emergence of debt-servicing problems reflects fiscal difficulties, placing an excessive burden of implicit taxation on the domestic financial sector can be viewed as yet another consequence of fiscal problems in emerging and developing economies. In this chapter, we will describe the fiscal effects of financial repression as well as techniques for the empirical measurement of the fiscal revenues derived from this set of policies toward the domestic financial sector.

The next section describes the components of financial repression, and Section II explains why financial repression has fiscal implications. The effects of financial repression on economic growth are considered in Section III, whereas Section IV takes up an important puzzle associated with financial repression: though we would expect financial repression to be harmful for economic growth, and the cross-country evidence is indeed consistent with this expectation, some of the fastest-growing emerging and developing economies in the world, the so-called miracle economies of East Asia, achieved their remarkable growth successes in the context of repressed financial sectors. Investigating how they did so helps to shed light on the specific channels through which financial repression affects growth.

I. THE INGREDIENTS OF FINANCIAL REPRESSION

We saw in Chapter 20 that in the early stages of financial development, commercial banks are likely to be the dominant domestic financial institutions because of their comparative advantage in overcoming financial-market imperfections. We also saw that the role of collateral in overcoming moral hazard problems, and of a well-functioning legal system in facilitating the enforcement of contracts, make formal financial institutions such as banks heavily dependent on the legal system. This dependence on the legal system makes formal banking activity highly visible to governments and makes banks relatively easy to tax. Historically, developing-country governments have tended to tax the banking system both directly and indirectly. Indirect taxation has been levied through the mechanism of financial repression. The set of policies characterized as financial repression includes the following.

1. Controls of Capital Inflows and Outflows

Under financial repression, domestic residents are typically not allowed to hold foreign assets, and domestic firms are not permitted to borrow abroad. The implication is that financially repressed economies are typically characterized by financial autarky and that foreign intermediaries cannot compete with the domestic financial industry. Thus, under financial repression, we can think of the domestic financial sector as operating in conditions similar to those of any other domestic productive activity that is insulated from external competition through a system of prohibitive tariffs or quantitative restrictions.

2. Restrictions on Entry into the Formal Financial Sector

Under financial repression, the domestic formal financial sector is not subject to free entry and exit, and many domestic banks may be publicly owned. These two conditions imply that the domestic financial sector does not typically operate under competitive conditions. Indeed, repressed banking sectors are typically dominated by a few banks, the largest of which are government owned.

3. High Reserve and Liquidity Requirements on Banks

Banks are required to keep high reserve ratios, either as vault cash or on deposit with the central bank, and such deposits typically do not earn interest. In addition, banks may be required to maintain liquidity ratios in the form of government securities, which, though they pay interest, typically yield a return much lower than would be required for banks to choose to hold them voluntarily. Through both mechanisms, banks pay an implicit tax and lose the freedom to allocate a large share of their portfolios to productive loans.

4. Interest Rate Ceilings on Bank Assets and Liabilities

The interest rates that banks can pay on deposits, and that they can charge for their loans, are legally controlled in repressed financial systems. This means that banks cannot compete with each other on price and cannot raise deposit rates to compete with nonbank intermediaries (informal financial markets or illegal capital flight) for resources. On the other side of their balance sheets, if interest rate ceilings on loans are binding, banks cannot allocate their loans on the basis of price and are thus forced to engage in nonprice rationing of credit.

5. Directed-Credit Restrictions on the Composition of Bank Asset Portfolios

In addition to having to set aside a substantial portion of their portfolios as required reserves and to meet liquidity requirements, banks often have limited discretion over the allocation of their remaining funds because they are typically forced to set aside designated shares of their lending portfolios for priority sectors or specific classes of firms, sometimes at preferential interest rates.

6. The Use of Bank Credit Ceilings as Instruments of Monetary Control

Under the circumstances described here, monetary policy has often been conducted by setting overall targets for total credit expansion by the domestic banking system, and then allocating this total among individual banks, thus restricting the amount of total credit that can be extended by each bank.

This set of policies obviously represents a severe set of restrictions on the entire process of domestic financial intermediation. It circumscribes the behavior of domestic borrowers and lenders, the number of firms that can engage in financial intermediation, and the behavior of firms that are already so engaged. In view of the important role of the domestic financial system in mobilizing and allocating resources described in the preceding chapter, the questions raised by such policies are what their implications for the functioning of the domestic economy might be, and why such a set of policies would be adopted in the first place.

From an economist's perspective, restrictions on financial intermediation such as those included in this package of policies would seem to preclude mutually advantageous trade in a variety of ways and would thus be presumed to have adverse effects on economic welfare. It is hard to see how, under such tight restrictions on behavior, the domestic financial sector would retain the incentive and flexibility to respond to profit opportunities created by changes in the economic environment in a way that would have the effect of lowering the external finance premium. As we saw in the preceding chapter, this is the mechanism that drives financial development and that enables the process of financial development to enhance economic welfare and facilitate economic growth.[2]

It would be inappropriate to jump to this conclusion too readily, however, because imperfections in credit markets, such as those described in the previous chapter, suggest that laissez-faire is not the appropriate standard of comparison to use in evaluating policies directed at the domestic financial system. These credit-market imperfections may create a valid role for some types of government interventions in the domestic financial system. However, as we also saw in the preceding chapter, the interventions called for are those that address those imperfections directly; that is, that *some* types of interventions may be justified by market imperfections does not mean that *any* intervention is justified. In the next chapter, we will look in much more detail at precisely what types of interventions may make sense in light of the credit-market imperfections that we have discussed.

As it happens, however, financial repression has not typically been motivated by the desire to address the problems created by credit-market imperfections for the process of financial intermediation, and thus there is no particular reason to believe that such policies would tend to improve the functioning of the domestic financial system in light of credit-market imperfections. In Section III, we review the arguments for the view that the policies associated with financial repression indeed do substantial economic harm, and we also examine some of the empirical evidence

[2] It is worth noting that the rationing of credit associated with restrictions on lending rates under financial repression may also have adverse effects on inequality because access to credit will often be restricted to the most advantaged members of society, even in the absence of corruption.

on this issue. Before doing so, however, we need to ask if financial repression has not typically been directed at the amelioration of the effects of credit-market imperfections, and if such policies may indeed be harmful, then why have they been adopted?

II. FISCAL ROOTS OF FINANCIAL REPRESSION

In this section, we will argue that the key reasons have been *fiscal* in origin. Financial repression affects both the expenditure and financing side of the public sector's budget. It permits the nonfinancial public sector to engage in industrial policies (i.e., to favor some economic activities over others) without the expenditure of fiscal resources and to borrow more cheaply than it could otherwise do. How does this work?

1. Financial Repression as a Fiscal Phenomenon

One way to view the set of policies that constitute financial repression is as follows: the restrictions on the behavior of bank customers and potential competitors (capital account restrictions and restrictions on entry into the domestic financial sector) have the effect of creating monopoly rents for the existing banks. Some portion of these rents is then effectively taxed away by the government through the restrictions that repression imposes on the behavior of the protected banks themselves.

For example, restrictions on capital inflows and outflows insulate the domestic financial system from external competition. Restrictions on capital outflows prevent the public from taking their savings abroad, thus creating a captive demand for the liabilities of the domestic financial system, while restrictions on inflows create a captive market for domestic bank loans. In effect, they make the demand curves for bank assets and liabilities less elastic by eliminating access to close substitute sources of funds as well as substitute saving vehicles. Restrictions on entry into the domestic financial industry, on the other hand, allow domestic banks to collude, thus exploiting their collective monopoly position in domestic financial intermediation.

The result of these restrictions is to create large potential wedges between bank borrowing and lending rates. Under financial repression, however, a substantial portion of this wedge is captured by the government rather than by the banks themselves. How does this work?

- High reserve and liquidity requirements create an artificially high demand from domestic banks for base money and for public-sector securities. The former increases the base on which the public sector collects seignorage revenue, whereas the latter reduces the interest rate that the public sector must pay on its securities, thus lowering its borrowing costs.

- Interest rate ceilings on loans directly reduce the cost for the nonfinancial public sector to borrow in the form of bank credit and advances from private banks.
- Directed credit policies permit the government to engage in *industrial policy* – that is, to favor specific economic activities over others – by granting them access to credit at subsidized rates rather than by paying them explicit subsidies that would need to be financed through the public sector's budget.

2. Financial Repression and Seignorage

To see how the first of these channels operates, we need to modify the public-sector solvency analysis of Part 3 of the book to take into account the existence of banks. Recall that the analysis of seignorage issues was based on a financial structure in which the central bank issued currency that was held by the public directly, and the public sector issued securities that were sold in an open market. In that context, seignorage revenue was given by $(\pi + n)m$, where m was the stock of central-bank liabilities. In Part 3, central-bank liabilities consisted of the stock of currency in the hands of the public. However, when banks are accounted for, the two concepts are no longer equivalent. Central-bank liabilities are referred to as the *monetary base*, or *high-powered money* (call it H, so m would be replaced by $h = H/PY$). The central bank's balance sheet thus becomes

$$SF_C^* + B_C = H \tag{21.1}$$

Money, on the other hand, consists of currency held by the public plus demand deposits (checking account balances at banks). The latter are the liabilities not of the central bank but of the commercial banks. The conceptual basis for this definition of money is that it consists of all assets that can be used directly to make payments. It is sometimes called *narrow money*, or *M1*, to distinguish it from definitions that also include assets that can easily be converted into means of payment.

What is the relationship between money defined this way and the monetary base? As we have already noted, commercial banks typically have to hold some fraction (call it rr) of their deposits as reserves. These can be held in the form of currency (a central-bank liability) or deposits at the central bank (which are, of course, also central-bank liabilities). Thus total central-bank liabilities H consist of currency held by the public (call this CUR) plus commercial bank reserves:

$$H = CUR + rrD \tag{21.2}$$

where D is total demand deposits. Money, on the other hand, is given by

$$M = CUR + D \tag{21.3}$$

To see what impact financial repression has on H (and thus on h and on seignorage revenue $(\pi + n)h$), we need to examine how it affects the demand for H, a simple form of which is given by equation (21.2). To inject more realism into the analysis, we need to take into account that other types of deposits are also subject to reserve requirements. Including reserve requirements on time deposits (savings accounts), for example, makes the demand for H

$$H = CUR + rrD + rr_T TD \tag{21.4}$$

where rr_T is the required reserve ratio on time deposits and TD is the stock of such deposits. Now all we need to do is specify the demands for CUR, D, and TD. To be concrete, consider the following asset-demand model[3]:

$$CUR/PY = f_C(\underset{-}{\pi}, \underset{-}{R_D}, \underset{-}{R_T}) \tag{21.5a}$$

$$D/PY = f_D(\underset{-}{\pi}, \underset{+}{R_D}, \underset{-}{R_T}) \tag{21.5b}$$

$$TD/PY = f_T(\underset{-}{\pi}, \underset{-}{R_D}, \underset{+}{R_T}) \tag{21.5c}$$

where R_D and R_T are interest rates paid by banks on demand and time deposits, respectively. These are standard portfolio demands similar to the ones we used earlier in the model of Part 2. Inflation is assumed to reduce the demand for all nominal financial assets. On the other hand, an increase in its own rate of return increases the demand for each asset, whereas an increase in the rate of return on a substitute asset reduces that demand. Substituting these asset demands into equation (21.2), we now have the following:

$$H = f_C(\underset{-}{\pi}, \underset{-}{R_D}, \underset{-}{R_T})PY + rr_D f_D(\underset{-}{\pi}, \underset{+}{R_D}, \underset{-}{R_T})PY + rr_T f_T(\underset{-}{\pi}, \underset{-}{R_D}, \underset{+}{R_T})PY \tag{21.6}$$

Now we can see what happens to H under financial repression. Recall that under financial repression, rr_D and rr_T would both tend to be high, whereas R_D and R_T would both be low. Thus we can see the following:

- An increase in either reserve ratio would increase the demand for H, given the values of the other variables.
- A reduction in R_T would tend to *reduce* the demand for H through the last term on the right (because it reduces the demand for time deposits) but to *increase* it through the first two terms (the demands for currency and demand deposits). What is the net effect? In principle, the effect is ambiguous. It depends on whether time deposits are better substitutes for assets held inside or outside the formal financial system. What might the latter consist of? Recalling the model of Part 2, the answer is primarily foreign assets, not included in the

[3] The model is from Anand and van Wijnbergen (1989).

preceding reformulation. But this is precisely what capital controls prevent from happening. Thus, with capital controls in place, reductions in TD would be likely to go primarily into CUR and D. Given that CUR has an implicit reserve requirement of unity (to hold a unit of currency, one has to hold a unit of high-powered money), H would tend to rise.

- A reduction in R_D would reduce demand for D and thus indirectly for H, but because demand deposits are transaction balances, these funds would likely go into the other asset that can be used as a means of payment, namely, CUR, which carries a higher reserve requirement. Again, the demand for H would rise if currency and demand deposits are close substitutes.

The conclusion is that high values of rr_D and rr_T, as well as low values of R_D and R_T, all of which are associated with financial repression, would be conducive to high values of H and thus to high values of seignorage revenue $(\pi + n)h$ for a given rate of inflation π.

3. Fiscal Revenues from Financial Repression

The second mechanism through which the government receives revenues from financial repression is through a reduction in its borrowing costs, either through lower interest rates on bank credit than would prevail otherwise or through mandatory holdings of government securities by domestic financial institutions. Because these securities are not held willingly, they must pay a lower interest rate than would be required for banks to hold them voluntarily. Giovannini and de Melo (1993) estimated the magnitude of the revenues derived from this source for a sample of 24 developing countries during the period 1972–1987, when financial repression was common. They estimated it by taking the difference between the ex post cost to the government of borrowing abroad (including capital gains or losses arising from exchange rate changes) and its cost of borrowing domestically in the same period, multiplied by the average stock of domestic debt held by the private sector in that period.

The estimates of financial repression tax revenues derived by Giovannini and de Melo (1993) are presented in Table 21.1. As is evident from the fourth column of the table, the implicit financial repression tax rate (the difference between the foreign and domestic interest rate, as a percentage of the former) was very high in some countries, with the implication that when domestic debt outstanding is also high, the financial repression tax can loom large compared to conventional sources of government revenue (shown in the third column). On average, for the countries in this sample, the financial repression tax amounted to about 2 percent of gross domestic product (GDP) and about 9 percent of tax revenue. Thus the magnitudes involved can be quite substantial.

Table 21.1. *Financial Repression Tax Revenue in 24 Countries*

	Tax Revenue as Percentage of GDP	Tax Revenue as Percentage of Total Tax Revenue	Implicit Tax Rate
Algeria	4.30	11.42	10.6
Brazil	0.48	1.57	13.4
Colombia	0.24	2.11	22.4
Costa Rica	2.33	12.76	25.1
Greece	2.53	7.76	16.0
India	2.86	22.38	11.0
Indonesia	0.00	0.00	0.0
Jamaica	1.38	4.74	7.4
Jordan	0.60	2.40	7.2
Korea	0.25	1.36	6.0
Malaysia	0.12	0.31	0.5
Mexico	5.77	39.65	45.8
Morocco	2.31	8.89	16.1
Pakistan	3.23	20.50	25.3
Panama	0.69	2.49	4.4
Papua New Guinea	0.40	1.90	5.6
Philippines	0.45	3.88	11.9
Portugal	2.22	6.93	15.8
Sri Lanka	3.40	19.24	14.5
Thailand	0.38	2.57	4.3
Tunisia	1.49	4.79	13.4
Turkey	2.20	10.89	55.8
Zaire	0.46	2.48	54.5
Zimbabwe	5.52	19.13	19.5

Source: Giovannini and de Melo (1993).

Chamley and Honohan (1990) estimated a broader concept of the magnitude of taxation of the financial sector, including not just the implicit tax on the interest rate on domestic government securities but also the inflation tax, the implicit tax on required reserves imposed by the fact that these reserves are remunerated at less than market rates, the implicit tax imposed by loan interest rate ceilings, and indirect taxes on the financial sector. They calculated the inflation tax as the product of the inflation rate and the stock of currency in the hands of the public, the implicit tax on required reserves as the gap between an estimated market interest rate and the rate of remuneration on reserves (usually zero) times the stock of reserves, the implicit tax on government borrowing as the gap between the market interest rate and the interest rate actually paid on government debt, and the implicit tax on lending to nongovernment borrowers as the gap between an estimated market interest rate and the controlled loan rate. They also took into account indirect taxes on financial intermediaries. Thus they estimated the volume of resources extracted from the financial system by using the following formula:

$$\begin{aligned} TAX = {} & (R^* - 0.01) * CURRENCY + (R^* - RRES) * RESERVES \\ & + (R^* - RTB) * GOVTBOR + (R^* - RTB + MARGIN) \\ & \times NONGOVTBOR + INDIRECT\ TAXES \end{aligned}$$

Here R^* is an estimate of the market-clearing risk-free interest rate that would prevail without interest rate ceilings, *RRES* is the rate of remuneration on reserves, *RTB* is the bank lending rate, *MARGIN* is an assumed risk premium for private borrowers, and *GOVTBOR* and *NONGOVTBOR* are the stocks of loans outstanding to the government and nongovernment sectors, respectively. They used several methods for estimating shadow market interest rates, including the rate on foreign borrowing (as in Giovannini and de Melo 1993) and a nominal rate calculated by adding an assumed real interest rate of 1 percent to the observed rate of inflation.

Chamley and Honohan (1990) calculated tax rates for five African countries during the decade 1978–1988. Their findings were similar to those of Giovannini and de Melo (1993). Specifically, they found that the taxes on the formal financial sector that they measured were in the range of 4–7 percent of GDP for Ghana, Nigeria, and Zambia during this period, and in the range of 2 percent of GDP for Cote D'Ivoire and Kenya, compared with explicit tax revenue in the range of 10–25 percent of GDP for most Sub-Saharan African countries. Most important, they found that the average tax collected was in all cases greater than the value added of the banking system, and in the case of the three high-tax countries, it was a multiple of banking system value added, even after excluding seignorage on currency in the hands of the public, a portion of taxation that does not fall on banks. These findings led them to conclude that

> by any reckoning, the financial sector has been very heavily taxed in comparison with other sectors. For instance, the average tax collected has in all cases exceeded the value added of the banking system. Even excluding the currency tax, which does not bear directly on the banking system, the average tax collected has been a multiple of the value added of the banking system in the three high-tax countries. (Chamley and Honohan, 1990, 16)

These results are supported by those of Ikhide (1992), who focused only on implicit taxation in the form of unremunerated reserve requirements in eight Sub-Saharan African countries. He found that such reserves were significantly higher in the eight countries in his sample than is typical in Organization for Economic Co-operation and Development countries and that the implicit tax on the financial sector just from this source ranged from about 1.5 percent of *GDP* in Tanzania to about 7.5 percent in Ethiopia. In five of the eight countries he examined, this amounted to more than one-quarter of government revenue.

III. FINANCIAL REPRESSION AND GROWTH

We saw in Chapter 20 that economic theory gives us reason to believe that an efficient domestic financial system can help to promote economic growth by improving the efficiency of resource allocation, by freeing up resources that would otherwise be used in intermediation to produce other goods and services, and by encouraging accumulation through the higher rates of return that such a system could offer to domestic savers. The empirical evidence reviewed in Chapter 20 was consistent with the proposition that financial depth is conducive to faster economic growth. In this section, we will review what theory and evidence have to say about the effects of financial repression on economic growth.

1. Financial Repression and Growth: Theory

To analyze the effects of financial repression on growth suggested by economic theory, we can go back to the small growth model that we used in the preceding chapter. Recall that in that model, the growth rate of productive capacity could be expressed as

$$\Delta Y/Y = A\sigma s$$

where A is a measure of total factor productivity, σ is an indicator of the efficiency of the financial system (in the sense of resources used by the sector), and s measures the ratio of investment to *GDP*. How is each of these components likely to be affected by the policies associated with financial repression?

a. Effects on Efficiency of Allocation

- *Restrictions on competition* may impair allocative efficiency because state-owned and protected banks will not have the competitive incentives to screen and monitor borrowers closely.
- The appropriation of funds by the public sector through the maintenance of *high reserve requirements* means that a portion of household saving will be channeled into government spending. To the extent that the government consumes these resources, σ will fall. To the extent that they are invested in public capital, the associated projects may not yield returns in excess of the foregone investment in the private sector.
- *Interest rate ceilings* on loans prevent the system from allocating capital to the most productive uses. Whether interest rates are controlled at below-market levels, banks have to screen prospective credit applicants. Interest rate ceilings prevent price-based allocation of funds (i.e., the weeding out of unproductive projects because they cannot yield a rate of return high enough to service the debt accumulated in financing them), leading to credit rationing, which may result in funds being allocated according to arbitrary criteria by individual

banks. Moreover, low interest rates in the formal financial system are likely to create an informal financial market. When formal and informal financial markets coexist and some firms have unlimited access to the formal market, the marginal product of capital in these favored firms will fall below the cost of borrowing in the informal system. Capital may thus be misallocated both among firms that have access to the formal market and between firms that have access to that market and those that do not.

- *Interest rate ceilings on deposits* may also cause funds to be misallocated through disintermediation, that is, by misallocating funds between the formal financial system and other types of financial intermediation. Their effect is to channel domestic saving into relatively unproductive investment by the savers themselves, into the informal financial system, or abroad.
- The use of *directed credit* forces institutions to lend to projects that may not meet a market test, that is, that may yield lower rates of return than alternative activities.
- Finally, the use of *bank credit ceilings* as an instrument of monetary control also reduces the efficiency of capital allocation by misallocating funds among banks. Ceilings on individual banks, applied uniformly, prevent lending from being reallocated to the banks that have the most productive investment opportunities available to them.

Because all these policies would tend to allocate resources in ways that fail to maximize social returns, we would expect their presence to result in a reduced value of total factor productivity in the economy.

b. Effects on the Costs of Intermediation

- *Restrictions on capital flows* as well as *restrictions on entry* limit the scope for competition in the formal financial system, removing an important incentive to reduce costs. This tends to divert to factors of production employed in the formal financial system resources that would otherwise have been intended for investment.
- Measures such as *directed credit*, or the *imposition of reserve and liquidity requirements*, in addition to distorting the allocation of funds, function like a tax on financial intermediaries. This diverts resources to the government that would otherwise have been intended for investment.
- Policies, such as *ceilings on deposit interest rates* or *credit ceilings on individual banks*, that reduce the rates of return offered to domestic savers in the domestic formal financial system and thus reduce the *scale* of formal domestic financial intermediation, compressing the size of the domestic financial system, tend to limit the size of scale economies as well as channel funds through the informal sector, which may be less efficient due to smaller economies of scale and scope. Both these measures can therefore be expected to increase the costs of operating the domestic financial system.

c. Effects through Returns to Accumulation

Finally, with a less efficient allocation of funds and higher costs of operation, the formal financial system would be forced to offer a lower return to savers under financial repression, even if deposit interest rate ceilings were not binding. In terms of the growth model, this may show up as a reduction in s (if savers respond to the lower rates of return available to them by saving less).

Recall that these channels of influence are not independent. In the basic model described earlier, for example, the adverse consequences of costs of financial intermediation for the growth rate depend on the volume of resources to be intermediated (s) as well as on the productivity of the investment foregone (A).

2. Financial Repression and Growth: Evidence

There is a substantial body of evidence examining the growth effects of financial repression. One strand of this evidence takes the cross-country approach that has become familiar from previous applications in this book. The methodology essentially involves estimating the determinants of capacity growth using a cross-country sample and testing for the independent effects of financial repression after controlling for other growth determinants. For example, Roubini and Sala-i-Martin (1992) use a real interest rate dummy, the reserve ratio, and inflation as indicators of financial repression. They control for the standard Barro-type growth determinants in cross-country regressions and test to see whether their proxies for financial repression enter the equation with the expected sign and with precisely estimated coefficients. They find that all their indicators of financial repression are negatively and significantly related to growth in a cross-country study.

There is also substantial case study evidence on this issue, much of it focusing on the experience of countries that have liberalized previously repressed financial systems. Unfortunately, because this research of necessity adopts a before-and-after methodology, failing to control for other determinants, it has proven to be controversial and inconclusive.

IV. FINANCIAL REPRESSION IN THE MIRACLE ASIAN ECONOMIES

Theory and cross-country evidence thus lead us to the conclusion that financial repression is harmful for economic growth. And indeed, many countries that have had heavily repressed financial systems have also tended to grow slowly. However, matters are not quite so straightforward because financial repression has not just been evident in countries that have been unsuccessful in sustaining extended periods of rapid economic growth. Some of the most successful economies in the world – the miracle economies of East and Southeast Asia – also pursued financial-sector policies during their high-growth periods that had many of the characteristics of financial repression. A handful of dynamic economies in this region have achieved what many other developing economies have sought: rapid, equitable, and reasonably stable growth sustained for long periods of time. Before leaving the topic of financial repression, then, it is worth asking the question, if financial repression is so harmful for growth, and if these countries indeed repressed their financial systems during their high-growth periods, how did they manage to grow so fast?

Understanding the contributions of policy to generating the East Asia miracle has long been a high priority for development economists. The interpretation of the experience of these countries, however, has been fraught with controversy. Among the unsettled issues, we know that these countries attained remarkable levels of saving and investment but do not know exactly how, and economists disagree about the contribution that gains in total factor productivity made to East Asian growth performance. The role of the financial system is central to both phenomena because, as we have seen, it both provides the incentives for private agents to accumulate resources and allocates those resources among competing investment opportunities. Here, too, however, lessons from the experience of East Asia have not been easy to distill. Specifically, to what extent can East Asia's accumulation and productivity performances be related to the financial policies pursued by these countries? Does the experience of the East Asian countries confirm the virtues or the vices of financial repression?[4]

1. Financial Policies in the Asian Miracle Economies

Japan was, of course, the first spectacularly successful economy in East Asia, followed by the four tigers of Korea, Taiwan, and the two city-states of Hong Kong and Singapore. Because of Japan's leadership position, it is useful to briefly review the role of the financial sector during Japan's economic takeoff before considering its common elements with the financial policies of other countries in the region.

The Japanese financial system underwent postwar restructuring in 1950–1955. The goal of the restructuring was essentially to organize the system to support the government's industrial policies. Parts of the system were oriented to providing long-term finance to encourage firms' taking advantage of dynamic economies of scale and to provide industrial infrastructure, whereas other parts were focused on combating unemployment by lending to traditional and low-productivity sectors (such as declining industries) and to support small firms, all while safeguarding the stability of the financial system itself. An important part of the new system was policy-based lending through the Trust Bureau of the Ministry of Finance, which managed the funds raised by the postal savings system. The latter had over 20,000 branches throughout the country. The growing modern sector of the economy was financed by the private financial system, whereas the declining traditional system was financed through the government at regulated rates.

[4] The classic (albeit controversial) reference on the East Asian growth experience is World Bank (1993).

Japan's private financial system was dominated by banks. Limitations were placed on the development of securities issues and secondary securities markets. Bond issuance was limited to government enterprises, public utilities, and a few other selected corporations, and bonds were sold through negotiation rather than competitive bidding. The domestic securities market was thus underdeveloped, with a limited role for direct lending. Restrictions on foreign capital inflows and outflows were in place throughout Japan's rapid growth period.

Markets for banking services in Japan were legally segmented (both functionally and geographically) because financial institutions were specialized by type and size of borrower as well as depositor. A small number of large wholesale banks were nationwide in scope and dealt with governments as well as large corporations. None of these banks were controlled by private shareholders. Smaller banks served local and regional markets. The system contained both private and government-owned banks, and their number remained roughly stable for the subsequent 40 years. Specialized institutions grew up or were created to serve small family businesses and agriculture, while governments created specialized institutions to serve exporters and priority sectors. Nationwide banks were limited as to the number of branches they could open, so markets for banking services were segmented geographically as well.

Commercial banks have also dominated the financial systems of Korea and Taiwan.[5] Bond and equity markets were thus underdeveloped in all three of these miracle economies. As in Japan, these markets tended to follow, rather than lead, the economic takeoff in Korea and Taiwan. Also as in Japan, during most of the period of rapid economic growth in each of these countries, the number of major banks was small, entry into the banking sector was restricted, and the capital account of the balance of payments was relatively closed. In all three countries, the ownership of banks was widely dispersed among industrial enterprises, other domestic institutions, and individuals. In Japan, they were controlled by managers, and in Korea, by the government. In Taiwan, until 1991, the major banks were owned and controlled by the government, while four smaller commercial banks were owned by private individuals.

Interest rates were controlled in all these countries, with large margins between borrowing and lending rates. Though financial repression in the form of interest rate ceilings was not as severe in these miracle economies as in many other developing countries, it was severe enough that all three countries at one time had (and Korea and Taiwan had until relatively recently) flourishing informal markets. In Japan, the formal system displaced a previously existing informal system, whereas in Korea and Taiwan, the informal system survived in parallel with the formal system, as a consequence of financial repression.

[5] For a description of the financial systems in these economies, see Park and Patrick (1994).

Directed credit was an important component of government policy in these economies. All three countries used the financial system to encourage exports and to promote specific sectors in the context of their industrial policies. However, the application of directed credit policies varied among these countries. In Korea, for example, credit policy involved much heavier subsidization than in Japan. Commercial banks had been nationalized in Korea during the early 1960s, and the central-bank law was changed to make it explicitly subordinate to the government. Both commercial and development banks were owned by the government and were involved in the administration of directed credit, responding to government directives to channel funds to priority sectors. Though controlled interest rates were doubled in 1965, this may have served primarily to transfer funds from informal markets to the banks and thus to increase the Korean government's control over credit allocation. In the 1960s, these credit allocations were directed at promoting exports, without much sectoral bias, and the performance of supported firms was closely monitored. In the 1970s, however, policy-based lending switched to supporting the "heavy and chemical industries" industrialization drive and, in the 1980s, to supporting declining industries, as in Japan. Policy loans amounted to about 60 percent of assets for deposit money banks throughout the period and mainly went to the manufacturing sector. Unlike Japan, Korea depended heavily on central-bank credit and bank deposits for mobilizing funds. In Japan, government sources of funds dominated, and foreign loans were also very important, but the latter were also allocated by the government.

As in Japan, banks were closely supervised in Korea and Taiwan. However, bank regulation tended to be concerned primarily with the safety of the system. Stability was achieved through limitations on competition, and cartel-like behavior among banks was not discouraged. Prudential supervision was inadequate by the standards of fully liberalized systems. Because public disclosure was required only of the small number of firms listed on stock exchanges, and because the reliability of financial data available even to insiders of private, unlisted companies was questionable, banks typically required loans to be heavily collateralized by real assets, and problems of assessing creditworthiness led to the prevalence of "relationship" banking in all three countries. Though legal deposit insurance was not present during the high-growth period in any of these economies, the banking system was perceived as implicitly insured by the government. In short, all three countries had very safe but not very competitive systems. In spite of this, portfolio allocation was traditionally very risk averse and, especially in the case of Taiwan, overly concerned with collateral. Bank failures were minimal until the second half of the 1990s.

2. Financial Repression and Growth in East Asia: Hypotheses

Financial repression and directed credit thus appear to have been an important part of the story in three of the most successful East Asian economies. These characteristics do not seem to suggest that the financial systems in these countries would have been a key force driving either high saving rates (by offering attractive returns) or high rates of growth of productivity (by allocating resources to dynamic sectors on the basis of market criteria). What does the evidence tell us about the role of financial-sector policies in the growth experience of these countries? In particular, did the miracle economies of Asia grow as fast as they did *despite* or *because of* financial repression?

The argument that they did so *despite* financial repression would go as follows: though financial repression is generally harmful and was actually harmful in these economies, it was not *very* harmful. The countries did not actually have the institutional mechanisms in place to operate a liberalized financial system (what these are is discussed in Chapter 22), and given that constraint, because of the way that repression was managed in these countries, they might actually have been better off with repressed systems than with liberalized ones. The aspects of the way that financial repression was managed that reduced its costs were the following:

- Financial repression was mild because the stable macroeconomic environment (low inflation) and competitive real exchange rate kept the gap between the controlled and market-clearing interest rates small, so savers did not face excessively distorted real interest rates. The World Bank (1993), for example, emphasized that financial repression in the miracle economies differed from that elsewhere in that it was moderate, that is, without very negative real interest rates; it was undertaken in context of financial stability rather than as the unintended consequence of rapid inflation; and bank regulators tended to squeeze spreads, ensuring that low deposit rates were passed on to borrowers.
- The system emphasized security of financial institutions, which may have been as important to savers as rates of return.
- A relatively small share of credit was influenced by policy-based lending (except for Korea). Moreover, directed credit policies discouraged lending for consumption, which encouraged saving, and encouraged lending based on market-friendly performance criteria (exports), which enhanced productivity. Systems in Japan, Korea, and Taiwan channeled resources into infrastructure and productive business uses and away from housing and consumer credit. By restricting credit to households and small firms due to the extreme risk aversion of managers of financial institutions, it may have indirectly stimulated saving by such agents. In addition, by allocating credit on the basis of export performance (in response to government directives), the regime was indirectly using a market-based test that may have particularly favored the most competitive and dynamic activities.

The argument that these countries grew fast *because of* financial repression is really a variation on the previous one that places much more emphasis on the positive contribution that may have been made to economic growth by the specific industrial policies that these countries implemented through the use of directed credit. It has two parts:

1. Liberalization would have been harmful, given the domestic institutional constraints, as mentioned above.
2. Industrial policy was growth promoting because of the role of dynamic scale economies, and policy-based lending was a superior way of implementing it.

The second point is what really distinguishes the second interpretation from the first. Because the countries grew so quickly, the question is whether the way that directed credit was managed in these economies minimized the harm it might conceivably have done, or whether it was actually a key ingredient of these economies' success.

This point has two components: that industrial policies were productive and that directed credit was a good way to implement them. As to the first, the case for industrial policies can be (and has been) based on a variety of arguments. For example, one view holds that if complementary industrial activities exhibit increasing returns, government industrial policy can play an important coordinating role. Alternatively, some observers of the East Asian miracle have argued for the presence of positive externalities in some activities such as exporting. Such externalities would justify a government role in the encouragement of exports.

But why would directed credit be a good way to implement such industrial policies? Among the arguments that have been offered are that because of asymmetric information, banks may tend to offer funds to firms that have collateral, available internal funds, or good track records rather than to those that have the best investment opportunities. This may create a role for the support of credit to small- and medium-sized firms. Alternatively, subsidized loans may provide the firms that benefit from them with greater incentives to perform than would outright subsidies, or loans may provide superior monitoring benefits than subsidies. Finally, the implicit financial repression tax used to fund such loans may be less distortionary than the relevant explicit taxes required to finance outright fiscal subsidies would have been.

3. Financial Repression and Growth in East Asia: Evidence

What light does the evidence shed on these interpretations of the role of financial repression in the growth experience of the Asian miracle economies? We can break the evidence down into several components.

a. Financial Deepening and Growth

First, consistent with the cross-country evidence discussed in Chapter 20, financial deepening actually accompanied real growth in Japan, Korea, and Taiwan (i.e., the ratio of financial assets, such as broad money, to *GDP* increased as incomes rose). Thus it proved possible to achieve high rates of expansion of the financial sector without liberalization of the financial system. This certainly is consistent with the view that the initial distortions were not excessively harmful and that these economies were successful despite financial repression simply because they did not overdo it.

The key question, however, is whether policy helped to ease the severity of financial repression, which fostered financial deepening and contributed to growth, or whether financial deepening was simply driven by fast growth. This is the familiar direction-of-causality problem that we have encountered before. There is some indication, however, that financial deepening was not purely the result of rapid growth but also responded to policies. This evidence is in the form of episodes in which the severity of financial repression varied due to events that could arguably be claimed to be exogenous. For example, the Korean ratio of financial assets to *GDP* was flat for a long period (1972–1978) when inflation was rapid and real interest rates were negative, despite the rapid growth in *GDP* at this time. Once policy changed to bring inflation under control and liberalize the financial sector somewhat, the ratio resumed its increase. Japan had a similar experience in 1965–1970, but Taiwan, which did not allow high inflation during its high-growth period (which began in the late 1950s), and had consistently higher real interest rates, did not. These episodes are at least suggestive of the possibility that changes in financial depth may been affected by factors that respond to policy in the short run – such as the level of domestic real interest rates – rather than simply responding to economic growth.

b. Financial Policy and Saving Rates

Second, as we have seen, private saving rates were very high in these countries, and they actually tended to increase over time. The difficulty is linking these high saving rates with policies directed to the financial sector. For example, there is no evidence that the increase necessarily had anything to do with policy-induced changes in the behavior of real interest rates in these economies. Indeed, the high saving rates in the region may have been due to a wide variety of factors. Prime candidates are rapid growth itself as well as demographic transitions that took place in the East and Southeast Asian regions.[6] They may also have been directly fostered by other policies. Governments encouraged saving through the pursuit of macroeconomic stability, through the way that banks were regulated, through restrictions on consumer credit, and through forced saving schemes.

Macroeconomic stability may have encouraged saving because, as we saw in Chapter 3, high inflation tends to be volatile inflation, making real interest rates

[6] Rapid growth can increase saving either through standard life cycle arguments that emphasize the higher share of income received by young savers in a growing economy or through habit effects (Carroll and Weil 1994).

often negative and unpredictable. This increases the risk associated with financial saving. Moreover, by protecting banks from competition, the miracle economies may have made them safer, encouraging saving in the formal financial sector. These countries also made saving through formal financial institutions more convenient to small savers. Postal saving systems not only in Japan, Korea, and Taiwan but also in Malaysia and Singapore were designed to attract small savers by offering greater security and lower transaction costs than the informal sector. Restrictions on consumer credit in the context of financial repression, coupled with taxes on luxury goods, may also have encouraged saving. Mandatory saving schemes undoubtedly played a role as well. In some countries (Japan, Singapore, and Malaysia), the implementation of such provident funds may have made a particularly large contribution to saving rates (however, whether these encourage total saving depends on the extent to which they replace saving that would have happened anyway). Finally, these economies tended to have high public saving rates compared to other developing countries, which increased national saving rates directly.

A different perspective on saving behavior in these economies is offered by Singh (1995), who argues that high saving rates in the miracle economies of Asia did not reflect particularly high levels of *household* saving but rather of *corporate* saving. These were the results of large corporate profits plus inducements and incentives for firms to retain earnings rather than passing them on to their shareholders. Large corporate profits were achieved by restrictions on domestic competition (in Japan and Korea) and through import protection (both in those countries as well as elsewhere in the region). Governments provided fiscal and other incentives for firms not to distribute these earnings as dividends, for example, by taxing dividends and not taxing retained earnings or capital gains.[7]

Some panel-data econometric evidence compiled by Dayal-Gulati and Thimann (1997) attributes high saving rates among the miracle economies to macroeconomic stability, public saving, the presence of mandatory saving schemes, and financial deepening. Unfortunately, it does not identify the role of policies in promoting financial deepening. In an overview of the issue of the determinants of high saving rates in these economies, Patrick and Park (1994) conclude that we do not at present know the extent to which increases in saving rates and their mobilization through the financial system were due to growth, reduced inflation, higher real interest rates, the spread of financial institution offices, or other variables.

[7] However, this argument would need to establish that households did not "pierce the corporate veil," that is, that they did not consider the savings of the corporations they owned to be their own savings and thus reduce their own saving as corporate saving increased. Otherwise, national saving would be unaffected by whether saving is done by corporations or by households. Singh does not address this issue.

c. Financial Policies and Investment Rates

Turning to investment, private investment was much higher in the Asian miracle economies than in other middle-income economies, though public investment rates have not been very different.[8] We have already seen that directed credit policies tended to favor investment rather than consumption. But does that mean that high private investment rates were driven by the supply of funds – and thus by policy-directed lending – rather than by the demand for funds? The answer is that we cannot draw this conclusion too readily because many other aspects of the policy environment also favored investment demand. Investment was in part fostered by macroeconomic stability and high growth itself. But it was also induced by secure property rights, complementary public investment in infrastructure, and policies that tended to reduce the cost of capital such as tax policies that favored investment and low tariffs on imported capital goods. The relative contributions of all these factors to the high investment rates in the region remain to be sorted out.

d. Financial Policies and the Efficiency of Investment

A separate issue concerns the efficiency of investment. This is a crucial issue in interpreting the roles of financial repression and directed credit because if directed credit policies allocated capital to relatively unproductive uses, we would expect this to show up in low rates of total factor productivity growth in the region.

Unfortunately, the relative contributions of capital accumulation and total factor productivity growth to the exceptionally high growth rates of the miracle economies have not proved easy to sort out. A well-known study of the miracle economies by the World Bank (1993) concluded that accumulation accounted for about two-thirds of growth in incomes per capita among these countries on average and growth in total factor productivity for the remaining third. The latter was still higher than in most other economies, however, both absolutely and as a share of output growth.

Two methodologies were used to address this issue in the World Bank study. First, cross-country regressions were estimated for 113 countries, with per capita growth estimated as a function of the ratio of the country's per capita income in 1960 to that of the United States, the investment-*GDP* ratio, the rate of primary and secondary school enrollments in 1960, and the rate of growth of the economically active population. This model accounted for about two-thirds of the actual growth in individual miracle economies, on average. Primary school enrollment accounted for most of the growth, with physical investment second and secondary enrollment third. Still, most regional differences in growth remained unexplained (the model yielded a positive regional dummy for the Asian miracle economies and negative

[8] However, public investment rates actually *rose* in these countries during the 1980s, in contrast to the experience of most other emerging and developing economies.

ones for Latin America and Sub-Saharan Africa). The World Bank concluded from the latter that the miracle economies must have been more successful in allocating resources to high-productivity activities because the growth residuals would have tended to capture this phenomenon.

The World Bank study also calculated total factor productivity (TFP) growth using a production function estimated for a sample of 87 countries from 1960 to 1989. The East Asian economies were found to have had high absolute levels of TFP growth. Hong Kong, Japan, Korea, Thailand, and Taiwan were in the top decile of countries in this regard. Indonesia, Malaysia, and Singapore were closer to the TFP growth rates of the high-income countries (about 1.5% per year) but were in the top third of all developing countries.

To see whether the miracle economies were catching up to international technological best practice, the study estimated growth residuals for each country using the estimated output elasticities of capital and labor for industrial countries as well as the average rate of TFP growth for such countries (1.5%) over 1960–1989 (the latter was taken to be a measure of the international rate of technological advance). The residual TFP growth for all the countries in its sample was then taken as a measure of the extent to which developing countries have been catching up to international technological best practice.

It found that TFP growth was in a fairly compact range around 1.5 percent per year for the rich countries, presumably reflecting the common international rate of technological change, so that residuals for these countries were close to zero. It also found that among the miracle economies, Japan, Hong Kong, Taiwan, and Thailand were catching up in TFP levels with the industrial countries, in the sense that all these countries had positive residuals (see also Page 1994). Korea was just keeping up with rich countries in its rate of TFP growth (i.e., it had a residual of approximately zero), while the investment-driven economies of Malaysia, Indonesia, and Singapore were falling behind. Still, even these countries were doing better than developing countries in other regions. For these economies, TFP growth provided a very small share of total output growth (TFP growth contributed less than 33%), while the others tended to look more like industrial countries (TFP growth contributed 30%–50% of total output growth).

The main conclusion from the World Bank study, as well as a subsequent study along the same lines by Page (1994), is that, though the growth story among the miracle economies is primarily one of accumulation, these economies were also good at allocation. On the basis of average world TFP growth, accumulation underpredicted growth for the miracle economies as a group. That these countries had higher rates of productivity growth than 70 percent of all countries suggests that not only were they accumulating but they were also using resources efficiently. How did they do so? The World Bank study argues that good fundamentals and limited price distortions were important. It concludes that the industrial policies adopted in many of these countries made no difference but that the export push adopted in all of them did, interacting with high levels of human capital (which facilitated acquisition of technological know-how) in a virtuous circle.

Collins and Bosworth (1996) also addressed the TFP growth controversy in the miracle economies. They based their interpretation on a study of 88 countries over the period 1960–1994. Using growth-accounting exercises, Collins and Bosworth (1996) found that growth in the miracle economies was mainly due to accumulation. They estimated TFP growth for the region over the whole period at about 1.1 percent, about the same as for industrial countries, excluding the United States, and only slightly higher than South Asia. Taiwan had particularly strong TFP growth, whereas TFP growth in the Philippines was particularly poor. By contrast, however, TFP growth was estimated to be negative in Africa and the Middle East and zero in Latin America.

These results were supplemented by cross-country regressions. Growth in GDP per worker was regressed on initial GDP relative to that in the United States, life expectancy in 1960, years of schooling in 1965, the change in the terms of trade, and the standard deviation in the terms of trade. In this regression, regional dummies (using East Asia as reference) proved to be statistically significant for all developing-country regions but not for industrial countries. When the dependent variable was expressed in terms of growth of capital per worker, regional dummies again proved to be significant, but not when the dependent variable was TFP growth. Collins and Bosworth concluded that East Asia was mainly an outlier with respect to accumulation, not with respect to TFP growth.

To see the channels through which policy may have affected growth, Collins and Bosworth (1996) added the budget balance relative to GDP, the standard deviation of the real exchange rate, and the Sachs-Warner measure of trade openness to this regression. They found that fiscal surpluses promoted growth through capital accumulation and that real exchange rate stability did so through productivity growth. Surprisingly, the Sachs-Warner trade measure, though strongly correlated with growth, was correlated with capital accumulation, not productivity growth. Inclusion of the policy variables tended to reduce the regional dummies by about one-third, so these policy differences explained about one-third of the otherwise unexplained growth gap between East Asia and the other developing-country regions.

Rodrik (1997) showed that estimates of the contribution to growth of total factor productivity would tend to be biased *downward* when the elasticity of substitution in production is less than unity, as long as technical change is biased toward saving (i.e., augmenting) the factor that is becoming relatively scarce (e.g., labor, when capital deepening is occurring), as seems plausible. Moreover, the bias is proportional to the extent of capital deepening, so it would be particularly severe for high-investing economies such as those in East Asia. In interpreting the growth experience of these countries, he also pointed out that an index of bureaucratic quality as well as the ratio of years of schooling to initial GDP per person tend to enter cross-country growth regressions significantly and that East Asia is superior to all other regions with regard to the latter. This implies a high initial ratio of skilled labor to the capital stock in these countries and thus a high return to capital that would account for the high subsequent rates of capital accumulation in the region.

Finally, Sarel (1997) expressed doubts about the contrarian view expressed by some economists that growth in East Asia has been almost entirely accumulation-driven because the results of such measurements depend critically on capital stock, factor share, and labor aggregation estimates that are questionable. His own estimates, using perpetual inventory methods for estimating the capital stock starting in 1900, produced high estimates of capital stock growth and labor force participation for the East Asian countries but also estimates of TFP growth that were very high for Korea, Taiwan, and Hong Kong (indeed, higher than in Japan) and not as high, but still respectable, for Singapore. He found that the proportion of growth explained by TFP growth was not different in the four tigers from that in the United States and Japan.

Overall, these studies seem to yield one important conclusion: whether the growth of total factor productivity was exceptional in the Asian miracle economies, there is little evidence to support the view that it was abysmally poor. Thus directed credit appears not to have created very severe distortions in the allocation of capital. The question is why not.

e. Management of Directed Credit Policies

Vittras and Cho (1995) ask precisely this question; that is, they investigate why directed credit policies were apparently successful in East Asia, though they have not been so elsewhere. They cite the roles of both economic and institutional factors in the region as contributing to this outcome. The economic factors include the pursuit of macroeconomic stability as well as the general development strategy adopted by these countries. This strategy featured export orientation, the encouragement of domestic competition, the reliance on the private sector, and the presence of a bias toward industrialization. Institutional factors include the creation of an effective monitoring system by the banks, the use of extensive consultation arrangements, and the development and propagation of credible development visions for these economies. They argue that the combination of macroeconomic stability with intense domestic competition, export orientation, and a reliance on the private sector promoted efficiency and provided objective criteria for monitoring on the part of banks. Moreover, the goals of policy-based lending were narrowly focused (on industrialization and export promotion) and were well coordinated with other policies.

Vittras and Cho (1995) draw lessons of two types from the experience of these countries with directed credit policies. What they call "good vision" lessons are that credit policies should have a small size and focus, should be of limited duration with clear sunset provisions, and should involve a low level of subsidy (to minimize distortions). In addition, they should aim at activities that generate positive externalities, such as those associated with industrialization and exports, and they should be based on a competitive private sector with internationally competitive operations. Finally, they should form part of a broader credible vision of economic development promoting growth with equity and involving a long-term strategy to develop a sound financial system operating on economic criteria. They also draw "good management" lessons. These include the principle that policy-based loans should be channeled through well-capitalized, administratively capable, and autonomous institutions; should be based on clear, objective, and easily identifiable criteria; should aim at good repayment records and low loan losses; and should be supported by effective mechanisms for consultation and communication between public and private sectors.

4. Financial Repression and Growth in East Asia: Lessons

What is clear from all this is that the management of financial repression in the miracle economies was special. Financial repression was managed in a way that, at the very least, prevented financial repression from posing a serious obstacle to growth and may even have permitted it to support growth in these countries. The key points were that the macroeconomic environment was stable, controlled nominal interest rates were not kept unduly low, and policy-based lending was used to allocate funds on the basis of reasonable performance indicators. This last was determined in part by the particular development strategy that these countries had adopted. It is unclear whether the net effect of all these components was positive or simply not very negative. The answer depends in part on whether the institutional mechanisms could have been put in place to make an alternative set of financial-sector policies feasible and on the view one takes about the contribution of industrial policies to the economic success of these countries. What is clear is that the relatively benign role of financial repression in these countries reflected a special set of circumstances that would be difficult to replicate elsewhere.

V. SUMMARY

We saw in the previous chapter that a well-functioning financial system could make potentially important contributions to economic welfare and growth. The emergence of such a system, however, can be influenced in important ways by government policy. These influences can be beneficial or harmful. Because of the many imperfections from which credit markets are likely to suffer in the real world, it is unlikely that the optimal set of policies toward the financial system would be one characterized by a laissez-faire stance on the part of the government. But this does not provide carte blanche for government intervention. A well-designed set of policies to promote the efficient functioning of the domestic financial system must be designed to target credit-market imperfections narrowly and efficiently. When policies toward the domestic financial system are designed with other goals in mind, the net effect of such policies may turn out to be harmful.

This chapter has explored this issue in the context of financial repression, a set of policies toward the financial system that arose out of fiscal motivations and that imposed severe restrictions on the behavior of the customers of domestic banks, their potential competitors, and the banks themselves. These policies were widespread among emerging and developing economies until relatively recent times and are still present in some developing countries. Theory suggests that policies of this type would tend to have harmful effects on economic welfare and growth, and we have seen that the evidence is consistent with this conclusion. Though under very favorable circumstances – such as those that prevailed in the Asian miracle economies – these effects may be mitigated, the evidence on this is unclear, and the relevant circumstances may in any case themselves be difficult to create.

The question, then, is how to make the transition from a repressed financial system to a well-functioning liberalized one, that is, how to conduct an appropriate financial reform. This involves identifying both what the residual role of the government should be in the liberalized financial system that ultimately emerges from the reform process and determining how to get from here to there, that is, the sequence of steps required to make the transition. The next chapter turns to these questions.

REVIEW QUESTIONS

1. What is financial repression? Describe the components of this set of policies toward the financial sector and how they fit together.
2. Explain how you would expect each of the components of financial repression described in this chapter to affect economic growth.
3. What are the mechanisms through which financial repression creates fiscal benefits for the government?
4. How was it possible for the East and Southeast Asian miracle economies to achieve sustained high growth rates despite their use of financial repression?
5. Explain the relevance of the controversy over the role of TFP growth in the Asian miracle economies to the issue of whether financial repression helped or hindered rapid growth rates in those economies.

EXERCISES

1. Explain each of the components of the financial repression tax as calculated by Chamley and Honohan (1990).
2. Using the fiscal solvency analysis of Chapter 9, evaluate whether a financial repression tax that increases the ratio of the adjusted primary surplus plus

seignorage to *GDP*, while at the same time reducing the economy's growth rate, would or would not ease the government's intertemporal budget constraint.

3. Use the growth model of Chapter 20 to identify the channels through which you might expect financial repression to have an adverse effect on a country's rate of economic growth.
4. Most macroeconomists believe that low interest rates stimulate investment. If that is so, why might we expect financial repression, which tends to be associated with binding interest rate ceilings, to be associated with *less* investment?
5. In evaluating the effects of financial repression on saving rates, how useful would it be to know that an increase in controlled deposit rates is associated with an increase in deposits in the domestic financial system? Explain.

REFERENCES AND FURTHER READING

Carroll, Christopher D., and David N. Weil (1994), "Saving and Growth: A Reinterpretation," *Carnegie-Rochester Conference Series on Public Policy*, vol. 40, pp. 133–192.

Chamley, Christophe, and Patrick Honohan (1990), "Taxation of Financial Intermediation: Measurement Principles and Applications to Five African Countries," working paper WPS 421, World Bank.

Collins, Susan, and Barry Bosworth (1996), "Economic Growth in East Asia: Accumulation versus Assimilation," *Brookings Papers on Economic Activity*, vol. 2, pp. 135–203.

Dayal-Gulati, Anuradha, and Christian Thimann (1997), "Saving in Southeast Asia and Latin America Compared: Searching for Policy Lessons," in J. Hicklin, D. Robinson, and A. Singh, eds., *Macroeconomic Issues Facing ASEAN Countries* (Washington, D.C.: International Monetary Fund, pp. 130-152.

Fry, Maxwell J. (1988), *Money, Interest, and Banking in Economic Development*, Baltimore: Johns Hopkins University Press.

Giovannini, Alberto, and Martha de Melo (1993), "Government Revenue from Financial Repression," *American Economic Review*, vol. 83, pp. 953–963.

Ikhide, S. I. (1992), "Financial Deepening, Credit Availability, and the Efficiency of Investment: Evidence from Selected African Countries," development research paper 2, UN Economic Commission for Africa.

McKinnon, Ronald I. (1973), *Money and Capital in Economic Development*, Washington, DC: The Brookings Institution.

Page, John (1994), "The East Asian Miracle: Four Lessons for Development Policy," in S. Fischer and J. Rotenberg, eds., *NBER Macroeconomics Annual 1994*, Cambridge, MA: MIT Press, pp. 219–268.

Patrick, Hugh T. (1994), "Comparisons, Contrasts, and Implications," in Yung Chul Park and Hugh T. Patrick, eds., *The Financial Development of Japan, Korea, and Taiwan: Growth, Repression, and Liberalization* (New York: Oxford University Press), pp. 325–372.

Patrick, Hugh T., and Yung Chul Park (1994), *The Financial Development of Japan, Korea, and Taiwan: Growth, Repression, and Liberalization*, Oxford: Oxford University Press.

Rodrik, Dani (1997), "TFPG Controversies, Institutions, and Economic Performance in East Asia," working paper 5914, National Bureau of Economic Research.

Roubini, Nouriel, and Xavier Sala-i-Martin (1992), "Financial Development and Economic Growth," *Journal of Development Economics*, vol. 39, pp. 5–30.

Sarel, Michael (1997), "Growth in East Asia: What We Can and Cannot Infer," *Economic Issues*, vol. 1, pp. 1–22.

Shaw, Edward S. (1973), *Financial Deepening in Economic Development*, New York: Oxford University Press.

Singh, Ajit (1995), "How Did East Asia Grow So Fast? Slow Progress toward an Analytical Consensus," working paper 97, United Nations Conference on Trade and Development.

Vittras, Dimitri and Yoon Je Cho (1995), "Credit Policies: Lessons from East Asia," policy research working paper 1458, World Bank.

World Bank (1993), *The East Asian Miracle*, Oxford: Oxford University Press.

22

Financial Reform

A natural response to the recognition of the adverse consequences that financial repression can have for growth is to attempt to undo these harmful policies. Because these policies take the form of restrictions on financial intermediation, their removal is referred to as *financial liberalization.* Many emerging and developing economies have indeed embarked on such a path over the past two decades. But if these restrictions are removed, the question arises of what – if anything – to replace them with. Should the government simply adopt a hands-off approach to the financial sector, or are there specific government policies that may enhance, rather than inhibit, the efficiency of financial intermediation?

It turns out that simply removing the harmful policies associated with financial repression overnight and adopting a laissez-faire approach toward the financial system may do more harm than good. In particular, it may generate new kinds of resource misallocations and may result in severe financial crises with important macroeconomic implications. As we saw in Chapter 20, credit-market imperfections create a role for public policy in a well-functioning financial system. The problem with financial repression is not so much the *fact* of public-sector involvement in regulating the financial sector as it is the *type* of public-sector involvement. Because what is required is a redefinition of the government's role, rather than the removal of any role for the government, a better term for the process of moving from a repressed financial system to a well-functioning one is *financial reform* rather than liberalization.

Two questions are posed by the process of financial reform. The first is what the appropriate role of public policy should be in a fully reformed system. In other words, if financial repression involves the wrong set of interventions, then what is the right set? The second question is how to progress from a repressed to a fully reformed system – i.e., how to conduct the *sequencing* of financial reform.

Sequencing may matter for several reasons. First, recall from the two previous chapters that financial repression has a fiscal dimension, and that the damage done by financial repression may depend on the macroeconomic circumstances in which the financial system operates. Thus, in formulating a strategy for reform, financial reformers will need to consider the interactions between macroeconomic conditions and the reformulation of financial-sector policies. Second, if a liberalized financial system requires an alternative set of public policies to support it, then the removal of repressive restrictions may be conditional on the implementation of these alternative policies. Finally, financial repression has many components – both domestic and external ones – and it may make a difference in what order these are removed. This chapter will take up each of these issues.

We will begin by considering some of the problems with a laissez-faire approach to domestic financial reform, specifically, the fact that an unregulated banking sector may be crisis prone. Section II examines potential responses of public policy to the problem of bank vulnerability. The external dimension of financial reform – liberalization of the capital account of the balance of payments – is the subject of Section III. Section IV then puts together the domestic and external dimensions of reform to consider the appropriate sequencing of reform measures. Finally, Section V provides some country case studies of domestic and external financial reform.

I. VULNERABILITY OF A LIBERALIZED BANKING SECTOR

Financial liberalization means the removal of the restrictions associated with financial repression. Accordingly, it takes the form of freeing interest rates to be determined by market forces, lowering or abolishing reserve and liquidity requirements, removing directed credit regulations, adopting *indirect* instruments of monetary control (central-bank intervention in securities markets or extension of credit to banks, instead of ceilings on the credit that can be extended by individual banks), opening up entry into the financial sector (including privatization), and opening up the capital account of the balance of payments. A laissez-faire approach to financial reform would have the government undertake all of these measures and then rely on competition and market efficiency to ensure the appropriate functioning of the deregulated financial system.

The fundamental problem with this approach, however, is that, as we saw in Chapter 20, the financial system essentially represents a market response to the dual problems of asymmetric information and opportunistic behavior. Under these circumstances, there is no presumption that unregulated competitive markets will be Pareto efficient. For example, information is *nonrivalrous* (its use by one agent does not prevent use by another) and *nonexcludable* (agents that have not paid for it cannot be prevented from acquiring it). Thus it has the characteristics of a *public good*, which is known to be undersupplied in competitive markets. Moreover, imperfect information also makes it less likely that financial markets will be perfectly competitive. Because the acquisition of information involves incurring fixed costs, the resulting scale economies make this likely to lead to noncompetitive outcomes.

From a purely microeconomic perspective, then, it is by no means obvious that a laissez-faire approach will result in an efficient and well-functioning financial system.

But as we shall now see, a fully liberalized system would also be vulnerable to other, potentially more serious types of problems, in the form of systematic misallocation of resources as well as vulnerability to systemic breakdowns in which many financial firms go out of business, severely disrupting financial intermediation in the economy. It is useful to classify these systemic breakdowns into two categories: *banking crises*, which are *solvency* problems (typically defined as a situation of generalized insolvency in the banking industry), and *bank panics*, which are *liquidity* problems (situations in which there is wholesale withdrawal of deposits from otherwise solvent banks). As we shall see, depending on the country's institutional and policy environment, a banking crisis can be accompanied by a bank panic, but it need not be, and a bank panic may or may not trigger a banking crisis.[1]

1. Banking Crises

Just as is true for any other borrower, moral hazard problems characterize the activities of financial intermediaries such as banks. Moral hazard problems in banking are caused by the principal-agent relationship between bank depositors and bank managers, particularly in the case of banks that have low net worth. Acting on behalf of the owners of banks, bank managers have an incentive to offer high deposit rates to attract large quantities of funds and then invest these in risky ventures because if such ventures are successful, bank owners get to keep the large gains that remain after paying off the bank's depositors, whereas if they are not, the owners of the bank lose very little of their own money. This is simply another example of the standard agency problems associated with borrowing and lending that we examined in Chapter 20.

As we saw there, one solution to this problem is for the lender to monitor the borrower and impose penalties for behavior that is not in the interest of the lender. Because the losers when risky ventures fail to pay off are the depositors, depositors have an incentive to monitor the conduct of bank managers to prevent such activities. They can penalize the managers by withdrawing their funds from the bank. Indeed, the liquidity of bank liabilities can be interpreted as a market-based response to moral hazard problems in banking. Liquid liabilities serve as a device through which depositors who learn of inappropriate behavior by bank managers acquire a means to penalize such activity at low cost to themselves (by withdrawing their funds) and thus are able to discourage it. This permits monitoring by depositors to discipline bank managers and, by reducing the scope of the moral hazard problem, may actually allow banks to borrow more cheaply.

[1] E.g., it would not be if deposits are explicitly or implicitly – and credibly – insured.

But the question is how effective this disciplining device is likely to be in practice. The incentives of depositors to monitor banks are weakened by several factors. First, the private, nontraded information generated by the banks in their loan activities makes the quality of their portfolios difficult (i.e., costly) for depositors to monitor; that is, using the language of Chapter 20, bank activities tend to be inherently opaque. Second, because the monitoring of bank activities by some depositors is enough to discipline the banks, and because monitoring is a costly activity, *all* depositors benefit if *some* depositors monitor. This means that monitoring would tend to be discouraged by a *free rider* problem in the sense that all depositors have an incentive to leave the monitoring to others. Finally, depositors may come to believe that in the event of a bank failure, the government would step in to bail them out by guaranteeing the value of their deposits. The government has an incentive to do this because, as we shall see later in this chapter, bank failures may tend to spread and become systemic, which creates the potential for substantial economic harm. While the government may publicly deny its willingness to do this, time-inconsistency problems such as those we studied in the context of central-bank behavior in Chapter 14 may make it impossible for the government to make this commitment credible. If depositors come to believe that their assets are not at risk, they will have no incentive to monitor.

The upshot is that there are ample reasons to expect that depositors, on their own, are unlikely to provide the appropriate amount of monitoring. If that is so, moral hazard will induce banks to take excessive risks (i.e., risks larger than their creditors would choose). This has important microeconomic and macroeconomic consequences.

From a microeconomic perspective, the decisions of bank managers acting as agents will tend to misallocate resources, favoring risky activities that maximize their expected private gains rather than the expected gains to society. These distortions in resource allocation harm the economy's growth performance just as do those arising from financial repression.

From a macroeconomic perspective, the prevalence of risky loans makes the values of bank portfolios excessively sensitive to the inherent volatility of the economic environment. The implication is that bank failures and generalized bank solvency crises will be more frequent than they would be if the influence of moral hazard were restrained. This is particularly the case if there are mechanisms that operate to spread bank failures from one bank to another. Bank panics provide just such a mechanism.

2. Bank Panics

Vulnerability to panic is inherent in banking. The reason has to do with the structure of bank portfolios. On the liability side of their balance sheets, by definition, banks offer liquid liabilities. As we have seen, the liquidity of their liabilities confers certain advantages to banks. In particular, it reduces the premium that they must pay for external finance by giving depositors greater leverage over the institution. On the other side of bank balance sheets, bank assets are not only of longer maturity than their liabilities (that is why banks are often said to engage in *maturity transformation*), but more important, their loans are not typically traded in secondary markets. The reason is that, as we have seen, these loans are information-intensive and so are difficult to sell to others for "lemons" reasons: asymmetric information about the quality of these loans creates an adverse selection problem in the sense that potential buyers of these loans may fear that banks trying to sell them are doing so precisely because the loans are of poor quality and thus would be unwilling to pay enough for them to compensate the banks for disposing of them. Without secondary markets for bank assets, banks cannot liquidate them quickly and at low cost.[2] Because their assets are thus much less liquid than their liabilities, banks are inherently illiquid institutions.

Why is this a potential problem? The reason is that it makes banks vulnerable to liquidity (confidence) crises. An important feature of the liquid liabilities of banks is that their depositors have access to them on a first come, first served basis (this is called the *sequential servicing constraint*). Suppose that, though information on the health of individual banks is costly for depositors to acquire, individual depositors can observe each other's actions at low cost. Then the loss of faith in the bank's soundness by any one depositor, causing him or her to withdraw funds from the bank, is likely to spill over to the others, who will tend to suspect that the depositor withdrawing funds has superior information about the bank and who will thus have an incentive to withdraw their funds as well. The incentive arises because withdrawing funds under these circumstances can yield a high return. By definition, it costs little to do so, and it may enable the depositor to avoid large losses arising from the sequential-service constraint. The latter means that in the event that the bank becomes insolvent, only depositors who remove their funds early will avoid losses. The likelihood of insolvency is increased by the fact that confidence crises can actually be self-fulfilling, in the sense that they can trigger insolvency of the bank through distress sales of the bank's assets. Thus individual depositors have to consider not just whether the actions of other depositors in withdrawing funds signal a preexisting solvency problem but also whether such actions may trigger withdrawals by other depositors that may actually *cause* a solvency problem. In this situation, banks become highly vulnerable to confidence crises.

There are various possible solutions to this problem. First, individual banks can address it directly by holding part of their portfolios in liquid, marketable assets. But this would leave less of their portfolios for banks to devote to the

[2] Banks' acquisition of nontraded rather than traded assets plays an important microeconomic role. Because it prevents others from free riding on banks' information acquisition, it allows banks to retain the profits from this activity and thus preserves their incentives to undertake it.

information-intensive loans that are their reason for existence, so this is not a socially desirable option if superior options are available that would achieve the same protection with less asset liquidity. One such option is an interbank market for reserves, which would allow individual banks to hold fewer reserves. In the United States prior to the founding of the Federal Reserve System, for example, banks issued clearinghouse loan certificates (a joint liability of all members of the clearinghouse) to their depositors in times of panic. Two problems with this approach are that asymmetric information may make this market dry up at times of crisis (banks may not be willing to lend to each other, any more than depositors are willing to lend to banks) and that it does not provide protection against a *generalized* liquidity crisis.

II. BANK VULNERABILITY AND PUBLIC POLICY

As we have just seen, a liberalized banking sector is potentially vulnerable to both banking crises and bank panics. While there are market mechanisms that can help to ameliorate these problems, these mechanisms suffer from certain shortcomings. In this section, we will examine how public policy can be used to address both types of vulnerabilities.

1. The Lender of Last Resort and Deposit Insurance

The shortcomings of market-based mechanisms for protection against bank panics could be avoided if there is an informed lender that can provide a large amount of resources to banks that are affected by a panic. This is the *lender-of-last-resort* function. The traditional prescription (articulated by Walter Bagehot in the classic *Lombard Street*) for how a lender of last resort should behave in times of crisis is that it should be prepared to lend in *generous amounts*, but at *penalty interest rates* and only on *good collateral*. The stipulation that lending should be in generous amounts serves to reassure depositors that the required amount of resources will be available, the imposition of a penalty interest rate is intended to restrict access to last-resort borrowing only to times of crisis, and the restriction requiring backing by good collateral is to make sure that lender-of-last-resort funding is used for liquidity rather than for solvency problems.

Though the lender of last resort could, in principle, be another bank, the possibility that bank panics can become generalized suggests that the role may best be performed by the central bank. Thus one role of public policy in a reformed financial system is to provide the lender-of-last-resort function.

But the lender-of-last-resort function has its own problems. From the perspective of the individual depositor, one shortcoming is that the function is exercised with discretion by the central bank, so the depositor is not sure to recover his or her assets in the event of a panic. This may leave an incentive for depositors to still "run" from the affected bank. From the perspective of policy makers, on the other hand, placing the lender-of-last-resort function with the central bank tends to complicate monetary policy. The reason, as we will see in the next chapter, is that a generalized panic will contract the money supply (by reducing the money multiplier), but it may also reduce the demand for money (because currency and demand deposits are imperfect substitutes, people may "run" into assets other than currency). This makes it difficult for the central bank to know how much liquidity it should inject into the economy at times of panic to avoid the emergence of disequilibrium in the money market.

A public-policy alternative that handles both problems is the institution of *deposit insurance*, usually in the form of a guarantee by a private or public institution that the value of deposits at participating banks are insured up to a certain amount. Banks pay a premium for insuring the value of their deposits, which is typically expressed as a percentage of the deposits covered and may or may not be related to the composition of the banks' portfolio. When depositors know that their claims are insured by such a mechanism, they have no incentive to run from a bank whose ability to meet its obligations is suspect. Because the removal of this incentive can be expected to reduce the incidence of bank panics, deposit insurance has the added benefit of making the life of monetary policy makers simpler than would be the case with only a lender of last resort in place.

But deposit insurance also has potentially harmful side effects. First, deposit insurance may aggravate the moral hazard problems in banking caused by the principal-agent relationship between bank depositors and bank managers, particularly for banks with low net worth. Once their deposits are insured, not only do depositors no longer have an incentive to run from potentially troubled banks but they also have no incentive to monitor the behavior of bank managers in the first place. The presence of insurance may thus induce bank managers to undertake riskier behavior than they would have otherwise because they will not be penalized by their depositors for doing so. Thus the presence of deposit insurance aggravates the moral hazard problem.

Second, deposit insurance may encourage adverse selection in entry into the banking industry. Risk-loving individuals would reap the largest benefit as bank owners from the opportunity to gamble with other people's money that arises from the insurance provided to bank depositors, especially when they do not have to take large stakes in their own banks. The likely result from both effects is that financial crises in which a large number of banks find themselves with negative net worth may be more likely in the presence of deposit insurance. The upshot is that deposit insurance may provide protection against bank panics (liquidity crises) at the cost of increasing exposure to bank crises (solvency crises).

However, the properties of deposit insurance schemes may matter for this outcome. Such schemes, for example, may be *explicit* (announced by the government and written into law) or *implicit* (i.e., everyone believes that depositors will be bailed out in the event of a crisis, though there is no formal provision for doing so).

Explicit schemes may have several advantages over implicit ones. For one thing, they allow the costs of insurance to be met by premia on bank deposits rather than by taxpayers. This makes it possible to make these premia risk based (i.e., based on the composition of banks' portfolios), which reduces the moral hazard problem by penalizing banks with riskier portfolios. The legal framework that they provide may also make it easier to close insolvent banks by obviating the need for expensive litigation with small depositors, again ameliorating the moral hazard problem (if banks are more likely to be closed down, bank owners are more likely to incur losses). Finally, they make it possible to limit the coverage of the protection, leaving more incentives in place for larger depositors to undertake monitoring activities. Thus an explicit scheme may create less moral hazard than an implicit one, through risk-based premia, through added credibility on bank closures, and by putting limitations on the government's liabilities.[3]

However, recent cross-country empirical evidence that we shall review in Chapter 26 (Demigurc-Kunt and Detragiache 1998) indeed suggests that the presence of deposit insurance has tended to increase the probability of banking crises. We shall see later that the solution is appropriate bank regulation and supervision. For now, though, we can simply note that even well-run banks are inherently vulnerable to panics and that this situation creates a role for public policy. Because the specific public policies that can address the problem of panics may themselves create (or aggravate) distortions, however, they need to be accompanied by other types of policy intervention.

2. Bank Regulation and Supervision

The additional intervention required involves *bank regulation and supervision*, that is, direct public oversight of bank activities for prudential purposes. These may be necessary to avoid solvency problems with or without deposit insurance, but they become more important if deposit insurance deters monitoring. What should bank regulation do? In principle, regulation has several objectives:

- **Avoid adverse selection in bank entry.** Bank regulation can achieve this goal by ensuring that individuals who are more likely to misuse banks (by taking excessive risk or engaging in related lending) do not get bank charters. This is the role of chartering regulations that require adequate capital investment, proscribe cross-ownership, and include character stipulations for bank ownership.
- **Align the incentives of bank owners with those of depositors.** This objective can be achieved by regulations ensuring that bank owners stand to make substantial

[3] Within the set of explicit schemes, fully funded schemes that offer extensive coverage and that are run by the government may actually create more moral hazard than schemes with partial coverage of deposits run by the private sector.

losses in the event of insolvency. This is the role of capital adequacy requirements and loan loss provisions.

- **Prevent excessive risk taking.** This is a more direct intervention, focused on asset management supervision. This means limiting bank holdings of excessively risky assets (such as stocks, real estate, or consumer credit), preventing lending to related parties, requiring diversification, and making sure that banks have appropriate loan evaluation procedures in place.[4]

In contrast with the lender-of-last-resort function and generalized deposit insurance, such measures imply substantial intervention with bank functions at the microeconomic level. But why should the government intervene in private transactions between depositors and their banks? There are essentially two reasons. First, as previously mentioned, the problems that call for these interventions arise from market failures created by asymmetric information and opportunistic behavior. Government intervention allows mutually advantageous trades (between depositors and banks) that would not otherwise take place, in the following sense: if bank managers could commit themselves to refrain from opportunistic behavior, depositors would entrust them with their money, and both would be happy. Because they cannot do so, depositors hold nonbank assets instead, and both are worse off. Thus government supervision in effect supplies a commitment mechanism for bank managers.

The other reason is because the failures of individual banks may create negative externalities, in the sense that they may affect the prospects of other banks. Because individual bank managers and depositors do not have the incentive to take these spillovers into account, there is a rationale for policy intervention. The problem is that banking crises can become systemic. While not all generalized banking crises need reflect bank-to-bank transmission – that is, they may also arise from general macroeconomic shocks that affect all banks (sometimes referred to as monsoon effects) – banking crises can be contagious. Channels of bank-to-bank contagion include the following.

- **Transmission.** The failure of one bank may affect the health of other banks through cross-bank claims or through the inability of depositors who lose access to their funds in the failed bank to service debt with other banks.
- **Pure contagion.** As we have seen, asymmetric information means that depositors are not well informed about the quality of bank loans. Suppose that it becomes known that some fraction of banks are insolvent, but not which ones, and that there is no mechanism for deposit guarantees in place. Then, as we saw earlier, in the absence of information about the solvency of a depositor's own bank,

[4] The supervision of bank portfolios mirrors the approach taken by insurance companies that cannot base premia on sufficiently fine risk categories.

it may make sense for that depositor to withdraw funds from his or her own bank when it is known that some banks are in trouble, just as it does when other individuals are seen to be withdrawing funds from the depositor's bank, because converting deposits into cash is relatively costless, and the expected return from doing so may be high.

The upshot is that the information problems that characterize the banking sector call for appropriate public-policy measures to address the vulnerability of a liberalized banking system to panics and solvency crises. These measures involve some combination of a lender-of-last-resort role for the central bank and a system of deposit insurance with a system for the regulation and supervision of banks specifically designed to address the moral hazard problems inherent in this type of activity. When these measures are not implemented, or when they are inadequate, a liberalized banking system will be vulnerable to panics and solvency crises, as we will see in Chapter 26.

III. CAPITAL ACCOUNT LIBERALIZATION

Up until this point, we have concentrated on the *domestic* dimension of financial reform; that is, our attention has been occupied by policies toward the domestic financial sector. Before looking at the reform process as a whole, we have to complement this with a consideration of the *external* dimension of reform. As we have seen, restrictions on capital inflows and outflows are an integral part of financial repression. Thus we have to factor into the analysis of the sequencing of reform the role of *financial opening* as well. This means that we will need to give separate consideration to the analytical case for capital account liberalization before we can turn to the sequencing of the various components of financial reform.

Though restrictions on capital flows are an important component of financial repression, in principle it is conceivable that the domestic financial system could be reformed without removing restrictions on capital movements. Thus a logical question to ask is whether, if the domestic financial system could be fully reformed without liberalizing the capital account, there would be a separate argument for opening up the capital account as well. In other words, is there a case for full capital-account liberalization in the context of a well-functioning domestic financial system?

1. The Effectiveness of Capital Controls

A possibility that we need to consider at the outset is that there may not be a meaningful choice to be made about capital account liberalization at all. It may simply not be possible to sustain effective capital account restrictions under certain circumstances. The effectiveness of controls on capital movements depends on factors that differ across countries as well as over time, and these factors can change exogenously and be the unintended result of various domestic policies. At the time that financial reform is undertaken in any given country, therefore, this combination of influences may make it impossible to maintain effective capital account restrictions.

The set of factors that influence the effectiveness of controls includes some that are international in scope, others that are of a domestic structural (i.e., slowly changing) nature, macroeconomic factors, and factors related to the design of the restrictions themselves. Factors that are international in scope include, for example, the state of the technology that is relevant for determining the magnitude of transaction costs in conducting financial arbitrage across international borders. Obviously, technological advances that reduce such costs make it more difficult to sustain effective restrictions on capital movements. Similarly, the international legal environment also matters. For example, the extent of international cooperation in reporting cross-border financial claims is likely to be an important determinant of the effectiveness of capital account restrictions. Capital flight will be discouraged if host countries declare the assets of the residents of capital-exporting countries to the taxing authorities in those countries. Finally, rates of return available in international capital markets affect the strength of the incentives for moving capital across international borders.

A second set of factors operates at a domestic structural level. The most obvious of these, of course, is the efficiency of the bureaucracy charged with administering capital account restrictions. Other domestic structural factors are of a type that may tend to change if financial reform is accompanied by market-oriented measures elsewhere in the domestic economy. Financial reform itself, for example, may alter the effectiveness of capital account restrictions because as the domestic financial sector becomes more competitive, innovative, and diversified in a reformed system, the opportunities for evading capital controls may multiply.

Changes elsewhere in the economy may also have this effect. For example, increased trade integration (a greater share of tradable goods in total domestic production and consumption) may make controls on capital movements more difficult to enforce. The reason is that an increase in the volume of trade increases possibilities for under- and overinvoicing exports and imports as well as for the use of payment leads and lags to effectively lend and borrow abroad. Similarly, an increase in the role of foreign direct investment in the domestic economy makes capital controls harder to enforce because they can be evaded through the use of transfer pricing by multinational enterprises.

A third set of factors affecting the effectiveness of capital account restrictions is the size of the domestic incentives (such as return differentials) motivating inflows or outflows. If effecting a capital inflow or outflow to carry out an arbitrage transaction involves incurring a fixed cost imposed by the control regime, then the deterrent effect of this fixed cost may be nullified if the prospective gains from the arbitrage operation significantly exceed it.[5] These incentives are generally the product of domestic macroeconomic phenomena, including monetary policy, perceived fiscal solvency, or perceived real exchange rate misalignment, among other phenomena.

Finally, the effectiveness of restrictions is very likely to depend on the design of the restrictions themselves. For example, the comprehensiveness of restrictions may affect their effectiveness. Controls that are not comprehensive – for example, that apply only to particular types of flows – can be evaded by changing the composition of flows.[6] The types of cross-border flows targeted by restrictions may also matter; that is, for a variety of reasons, controls on capital outflows may differ in effectiveness from controls on inflows, other things equal.[7]

Although economists tend to have strong views about the potential effectiveness of capital controls, the empirical evidence on the issue is problematic. Restrictions on capital flows can be intended to restrict the size of net capital flows and/or to affect their composition. The evidence differs on their ability to attain these two objectives. We can summarize it as follows.[8]

First, with regard to effects of capital account restrictions on the *magnitude* of capital flows, the evidence indicates that countries have at times been successful in using controls to preserve short-run monetary autonomy, driving temporary wedges between domestic and foreign interest rates.[9] But there is little evidence that they can sustain such wedges permanently – at least under industrial-country conditions. How large the attainable wedge is and how long it can be sustained are likely to depend on the factors influencing the effectiveness of controls (such as those listed earlier), though little systematic research appears to have been conducted to date on the role of the factors that may render controls effective or ineffective. Overall, controls appear to have been able to preserve a limited degree of monetary autonomy temporarily, even in the fairly sophisticated financial systems of many industrial countries. This suggests that generic capital account restrictions can

5 This means, in particular, that the size of capital flows in the presence of controls is not a reliable indicator of the effectiveness of controls because large flows may simply indicate large ex ante arbitrage opportunities.

6 Notice that though this would make controls ineffective in preserving monetary autonomy (or limiting the total magnitude of flows), it would not necessarily make them ineffective if their purpose is to alter the composition of flows.

7 One argument for this is that the residence of the capital-importing or capital-exporting agent may matter. Domestic residents may be more prepared to evade restrictions than foreign agents, making controls on outflows less effective than controls on inflows.

8 This evidence does not necessarily speak clearly. One difficulty is that of attribution. E.g., tests of monetary autonomy based on interest parity conditions in the presence of capital controls may be more informative if autonomy fails than if it holds because in the latter case, natural barriers (e.g., differing perceived risk characteristics of domestic and foreign assets) may be operative. Even when changes in the intensity of controls are examined to minimize such problems, it becomes necessary to control for changes in other conditions such as in the stance of monetary policy.

9 See Dooley (1995) as well as Magud and Reinhart (2007).

remain effective after domestic financial reform. But the evidence from industrial countries also indicates that controls have not been able to prevent the emergence of large capital outflows and inflows when prospective arbitrage profits have been large.

In the emerging-economy context, by contrast, the large and persistent parallel exchange-market premia that existed in the presence of the capital account restrictions associated with financial repression, as well as the reduction of such premia in response to changes in foreign exchange regulations, suggest that controls can affect cross-border capital flows under a wide variety of country circumstances. That substantial inflows have followed the removal of controls in several cases (e.g., Korea in 1992), and that inflows have slowed after controls were reimposed (Chile in 1991 and Malaysia in 1994), suggests that controls *can* work, at least in some cases, and at least temporarily. On the other hand, as in industrial countries, episodes of extensive capital flight in the presence of restrictions on the outward movement of capital indicate clearly that controls cannot prevent large capital movements when perceived arbitrage margins are large.

There has been a substantial amount of more systematic country-specific research on the effectiveness of capital account restrictions in emerging economies during recent years. Some of this evidence is described in Box 22.1. Overall, it can be summarized as providing weak evidence for the effects of controls on the total magnitude of inflows but stronger evidence for effects on their composition.

In addition to this country-specific work, there is a limited amount of cross-country research on this issue. Montiel and Reinhart (1999), for example, tested for the effects of capital account restrictions in a group of emerging economies by estimating a set of fixed-effects panel regressions explaining the volume and composition of various types of capital inflows for a 15-country panel of such economies, using annual observations over the period 1990–1996. Inflows were explained as a function of a variety of macroeconomic variables, including the intensity of sterilization in each country, the severity of capital account restrictions, the level of international interest rates (measured as the yield on three-month U.S. Treasury bills) and a so-called tequila effect dummy to capture possible effects on capital inflows of the Mexican financial crisis at the end of 1994.

The results were consistent with those of the country-specific studies cited earlier. In various specifications, the coefficients on the capital control proxy, though consistently of the theoretically predicted sign, proved to be measured with a relatively low level of precision. However, when the dependent variable measured the *composition* of capital inflows, the explanatory variable measuring the intensity of controls proved to be associated with a significantly lower share of short-term flows and portfolio flows in total capital inflows as well as a higher share of foreign direct investment. The authors concluded that though there was some weak evidence that explicit capital inflow restrictions, and "prudential measures" affecting capital inflows (usually limiting banks' foreign exchange transactions or foreign

Box 22.1. Effectiveness of Capital Controls in Emerging Economies: Country Studies

Brazil

Using monthly data from January 1988 to December 1995, Cardoso and Goldfajn (1997) found that changes in an index of various Brazilian capital account restrictions had short-run impacts on both the volume and composition of private capital inflows. The impact peaked after about six months, however, and died out thereafter.

Chile

Quirk and Evans (1995) and Le Fort and Budnevich (1996) found that the maturity composition of capital flows into Chile was altered in favor of longer maturities after the imposition of capital controls in the form of unremunerated reserve requirements (URR).

Valdes Prieto and Soto (1996) found that an implicit tax variable designed to capture the effects of the URR did not affect aggregate flows of short-term capital in Chile, though it did affect at least one important component of such flows. Edwards (1998) found that the controls may have reduced the speed of convergence of Chilean interest rates toward uncovered interest-parity levels, increasing Chile's degree of monetary autonomy. Nadal-DeSimone and Sorsa (1999) as well as Cowan and De Gregorio (2005) provide more recent overviews.

Colombia

Cardenas and Barreras (1996) estimated capital-flow equations with interest differentials adjusted by an estimate of the implicit tax associated with the unremunerated reserve requirements on capital inflows imposed by Colombia during part of their sample period. They found that the tax term had no impact on capital flows. On the other hand, the tax term did tend to affect domestic interest rates as well as the parallel exchange market premium. They also found a substantial change in the structure of Colombia's external borrowing after these controls were implemented. They concluded that controls in the form of unremunerated reserve requirements had little effect on the total magnitude of flows but altered their composition.

Malaysia

Quirk and Evans (1995) as well as Reinhart and Dunaway (1996) concluded on the basis of before-after comparisons that severe controls on capital inflows imposed in Malaysia in 1994 may have affected both the magnitude and composition of inflows in the short run.[10]

[10] It may be worth emphasizing that testing the effectiveness of capital account restrictions requires controlling for changes in the degree of sterilization because a loosening of monetary policy accompanying an intensification of capital account restrictions (the cases of Chile and Malaysia) could mistakenly attribute any changes in the volume and composition of capital flows to the change in restrictions rather than to the change in monetary policy. Conversely, a tightening in monetary policy at the time when the taxes or controls are introduced (Brazil) could undermine the effectiveness of the controls by raising domestic interest rates to levels where either domestic assets remain attractive even on an after-tax basis or by providing an incentive to circumvent the new controls.

exchange exposure), affected the total magnitude of flows in their sample countries, such measures seemed to be more effective in altering the composition of capital inflows than in reducing their overall magnitude.

Overall, then, the evidence from emerging economies on whether capital account restrictions have typically been effective in altering the total magnitude of capital flows is mixed. At best, there is weak evidence of their effectiveness. However, there appears to be much more consensus that controls may be effective in altering the composition of flows.

2. Optimality of Capital Account Restrictions

Even if capital restrictions *can* remain effective, that does not mean that they should be retained. In discussing the liberalization of the domestic financial system, we saw that a well-functioning liberalized financial system could be expected to confer certain benefits on the economy but that laissez-faire was unlikely to be the best policy; that is, a role for public policy would remain. We can perform a similar analysis for the external component of liberalization, identifying the benefits of capital account openness and then considering whether there is a case to be made for any type of policy intervention in this area.

The argument for liberalizing capital flows is that the enhanced financial integration with world capital markets that would follow from removing restrictions on capital movements can yield a number of economic benefits[11]:

- Enhanced financial openness can increase national wealth by allowing domestic firms to finance investment projects with rates of return greater than the costs of foreign borrowing or by permitting domestic savers to invest in higher-yielding assets than may be available in the domestic economy.
- It may also permit an improved intertemporal allocation of consumption for the domestic economy. This is because international borrowing and lending allows the domestic economy to divorce the path of its consumption expenditure from that of its income, allowing it to allocate its consumption over time in ways that may better suit the preferences of domestic residents.
- Financial openness gives domestic residents access to gains from portfolio diversification, reducing the overall risk in their portfolios without necessarily sacrificing expected return.
- Financial openness can support the process of financial reform by promoting the efficiency of financial intermediation. It can do this by increasing competition for domestic financial intermediaries.
- Finally, financial openness can impose macroeconomic discipline on governments by reducing their ability to tax the financial system through explicit

[11] For a recent overview, see Kose et al. (2006).

or implicit taxes. As we have seen before, one motivation for retaining capital account restrictions under financial repression was precisely to avoid this discipline.[12]

Because the retention of effective controls on capital movements would imply forgoing – or at least reducing – these potential benefits, the view that restrictions should be retained faces a serious burden of proof. Perhaps the best argument for restricting the scope of financial openness is as a means of counteracting distortions in the domestic economy that are not taken into account in the list of benefits from liberalization enumerated earlier.

One argument for retaining controls over capital *inflows*, for example, is that there may be externalities associated with the act of foreign borrowing. If the risk premium faced by individual domestic borrowers in international capital markets, for example, depends on the country's *total* stock of outstanding external debt, then each individual's act of borrowing abroad imposes higher borrowing costs on other domestic residents. Because individual borrowers would not take this effect into account if external borrowing were conducted in a decentralized fashion, the country as a whole would tend to overborrow. In this situation, there is a case for permanent intervention by the government to restrict external borrowing, in the form of a small Pigouvian tax on external borrowers that would cause individual domestic borrowers to take into account the effects of their own borrowing on the premium faced by other domestic agents.

Alternatively, capital controls could be justified as permanent second-best instruments to offset the effects of purely *domestic* microeconomic distortions that cannot be removed. The preferred (first-best) policy would always be to remove the distortion, but where this is not possible, a case can be made for controls on second-best grounds.

For example, external overborrowing may arise from moral hazard problems, if domestic borrowers are induced to borrow more because they believe that they will be bailed out by the government in the event that they are unable to repay. Prudential restrictions on external borrowing and lending by domestic financial institutions can plausibly be warranted on these grounds. These restrictions on capital movements have a clear second-best rationale because they are directed at a specific distortion: the inability of the government to credibly commit to removing certain implicit guarantees.

Similarly, temporary restrictions on capital movements may be justified when the public is induced to borrow or lend abroad because it disbelieves a policy commitment that the government is actually determined to carry out. For example, if the public does not believe that an exchange rate–based stabilization will be

[12] It may be worth pointing out that, to the extent that effective capital account restrictions can be difficult to reimpose once they have been removed, this means that the removal of such restrictions can be used by the government in the same way as the granting of independence to the central bank, i.e., as a signal of its good future fiscal intentions. See Bertolini and Drazen (1997).

sustained, its expectation that a devaluation is forthcoming may cause it to move funds abroad. The resulting increase in domestic interest rates may actually make the stabilization harder to achieve. In this case, the distortion that motivates the capital flow is the absence of a precommitment mechanism for the authorities that could make their intentions credible. If credibility is not achievable, temporary capital account restrictions may be warranted, again on second-best grounds.

Finally, restrictions on capital flows have been defended on more general grounds of macroeconomic stabilization: to insulate the domestic economy from external financial shocks and to preserve monetary autonomy, thereby permitting the domestic monetary authorities to use the exchange rate and money supply as separate policy instruments.

The first goal, however, is questionable. Its justification requires that the government be better informed about the duration of financial shocks than the private sector or that the transmission of such shocks to the domestic economy tend to aggravate the effects of some previously existing distortion. Restrictions designed to achieve this purpose are hard to justify as a general proposition. With regard to the objective of preserving some degree of monetary autonomy, the benefits to be gained for macroeconomic management are large only if other stabilization instruments are not available. If fiscal policy is sufficiently flexible to be used for stabilization purposes, for example, then price stability and satisfactory current account performance can be pursued through the combined use of fiscal and exchange rate policy. The first-best solution to this problem may thus be to increase the flexibility of fiscal policy rather than to adopt capital controls. At best, therefore, these considerations provide an argument for retaining restrictions as a transitory device, until the fiscal system can be reformed in such a way as to make fiscal policy an effective stabilization tool.

In short, there may indeed be legitimate arguments for retaining some forms of mild restrictions on capital movements for extended periods, even after the domestic financial system has been reformed, or even for retaining or reimposing much stronger restrictions temporarily under special circumstances. But it is hard to come up with defensible arguments for the retention of the very severe forms of capital controls that typically accompany domestic financial repression. Thus, in analyzing reform, the general case should probably be taken to be one in which domestic financial reform is accompanied by at least some liberalization of capital movements. A general reform strategy, therefore, should include a capital account liberalization component. Next, we will consider how this component might fit into the general reform process.

IV. THE SEQUENCING OF FINANCIAL REFORM

The next step is to consider whether there is anything we can say about how the *process* of reforming the domestic financial system might be carried out. Though theory tells us that a well-functioning financial system can make an important contribution to fostering economic growth, the process of reforming the financial system may prove to be a difficult one. We have already seen that without the necessary institutional preconditions in place, financial liberalization can go wrong, and not only may the favorable growth outcomes expected from the process not be achieved but the financial system may itself become a source of macroeconomic problems that undermine economic growth. More generally, because the distortions associated with different components of financial repression are likely to interact with each other as well as with the macroeconomic environment, the sequence in which financial reform measures are adopted may affect how the economy performs during the reform process. The question, then, is what considerations determine the optimal sequence in which measures should be implemented during the process of financial reform?

The concern with the sequencing of reform (a term that has come into wide use as a verb) essentially reflects the view that the conditions under which domestic and external financial liberalization are undertaken may ultimately determine whether they yield the expected economic benefits and that some of these conditions, at least, are under the control of the authorities conducting the reform process.

For example, some observers have argued that freeing domestic interest rates when bank supervision is weak and the domestic macroeconomic situation is unstable is a recipe for disaster, even if the capital account remains closed.[13] Macroeconomic instability increases the variance and covariance of a given set of projects in banks' portfolios, and as we have seen, weak supervision in the context of free (or underpriced) deposit insurance creates incentives for banks to endogenously increase the riskiness of the lending they choose to undertake. A strategy of paying high interest rates to attract deposits, and then investing these funds in high-risk projects with low expected returns, may maximize expected returns to the owners of the banks, though not to society as a whole. When banks fail, the combination of macroeconomic crisis and lost information capital contributes to lower- and poorer-quality investment and reduced growth. As we will see, the evidence to be reviewed in Chapter 26 about the causes of banking crises around the world is consistent with this view. But this would suggest that freeing domestic interest rates should await the establishment of macroeconomic stability as well as the appropriate regulatory institutions for the banking system – in other words, that the process of financial reform should be sequenced.

Indeed, there is a reasonably broad consensus among economists concerning the order in which the major steps in reforming the financial sector are best undertaken. In brief, this consensus takes the view that macroeconomic stabilization should precede domestic financial reform and that domestic financial reform should precede the liberalization of the capital account of the balance of payments. Why should this be?

[13] See Villanueva and Mirakhor (1990).

1. The Sequencing of the Broad Components of Reform

Recall from Chapter 21 that financial repression has a strong fiscal motivation and that the revenue that governments collect from taxing the financial system through repression can be substantial. This means that financial liberalization without a fiscal adjustment is likely to increase the rate of inflation. Because higher inflation may be harmful for growth, financial reform undertaken to stimulate growth could be self-defeating without a previous fiscal adjustment. Moreover, as we saw in Section II, safeguards against moral hazard need to be in place before the banking system is liberalized. One such safeguard is the adequate capitalization of banks, which may require a substantial injection of government funds to absorb the bad loans of the banking sector before it is liberalized. In other words, a fiscal adjustment may be needed to generate the resources with which to reform the banking sector. The implication is that, for both reasons, weaknesses in the government's budget have to be addressed before financial repression can be eliminated; that is, *stabilization should precede domestic financial reform.*

Regarding domestic financial reform itself, we also saw in Section III that before the domestic financial system is liberalized, institutional policies should be put in place to protect against liquidity crises and moral hazard lending. This means, on one hand, that the lender-of-last-resort function and/or mechanisms for deposit insurance should be clarified, and on the other hand, that adequate prudential systems need to be established, and banks well capitalized, before bank activities are liberalized. This means in particular that if previously repressed banks have their capital reduced by nonperforming loans, these institutions need to be recapitalized before they are freed to make borrowing and lending decisions on their own.

Together, these observations suggest that the early stages of the reform process should emphasize fiscal reforms, institutional reforms to create an appropriate regulatory structure for the domestic financial system, and the recapitalization of domestic banks to the extent necessary. While these measures are implemented, restrictions on the activities of domestic banks should remain in place. Once they are completed, the domestic financial sector can be liberalized, permitting competition among financial institutions through the setting of interest rates, allocation of funds, and entry into as well as exit from the financial sector.

Where does the liberalization of the capital account fit into this sequence? In the context of the discussion of sovereign risk premia in Chapter 10, we saw that investor confidence in the sustainability of a fiscal policy regime that would safeguard the value of their assets is necessary to prevent capital flight when the capital account is (de facto) open. Moreover, as we will see in Chapter 24 (but as already mentioned in the preceding section), adequate flexibility of domestic macroeconomic policy instruments is required to counteract the effects of capital movements when the capital account is open. If fiscal consolidation and the credibility of the government's long-run fiscal policy stance are not achieved before this opening takes place, it may later prove impossible (because of creditor reactions) to adopt temporarily looser fiscal policy in response to contractionary shocks such as an increase in external interest rates. Thus, both to avoid sovereign risk problems and to maximize the room for fiscal policy to maneuver, stabilization should precede financial opening.

Similarly, as we saw in the previous section, distortions in resource allocation through the domestic financial sector should be removed to the greatest possible extent before restrictions on capital flows are removed, so foreign resources are not misallocated by the domestic financial system (which could result in *immiserizing capital inflows*, i.e., foreign borrowing that yields a return lower than the cost of the funds). The implication is that *domestic financial reform should precede financial opening*. The upshot is that capital account liberalization should be undertaken as the last stage in the reform process.

Putting these observations together yields the prescribed sequence for the broad components of reform: macroeconomic stabilization, the implementation of prudential mechanisms, and the capitalization of banks should come first, followed by the liberalization of the domestic financial sector, followed finally by capital account liberalization. This sequence has come to command a wide consensus among economists who have studied the reform process. The appendix to this chapter provides a sample of the reasoning offered by several knowledgeable observers in support of this sequence, which has become the conventional wisdom on the subject.

2. Specific Reform Measures

Though the preceding discussion provides some broad guidelines, it falls short of providing a specific sequence for the reform process because each of the three broad areas of reform itself comprises several components. Not only can these components themselves be sequenced in a variety of ways but considering them separately may cause us to reconsider the conventional sequence as it applies to the role of a specific component. As an illustration of the latter, in this section, we will consider an analysis by Fischer and Reisen (1990) of the roles of specific components of the capital account of the balance of payments.

Fischer and Reisen (1990) take a somewhat more nuanced view of the role of capital account liberalization in the general sequence of financial reform than that which appears in the conventional analysis presented earlier. They decompose capital account liberalization into several parts and argue that different components of the capital account should be liberalized at different stages of the reform process.

According to Fischer and Reisen (1990), liberalization of foreign direct investment (FDI) and trade finance should come early in the reform process, for two reasons. First, they argue that openness to FDI and to international trade are essential for development. The argument for the beneficial effects of these flows on development is that on one hand, positive spillovers emanate to the rest of the domestic economy from foreign direct investment, and on the other hand, that the availability of trade finance helps the economy to reap the benefits of commercial openness. Second, because these types of resource flows are not intermediated through the domestic financial system, they are not subject to misallocation as the result of any distortions that may exist in that system.

Beyond this, Fischer and Reisen (1990) take fiscal consolidation to be the most important next step, for the reasons mentioned earlier (i.e., because it is needed to permit the government to do without revenues from financial repression and to provide a stabilization instrument when the capital account is opened as well as because a strong fiscal position is needed to cope with bad loan problems in the reforming financial sector).

Next, they would implement steps to reform the domestic banking system. They argue that this should precede financial opening because otherwise, financial opening will lead to high domestic interest rates for moral hazard reasons. They take this to mean the removal of excessive bad loans to increase the franchise value of banks, the strengthening of prudential regulation and supervision, the establishment of legal and accounting systems to cope with systemic risks, and the enforcement of competition to foster allocative efficiency in the financial sector. Because these measures take time to implement, they should be enacted early in the reform process.

After macroeconomic stability is achieved, the requisite institutional mechanisms are in place for the domestic financial sector, and the bad loan problem is resolved, domestic interest rates can be freed. Under these conditions, overintermediation (i.e., moral hazard-driven borrowing and lending) should not be a problem. At the same time, the government should take steps to foster deepened securities markets. Once high-yielding domestic instruments are in place and there is no debt-overhang problem to trigger capital flight, capital *outflows* can be liberalized. When domestic financial reform has been completed (interest rates have been freed, bad loans removed, and reserve requirements lowered), the entry of foreign banks into the domestic financial system can be permitted.

The final step in the reform process for Fischer and Reisen (1990) should be the liberalization of short-term capital inflows. This should happen only after free entry has increased bank competition and enhanced credit-market integration, after bad loan problems have been resolved and banks have accumulated experience exercising independent credit judgment, after prudential regulation has proven adept at preventing distress borrowing, and after stabilization has succeeded in producing low domestic interest rates.

V. COUNTRY CASE STUDIES

The preceding analysis suggests that the effects of financial reform should depend on how it is carried out. Unfortunately, it is not easy to bring evidence to bear on this issue. A study by King and Levine (1993), for example, considered the effects of financial reforms from a variety of perspectives. For instance, they looked at the behavior of several financial indicators before and after financial reforms in five countries, finding that all four of their indicators of financial depth increased and that the ratio of currency to demand deposits fell in Argentina, Chile, Indonesia, Korea, and the Philippines. During all of the reform episodes that they examined, except one, the real interest rate also rose after the reform. They did not try to link these phenomena with subsequent growth performance. King and Levine also looked at the relationship between their financial indicators and the success of World Bank adjustment lending, on the hypothesis that structural adjustment measures are more likely to promote growth under a well-functioning financial system and that a given set of growth-enhancing measures is more likely to be successful if they are accompanied by a financial reform. They found that intensive-adjustment-lending countries with better initial financial depth tended to grow faster during the subsequent five years. This relationship held up after correcting for other growth determinants.

But of course, this kind of evidence has little to say about the effects of alternative reform *strategies*, that is, about the sequencing of reform. We have already seen some indirect evidence on this issue in the previous chapter, where it was suggested that financial liberalization in the context of an inappropriate domestic institutional environment tended to be associated with financial crises. More detailed evidence on the effects of financial reform strategies tends to be in the form of country case studies. The methodology is descriptive, and conclusions about the effects of reform strategies are typically based on before-and-after examination of the data, thereby failing to control for other influences on the variables of interest.

An extensive older literature of this type exists analyzing the financial liberalization experience of the Southern Cone countries in South America during the late 1970s.[14] This literature's conclusions are that the liberalization effort there – comprising the removal of restrictions on the domestic financial sector as well as the opening up of the capital account – failed because the process was undertaken in an unstable macroeconomic setting and with an unsatisfactory domestic regulatory framework. In other words, financial liberalization was premature in the sense that it preceded other steps that should actually have been taken earlier in the reform process. Financial crises with severe macroeconomic implications soon followed in each of the three countries involved. Thus the Southern Cone experience supports the policy prescription that the achievement of domestic macroeconomic stability (the centerpiece of which is a sustainable fiscal position, as we have seen before) and the implementation of an appropriate regulatory framework for the domestic financial system should precede the removal of restrictions on domestic financial institutions and the opening of the capital account.

[14] See, e.g., the classic study by Diaz Alejandro (1985). Indeed, economists' research interest in the optimal sequencing of reform was essentially triggered by the experience of these countries.

Because this experience is more familiar (and because we will return to it in Chapter 27, in this section, we will briefly examine the financial reform experience among emerging economies elsewhere – specifically, in Asia and Sub-Saharan Africa – at the times when they began their transitions from financial repression, that is, during the early years of their reform processes.

1. Financial Reform among Asian Economies

Several countries in Asia, including the miracle economies, undertook reform of their financial systems during the decades of the 1970s and 1980s. The experience of these countries is less well known than the earlier ones of the Southern Cone countries.[15] In Chapter 27, we will consider the role of financial reform in the financial crisis that broke out in Southeast Asia in mid-1997. For now, we will focus on the cases of Japan, Korea, and Taiwan during an earlier period. Overall, we can summarize the main outlines of the experiences of these countries as follows:

1. Among the three miracle economies reviewed in more detail in Chapter 21, Japan in the mid-1970s, as well as Korea and Taiwan in the early 1980s, decided to liberalize what had been heavily regulated, domestically closed, and moderately repressed financial systems. Reform of the financial sector was undertaken in the context of a stable domestic macroeconomic environment. Growth was rapid, inflation was relatively low, and all the governments involved were in relatively strong fiscal positions. All three systems moved from highly regulated, moderately repressed states to less regulated, more competitive, market-based systems during the course of this process.
2. The *scope* of reform tended to be broad in these countries. The easing of controls over interest rates was accompanied by measures to promote competition in the financial sector, such as increased freedom of entry by domestic banks, expanding the scope of permissible activities for different institutions and relaxing restrictions on entry by foreign banks. The supervisory framework for the financial system was typically strengthened by centralizing supervisory responsibilities, developing and unifying the regulatory framework (as well as extending it to nonbanks), and in some cases, providing explicit deposit insurance. Money markets were fostered by creating new instruments (such as central-bank and government securities) with flexible interest rates.
3. However, the *pace* of liberalization was gradual and sometimes partial. Liberalization was slow and piecemeal in Japan, Korea, and Taiwan by comparison with the experience of the Southern Cone. All three countries tended to follow the conventional sequencing of reform, with real sector reform first, then domestic financial reform, then capital account liberalization. The partial nature of

[15] An overview of financial liberalization in several of these countries during the decade of the 1980s is provided by Tseng and Corker (1991) as well as World Bank (1994a).

reform is evidenced, for example, by the fact that interest rates were not always fully liberalized. Sometimes (as in Korea), managed rates were simply adjusted more frequently. Also, though total credit controls were eliminated, sectoral credit allocation requirements and selective rediscounting remained in place for some time. The deregulation of deposits began in Japan in 1981, when new types of instruments were permitted for commercial banks, trust banks, long-term credit banks, and the postal savings system. These still had regulated rates, but new instruments with market-determined rates were introduced at about the same time. However, deregulation of small time and demand deposits was put off at that time. Interest rate controls were eventually lifted in Japan, and international transactions were liberalized. In Taiwan, the pace and intensity of reform increased in the mid-1980s, and by 1989, deposit and loan rates were free. In 1990, 16 new private banks were licensed, and some government banks were privatized. The liberalization of capital flows started in the 1990s. Finally, in Korea, interest rate deregulation began only in 1991, on a fixed schedule. The intended pace of capital account liberalization was slower in Korea than in the other two countries.

4. The financial results of the early stages of this gradual liberalization process did not tend to be disruptive, as had occurred in the more rapid liberalizations of the Southern Cone. The gradual liberalization of interest rates tended to increase nominal rates but not necessarily enough to establish very large positive real rates. Positive real rates were attained by changes in nominal rates and, especially in Korea, through a reduction in the rate of inflation. Yet in each of these countries, positive real interest rates followed liberalization, and financial deepening occurred at the same time. Whether liberalization was accompanied by desirable growth effects is another matter entirely. The difficulty, of course, is in controlling for other factors that were influencing macroeconomic performance in the liberalizing countries at the same time.

2. Financial Reform in Sub-Saharan Africa

Financial liberalization has been undertaken in Africa much more recently (since the mid-1980s). The adoption of some form of liberalization or bank restructuring measures became fairly widespread after that time.[16] A partial list of the early liberalization measures is provided in Box 22.2. In addition to the countries listed in the box, interest rates had been freed by the early 1990s in Burundi, Madagascar, Mauritania, and Zambia. Many other countries moved from setting rates to setting minimum deposit and maximum lending rates at the same time, or to regulating

[16] World Bank (1994b) lists a total of 19 Sub-Saharan African countries that undertook some form of liberalization measures in the financial sector from the mid-1980s to the mid-1990s.

Box 22.2. Financial Liberalization in Sub-Saharan Africa

The Gambia (September 1985)

Ceilings on interest rates were removed in September 1985, an auction system for issuing treasury bills was introduced in July 1986, and quantitative controls on credit were removed in September 1990.

Nigeria (July 1987)

Directed credit restrictions were relaxed over the period 1983–1987 by increasing the sectoral aggregation of directed credit allocations. On July 31, 1987, the central bank removed interest rate controls and raised both the treasury bill and rediscount rates by 4 percentage points, to 15 and 14 percent. In November 1989, an auction system was instituted for treasury bills and certificates, but the central bank retained a reservation price.

Ghana (September 1987)

Ceilings on interest rates were removed, while the removal of quantitative credit controls was scheduled for 1992.

Malawi (April 1988)

Ceilings on interest rates were removed, and quantitative credit ceilings were eliminated in January 1991.

Uganda (July 1988)

On July 1, 1988, an increase of 10 percentage points was announced on most interest rates.

Benin and Cote d'Ivoire (October 1989)

The BCEAO abolished its preferential discount rate, but bank interest rates remained subject to regulation. Cote d'Ivoire is also a member of the BCEAO, so it was affected by these liberalizing measures.

Cameroon (October 1990)

The BEAC (French acronym for the Bank of Central African States) eliminated its preferential lending rates, simplified its interest rate structure, and increased its power to determine interest rate policy with the intention to move toward greater flexibility in rates.

Tanzania (July 1991)

The system of fixed interest rates and fixed differentials was replaced by a single maximum lending rate of 31 percent on July 25.

Kenya (July 1991)

Interest rate ceilings were removed.

Source: Turtleboom (1991) and Galbis (1993) spreads. Interest rate liberalization has recently been undertaken throughout the continent.

However, low positive real interest rates were *not* the rule after liberalization in Sub-Saharan Africa, as they had been in the Asian miracle economies even *before* liberalization. Post-liberalization, Sub-Saharan African countries tended either to continue to have the negative interest rates that were typical under financial repression or, in the case of those with fixed exchange rates and overvalued currencies, very high positive rates. In part, this may have reflected the incomplete and piecemeal nature of the liberalization in these countries. In the case of the Central Bank of the West African States (BCEAO) of the West African Monetary Union, for example, the abolition of the bank's preferential discount rate in October 1989 left interest rate controls in place. The same was true of the Central Bank of the Central African States, which abolished its preferential discount rate in October 1990. Though Nigeria removed interest rate controls in July 1987 and introduced treasury bill auctions in November 1989, the government continued to exert influence over interest rates emerging in such markets through the reservation prices set by the central bank. Some adjustments to previously fixed interest rates were recorded in Tanzania in 1991 and Uganda in 1988, but neither country moved immediately to flexible interest rate determination. Other possible explanations for the behavior of real interest rates after liberalization are lack of competition due to the small number of banks and the continued role of government-owned institutions.[17]

In addition to the behavior of real interest rates, widening post-liberalization spreads between lending and deposit rates appear to have been a problem in several Sub-Saharan African countries that liberalized interest rates during the 1980s (Gambia, Ghana, Malawi, and Nigeria, according to Turtleboom (1991)). Again, lack of competition is a possible explanation, as is the continued presence of nonperforming loans.

Regarding the latter, the experience with bank restructuring also appears unfortunately not to have been favorable. Recapitalization operations tended to be expensive, recovery of nonperforming assets was poor, and restructuring operations had to be repeated in several cases (World Bank 1994). In Ghana, for example, nonperforming assets restructured under a plan undertaken in 1989 amounted to 41 percent of total bank credit to the nongovernment sector (Kapur et al. 1991). Overall, then, the reform process did not proceed smoothly in Sub-Saharan Africa.

VI. SUMMARY

As we have seen in previous chapters, the economy pays a high cost when the government finances itself through the domestic financial system, by relying on

[17] World Bank (1994b) reports that the number of commercial banks in Sub-Saharan Africa in which the government retains a majority share fell from 108 in 1982 to 89 in 1992.

financial repression. Where such a regime is in place, financial reform is appropriate – there are ways to raise revenue that are less costly to the economy. However, the alternative to financial repression is not necessarily a laissez-faire policy stance toward the financial system. As we have seen in this chapter, insufficient government involvement can also be harmful. The appropriate role of government in the financial sector has both institutional and macroeconomic dimensions. We can summarize this role as follows.

First, from an institutional perspective, we saw in Chapter 20 that the most basic requirement is for the government to create an appropriate legal framework within which the financial sector can function. This includes clear and well-established property rights, appropriate accounting and disclosure standards, and a judicial system that functions effectively to enforce contracts and punish fraud. Such a framework facilitates the use of collateral and lowers the costs of monitoring as well as those of contract negotiation and enforcement. These circumstances combine to reduce the premium for external finance and thus expand the role of financial intermediation.

Second, as we have seen in this chapter, the government needs to ensure the existence of a financial safety net to avoid liquidity crises. Even in a deregulated financial system, it may be impossible for the government to fail to operate a financial safety net in the form of an implicit deposit insurance system. The likelihood of financial panics under imperfect information, and the role of banks in the payments mechanism, probably mandate this result. Making the deposit-insurance system explicit and charging banks a fee in accordance with the riskiness of their portfolios may be the best way to avoid moral hazard problems.

However, if deposit insurance is improperly priced, even a competitive financial system could malfunction after liberalization, unless it is adequately supervised, due to moral hazard problems. Thus the third responsibility of the government in the institutional realm is to provide an adequate supervisory and monitoring framework to prevent collusion as well as to avoid the excessive risk taking associated with moral hazard problems. The supervisory function would complement the financial safety nets mentioned earlier, inducing liberalized institutions to behave in a prudent fashion by monitoring their portfolios, ensuring that appropriate provisions are made against suspect loans and that bank capital is adequate.

A separate dimension to the government's regulatory function is the prohibition of collusion, as suggested by the results of financial liberalization in Sub-Saharan Africa. In the absence of foreign competition (capital mobility), freeing the domestic financial system could result in noncompetitive behavior, with low deposit rates and high lending rates providing substantial profit margins to colluding institutions. This would be more likely in countries with limited access to external finance, strong domestic banking associations, and limited supervisory mechanisms.

Finally, the government's role also has a macroeconomic dimension, with a narrow budgetary component and a broader component related to macroeconomic management. The budgetary component arises because if other forms of revenue are not in place to replace those that the government was previously receiving from financial repression, inflation is the likely result. As we have seen, for a variety of reasons, the emergence of inflation may undermine the very growth process that financial liberalization is intended to promote. Thus fiscal adjustment is an important component of financial liberalization to replace lost revenue from financial repression.

In addition to this, as we saw in Chapter 20, the government has a broader role through its macroeconomic management more generally to ensure a stable macroeconomic environment. In Chapter 20, we interpreted this as meaning an environment in which financial intermediation was not confronted with avoidable macroeconomic uncertainty because macroeconomic volatility tends to increase the loan evaluation and monitoring costs that contribute to a higher external finance premium. This also involves avoiding subjecting the domestic economy to large macroeconomic shocks that could trigger a costly and disruptive banking crisis in a weak and vulnerable domestic financial system.

The easing of restrictions on capital movements is an important part of the financial reform process because, just as a well-functioning domestic financial system holds the promise of substantial benefits in economic welfare and growth, a relatively open capital account does so as well. Though there may be valid arguments for retaining (or imposing) some types of mild restrictions on capital movements, except in extreme (and temporary) circumstances, these fall far short of the near-absolute barriers on international capital flows that tend to be associated with financial repression.

One valid argument for capital account restrictions is that they may serve as a useful second-best policy when there are important domestic distortions in the real or financial sectors. This implies that opening up the capital account is probably best deferred until domestic financial reform is complete. Given that domestic financial reform itself is more likely to be successful when domestic macroeconomic stability has been achieved, featuring specifically a sustainable fiscal stance without reliance on financial repression "taxes," this suggests a sequence of reforms led by macroeconomic stabilization, followed by domestic financial reform and including capital account liberalization only as a final step. Domestic financial reform itself needs to proceed in stages, with an appropriate regulatory framework and adequate bank capitalization in place to ameliorate moral hazard problems before restrictions on bank activities are removed. Such a sequence has indeed come to represent conventional wisdom among economists, though a more nuanced view might allow for a removal of restrictions on certain types of capital flows (e.g., foreign direct investment and trade credits) much earlier in the process. The experience of a wide range of emerging economies tends to support the conventional wisdom.

APPENDIX 22.1. OBSERVATIONS ON SEQUENCING OF FINANCIAL REFORM

As indicated, the sequence of reform in which the achievement of macroeconomic stability, the implementation of an appropriate regulatory and supervisory framework, and the capitalization of banks all precede domestic financial liberalization, which in turn precedes the opening of the capital account, has come to represent the conventional wisdom on the optimal sequencing of reform. This appendix illustrates this by describing the policy prescriptions of three knowledgeable observers.

1. Turtleboom (1991)

Turtleboom (1991), for example, proposes a four-step sequence of reform:

1. Proceed simultaneously in restoring macroeconomic equilibrium and restructuring or liquidating insolvent financial institutions. The latter process should be initiated early because it is likely to take time.
2. Introduce indirect instruments of monetary control with freely determined interest rates such as treasury bills sold at auction. At the same time, establish supervisory guidelines for banks regarding loan classification, provisioning for bad debt, interest rate capitalization, capital adequacy, and limits on portfolio concentration.
3. Increase competition among banks by granting more bank licenses, permitting the entrance of foreign banks, and privatizing government-owned banks.
4. As a final step, remove interest rate controls and directed credit ceilings.

Notice that this sequence involves putting the institutional and macroeconomic conditions in place before liberalization proper (in the form of step 4) is attempted but that no specific attention is paid to the role of capital account liberalization.

2. McKinnon (1992)

In a book titled *The Order of Economic Liberalization*, Ronald McKinnon considered both the issues of domestic financial reform and capital account liberalization in the context of a broader reform sequence for the previously centrally planned economies of the former Soviet bloc.

The first step advocated by McKinnon is to get the fiscal deficit under control, with an important contribution to be made by a broad-based tax system that raises revenue from both households and enterprises. He argues that macroeconomic stabilization would open up domestic capital markets by moving interest rates in the banking system to positive real levels. In McKinnon's view, the accumulated experience of many countries supports the key role of positive real interest rates in improving the efficiency of investment. He argues, however, that raising real interest rates to positive levels is more effective (in terms of growth promotion) if done through price-level stabilization rather than through raising nominal interest rates. An increase in money demand would be an important growth-enhancing consequence of price-level stabilization brought about by fiscal means. If anything, nominal interest rates should continue to be controlled – and if necessary, *reduced* – during stabilization to keep real rates from becoming too high.

Once macroeconomic stability is achieved and domestic capital markets are opened up by the prevalence of positive real interest rates, the *current* account of the balance of payments can be liberalized, and domestic prices can be freed. Exchange rates should be unified before quantitative controls are lifted on exporting and importing, but once this is done, the advocated sequence of trade liberalization is a conventional one; that is, quantitative restrictions should be replaced with "equivalent" tariffs, and these should be reduced gradually and predictably, beginning with the highest tariffs, to a low common tariff rate. McKinnon argues that the exchange rate should not be used as a nominal anchor. Instead, the authorities should set the rate of depreciation equal to the difference between domestic and foreign inflation rates, thereby pegging the real exchange rate.

McKinnon would put decentralization and privatization of the banking system fairly late in the reform process, citing the problem of moral hazard created by deposit insurance in the presence of macroeconomic risk that cannot be diversified away. Moreover, he would implement capital account convertibility as the very last step in economic liberalization. He views capital outflows as harmful because they reduce the base of the inflation tax, thereby increasing the rate of inflation required to finance a given fiscal deficit, and he views capital inflows as potentially harmful as well because they result in real exchange appreciation.

3. Williamson (1991)

Williamson shares the view that capital account liberalization should be undertaken at a very late stage of the reform process. He cites a variety of arguments in the reform sequencing literature for liberalizing capital inflows late:

1. As noted by McKinnon, capital inflows may result in an appreciation of the real exchange rate, which would undermine competitiveness of tradable goods, impairing export-led growth.
2. Permitting capital inflows before fiscal discipline is established may permit the maintenance of unsustainable fiscal deficits.
3. Capital that flows in before trade is liberalized may go into the wrong industries because its allocation will be responding to distorted relative prices.
4. Capital that flows in before the financial system is liberalized may go into the wrong investments.

Thus, according to Williamson, a good time to liberalize inflows is when nontraditional export industries are firmly established, fiscal discipline is in place, and both the import regime and the domestic financial system have been liberalized.

With regard to capital outflows, Williamson argues that the removal of controls on outflows should not be viewed as part of a strategy for fine-tuning the net flow of capital because it is not clear that net inflows will actually decrease (we will consider this issue in Chapter 24). He argues that because of the impossible trinity, controls should not be dismantled until a new policy instrument (presumably fiscal policy) becomes available. He also suggests that this policy should not be implemented until the country has been accepted as a member of the community of market-oriented democratic states that can be relied on to maintain the policy in the face of difficulties and until participation in a multilateral tax information-sharing agreement becomes possible. Finally, Williamson maintains that because liberalizing the capital account essentially privatizes decisions concerning the allocation of foreign investments, a good time to liberalize is when substantial foreign assets have been accumulated. Thus his criteria essentially involve waiting until an open capital account can be considered to be a policy regime that investors regard as permanent, until the ability to manage aggregate demand is enhanced by a measure of fiscal flexibility, and until arrangements are in place to limit the erosion of the tax base through capital outflows.

REVIEW QUESTIONS

1. Describe the difference between solvency and liquidity crises. What is the relationship between these two types of banking crises?
2. Why are banks particularly vulnerable to solvency and liquidity crises compared, say, to nonfinancial firms?
3. Explain how the availability of a lender of last resort or deposit insurance can help mitigate the risk of liquidity crises. What are the pros and cons of these two types of public policies?
4. What potential benefits are available to an economy from liberalizing the capital account of its balance of payments? Why might these benefits not be achieved in practice?
5. Why might the sequencing of macroeconomic and financial reform matter? Make a case for a particular sequence of reform.

EXERCISES

1. Indonesia liberalized the capital account of its balance of payments in 1967, after it had achieved some macroeconomic stability but while its domestic financial system was still repressed. What effects would you expect capital account liberalization to have had in that context?

2. Would it make sense for a country to have a lender of last resort if it has already implemented a deposit insurance scheme? Explain why or why not.
3. What do we mean by the terms *overborrowing* and *underborrowing*? Using diagrams and examples, describe how such situations could come about. Does the emergence of overborrowing or underborrowing justify the maintenance of financial autarky? Explain why or why not.
4. Chile, which is usually taken to be a well-managed developing country, imposed restrictions on capital inflows in the form of unremunerated reserve requirements in the 1990s. Can you provide a rationale for the imposition of such controls that is consistent with the view of Chile as a well-managed economy?
5. The last section of this chapter described the contrasting liberalizing experiences of several Asian miracle economies and a large number of economies in Sub-Saharan Africa. On the basis of the institutional and policy requirements for a well-functioning financial system described earlier in the chapter, can you speculate about the reasons for some of these contrasts?

REFERENCES AND FURTHER READING

Bertolini, Leonardo, and Allan Drazen (1997), "Capital Account Liberalization as a Signal," *American Economic Review*, vol. 87, pp. 138–154.

Caprio, Gerard, Izak Atiyas, and James A. Hanson, eds. (1994), *Financial Reform: Theory and Experience*, Cambridge: Cambridge University Press.

Caprio, Gerard, Patrick Honohan, and Joseph Stiglitz, eds. (2001), *Financial Liberalization: How Far, How Fast?* Cambridge: Cambridge University Press.

Caprio, Gerard, Patrick Honohan, and Dimitri Vittas (2002), *Financial Sector Policy for Developing Countries: A Reader*, Washington, DC: World Bank.

Cardenas, Mauricio S., and Felipe Barrera (1993), "Efectos Macroeconomicos de los Capitales Extranjeros: El Caso Colombiano," working paper 147, Inter-American Development Bank.

Cardoso, Eliana, and Ilan Goldfajn (1997), "Capital Flows to Brazil: The Endogeneity of Capital Controls," working paper WP/97/115, International Monetary Fund.

Cowan, Kevin, and Jose De Gregorio (2005), "International Borrowing, Capital Controls and the Exchange Rate: Lessons from Chile," working paper 322, Central Bank of Chile.

Demigurc-Kunt, Asli, and Enrica Detragiache (1998), "The Determinants of Banking Crises in Developing and Developed Countries," *IMF Staff Papers*, vol. 45, pp. 81–109.

Diaz-Alejandro, Carlos (1985), "Good-Bye Financial Repression, Hello Financial Crash," *Journal of Development Economics*, vol. 19, pp. 1–24.

Dooley, Michael (1995), "A Survey of Academic Literature on Controls over International Capital Transactions," *IMF Staff Papers*, vol. 43, pp. 639–687.

Edwards, Sebastian (1998), "Capital Flows, Real Exchange Rates, and Capital Controls: Some Latin American Experiences," working paper 6800, National Bureau of Economic Research.

Fischer, Bernhard, and Helmut Reisen (1990), *Financial Opening: Why, How, When*, San Francisco: International Center for Economic Growth.

Galbis, Vicente (1993), "High Real Interest Rates under Financial Liberalization: Is There a Problem?" working paper WP/93/7, International Monetary Fund.

Kapur, Ishan, Michael Hadjimichael, Paul Hilbers, Jerald Schiff, and Philippe Szymczak (1991), "Ghana: Adjustment and Growth, 1983–91," occasional paper 86, International Monetary Fund.

King, Robert G., and Ross Levine (1993), "Finance, Entrepreneurship, and Growth: Theory and Evidence," *Journal of Monetary Economics*, vol. 32, pp. 513–542.

Kose, M. Ayhan, Eswar Prasad, Kenneth Rogoff, and Shang-Jin Wei (2006), "Financial Globalization: A Reappraisal," working paper 12484, National Bureau of Economic Research.

Le Fort, Guillermo V., and Carlos Budnevich (1996), "Capital Account Regulations and Macroeconomic Policy: Two Latin American Experiences," in *Studies on International Monetary and Financial Issues for the Group of Twenty-four, Report to the Group of Twenty-four*, unpublished manuscript.

Magud, Nicolas, and Carmen Reinhart (2007), "Capital Controls: An Evaluation," in S. Edwards, ed., *Capital Controls and Capital Flows in Emerging Economies: Policies, Practices, and Consequences*, Chicago: University of Chicago Press, pp. 645–674.

McKinnon, Ronald I. (1992), *The Order of Economic Liberalization*, Baltimore: Johns Hopkins University Press.

Montiel, Peter J., and Carmen M. Reinhart (1999), "Do Capital Controls and Macroeconomic Policies Influence the Volume and Composition of Capital Flows? Evidence from the 1990s," *Journal of International Money and Finance*, vol. 18, pp. 619–635.

Nadal-De Simone, Francisco, and Piritta Sorsa (1999), "A Review of Capital Account Restrictions in Chile in the 1990s," working paper WP/99/52, International Monetary Fund.

Quirk, Peter J., and Owen Evans (1995), "Capital Account Convertibility: Review of Experience and Implications for IMF Policies," occasional paper 131, International Monetary Fund.

Reinhart, Carmen M., and Steven Dunaway (1996), "Dealing with Capital Inflows: Are There Any Lessons?" Research for Action Working Paper 28, United Nations University-World Institute for Development Economics Research.

Tseng, Warda, and Robert Corker (1991), "Financial Liberalization, Money Demand, and Monetary Policy in Asian Countries," occasional paper 84, International Monetary Fund.

Turtleboom, Bart (1991), "Interest Rate Liberalization: Some Lessons from Africa," working paper WP/91/121, International Monetary Fund.

Valdes Prieto, Salvador, and Marcelo Soto (1996), "New Selective Capital Controls in Chile: Are They Effective?" unpublished manuscript, World Bank.

Villanueva, Delano, and Abbas Mirakhor (1990), "Strategies for Financial Reforms," *IMF Staff Papers*, vol. 37, pp. 509–536.

Williamson, John (1991), "On Liberalizing the Capital Account," in Richard O'Brien, ed., *Finance and the International Economy*, vol. 5, New York: Oxford University Press, pp. 136–146.

World Bank (1994), *Adjustment in Africa: Reforms, Results, and the Road Ahead*, New York: Oxford University Press.

23

The Benchmark Model with Banks

Chapter 20 emphasized that financial intermediation is a vital productive activity and is often performed by specialized firms that act as middlemen between borrowers and lenders. As we saw there, commercial banks are among the most important such firms in all countries, even in those where other forms of financial intermediation are well established. But in most developing countries, where securities markets are often rudimentary, commercial banks tend to dominate in performing the financial intermediation function, at least in the formal sector of the economy.

Despite this key role of banks, the analytical model that we developed in Part 2 of this book was based on an economy in which all financial intermediation was conducted through anonymous securities markets rather than through banks. Accordingly, that model analyzed monetary policy and monetary transmission in a context in which there were only two types of domestic financial assets: currency and interest-bearing securities issued by the domestic government. Because there were no banks in the model, there was no role in the model for bank deposits or bank loans. Having explored the *microeconomic* role of banks in the preceding chapters, however, we can now consider a more realistic framework for the domestic financial system, one in which banks play a central role as financial intermediaries in the domestic economy. This chapter will therefore examine how monetary policy works in a liberalized bank-dominated financial system, that is, a financial system in which banks are the dominant form of financial intermediaries but in which the banking system is not repressed so there are no interest rate restrictions or directed credit.

The structure of the chapter is as follows. In the next section, we describe the financial system in our model economy by examining how the balance sheet constraints and asset-demand functions of Chapter 6 have to be modified to allow for the role of bank assets and liabilities as well as by specifically describing the behavior of banks. Section II then demonstrates how equilibrium in this seemingly much more complicated economy can – just as with the simpler framework of

Chapter 6 – be summarized in a single bond-market equilibrium condition. Section III uses this equilibrium condition to analyze how monetary policy works in this environment and considers the additional complexities that this setting introduces into the formulation of monetary policy.

I. THE FINANCIAL SECTOR WITH BANKS[1]

As we did in Chapter 6, we will analyze the domestic financial sector in two steps. First, we will describe the financial balance sheets as well as the asset-demand and supply functions of each of the agents that participate in domestic financial markets. In doing so, as in Chapter 6, we will need to specify how the domestic economy is connected financially to the rest of the world. Our model will contain five different types of economic agents: the domestic nonbank private sector (which includes households and firms, also referred to as the *nonbank public*), commercial banks, the central bank, the government, and the rest of the world. Regarding financial connections between the domestic economy and the rest of the world, we will assume that the domestic private sector and commercial banks both can hold foreign assets and that foreigners can hold claims on the domestic economy, so the economy is financially open. To focus on the most general case, claims on the domestic economy and foreign assets are viewed both by domestic and by foreign residents to be imperfect substitutes for each other.

1. The Nonbank Private Sector

As in Chapter 6, we will denote the financial net worth of the domestic nonbank public as W_P. The nonbank public's financial net worth is the difference between its financial assets and its financial liabilities. In Chapter 6, the nonbank public could hold three assets: money (M), domestic government bonds (B_P), and foreign bonds (SF_P^*), so its financial net worth was given by equation (6.1), reproduced here:

$$W_P = M + B_P + SF_P^* \tag{6.1}$$

Money essentially consisted only of currency, which was a liability of the central bank. Now that we are allowing for the existence of banks, however, we have to make two changes. First, we need to make a distinction between the liabilities of the central bank and the money held by the nonbank public. The liabilities of the central bank, which are referred to as the *monetary base* (M), or as *high-powered money* (for reasons that we will see later), can either be held as currency by the nonbank public (which we will call CUR) or as reserves by the commercial banks, to be discussed further later. The money held by the nonbank public, on the other

[1] More advanced versions of the model developed in this section are presented in Agenor and Montiel (2008a, 2008b).

hand, can consist not only of currency but also of checking accounts (called *demand deposits* because they are available on demand) at commercial banks, denoted D. The sum of currency in the hands of the nonbank public and deposits at commercial banks is usually referred to as *M1*. Second, in contrast to Chapter 6, we will now also allow the nonbank public to incur financial liabilities by taking out loans from the banking system, which we denote L. Notice that the nonbank public's total holding of financial assets can exceed its financial wealth by the amount of those liabilities. With these two changes, the nonbank public's financial wealth is given by

$$W_P = (CUR + D) + B_P + SF_P^* - L \tag{23.1a}$$

The next step is to describe how the nonbank public allocates its financial portfolio among the various assets that it can hold and loans from the banks. To do so, we will first make a simplifying assumption: all the agents in our model consider loans to the nonbank public to be *perfect substitutes* for government bonds. In other words, the liabilities of the nonbank public and those of the government carry the same perceived risk. This is obviously not realistic (and in fact, it could obviate the need for banks as special financial intermediaries), but it will help us isolate exactly how the introduction of banks affects the model of Part 2 by avoiding combining the introduction of banks with that of an entirely new asset. Under this assumption, it is useful to rewrite equation (23.1a) as

$$W_P = (CUR + D) + (B_P - L) + SF_P^* \tag{23.1b}$$

Comparing this with equation (6.1), you can now see that we have merely replaced M by $(CUR + D)$ and B_P by $(B_P - L)$. Notice that because loans to the nonbank public and domestic government bonds are perfect substitutes, they must both carry the same interest rate R.

Currency is typically held for convenience in making small transactions. Because we will not distinguish in our model between large and small transactions, it is reasonable to assume that their relative importance is unchanging. In that case, the ratio of the amount of currency that the nonbank public holds to the amount of money that it holds in the form of demand deposits should also be constant, and we can write the demand for currency as

$$CUR = \gamma D \tag{23.2}$$

where γ is a positive constant.

The demand for bank deposits, on the other hand, can be taken to be determined in the same way that we determined the demand for money in Chapter 6: the real demand for deposits is an increasing function of income and a decreasing function of the rate of return on competing assets. As in Chapter 6, we will assume that the relevant competing asset is the other domestic asset that the nonbank public can hold. The competing asset in this case is domestic government bonds and bank loans because the alternative to holding money is to acquire government bonds or pay back loans, and the nonbank public is indifferent between the two. Thus the opportunity cost of deposits is the nominal interest rate on domestic government bonds, which we have been calling R. Another difference compared to the demand for money in Chapter 6 is that, unlike the cash to which we were referring when considering the demand for money in that chapter, deposits at banks typically *do* pay interest, and we would expect the demand for deposits to be an increasing function of the interest rate paid on them. Denoting the deposit interest rate as R_D and the *real* demand for deposits as d, we can therefore write the nominal demand for deposits as

$$D = Pd(\underset{+}{R_D}, \underset{-}{R}, \underset{+}{Y}) \tag{23.3}$$

Turning next to the demand for foreign bonds, we will describe it exactly as we did in Chapter 6: the demand for foreign bonds is a fraction of nonmonetary wealth that depends negatively on the differential between the rates of return on domestic bonds and loans and foreign bonds:

$$SF_P^* = f(R \underset{-}{-} R^*)[W_P - (CUR + D)] \tag{23.4}$$

Because loans and government bonds are perfect substitutes, there is no separate demand for either of them on the part of the nonbank public. Instead, the nonbank public has a demand for the net asset $(B_P - L)$. This demand is a residual after the nonbank public has satisfied its other asset demands; that is,

$$B_P - L = W_P - (CUR + D) - SF_P^*$$

Using equation (23.4), this becomes

$$\begin{aligned} B_P - L &= [W_P - (CUR + D)] - f(R, R^*)[W_P - (CUR + D)] \\ &= (1 - f(R - R^*))[W_P - (CUR + D)] \\ &= b(R - R^*)[W_P - (CUR + D)] \end{aligned}$$

where $b(R - R^*) = 1 - f(R - R^*)$. Finally, using equations (23.2) and (23.3) to substitute for CUR and D, we can write

$$B_P - L = b(R \underset{+}{-} R^*)(W_P - (1 + \gamma)Pd(\underset{+}{R_D}, \underset{-}{R}, \underset{+}{Y}) \tag{23.5}$$

It may be worth pausing to call attention to several properties of this function:

- An increase in the loan interest rate R increases the nonbank public's demand for government bonds net of loans for two reasons: (1) when the domestic interest rate R increases, the nonbank public reduces its holdings of foreign bonds to engage in some combination of buying domestic government bonds and paying back some of its bank loans and (2) an increase in the domestic interest rate also induces the nonbank public to reduce its demand for money to buy bonds and repay loans.

- Increases in the deposit rate or domestic real gross domestic product (GDP) have the opposite effect as an increase in the bond interest rate. They reduce the nonbank public's demand for bonds and increase its demand for loans by inducing it to hold more money. Because the nonbank public's wealth is predetermined, it can finance larger holdings of money only by drawing down its stock of other assets or incurring more debt with the banking system.
- Finally, increases in the foreign interest rate R^* or in the amount of currency that the nonbank public wishes to hold for each unit of deposits (γ) reduces its demand for domestic bonds and increases its demand for loans because to hold more foreign bonds or more domestic currency with an unchanged level of wealth, it is forced to sell government bonds and/or incur more debt.

2. Commercial Banks

Next, consider the role of banks. Banks accept deposits from the nonbank public, and they either hold those resources in the form of reserves (RES) or acquire interest-bearing assets, which can consist of domestic government bonds, loans to the nonbank public, or foreign bonds. We will use the symbol B_B to denote banks' stock of government bonds and F_B^* to denote the foreign-currency value of their holdings of foreign bonds. We will assume that banks have no net worth. That means that we can write the banking sector's balance sheet as

$$B_B + L + SF_B^* + RES = D \tag{23.6}$$

Because banks hold reserves to cope with the possibility of unexpected deposit withdrawals, we will assume that reserve holdings are a constant fraction rr of the deposits that banks have outstanding:

$$RES = rrD \tag{23.7}$$

Using equations (23.6) and (23.3), this means that banks' demand for interest-bearing assets can be expressed as

$$B_B + L + SF_B^* = (1 - rr)Pd(R_D, R, Y) \tag{23.8}$$

To keep things simple, we will assume that banks allocate their portfolios among interest-bearing assets in the same way that the nonbank public does. Thus their demand for domestic bonds and loans is given by

$$B_B + L = b(R - R^*)(1 - rr)Pd(R_D, R, Y) \tag{23.9}$$

leaving $(1 - b)(1 - rr)Pd$ to be allocated to foreign bonds. Because government bonds and loans to the nonbank public are perfect substitutes, banks are indifferent how they allocate their portfolios between B_B and L.

We will assume that banks are perfect competitors, and as such, they make zero profits. For this to be so, because for every dollar of deposits that they collect, they can only invest $(1 - rr)$ dollars in government bonds, loans to the nonbank public, or foreign bonds, the relationship between their lending and deposit interest rates must be given by

$$(1 - rr)(bR + (1 - b)R^*) = R_D \tag{23.10}$$

The bank earns a rate of return $bR + (1 - b)R^*$ on its portfolio, a weighted average of the interest rate on domestic bonds and that on foreign bonds, where the weight b is the share of the bank's portfolio invested in domestic bonds and loans. Notice that if the rate of return on domestic bonds exceeds that on foreign bonds, that is, $R > R^*$, an increase in R must increase the deposit rate, both directly and indirectly (the latter by increasing the share of the higher-interest asset in the bank's portfolio). This raises an interesting issue. Substituting equation (23.10) in the nonbank public's demand for deposits (equation (23.3)), we can write that equation as

$$D = Pd((1 - rr)(bR + (\underset{+}{1} - b)R^*, \underset{-}{R}, \underset{+}{Y})$$

Notice that an increase in the domestic interest rate R has conflicting effects on the demand for deposits: while it increases the opportunity cost of deposits, which reduces the demand for deposits, it may simultaneously increase the deposit rate (and would definitely do so if $R > R^*$), which would *increase* the demand for deposits. Which effect dominates? Again, to stay as close as possible to our analysis of Chapter 6, we will assume for now that the opportunity cost effect dominates (but see Section III); that is, an increase in the interest rate R *reduces* the demand for deposits. Under this assumption, an increase in the domestic interest will reduce the demand for money, as we assumed in Chapter 6. This being so, we can write the demand for bank deposits as

$$D = Pd(\underset{-}{R}, \underset{+}{Y}, \underset{-}{rr}; \ldots) \tag{23.11}$$

3. The Central Bank

The balance sheet of the central bank can be described as in Chapter 6, with the added wrinkle that we have already encountered: in the presence of commercial banks, the liabilities of the central bank (the *monetary base*), which we called M in Chapter 6, do not all circulate as currency in the hands of the public. Some of that currency is instead held in commercial bank vaults as reserves or is deposited (again as reserves) by commercial banks at the central bank. From the perspective of our model, it makes no difference whether commercial banks hold currency as vault cash or deposit it at the central bank, so we shall make no distinction between these two ways for banks to keep their reserves. We can therefore write the central bank's balance sheet as

$$\begin{aligned} SF_C^* + B_C &= M \\ &= RES + CUR \end{aligned} \tag{23.12}$$

As in Chapter 6, the central bank does not make portfolio decisions but instead makes decisions about exchange rate and monetary policies.

4. The Government

We assume that the government's financial net worth (W_G) is equal to the negative of the value of the stock of government bonds outstanding, B:

$$W_G = -B \tag{23.13}$$

As in Chapter 6, the government has no portfolio allocation decisions to make.

5. The Rest of the World

The final sector in our model is the rest of the world. Recalling that the domestic-currency value of the domestic economy's international investment position is denoted $S \cdot IIP^*$ and that foreign holdings of domestic bonds are given by B_F, we can write the rest of the world's balance sheet exactly as in Chapter 6, except that we now account for commercial bank holdings of foreign assets:

$$-S \cdot IIP^* = B_F - S(F_C^* + F_P^* + F_B^*) \tag{23.14}$$

Notice that we are assuming here that, although the rest of the world holds domestic government bonds, it does not make loans directly to the domestic private sector but only through domestic banks.

Finally, we need to specify the rest of the world's demand for domestic bonds. For present purposes, all we need is to determine how the domestic interest rate affects the rest of the world's demand for domestic bonds. We will write the foreign demand for domestic bonds as

$$B_F = B_F^D(\underset{+}{R - R^*}, \ldots) \tag{23.15}$$

where the ellipses represent all the other factors that may affect the foreign demand for domestic bonds.

II. FINANCIAL-MARKET EQUILIBRIUM

The financial-market model described in the last section contains four distinct assets: monetary base, domestic bonds and loans, bank deposits, and foreign bonds.

In financial-market equilibrium, the demand for each of these assets must equal their supply. To see how financial-market equilibrium is determined, it is useful to think of each of the five sectors in the model (the domestic nonbank public, domestic banks, the central bank, the government, and the rest of the world) as *demanding* assets to hold on its own account and *supplying* liabilities as assets for others to hold. Using superscripts to indicate asset demands and supplies, we can rewrite the five sectoral balance sheets as

$$W_P = CUR^D + D^D + SF_P^{*D} + \left(B_P^D - L^S\right) \tag{23.16}$$

$$0 = (B_B + L)^D + SF_B^{*D} + RES^D - D^S \tag{23.17}$$

$$0 = SF_C^{*D} + B_C^D - RES^S - CUR^S \tag{23.18}$$

$$-S \cdot \mathrm{IIP}^* = B_F^D - S(F_P^* + F_C^* + F_B^*)^S \tag{23.19}$$

$$W_G = -B^S \tag{23.20}$$

Adding these together, recalling that $W_P + W_G = S \cdot IIP^*$, and rearranging, we have

$$\begin{aligned} 0 = {} & [CUR^D + RES^D - CUR^S - RES^S] \\ & + \left[B_P^D + B_B^D + B_C^D + B_F^D - B^S + L^D - L^S\right] + [D^D - D^S] \\ & + S\left[F_P^{*D} + F_C^{*D} + F_B^{*D} - (F_P^* + F_C^* + F_B^*)^S\right] \end{aligned}$$

This equation simply states that the sum of the excess demands for the four types of assets must always equal zero. In other words, the four asset markets are not independent. If any three of them clear (and therefore have excess demands equal to zero), the fourth one must clear as well (i.e., its excess demand must also be zero).

But for two of the assets in our model, there can be no excess demand because the supply of those assets is perfectly elastic. This is true of the supply of bank deposits because banks simply set the deposit rate and accept all deposits that the private sector wants to create. It is also true of foreign bonds: because the domestic economy is small, it faces a perfectly elastic supply of foreign bonds at the interest rate R^*. This means that we must always have $[D^D - D^D] = S[F_P^{*D} + F_C^{*D} + F_B^{*D} - (F_P^* + F_C^* + F_B^*)^S] = 0$. This being so, the previous equation becomes

$$\begin{aligned} 0 = {} & [(CUR^D + RES^D) - (CUR^S + RES^S)] \\ & + \left[\left(B_P^D + B_B^D + B_C^D + B_F^D + L^D\right) - (B^S + L^S)\right] \end{aligned}$$

The importance of this result is that our model only contains *one* independent asset-market equilibrium condition: if the market for domestic bonds and loans clears, then the market for the monetary base must do so as well. We therefore can investigate the determination of domestic financial-market equilibrium by looking at either of the two markets.

To facilitate comparison with Chapter 6, we will focus on equilibrium in the market for domestic bonds and loans. We can write the equilibrium condition for this market in the form

$$B^S + \left(L^S - B_P^D\right) = \left(B_B^D + L^D\right) + B_F^D + B_C^D$$

The left-hand side of this equation is the net supply of domestic bonds and loans by the government and the nonbank public together. The first term on the right-hand side is the demand for bonds and loans by the commercial banks, the second term is the demand for domestic bonds by foreigners, and the last term is the demand by the central bank. Alternatively, substituting from equations (23.5), (23.8), and (23.14), we can write the equilibrium condition in the bond and loan market as

$$\begin{aligned} B - B_C = {} & b(R - R^*)(W_P - (1+\gamma)Pd(R, Y, rr) \\ & + b(R - R^*)(1 - rr)Pd(R, Y, rr) + B_F^D(R - R^*, \ldots) \end{aligned} \tag{23.21}$$

Recall from Chapter 6 that the economy's level of real output Y depends on the domestic price level through the short-run aggregate supply function. Holding P (and therefore Y) constant, this equation therefore contains a single endogenous variable: the domestic interest rate R. Our primary interest in this chapter is to see how the domestic interest rate is affected by changes in the central bank's monetary policy variable, B_C.

As in Chapter 6, we can proceed graphically. The left-hand side of equation (23.21) is the supply side of the market for domestic bonds. It represents the supply of domestic government bonds available for the domestic nonbank public, domestic banks, and foreigners to hold. The supply of bonds is equal to the cumulative amount that the government has issued, B, minus the stock of bonds held by the central bank, B_C. The latter is the key monetary policy variable in the model. Because the supply of bonds does not depend on the domestic interest rate, we can depict it as the vertical line B^S in Figure 23.1.

The first two terms on the right-hand side, in turn, represent, respectively, the demand for government bonds and the supply of loans by the domestic private sector as well as the demand for government bonds and loans from banks. Because the supply of loans from the private sector and the demand for such loans cancel out, the sum of the first two terms on the right-hand side is equal to the domestic demand for government bonds. The third term is the demand for such bonds from the rest of the world.

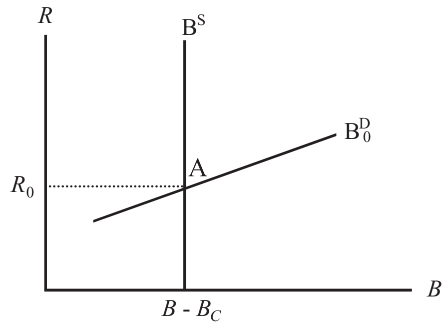


Figure 23.1. Equilibrium in the domestic bond market

Now consider the effect of an increase in the interest rate on domestic government bonds on the domestic demand for such bonds. As R increases, there are several effects:

- For a given amount of nonmonetary wealth, the nonbank public switches from foreign bonds to domestic bonds and debt repayment, thereby increasing demand for the latter.
- The nonbank public switches from deposits and currency to nonmonetary wealth and allocates a fraction b of its additional nonmonetary wealth to an increased demand for domestic bonds and debt repayment.
- As the nonbank public reduces its demand for deposits, banks have fewer resources to lend and therefore *decrease* their demand for domestic bonds and private debt.
- Foreigners switch from their own bonds to domestic ones, increasing the demand for domestic bonds.

Of these four effects, all but the third result in an increased demand for domestic bonds as the interest rate on such bonds increases. To see what the net result of these effects must be, notice that the second and third effects both arise from the same source: a switch from deposits and currency to interest-bearing assets by the nonbank public. Suppose that as the result of the increase in the domestic interest rate, the nonbank public reduces its real demand for bank deposits, say, by d_R. Consequently, it reduces its total nominal demand for money by $(1+\gamma)Pd_R$ and uses the resources freed up to hold domestic government bonds, repay bank loans, and hold more foreign assets. Its increased demand for domestic government bonds and loan repayment is given by $b(1+\gamma)Pd_R$. This is the second of the effects listed earlier. At the same time, because the nonbank public holds fewer deposits, banks have fewer resources to lend. This is the third effect. They reduce their demands for government bonds and supply of loans to the private sector by $b(1-rr)Pd_R$.

The total effect on the demand for domestic bonds arising from these two effects together is therefore given by

$$[b(1+\gamma) - b(1-rr)]Pd_R$$

Notice that this expression must be positive but smaller than the nonbank public's increased demand for domestic bonds and loan repayment arising from its reduced demand for deposits. Why is that?

To answer this, assume initially for simplicity that $\gamma = 0$. In this case, all money is held in the form of deposits at banks. When the nonbank public retains an additional \$1 of wealth in nonmonetary form, it allocates \$$b$ to domestic bonds and debt reduction and \$$(1 - b)$ to holding foreign bonds. When it holds that \$1 in monetary form, though, the decision about how that \$1 will be allocated is made by banks. Bank allocate \$$b(1 - rr)$ to domestic bonds and loans, \$$(1 - b)(1 - rr)$ to foreign bonds, and \$$rr$ to cash. The total demand for bonds *falls* in that case because banks divert some of that demand into demand for cash. Because this outcome is reversed when market interest rates rise, causing the nonbank public to hold *less* of its wealth in the form of bank deposits, the total demand for domestic bonds rises, but by less than the demand by the nonbank public, as the result of the partial offset in the form of reduced bank demand.

As you can see, the effect of $\gamma > 0$ is to *increase* the marginal effect of a change in the nonbank public's demand for deposits on the private demand for bonds and debt reduction, while leaving the marginal effect of a change in their deposit base on banks' demand for bonds and loans unchanged. This *magnifies* the net effect of a change in the demand for deposits on the demand for bonds. Intuitively, the reason is that when $\gamma > 0$, a reduction in the nonbank public's demand for deposits is supplemented by a reduction in its demand for currency, freeing up that many more resources for holding bonds and paying down debt.

This reasoning implies that an increase in the interest rate on domestic bonds must increase the demand for bonds, as in Chapter 6. Thus the demand-for-bonds curve, labeled B^D in Figure 23.1, must be upward sloping.

III. MONETARY POLICY WITH BANKS

How does monetary policy work in this setting? The answer is that, under the assumptions we have made so far, it actually works very similarly to what we saw in Chapter 6.[2] That is why it did not do much violence to reality to develop the model of Part 2 in the simpler context in which all financial intermediation was conducted

[2] The transmission of monetary policy in the presence of banks becomes more complicated under more realistic assumptions and can be strongly influenced by the structure of the domestic financial system. For discussions in the context of advanced economies, emerging economies, and developing countries, respectively, see Gerlach and Smets (1995), Kamin et al. (1998), and Mishra et al. (2010).

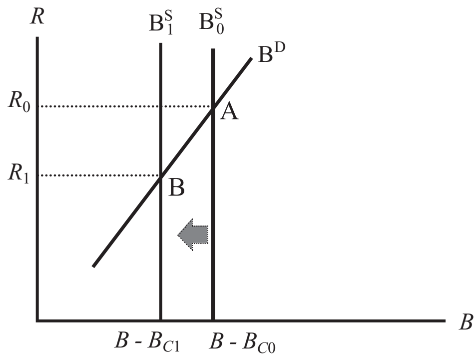


Figure 23.2. Expansionary monetary policy

through securities markets, even though that setting is particularly unrealistic in the context of emerging and developing economies. However, it is worth considering what happens when some of the assumptions made earlier do not hold. We will do so in this section. We will also consider the effects of certain shocks that had no counterpart in the earlier version of our model without banks.

1. Monetary Policy in the Standard Case

To examine the effects of monetary policy, let us take the concrete case of a monetary expansion, in the form of an increase in central-bank holdings of domestic bonds, B_C. You can think of this as happening in a context in which the government holds periodic auctions of bonds and the central bank simply enters the auctions as a buyer, along with banks and possibly individuals. Alternatively, you can simply think of it as the central bank buying bonds in a secondary market. In either case, the effect of such a policy is straightforward: it simply reduces the stock of domestic bonds that the public (nonbank, bank, and foreign) has to be induced to hold, thereby shifting the bond supply curve to the left in Figure 23.2, to a position such as B_1^S. Because the public's bond demand curve is unaffected by this action, to clear the domestic bond market, the equilibrium interest rate on domestic bonds must fall, moving the bond market from a point such as A to one such as B.

Consider what else the policy implies. Because the domestic interest rate has fallen, the private sector's demand for bank deposits must increase. Because $M1 = (1 + \gamma)d$, the money supply increases as well. The lower interest rate on domestic bonds and loans induces the nonbank public to switch out of domestic bonds and into not just bank deposits but also foreign bonds. Banks switch out of domestic bonds and loans and into foreign bonds. Finally, foreigners also switch out of domestic bonds and into their own bonds. Because all three types of agents switch into foreign bonds, the policy is therefore associated with a capital outflow. But the change in banks' demand for domestic bonds and loans is ambiguous. On one hand, their switch from domestic bonds and loans to foreign bonds reduces their demand for domestic bonds and loans, but on the other hand, they undergo an increase in the size of their deposit base as the result of increased household deposit demand, which increases their demand for domestic bonds and loans.

2. Relaxing Some Simplifying Assumptions

Two assumptions that we have made so far ensured that the bond demand curve had a positive slope: that the demand for bank deposits was inversely related to market interest rates and that banks allocated their assets between bonds and loans and foreign bonds in the same manner that the nonbank public allocated its nonmonetary wealth. Neither of these assumptions is necessary for the bond demand curve to have a positive slope, but together, they are sufficient to ensure that it does so.

Actually, we are on pretty safe ground in assuming that the bond demand curve has a positive slope – and therefore that monetary policy would work as described in our model with banks – because a positive slope is required for the bond market to be well behaved in a very specific sense. It is sensible to assume, for instance, that an excess demand for bonds would drive the bond interest rate down because when bonds are in short supply, prospective buyers of bonds should be willing to compete to acquire them by being willing to accept lower interest rates on the bonds they acquire. Similarly, when there is an excess supply of bonds, the market interest rate should rise because the government will only be able to induce more lenders to hold its bonds by offering them more attractive rates of return. But if this is so, then for the bond market to be stable – that is, for it to return to equilibrium after the bond supply and/or demand curves shift – an above-equilibrium interest rate should be associated with an excess demand for bonds and a below-equilibrium rate with an excess supply. As you can easily verify, though, because the bond supply curve is vertical, for this to be true, the bond demand curve B^D must indeed have a positive slope.

This means that we can dispense with the two assumptions mentioned earlier and still analyze the bond market as we did previously. What would happen if these assumptions did not hold?

a. Market Interest Rates and the Demand for Deposits

The assumption that the demand for deposits is inversely related to market interest rates made the demand for money behave in conventional fashion in the model with banks. The assumption is a reasonable one, but as we have just seen, it is not necessary for the bond market to be well behaved. The assumption implies that when the interest rate on domestic bonds rises, the higher returns on domestic bonds and the higher cost of loans induce the nonbank public to switch into domestic bonds and debt repayment and away from both bank deposits and foreign assets. But it could have been otherwise. Because banks increase the deposit interest rate when they can earn a higher return on their assets, an increase in the interest rate on domestic bonds and loans could, for example, cause the nonbank public to move funds out of foreign bonds and into both domestic bond and debt repayment and bank deposits. In this case, the demand for bank deposits would be an *increasing* function of the market interest rate. For this to be true, bank deposits would have to be relatively more sensitive to their own interest rates and/or less sensitive to the interest rate on competing assets than we assumed previously.

What would this imply for the preceding bond-market analysis? In that case, though the increased supply of deposits to banks would increase their demand for bonds, this would be more than offset by the reduced demand for bonds in favor of deposits by the nonbank private sector, for the reasons explained in the preceding section. This would partially offset the shift into domestic bonds and out of foreign bonds by the domestic nonbank sector, domestic banks, and the rest of the world. The upshot would be that the demand for bonds would be *less* sensitive to the interest rate than we previously assumed. This means that the B^D curve would be *steeper*, and consequently, that the central bank's monetary expansion would have a *larger* effect on the equilibrium interest rate. In other words, under these conditions, the impact of monetary policy, in the form of a given-sized change in B_C on the market interest rate, would be larger. Thus our assumption that the demand for bank deposits falls when market interest rates rise can now be seen as weakening the effects of monetary policy on the equilibrium real interest rate by making the demand for bonds more elastic with respect to changes in the bond interest rate.

b. Bank Portfolio Allocation

We assumed earlier that banks make portfolio choices between domestic bonds and loans, on one hand, and foreign bonds, on the other, in the same way that the nonbank public does. But this, of course, need not be true: the shares of bank portfolios devoted to the two kinds of assets may be quite different from those chosen by the nonbank private sector, and in principle, banks could consider domestic bonds and loans either closer or less close substitutes for foreign bonds than does the nonbank public.

Suppose, for example, that bank holdings of foreign bonds can be positive or negative, whereas private-sector holdings can only be positive; that is, banks have access to foreign borrowing at the world interest rate, while the nonbank public does not. In that case, if banks are net borrowers from the rest of the world, their holdings of domestic bonds and loans may actually *exceed* their deposit base because they fund their acquisition of bonds and loans partly by attracting deposits and partly by borrowing abroad. Consider now the effects of an increase in domestic interest rates and assume once again that higher interest rates reduce the demand for bank deposits by the nonbank public. What will happen to the demand for domestic bonds and loans by banks? The answer depends on how banks' foreign borrowing responds. The higher interest rate on domestic bonds and loans gives banks an incentive to increase their ratio of foreign borrowing to deposits, but if that increase is small, because their deposit base contracts, they may actually *reduce* their demand for bonds and loans. Conversely, if the ratio of foreign borrowing to deposits expands sufficiently, then their demand for bonds and loans could actually *increase*.[3] Again, the shape of the B^D curve would be affected, with the curve being flatter the more banks expand their foreign borrowing in response to an increase in domestic interest rates. This means that the effects of monetary policy actions on the domestic interest rate are weaker the greater the responsiveness of banks' foreign borrowing to changes in the interest rate on domestic bonds.

The moral of the story is that in the presence of banks, the shape of the B^D curve depends both on the properties of the nonbank public's demand for bank deposits and on how banks manage their portfolios of domestic bonds, loans, and foreign assets. To achieve a specific target for the domestic interest rate or the money supply, in a bank-inclusive world, the central bank needs to be informed about these important aspects of private-sector behavior.

IV. FINANCIAL-MARKET SHOCKS

There are two other aspects of the portfolio decisions of the nonbank public and banks that we have not yet considered but that also have important implications for monetary policy: the nonbank public's demand for cash and banks' demand for reserves. These depend, respectively, on the parameters γ and rr. As we will see later, changes in these parameters affect the position of the B^D curve and therefore the equilibrium value of the domestic interest rate for a given value of the central bank's holdings of domestic government bonds B^C.

Going back to equation (23.21), it is easy to see how a change in the nonbank public's desired currency-deposit ratio γ would affect its demand for bonds. An exogenous increase in γ would cause the nonbank public to increase its total demand for money, other things equal. Because its total financial wealth is fixed, an increase in its demand for money implies a reduction in its demand for bond and debt repayment as well as for foreign bonds. The implication is that the bond demand curve would shift to the left, as in Figure 23.3, and the equilibrium value of the domestic interest rate would rise.

The effects of an increase in banks' reserve ratio rr are a little more complicated to describe. First, for a given value of the domestic bond and loan interest rate, an increase in the reserve ratio would induce banks to lower their deposit rates because deposits become less profitable when a larger share of each \$1 collected in deposits is held in non-interest-bearing form. This reduced demand for deposits implies a reduced demand for money on the part of the nonbank public and a

[3] The ratio of foreign loans to deposits would increase slightly if foreign loans and domestic bonds and loans are considered by banks to be poor substitutes and would expand significantly if they were considered to be close substitutes.

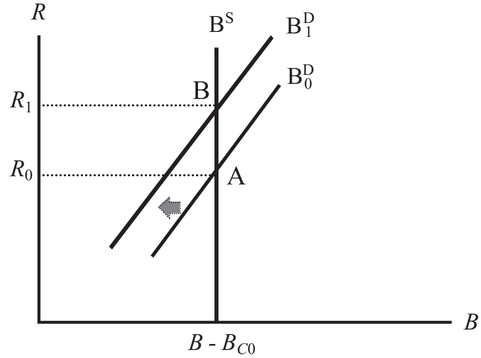


Figure 23.3. Bond-market effects of an increase in γ

larger stock of nonmonetary wealth for it to allocate to bond and loan repayment and foreign bonds. Because part of this increase in the nonmonetary component of wealth would be held in the form of domestic bonds, this effect induces an increase in the demand for bonds by the nonbank public.

At the same time, however, the contraction in the banks' deposit base causes them to curtail their demand for bonds and loans. If banks allocate their portfolios in the same way as the nonbank public does, this partly offsets the increased demand for bonds by the nonbank public. However, the offset is only partial, as we have seen, because the nonbank public's nonmonetary wealth *increases* by $\$(1 + \gamma)$ for each \$1 decline in its demand for deposits, while the banks' investable portfolio *decreases* by only $\$(1 - rr) < \$(1 + \gamma)$. The total size of this net positive effect on the demand for domestic bonds depends on how sensitive the demand for bank deposits is to the deposits' own rate of return. The effect will be larger the more sensitive the demand for deposits is to the deposit rate.

Offsetting this positive effect on bond demand, however, is a negative effect arising from the diversion of bank resources into holding reserves rather than bonds and loans or foreign assets. To the extent that banks contract their holdings of domestic bonds as the result of their increased demand for reserves, the demand for domestic bonds must fall.

The net result of these two effects depends on how sensitive the demand for bank deposits is to changes in the deposit rate. If the demand for deposits is not very sensitive to the deposit rate, the first (positive) effect on bond demand will be weak, and the second (negative) effect will dominate, causing an increase in banks' reserve ratios to shift the bond demand curve to the left, as in the case of the increase in the nonbank public's demand for currency.

It is worth noting that the two shocks just described are of much more than just academic interest. Unpredictable shifts in the currency-deposit ratio or the reserve ratio substantially complicate matters, for example, for central banks that are conducting monetary policy by targeting a monetary aggregate. The reason is that, holding B^C constant, such shocks affect more than one determinant of the money supply. More important, as we will see in Chapter 26, banking crises tend to induce switches into currency and away from deposits on the part of the nonbank public as well as into reserves and away from interest-bearing assets on the part of banks. As we have just shown, such shocks are likely to increase the equilibrium value of the domestic interest rate in the absence of a central-bank response.

V. SUMMARY

In this chapter, we have investigated how to incorporate commercial banks as financial intermediaries into the benchmark macroeconomic model that we developed in Part 2 of this book. We saw that under fairly general and intuitive assumptions, the analysis of financial-market equilibrium was quite similar to that conducted in the simpler framework of Chapter 5. However, the information requirements imposed on the central bank are more demanding because it has to understand not just how the nonbank public reacts to policy changes but how banks behave as well. There are also sources of shocks (specifically, shifts in the demands for currency and bank reserves) in this more general form of the model that were not present in Chapter 6. These shocks tend to be particularly important in the context of banking crises, as we will see in Chapter 26.

Although we did not rederive the BB curve and show how it interacts with the economy's GM curve to determine short-run macroeconomic equilibrium, this is straightforward to do once one recognizes that, under the second of our two standard assumptions, an increase in the domestic price level shifts the B^D curve to the left because it increases the nonbank public's demand for deposits, thus decreasing the aggregate demand for domestic bonds. The implication is that an increase in the domestic price level must be associated with an increase in the value of the interest rate required to clear the domestic financial markets. This positive association between the price level and the interest rate traces out a BB curve that is very similar to the one we derived in Chapter 6. Because the GM curve remains unchanged, short-run macroeconomic equilibrium can be analyzed very similarly to how it was done with the simpler financial structure of Chapter 6.

REVIEW QUESTIONS

1. What is the effect of the introduction of banks on the menu of financial assets available for the nonbank private sector to hold?
2. What effect would you expect an increase in the interest rate on domestic government bonds to have on the nonbank private sector's demand for bank deposits?

3. If the demand for bank deposits on the part of the nonbank private sector is an increasing function of the interest rate on government bonds, an increase in that interest rate would reduce the amount of resources that commercial banks have available to invest in government bonds. That being so, how is it possible that an increase in the interest rate on government bonds would actually increase the demand for such bonds?
4. Our financial model with banks contained four distinct financial assets: base money (currency plus reserves), bank deposits, domestic government bonds and bank loans, and foreign bonds. How, then, was it possible for us to analyze domestic financial-market equilibrium by looking at the market just for government bonds and bank loans?
5. Explain the effects of increases in the nonbank public's desired ratio of currency to deposits and in banks' desired reserve ratio on the equilibrium interest rate in the bond and loan market, holding the price level constant.

EXERCISES

1. Use the bond-market equilibrium diagram with banks to show how the BB curve can be derived in the bank-dominated economy.
2. Suppose that in response to a financial panic, an economy's nonbank public increases its demand for currency, while the economy's commercial banks increase their demand for reserves. Using GM and BB curves, explain what effect you would expect a financial panic to have on the domestic interest rate and real GDP in an economy that suffers such a panic.
3. You are the governor of the central bank of a country that maintains an inflation target and implements it with a Taylor rule. Commercial banks are the main financial intermediary in your country, but there are also well-functioning markets for domestic government bonds. What instruments would you have at your disposal to implement the Taylor rule, and how would you manage those instruments in response to shocks to the demand for bonds that, in the absence of central-bank action, would affect the domestic price level or real GDP only through their effects on the domestic interest rate?
4. Using the bond and loan market equilibrium diagram, show how the central bank should respond to a generalized bank panic that increases the economy's currency-deposit ratio if the central bank wants to prevent the panic from influencing real *GDP*.
5. How would the financial-market equilibrium in our model be affected if only banks – and not the nonbank private sector – are able to hold foreign assets? How would this change in our model affect the impact that a given change in monetary policy would have on the equilibrium value of the domestic interest rate, holding the price level constant?

REFERENCES AND FURTHER READING

Agenor, Pierre-Richard, and Peter J. Montiel (2008a), *Development Macroeconomics*, 3rd ed., Princeton, NJ: Princeton University Press.

Agenor, Pierre-Richard, and Peter J. Montiel (2008b), "Monetary Policy Analysis in a Small Open Credit-Based Economy," *Open Economies Review*, vol. 19, pp. 423–465.

Gerlach, Stefan, and Frank Smets (1995), "The Monetary Transmission Mechanism: Evidence from the G-7 Countries," discussion paper 1219, Centre for Economic Policy Research.

Kamin, S., P. Turner, and J. Van't dack (1998), "The Transmission Mechanism of Monetary Policy in Emerging Market Economies: An Overview," in *The Transmission of Monetary Policy in Emerging Market Economies*, Bank for International Settlements policy paper 3, BIS, pp. 5–65.

Mishra, Prachi, Peter Montiel, and Antonio Spilimbergo (2010), "Monetary Transmission in Low-Income Countries," working paper WP/10/223 International Monetary Fund.

Morsink, James, and Tamim Bayouimi (2001), "A Peek inside the Black Box: The Monetary Transmission Mechanism in Japan," *IMF Staff Papers*, vol. 48, pp. 22–56.

24

Coping with Capital Inflows

As indicated in Chapter 22, many emerging and developing economies have embarked on the road to financial reform during the past two decades. We also saw in that chapter that an important component of such reform has been the liberalization of the capital account of the balance of payments. Restrictions on capital movements have been removed or greatly weakened in many of these economies, in a process that gathered force around the world during the late 1980s and early 1990s. This process of financial reform and capital account liberalization was part of a much broader reorientation of economic policies among emerging economies toward a much more market-friendly disposition. In the international financial arena, this switch in policy regimes marked the transition from the debt crisis conditions that many emerging and developing economies experienced during the 1980s (see Chapter 25) to a situation in which several emerging economies became the recipients of large amounts of private capital from the rest of the world, surpassing in magnitude the flows associated with the 1974–1981 inflow episode that preceded the international debt crisis of the early 1980s. Capital inflows grew rapidly during the decade of the 1990s but stabilized in the wake of the succession of financial crises that afflicted emerging economies during the late 1990s and early 2000s, before accelerating once again after 2002 and collapsing during the international financial crisis of 2008. Figure 24.1 shows the evolution of capital flows to emerging and developing economies over the entire 1970–2008 period.

Ironically, just as the scarcity of external funding posed serious macroeconomic problems for the heavily indebted emerging economies during the 1980s, the large inflows of private capital that many of these economies began to receive during the first half of the 1990s, and that they began to receive once again after 2002, were also viewed as a policy problem. Indeed, the capital inflow "problem" became a serious concern for economists and emerging-economy policy makers during the first half of the decade of the 1990s and again after 2002.

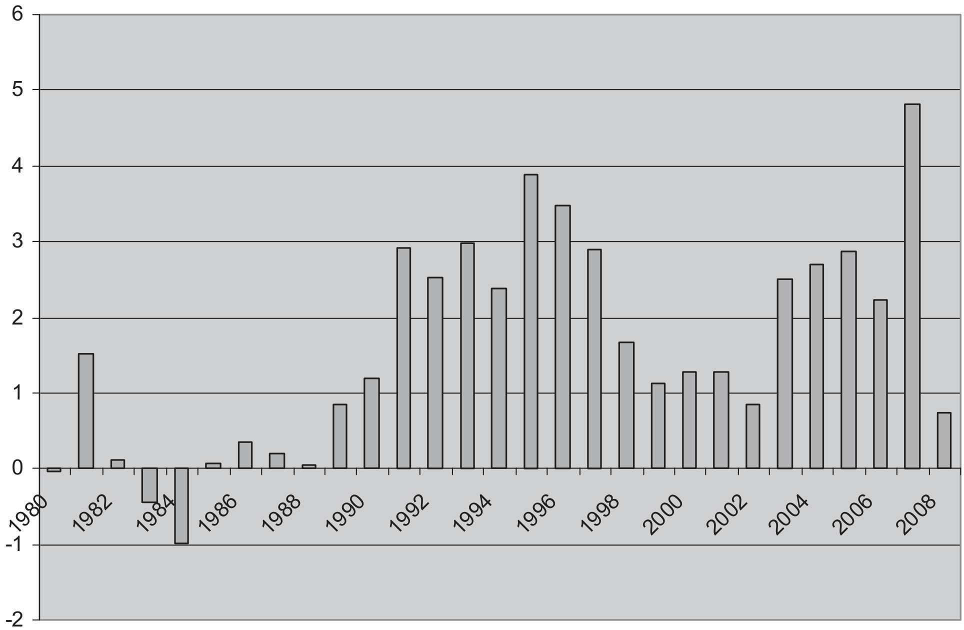


Figure 24.1. Total capital inflows to emerging and developing economies, 1980–2008 (as a percentage of GDP)

Section I of this chapter analyzes the factors driving such capital flows during these two episodes.[1] Their implications for economic welfare in the recipient countries are considered from a microeconomic perspective in Section II. Section III examines the effects of these flows on the domestic macroeconomic equilibrium in the recipient countries as well as the possible policy responses available to them. Country experiences with the management of capital inflows will be reviewed, focusing on comparisons of policy responses in Latin America, as well as in East and Southeast Asia during the inflow episode of the early 1990s, to draw lessons about appropriate strategies for the macroeconomic management of capital inflows in a newly liberalized environment. It is notable that each of the three last major inflow episodes (including that of 1974–1981) has been followed by major international financial crises. Accordingly, Section IV examines the implications of capital inflow episodes for the vulnerability of the recipient countries to financial crises, which are the subject of the next part of the book.

I. FACTORS DRIVING CAPITAL INFLOWS

As Figure 24.1 illustrates, the history of capital flows to developing countries has tended to be an episodic one. The period of low international real interest rates in

[1] The earlier 1974–1981 inflow episode is somewhat different and is considered in the next chapter in the context of the subsequent international debt crisis.

the 1970s triggered a reactivation of private capital flows into emerging economies, which had been minimal since the Great Depression. These flows dried up for many countries during the international debt crisis after 1982. The post-liberalization capital inflow episode of the early 1990s actually dates from about 1987–1988, when substantial amounts of private capital began to flow into Thailand. Inflows picked up during 1989 in Malaysia and during 1990 in Indonesia. The surge of capital inflows started slightly later in Latin America. The break in the capital inflow experience of this region came in 1990, when Chile and Mexico began to receive large amounts of capital. By 1992–1993, capital inflows to emerging economies had become comparable, relative to the size of the recipient economies, to their pre-1982 levels. After a pause that coincided with a tightening of U.S. monetary policy in 1994, the magnitude of flows picked up again in 1995–1996, shrugging off the Mexican crisis. Net private capital inflows slowed after the Asian crisis of 1997. There was a gradual recovery in the first half of 1998, but a collapse occurred once again after the August 1998 Russian default. A new period of low international interest rates and relative international prosperity led to a dramatic increase in inflows after 2002, which ended abruptly with the outbreak of the international financial crisis in 2008.

1. Characteristics of Capital Flows to Emerging Economies in the Early 1990s and Late 2000s

As suggested by the preceding description, the timing of the return of capital flows to emerging economies during the decade of the 1990s did not tend to be uniform across countries. Initially, the surge of new capital was primarily an East Asian and Latin American phenomenon. The phenomenon subsequently became more widespread, however, reaching South Asia as well as Sub-Saharan Africa by 1993. Nonetheless, whether expressed in absolute dollar values or scaled by exports or gross domestic product (GDP), inflows clearly tended to head disproportionately to countries in East and Southeast Asia and South America. Within each of these regions, flows tended to be concentrated in several large emerging economies. Over the period from 1989 to mid-1993, for example, 85 percent of all portfolio flows to East Asia were accounted for by China, Indonesia, Korea, and Thailand, while Latin America, Argentina, Brazil, Mexico, and Venezuela accounted for almost 95 percent of portfolio flows over the same period (Gooptu 1994). In all, 95 percent of all private capital inflows during 1988–1995 went to 21 countries. The most recent inflow episode, by contrast, was much more widespread geographically, with flows going not just to Asia and Latin America but also to emerging European and other emerging-market economies (see International Monetary Fund 2007).

During the decade of the 1970s, capital flows to emerging economies largely took the form of syndicated bank loans to governments. By contrast, equity instruments, both in the form of direct foreign investment and portfolio investments, played an important role in the 1990s (Figure 24.2). Syndicated bank loans were relatively

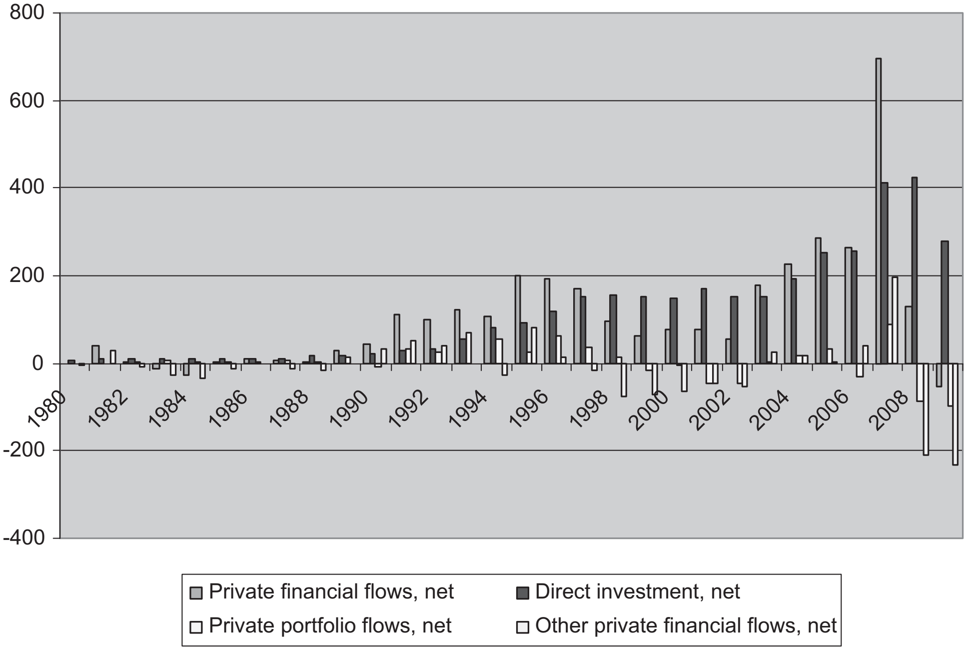


Figure 24.2. Composition of net private capital inflows to emerging and developing economies, 1980–2008 (in US$ billion)

unimportant in the early 1990s, reflecting the fact that new types of lenders other than banks became involved in emerging economies. Institutional investors (mutual and pension funds), for example, played a major role in the 1990s. As an indication of their growing impact during this period, in 1986, there were 19 emerging-market country funds and 9 regional or global emerging-market funds. By 1995, however, there were 500 country funds and nearly 800 global or regional emerging-market funds. The combined assets of all emerging-market funds rose from $1.9 billion in 1986 to $132 billion in the middle of 1996. The most recent episode, by contrast, has been dominated, at least on a net basis, by foreign direct investment (FDI) flows, as portfolio capital inflows have been offset by outflows to a much greater extent than during the episode of the 1990s, consistent with a much increased degree of financial integration among the recipient countries, which would be expected to facilitate two-way capital flows.

These trends away from commercial bank lending and toward portfolio and equity investment were geographically widespread. However, significant disparities in the composition of flows across regions were also apparent, at least during the early years of the inflow episode of the 1990s. For example, in Latin America, portfolio flows accounted for the majority of the new inflows until 1996–1997, whereas in East Asia, foreign direct investment was dominant until 1995–1996, when bank lending became much more important. By contrast, in the most recent episode, the dominance of FDI flows was evident in every region.

Not surprisingly, in view of the fact that the last two capital-inflow episodes followed a market-friendly reorientation of policies and financial reform in the recipient countries, the sectoral composition of capital inflows in these episodes was markedly different from that of the 1970s. Though governments were the primary emerging- and developing-country borrowers in the 1970s, in the more recent inflow episodes, borrowers in the recipient economies tended to be private agents rather than the public sector. The sectoral identity of the borrower presents a stark contrast between the recent inflow episodes and that of the 1970s, again consistent with an increased degree of financial integration in the recipient countries.

2. Pull Factors, Push Factors, and Changes in Financial Integration

Why did capital begin to flow back into emerging economies in such magnitudes in the early 1990s? The debate on this question among economists began almost as soon as the new inflows materialized. Leading explanations have typically been classified into three categories: push factors, pull factors, and changes in the degree of financial integration. Pull factors are those that originate in the borrowing countries. They represent improvements in the risk-return characteristics of assets issued by emerging-economy borrowers that help to attract capital from abroad. Push factors, on the other hand, originate in the lending countries. They drive capital out by reducing the attractiveness of lending to industrial-country borrowers.

As we saw in the model of Part 2, factors of either or both types could influence capital flows as long as the recipient economy is financially open, without any change in the degree of that economy's integration with world capital markets. The third factor that may have been at work during this period, however, was precisely a change in the degree of integration of emerging economies with such markets. As we have seen before, the degree of integration depends on the costs of cross-border borrowing and lending, and these fell sharply during this period for various reasons, including reduced information, transportation, and communication costs; changes in regulations in creditor countries that increased access to their financial markets by emerging-economy borrowers; and the removal of capital account restrictions in many emerging economies in the context of financial reform.

The potential roles of these three factors are illustrated in Figure 24.3, which shows the supply and demand curves for external funds from an emerging economy, as a function of the interest rate on such funds R. The gap between the notional supply curve S and the "effective" supply curve S' represents the cost per dollar borrowed of moving funds from the lending to the borrowing country, a measure of the degree of financial integration. The point where the effective supply curve and the demand curve intersect represents the equilibrium value of capital inflows. This value could be increased by pull factors, which would be represented by rightward shifts in the demand curve for funds; by push factors, which would tend to shift the notional supply curve S to the right, bringing the effective supply curve S' along

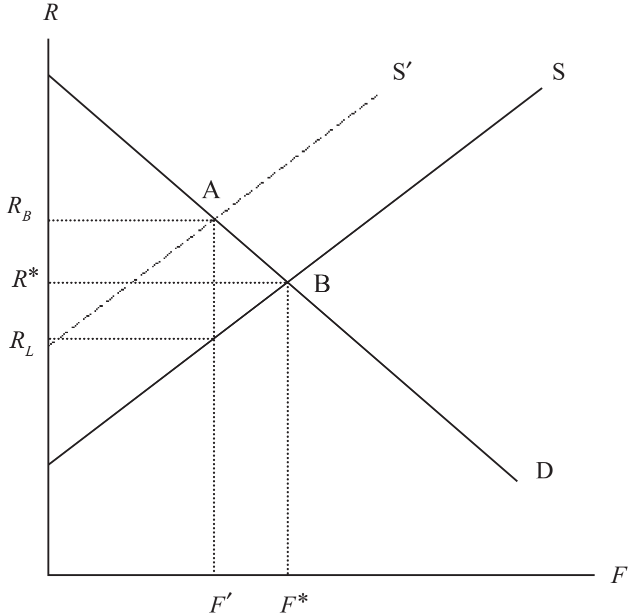


Figure 24.3. Factors driving capital inflows: A simple decomposition

with it, as long as the degree of financial integration remained unchanged; and by a change in the degree of financial integration, which would shift the S' curve down, closer to an unchanged position of the notional supply curve S. Notice that if the cost of moving funds is sufficiently high, S' could intersect the vertical axis above the point where the demand curve crosses it, meaning that the equilibrium value of capital inflows would be zero. This is the capital-constrained regime that prevailed for the heavily indebted countries during the period of the international debt crisis in the 1980s, but a move from a constrained to an unconstrained regime could have resulted from changes in any of the three factors we have discussed.

What factors contributed to the transition from a constrained to an unconstrained regime in the early 1990s? Unfortunately, despite a substantial amount of research that has been done on this issue, no consensus has emerged on the relative roles of push factors, pull factors, and changes in the degree of financial integration. Some of the most visible studies that addressed this issue are reviewed in the appendix to this chapter. Much of that evidence supported the view that push factors – specifically, declines in asset returns in the industrial countries – played an important role in reviving capital inflows, and the association of the more recent inflow episode with low international interest rates supports the view that push factors are generally important. However, the sharp differences noted earlier in capital inflow levels across countries in the episode of the 1990s point to the importance of specific country characteristics for foreign capital absorption. Whereas push factors may account for simultaneous rightward shifts in the supply curves facing many emerging economies, thus explaining the timing, and perhaps the overall magnitude, of capital inflows, pull factors that determine the position of the demand curve for funds in each country must nevertheless play an important role in explaining the geographic distribution of flows.

II. WELFARE IMPLICATIONS OF PUSH VERSUS PULL

It may seem somewhat surprising (and perhaps paradoxical) that if capital *outflows* were perceived as a policy problem during most of the 1980s, capital *inflows* to the same countries would also be perceived that way just a few years later. So a logical question to pose in interpreting the two recent capital-inflow episodes from a policy perspective is the following: should the arrival of large capital inflows in emerging economies be considered a good thing or a bad thing?

The answer is a familiar one: it depends. Whether the arrival of capital inflows represents a positive or negative development depends on what is driving the inflows as well as the characteristics of the economy receiving them.

It may be tempting to take the view that flows attracted to the recipient country by domestic pull factors do not present a policy problem because they represent a restoration of creditworthiness – that is, the rectification of the adverse conditions that emerged in the context of the debt crisis – whereas flows pushed out of the source countries are an external shock that can easily be reversed and thus may call for a domestic policy response. This would be incorrect, however. The distinction between pull and push factors does not necessarily have normative content. None of the problems that may be associated with the arrival of capital inflows depend in any simple way on whether those flows were pushed or pulled.

What are those potential problems? The policy challenge posed by the arrival of capital inflows emerges in four independent forms:

1. avoiding immiserizing external borrowing (making sure that funds are allocated correctly within the domestic economy)
2. avoiding macroeconomic overheating (destabilization associated with large inflows)
3. avoiding a loss of competitiveness
4. avoiding vulnerability to sudden outflows (destabilization associated with large outflows) and financial crises

In this section, we will consider the first of these, leaving the other three for the sections that follow.

Immiserizing external borrowing can arise because decisions about the amount of resources borrowed, and their allocation among competing domestic uses, are undertaken in a decentralized fashion rather than by a benevolent social planner. This means that in the presence of distortions, resources may be misallocated, yielding a social return that is lower than the cost of funds. There are two types of distortions that can result in immiserizing external borrowing:

1. *static* distortions, which cause a given amount of external resources to be allocated inefficiently among competing domestic uses, thus causing those resources to yield a social rate of return lower than the external cost of capital
2. *dynamic* distortions, which result in overborrowing; that is, even if resources are allocated to their most productive domestic uses, the volume of resources absorbed pushes the marginal rate of return below the cost of capital

We have already seen examples of these in previous chapters. For instance, static distortions can arise when the state borrows externally and allocates those resources directly according to political criteria, as happened in the case of the policy-based lending (directed credit) that arises under financial repression, where the state allocates external resources indirectly (Chapter 21). Static distortions can also arise in the real sector, for example, when the international trade regime – and thus domestic resource allocation – is distorted by excessive protection. We saw examples of dynamic distortions that encouraged excessive foreign borrowing in Chapter 22. These included, for example, "incredible" domestic reforms. Some policies may create both types of distortions. As discussed in Chapter 22, for example, premature financial liberalization, which may involve implicit or explicit guarantees to domestic financial intermediaries without appropriate regulatory safeguards, would have this effect by encouraging moral hazard–driven borrowing and lending for excessively risky activities by these intermediaries.

What is the relationship between the roles of push and pull factors and immiserizing external borrowing? The answer is that there is very little connection between them. Both types of factors can be welfare enhancing or welfare reducing.

Pull factors can improve the welfare of the domestic economy if they emanate from a favorable exogenous shock in an undistorted domestic economic environment, for example, a positive domestic productivity shock or the discovery of new natural resources. If this is what attracts capital inflows, those inflows would be welfare enhancing because their role would be to permit domestic residents to smooth over time the added consumption that their increased wealth would make possible. Alternatively, pull factors arising from the removal of a previously existing distortion in the domestic economy, with the effect of leaving an undistorted domestic economic environment, would also attract welfare-improving inflows. An example would be the removal of a policy-induced gap between private and social rates of return on domestic investment such as would occur through the elimination of a debt overhang (see Chapter 25) or the removal of domestic interest rate ceilings in the context of a comprehensive financial reform (Chapter 22).

But pull factors can also be welfare reducing. They would have this effect if they represented the creation of a new distortion in an otherwise undistorted environment (e.g., the incredible reforms alluded to before) or if they emerged as the result of the aggravation of a preexisting distortion. With respect to the latter, Dooley (1997) has stressed, for example, that improved creditworthiness of the domestic public sector may make previously existing inappropriate guarantees more credible because the government would have an increased ability to make them effective. Another example, taken from the sequencing analysis of Chapter 22, is the removal of capital controls when the financial system has not been reformed.

Push factors can also be either welfare enhancing or welfare reducing. An example of a welfare-enhancing push factor is the arrival of a favorable external shock in the context of an undistorted domestic economic environment, for example, a permanent or transitory reduction in the world real interest rate for an economy that is a net international debtor or a favorable terms of trade shock. Welfare-reducing push factors, on the other hand, would include favorable external shocks in the presence of domestic distortions such as lower world interest rates in a financially open economy with a domestic financial system characterized by important distortions or a favorable external shock when the domestic private sector cannot distinguish whether it is of a permanent or transitory character.

The policy implications associated with immiserizing external borrowing are clear. In each of the welfare-reducing cases just discussed, the appropriate policy response is to remove the distortions and allow the free inflow of capital. On the other hand, if the distortions are in place and cannot be removed, it may (but also may not) be preferable to resist financial integration until the distortions are removed. If even the suboptimal return available in the presence of domestic distortions exceeds the cost of funds, the country may be better off by borrowing abroad, even in the presence of domestic distortions.

III. MACROECONOMIC OVERHEATING: DESTABILIZING INFLOWS[2]

The arrival of capital inflows has often represented a large macroeconomic shock to the host countries in the form of an excessive expansion of aggregate demand. Digesting resources of this magnitude without macroeconomic overheating would present a policy challenge no matter what triggered the inflows. Thus, once again, there is no mapping from the domestic or external origin of the triggering factor to the policy challenge. What is the nature of the policy problem in this case?

1. Inflows and Overheating: The Transmission Mechanism

In the previous section, we saw how capital inflows could reduce the *level* of domestic income, through immiserizing external borrowing. But the arrival of large inflows could also *destabilize* income. The specific mechanism through which they would do so depends on the exchange rate regime used by the domestic economy. Consider

[2] This section draws on Montiel (1996).

the fixed exchange rate model of Part 2, and assume that inflows are triggered by a push factor in the form of a reduction in the world rate of interest. Then, recalling our short-run model from Chapter 7, it is easy to see that this change in the external financial environment represents an expansionary shock to the domestic economy. In terms of the model, the BB and MM curves shift to the right, the demand for domestic securities increases, and capital inflows resulting from portfolio reallocations on the part of both foreign and domestic agents cause the domestic money supply to expand. The resulting policy challenge takes the form of macroeconomic *overheating*, that is, of an excessive expansion of aggregate demand, causing an appreciation of the real exchange rate and deterioration in net exports.

The mechanism through which inflows would have this effect is as follows: with a predetermined exchange rate, large capital inflows are likely to generate an overall balance of payments surplus. To avoid an appreciation of the nominal exchange rate, the central bank would have to intervene in the foreign exchange market to buy the excess supply of foreign currency at the prevailing exchange rate. Ceteris paribus, this would result in an expansion of the monetary base. Base expansion would lead to growth in broader monetary aggregates, which would fuel an expansion of aggregate demand. This, in turn, would put upward pressure on the domestic price level. With the nominal exchange rate fixed, rising domestic prices would imply an appreciation of the real exchange rate. Schematically, we can depict the transmission mechanism as follows:

Inflows ↑ → Capital account ↑ → Reserves ↑ → Monetary base ↑ → Money supply ↑ →

Aggregate demand ↑ → Domestic price level ↑ → Real exchange rate ↓ → Trade deficit ↑

↘ ↗

Real output ↑

2. Policy Intervention: A Menu

This causal chain can be broken at various points by policy intervention. One useful way to organize the menu of policies available to the authorities to resist the emergence of overheating is thus according to where the intervention occurs along the chain of transmission described earlier. Accordingly, policy interventions can be classified as follows:

- Policies designed to restrict the net inflow of capital, either by restricting gross capital inflows or promoting gross capital outflows. Such policies include the imposition of administrative controls on capital inflows as well as the elimination of a variety of restrictions on capital outflows. They may also include the widening of exchange rate bands with the intention of increasing uncertainty, as we saw in Chapters 18 and 19.

- Policies that seek to restrict the net *foreign exchange* inflow (reserve accumulation) by encouraging a current account offset to a capital account surplus. Trade liberalization and nominal exchange rate appreciation would have this effect. As shown in Chapter 17, in the limit (floating exchange rates), the latter could avoid any foreign exchange accumulation whatsoever and actually *reverse* the effects of capital inflows on aggregate demand.
- Policies that accept the reserve accumulation associated with a balance of payments surplus, but attempt to ameliorate its effects on the monetary base. These amount to sterilized intervention, as well as attempts to limit recourse to the central bank's discount window.
- Policies that accept an increase in the base, but attempt to restrain its effects on broader monetary aggregates. Increases in bank reserve requirements, as reflected in the reserve ratio *rr* in Chapter 23, are examples of such policies.
- Policies that accept a monetary expansion, but attempt to offset expansionary effects on aggregate demand that could result in inflation and/or real exchange rate appreciation. This refers essentially to fiscal contraction.

In the rest of this section, we will examine some of the considerations involved in framing a policy response to avoid macroeconomic overheating. We will begin by considering the pros and cons of individual policies in the preceding list, and then we will turn to overall policy strategies.

3. Restrictions on the Magnitude of Gross Inflows

As we have already seen, the key requirement for controls on capital inflows to improve welfare is the presence of a distortion that creates an excessive level of foreign borrowing. When such a distortion exists and cannot be removed, restrictions on capital inflows may represent a first-best (when the act of foreign borrowing itself creates externalities) or second-best (when the negative welfare consequences of a new or preexisting domestic distortion that cannot be removed are magnified by external borrowing) policy. Alternatively, such restrictions could be justified on macroeconomic grounds as providing insulation from foreign financial shocks when other stabilization instruments are not available or are too costly to use; as preserving monetary autonomy; or as useful tools for liquidity management, that is, altering the composition of external liabilities in favor of longer maturities. Recall that in Chapter 22, however, we argued that these macroeconomic justifications for capital account restrictions were subject to important limitations.

Aside from these, the use of this instrument raises serious questions of feasibility. As we saw in Chapter 22, the evidence that restrictions on capital inflows are effective in altering the total magnitude of such flows is weak. Though there is stronger evidence that restrictions on gross flows can alter their composition, this may matter for the problem of vulnerability (to be described in the next section), but not for that of overheating.

4. Encouragement of Gross Outflows

The optimality and feasibility issues that arise in the context of the restriction of capital inflows also arise when the issue is the liberalization of gross outflows. Some of the benefits that can be expected from liberalizing capital outflows were discussed in Chapter 22. It is unclear, however, whether liberalizing gross outflows can help to address the macroeconomic overheating problem by reducing the volume of *net* capital inflows. The reasons are several. First, in parallel with the case of liberalizing inflows, restrictions on *outflows* may not be effective. Second, even if such restrictions are potentially effective, they may not be binding at the time they are removed; that is, residents may prefer to keep their money at home. Finally, facilitating gross outflows may not succeed in reducing *net* capital inflows.

Evidence on the effectiveness of outflow restrictions was mixed prior to the inflow episode of the 1990s. Mathieson and Rojas-Suarez (1993), for instance, tested whether capital flight responded to economic fundamentals in the same manner in countries with and without strong restrictions on capital outflows, with mixed results. They found that though fundamentals continued to influence capital outflows even in the presence of controls, responses to fiscal imbalances were slower, and those to default risk were weaker, in countries with strong capital controls. Overall, they concluded that controls on outflows tended to prove ineffective in stemming capital flight.

That restrictions on outflows would not be binding in the midst of an inflow episode would not be surprising if domestic residents responded to the same economic incentives that cause foreigners to bring capital into the country. Indeed, much of the inflow episode may represent the repatriation of capital by domestic residents.

Finally, there are various reasons to suspect that even if outflow restrictions are effective and binding, their removal may not have the desired effect of reducing net inflows because the very act of removing such restrictions may attract additional inflows. Laban and Larrain (1993), for example, have pointed out that the presence of effective controls on outflows renders inflows irreversible. If future policies affecting the return on loans to domestic agents are uncertain, the option to keep funds abroad while the uncertainty is resolved becomes valuable, and foreign creditors may thus refrain from lending in this situation. Removing the outflow restrictions eliminates the irreversibility and thus increases the relative return on domestic lending by eliminating the value of the option to wait. Similarly, Bartolini and Drazen (1997) have argued that because controls on outflows are often maintained for fiscal reasons (recall from Chapter 21 that such controls facilitate the collection of financial repression taxes), their removal is interpreted by foreign investors as a signal that future capital taxation is less likely, thereby inducing capital inflows.

What is clear from country experience is that substantial inflows followed the removal of restrictions on outflows in many emerging economies, an experience that has also characterized industrial countries.[3] Whether the removal of outward restrictions in the episode of the early 1990s – when such restrictions were still in wide use – diminished net inflows to any significant extent is impossible to say on the basis of present evidence, but it is clear that they did not represent a complete solution to the inflow problem anywhere that they were used. The observation just mentioned – that the inflow episode was often preceded by capital account liberalization, including the removal of restrictions on outflows – is certainly consistent with the view that the removal of restrictions on outflows simply attracts additional inflows.

5. Trade Liberalization

From a macroeconomic perspective, trade liberalization lowers the domestic-currency price of importable goods directly and may lower the price of nontraded goods indirectly (by depreciating the equilibrium real exchange rate).[4] To the extent that it induces a reduction in the current account balance, it absorbs some of the foreign exchange generated by the capital inflow, easing monetary pressures as well. The most controversial issue that arises with respect to trade liberalization as a means to restrict the net inflow of foreign exchange concerns its potential efficacy.

First, trade liberalization may not reduce the current account balance. Because the current account is the difference between domestic saving and investment, the effect of trade liberalization on the current account depends on how saving and investment are affected. Both theory and evidence suggest that the effects of trade liberalization on the current account are ambiguous, depending on a host of structural characteristics of the domestic economy as well as on the nature of the liberalization program. The former include the importance of nontraded goods, sectoral factor intensities, the nature of accompanying fiscal policies, and the extent of labor market rigidities. The latter include the incidence of tariffs (whether they fall on intermediate or final goods) and their projected future paths.

Ostry (1991), for example, shows that if temporary tariffs on intermediate goods are reduced, and tradable goods are more intensive in both intermediate and capital goods than nontradables, then the effect of the liberalization program will be to increase saving and reduce investment, thereby unambiguously improving the trade balance. The reduction in tariffs on intermediates will result in a short-run real appreciation as the traded goods sector expands, absorbing resources from the

[3] See Bartolini and Drazen (1995).

[4] See Mussa (1986).

nontraded sector. This real appreciation will cause agents to expect a larger real depreciation over time because future trade policy is left unaffected. Consequently, the real interest rate rises, and consumption tilts toward the future, increasing domestic saving. In turn, the increase in future consumption causes a future real appreciation that, relative to the undisturbed equilibrium, shifts capital from the traded to the nontraded sector in the future. Because the traded sector is relatively capital intensive, the implication is a reduction in today's aggregate investment. With saving higher and investment lower, the trade balance unambiguously improves.

Though this example may appear contrived, it merely illustrates the general principle that it is indeed quite possible in theory for a trade liberalization to improve the trade balance. The experience of liberalizing countries, as summarized, for example, in Thomas et al. (1991), suggests that this result is more than a theoretical curiosity.

Second, even if trade liberalization produces a trade deficit, the associated efficiency gains may trigger new capital inflows.

6. Exchange Rate Flexibility

The potential inflationary implications of capital inflows can be completely avoided by refraining from intervention in the foreign exchange market. Permitting a (temporary) appreciation of the nominal exchange rate in response to a favorable external interest rate shock (by restricting the scale of foreign exchange intervention) will dampen and possibly reverse the expansionary effect of the foreign shock on domestic aggregate demand by appreciating the real exchange rate. Indeed, a capital inflow arising from a reduction in external interest rates becomes a *deflationary* shock under fully flexible exchange rates, as we showed in Chapter 17. This outcome will be desirable if domestic macroeconomic conditions are such that policy makers seek to avoid stimulating aggregate demand. Thus, to the extent that capital inflows are permitted to materialize, the desirability of foreign exchange intervention depends in part on the requirements for macroeconomic stability.

The trade-off, however, concerns the implications for domestic resource allocation. If the authorities allow the nominal exchange rate to appreciate in response to capital inflows, the profitability of the traded goods sector will obviously be affected adversely. Aside from possible political economy considerations, policy makers may have two reasons to be concerned with this outcome.

First, if the capital inflow is believed to be temporary, an appreciation of the official exchange rate may tend to aggravate the effects of any previously existing domestic distortions biasing domestic resource allocation away from the traded goods sector (and causing the true social value of foreign exchange to exceed its official value). Second, with temporary capital inflows, the associated real exchange rate appreciation will also be temporary, and any costly resource reallocations induced by changes in relative sectoral profitability between the traded and nontraded goods sectors would later have to be reversed. Because such costs represent fixed costs from the perspective of private agents, the associated resource reallocations would not be undertaken unless the incentives for doing so were perceived to be long lasting. Because private agents will find it in their best interest to avoid the costs of transitory resource reallocation, the noise introduced into relative price signals by allowing excessive nominal exchange rate variability may reduce the efficiency of resource allocation.

Third, the preceding discussion treats the exchange rate as an instrument of short-run stabilization policy. However, the exchange rate also plays another role in small, open economies – that of a nominal anchor. We saw that role of the exchange rate in the analytical model of Part 2 as well as in Chapters 13 and 18. Allowing the exchange rate to adjust means adopting a different nominal anchor. If this is less desirable (e.g., because money demand is unstable, and the prerequisites for inflation targeting are not met in the economy in question), then allowing the exchange rate to pick up some of this pressure has a cost.

On the other hand, as discussed in Chapter 18, *nominal* exchange rate instability may be a good thing because it increases the uncertainty faced by potential foreign investors, and an increase in uncertainty may be desirable to discourage capital inflows that are driven, at least in part, by moral hazard considerations.

What was the experience with exchange rate policy during the early 1990s among countries experiencing large inflows? Several countries (e.g., Chile and Colombia) undertook revaluations in response to capital inflows. A second group of countries (Chile, Colombia, Indonesia, and Mexico) adopted exchange rate bands to allow more flexible nominal exchange rates, but full flexibility tended to be adopted only after crises (e.g., in Mexico in 1994, Southeast Asia in 1997, and Brazil in 1998). Moreover, despite the operation of the transmission mechanism described previously, many countries managed to avoid real appreciation over the course of the surge episode. This would seem puzzling if the only shock at work is the one that triggered the inflow. The role of fiscal policy responses in generating these outcomes is considered later. Finally, the link between real appreciation and the emergence of current account deficits was certainly not airtight among the countries receiving capital inflows. On one hand, the emergence of large current account deficits during the inflow episode was not restricted to countries that experienced real appreciations (Malaysia and Thailand both had large adverse movements in the current account balance with stable real exchange rates). On the other hand, some of the countries that experienced very substantial real exchange rate appreciation in the early 1990s (Argentina and Mexico, in particular) also exhibited very large current account deficits.

The situation was rather different in many of the larger emerging economies during the episode of the 2000s, in that many of those countries had adopted managed floats by the time the inflow episode materialized, in contrast with the fixed rates that prevailed in the early 1990s. Yet the outcome was not dissimilar.

Avoiding a real exchange rate appreciation was a high priority for most of these countries, so they intervened heavily in the foreign exchange market. Consequently, they accumulated very large amounts of foreign exchange reserves and exhibited relatively limited real appreciation. By contrast, in some of the smaller emerging economies in Europe, which had maintained much more rigid nominal exchange rates, the episode of the late 2000s resulted in more substantial real appreciation.

7. Sterilization

We saw in Part 2 that in the extreme case that domestic and foreign assets are perfect substitutes, sterilization becomes impossible. We also saw that even if it is possible, sterilization may, under certain circumstances (concerning the shape of the GM and BB curves), tend to magnify the size of capital inflows because it prevents the domestic interest rate from falling to restore portfolio equilibrium. Finally, we saw that sterilized intervention has quasi-fiscal costs because the central bank exchanges high-yielding domestic assets for low-yielding reserves. The magnitude of these costs will be greater the higher the degree of capital mobility and the larger the gap between domestic and foreign rates of return. Thus the fiscal feasibility of this policy is an issue.

But sterilization faces other limitations in practice that are not completely captured in the model we developed in Part 2. Those reflect the fact that, even if sterilization succeeds in limiting domestic monetary expansion, it may not succeed in insulating the domestic economy from the effects of capital inflows. This would be true under two sets of circumstances:

1. If domestic interest-bearing assets are perfect substitutes among themselves, as assumed in the model of Part 2, insulation would fail if the shock that triggers the inflows affects domestic money demand. In this case, with shifting money demand but fixed supply, domestic interest rates would change. This can be verified in the model of Part 2.
2. On the other hand, if domestic interest-bearing assets are *imperfect* substitutes among themselves (a feature not captured in our model), then a capital inflow may be associated with a shift in the composition of demand for domestic interest-bearing assets as well as with an increase in the total demand for such assets. In this case, unless the composition of domestic assets emitted in sterilization operations matches that demanded by creditors, the structure of domestic asset returns would be altered.

In practice, sterilized intervention has tended to be the most widely used stabilization tool in response to capital inflows in both inflow episodes because it possesses the great advantage of flexibility. An important lesson is that sterilization proved to be possible despite the capital account liberalization that many of the recipient countries had previously undertaken and despite the large magnitude of inflows that they received. Indeed, many countries registered an *increase* in domestic interest rates over the period of sterilization in the episode of the 1990s. Thus the removal of capital account restrictions in the context of financial liberalization did not imply that emerging economies moved into a regime of perfect capital mobility. It is likely, however, that "offset coefficients" associated with sterilization have been increasing over time.

On the other hand, many countries did not sustain policies of aggressive sterilization consistently over the inflow period. This clearly suggests that sterilization was not a panacea, despite its widespread appeal among the countries that experienced inflow surges. In the absence of data on quasi-fiscal costs, the importance of fiscal rigidities that made the quasi-fiscal costs of sterilizing too burdensome cannot be dismissed as an explanation, but it is clear that shifts in domestic economic circumstances that made lower interest rates more attractive have also played a role.

Finally, the effectiveness of sterilization in insulating economies from the effects of external financial shocks during the inflow episode is open to question. By and large, sterilization does not seem to have completely insulated the recipient economies from the effects of capital inflows. Asset markets, in particular, typically recorded massive increases in value during both surge periods. This is consistent with an imperfect-substitutability story in which foreign creditors demand domestic financial assets different from those issued by the central bank in the course of its sterilization operations.

8. Policies to Influence the Money Multiplier

If, for fiscal or other reasons, sterilization is incomplete, the implication of a foreign exchange inflow is an expansion in the monetary base. Monetary expansion can still be avoided by a commensurate reduction in the money multiplier (the ratio of M1 to the monetary base) achieved through an increase in reserve requirements or other restrictions on credit expansion by the banking system. Measures directed at decreasing the money multiplier by increasing reserve requirements may indeed avoid quasi-fiscal costs, but as we saw in Chapter 21, they do so through implicit taxation of the banking system. The economic implications of this tax depend on how the tax burden is ultimately shared among bank shareholders, their depositors, and their loan customers. Nonetheless, as we saw previously, the likely effect of this policy is to shrink the domestic financial system, an outcome that runs counter to the financial reform that has been undertaken in most reforming economies and that may have adverse implications for economic growth.

Moreover, such measures may be of limited effectiveness. Increases in reserve requirements may have little effect, for example, if banks are already holding excess reserves. If reserve requirements are changed selectively for different components of banks' liability portfolios, then their effects could be evaded as bank creditors shift to assets not affected by changes in reserve requirements. Finally, even if changes in reserve requirements are applied broadly across bank liabilities, domestic credit expansion could materialize through nonbank institutions (*disintermediation*). The scope for doing so – and thus for avoiding an increase in domestic aggregate demand – depends on the sophistication of the domestic financial system.

9. Fiscal Contraction

If domestic monetary expansion is not avoided, or if an expansionary financial stimulus is transmitted outside the banking system, the stabilization of aggregate demand will require a fiscal contraction. Feasibility and optimality issues arise in this context as well.

With respect to feasibility, fiscal policy may simply prove too inflexible to be available as a tool to respond to fluctuations in capital movements. The budgetary process in most countries may not be able to respond sufficiently quickly, and lags in response may indeed aggravate the stabilization problems created by volatile capital movements. Beyond this, political pressure may prevent fiscal contraction during the good times associated with inflow episodes (Chapter 11). But even if fiscal policy can be changed, the desired effects on domestic demand (and thus on the real exchange rate) will be forthcoming – that is, the policy will be effective – only if expenditure cuts fall on domestic goods.

Whether it is optimal for fiscal policy to respond in this way is also at issue. We saw in Part 3 of this book that expectations of future fiscal policy have potentially important implications for macroeconomic performance. This raises the question of whether fiscal policy should be designed to anchor long-run expectations of inflation and taxation, or should be guided by countercyclical objectives. Of course, in principle, these goals are not mutually exclusive because short-run deviations from the medium-term fiscal stance can be designed to respond to transitory shocks and thus achieve stabilization objectives. As discussed in Chapter 11, the problem is that if government credibility is lacking, adherence to the medium-term stance in the face of shocks may be the surest way to achieve it. In a nutshell, the issue is whether the achievement of fiscal credibility is compatible with the adoption of feedback rules for fiscal policy.[5]

A separate optimality issue is that if the stabilization objective is adopted, changes in marginal tax rates in response to temporary capital inflows should be avoided because fluctuations in such rates would distort the intertemporal choices made by private economic agents.

In practice, fiscal policy did not prove to be a very flexible instrument in responding to inflows either in the early 1990s or in the mid-2000s. After the fiscal adjustments that many emerging economies made during the late 1980s, few of them

[5] Notice that if such a rule were to be applied symmetrically, it would imply that capital outflows should elicit an *expansionary* fiscal response.

found it possible to engage in additional fiscal tightening in response to inflows, and where additional fiscal tightening took place, the changes in the fiscal stance were not typically large compared with previous fiscal adjustments in the countries concerned. This may reflect a variety of factors, including so-called stabilization fatigue arising from the substantial fiscal adjustment that many countries had already undertaken prior to the inflow episode or political economy considerations that make it difficult to undertake fiscal austerity when external constraints on the economy are not perceived to be binding. Similarly, the growth of real government spending accelerated in many recipient countries during the inflow episode of the mid-2000s (International Monetary Fund 2007).

Whatever the reason for the nature of the fiscal response, however, the absence of additional fiscal tightening may have played an important role with regard to outcomes for the real exchange rate. Real appreciation was avoided, for example, in all the East Asian countries that tightened fiscal policy in response to inflows in the early 1990s (World Bank 1997). The frequency of real appreciation elsewhere supports the implication of theory that in the presence of capital inflows, the avoidance of real appreciation requires a fiscal contraction to free up the requisite supply of domestic goods without a relative price change.

Nonetheless, tighter fiscal policy was not sufficient to avoid real appreciation. Real appreciation accompanied fiscal tightening during the early 1990s in Argentina and Egypt, for example, but both these countries were in the midst of stabilizing from high inflation, and it is likely that the behavior of the real exchange rate reflected inflation inertia. If this interpretation is correct, the implication is that real appreciation would have been more severe in these countries if fiscal policy had been looser.

10. Policy Strategies

Up to this point, we have concentrated on some of the considerations involved in implementing specific individual policies to combat the macroeconomic overheating that may be caused by the arrival of capital inflows and have reviewed some country experience with some of these policies. Individual policies obviously have pros and cons associated with their use, and if the objective is to avoid macroeconomic overheating, then it is obvious that there are many ways to combine the policies described earlier to achieve that objective. How did the combination of policies actually chosen by emerging economies to combat overheating affect their macroeconomic performances during capital-inflow episodes? Before leaving the problem of overheating, it is useful to consider whether there are lessons to be learned from country experiences with alternative policy combinations, that is, about alternative *policy strategies.*

A study by the World Bank (1997) considered precisely this issue for the episode of the early 1990s. In a sample of 21 countries accounting for about 90 percent of all private capital flows to developing countries during the period 1990–1994, it found the following:

- Concerning the extent of macroeconomic overheating that actually materialized, the vast majority of the countries in the sample used by the study managed to avoid an acceleration of inflation during their inflow periods, compared to the pre-inflow period. Moreover, though increases in current account deficits were widespread among the sample countries, they were not large, and about half the improvement in the capital account resulted in reserve accumulation rather than increases in current account deficits. The study concluded that despite receiving very large capital inflows, countries had been fairly successful at avoiding overheating.
- Country experience tended to differ across various dimensions: as we have already seen, some countries managed to avoid real exchange rate appreciation, whereas others did not. Countries differed as well with regard to changes in the composition of absorption (total spending by domestic residents). In some cases, the increase in absorption relative to income was dominated by investment and, in other cases, by consumption.
- Countries tended to use a wide range of the policies described earlier rather than relying heavily on a small subset of them. Sterilized intervention was the most common policy employed to resist overheating. Tighter fiscal policy was also featured in some countries, but it was unclear whether this was a response to the arrival of capital inflows or a continuation of previously undertaken fiscal consolidation.

The study found strong cross-country positive correlations during the inflow period between the extent of real exchange rate appreciation and both the increase in the current account deficit and the change in the share of consumption in absorption. It also found a weaker positive association between the increase in the share of investment in absorption and the change in economic growth. Though in all of these instances, the correlations could be interpreted as the result of causation operating in either direction, the study found a strong association between the extent of fiscal contraction during the inflow period and the increase in the share of investment in absorption. It thus concluded that a tight money–loose fiscal strategy to resist overheating was associated with faster growth and less real appreciation than an alternative tight money–relatively loose fiscal strategy. Countries following such a strategy included Thailand, Indonesia, Malaysia, and Chile.

IV. VULNERABILITY

The third potential problem associated with the arrival of capital inflows is the likelihood that such inflows will be reversed suddenly, as happened in the early 1980s at the onset of the international debt crisis. In other words, the increased financial integration that follows financial reform, and that may initially be associated with the arrival of capital inflows, may also make the country vulnerable to capital account crises, in the form of the sudden reversals of these capital flows. This raises two questions: (1) what determines the probability of such capital-flow reversals? and (2) what are the macroeconomic implications of such reversals? These issues are the subject matter of Chapter 27. The purpose of this section is to provide an overview of the first of these topics, both to link it with what has been said earlier in this chapter and to provide a transition to the material of Part 7 of the book.

1. Sources of Vulnerability

When is a reversal of capital inflows likely to happen? The answer is similar to what one might say about the determinants of the likelihood of withdrawals from banks by their depositors. Economic agents (both foreign and domestic) that have claims on the domestic economy are likely to liquidate those claims – causing a reversal in capital inflows – when they come to believe that the value of their assets is in danger of becoming impaired. In turn, this is likely when the following occur:

- the country's fiscal solvency comes into question
- its exchange rate is perceived to be overvalued
- the domestic financial system is perceived to be fragile
- the economy's public sector is highly illiquid

The first three situations make outflows more likely, whereas the last makes a crisis more likely if there are outflows and makes self-fulfilling crises more probable.

We will see in the next chapter that prospective insolvency of the public sector in the heavily indebted developing countries was the key factor behind the international debt crisis of the 1980s. It happened that at the time, most of the external debt outstanding was owed by the public sectors of these countries to consortia of foreign banks in the form of *syndicated loans* and was denominated in foreign exchange. However, doubts about the solvency of the domestic public sector would impair the value of claims on private agents as well, if only because of the likelihood of future taxation on those claims as the public sector tries to bring its fiscal accounts into balance. This is the sovereign risk analysis that was the focus of Chapter 10. Thus public-sector solvency problems would tend to impair the value of all claims on the domestic economy, as discussed in that chapter, regardless of the identity of the domestic debtor or the currency of denomination.

As we saw in Chapter 16, an overvalued domestic real exchange rate means that a future real exchange rate adjustment is in the works, unless the domestic authorities are prepared to alter some of the real determinants of the equilibrium real exchange rate to make the actual real exchange rate an equilibrium one. When such an adjustment is in the offing, creditors who hold claims denominated in domestic currency face a potentially large loss of value on their assets and have an incentive to convert them into assets denominated in other currencies. Again, such a situation would tend to lead to a reversal of capital flows, in the form of an exchange rate crisis.

Fragility in the domestic financial system (defined as in Chapter 22 as vulnerability to a loss of capital in response to even mild shocks) clearly impairs the value of claims on the financial system, unless those claims are insured, and may do so even if they are insured if the insurance fund (or the government's support) is limited. Moreover, to the extent that a potential banking crisis has adverse real effects on the economy as a whole, as discussed in Chapter 22, the emergence of such a crisis would impair the value of all claims on the domestic economy, not just claims directly on the financial system. Again, wholesale withdrawal of claims on the domestic economy is likely if such a situation emerges.

As we shall see in Chapter 27, the combination of the second and third of these factors may have been behind the Chilean crisis of 1982, the Mexican crisis of 1994, and the Asian crises of 1997. Aside from the international debt crisis of the early 1980s and the Argentine crisis of 2001–2002, these were the most severe episodes of capital-flow reversals among emerging economies over the past two decades.

Finally, the composition of the public sector's portfolio may also provide a trigger for a capital-flow reversal, even if the public sector is solvent. If the public sector's liabilities are liquid, but its assets are not, then the sector as a whole is in the same position as the individual banks described in Chapter 22; that is, a wholesale withdrawal of claims could make the public sector insolvent in the sense that it may prove unwilling to meet its obligations because any attempt to do so by liquidating assets, cutting spending, or raising taxes could do unacceptable damage to the domestic economy. If this situation prevails, then there may be two possible equilibria: one in which the public-sector creditors who hold liquid assets do not expect a crisis, and continue to renew these claims as they come due, so no crisis happens, and another in which these creditors *do* expect a crisis, so they fail to renew their claims, thereby triggering a crisis that fulfills their expectations. We will return to this analysis of multiple equilibria created by the structure of the public sector's balance sheet when we look at the Mexican crisis in more detail in Chapter 27.

2. Sources of Vulnerability: Evidence

In recent years, there has been an outpouring of research on the determinants of capital account reversals. We will come back to this evidence in our discussion of currency crises in Chapter 27. For now, however, it may be useful to review the findings with respect to the determinants of vulnerability to capital-flow reversals of the World Bank (1997) study discussed in the previous section.

To see what kinds of policies were associated with vulnerability in the early 1990s, the study examined how well a sample of 13 emerging economies had weathered the Mexican crisis during 1995. Countries were classified into two groups: those that were strongly affected by the Mexican crisis (and were thus judged to be vulnerable) and those that were not. The judgment was made on the basis of two criteria: whether the countries in the sample had sustained private capital inflows at previous levels during 1995 and whether equity prices in their domestic stock markets were stable in that year. On the basis of these criteria, the strongly affected group included Argentina, Brazil, India, Pakistan, Turkey, Venezuela, and Mexico. The less affected group consisted of Chile, Colombia, Indonesia, Korea, Malaysia, and Thailand.

The characteristics that distinguished the two groups were taken to be indicators of vulnerability. Over the period leading up to the Mexican crisis (1988–1993), the less affected group was found to have had higher growth of real output, a higher share of investment in GDP, lower inflation, substantially less real exchange rate appreciation (20% on average versus 3.5%), much smaller fiscal deficits (a surplus of 0.6% of GDP on average versus a deficit of 4% of GDP), and a smaller debt-export ratio.[6]

The interpretation given to these results was that vulnerability to capital-flow reversals was greater when the real exchange rate was out of line, when government debt obligations were large, when fiscal adjustment was perceived as politically or administratively infeasible, and when the country's growth prospects were perceived to be weak. This interpretation is consistent with the factors listed earlier as being conducive to capital-flow reversals.[7]

V. SUMMARY

Whether episodes of large capital flows to emerging economies have been driven primarily by push factors, pull factors, or some combination of the two, it is likely that the market-oriented reforms undertaken by these countries in the late 1980s – especially the financial reforms that resulted in more open capital accounts – left these countries much more integrated with international capital markets than they had been previously. The experience of these countries has shown that this enhanced financial integration presents both opportunities and pitfalls. In Chapter 22, we saw that enhanced financial integration can confer many benefits on the domestic economy. In this chapter, we have seen that increased integration also poses serious policy challenges, even when capital is flowing into the country rather than out.

[6] It is notable that these characteristics did not include the size of their current account deficits.

[7] There is an interesting application here to the policy mix issue of the previous section. The World Bank study concluded that a policy mix that assigns the real exchange rate to the achievement of external balance (an appropriate outcome for the current account) and fiscal policy to internal balance (full employment with low inflation) seems to be conducive to the avoidance of vulnerability.

A possible response to this situation is for emerging economies that choose to live less dangerously to retain or reimpose restrictions on capital movements. As we have seen previously, weak restrictions on capital movements may indeed be justified on either first-best or second-best grounds. In practice, however, their effectiveness is questionable. Such restrictions may be both useful and effective in changing the composition of inflows, and if they can indeed affect total flows, they may also be useful as a transitional device to preserve monetary autonomy while more fiscal flexibility is achieved. But controls on capital movements may only preserve monetary autonomy for a short time, and as financial integration proceeds in emerging economies, that time may be getting shorter.

In the alternative that enhanced financial integration is embraced by domestic policy makers, it is clear that the premium on high-quality domestic economic institutions as well as on high-quality domestic macroeconomic management increases sharply. If domestic economic institutions and the quality of domestic macroeconomic management are inadequate, financial integration can lead to lower and less stable income than under financial autarky. Of particular importance are fiscal and monetary policies, exchange rate policies, and financial-sector policies, the subjects of Parts 3–6 of this book.

We have seen that responsible and flexible fiscal management is a central component of good macroeconomic management under conditions of high financial integration. This can help avoid overheating, reduce vulnerability to capital-flow reversals, and more generally help to cope with the volatility that capital account openness can introduce to the domestic macroeconomic environment.

In avoiding overheating in response to capital inflows, we saw in this chapter that the fiscal-monetary mix adopted in the recipient economy seems to matter. A tight fiscal–loose monetary mix of policies tends to favor domestic investment over consumption and to generate both less real appreciation and smaller current account deficits in response to capital inflows. This mix of policies also seems to have been more successful than the alternative loose fiscal–tight money mix in avoiding vulnerability to capital flow reversals during the first half of the 1990s.

With regard to coping with the volatility of domestic income that can emerge with a high degree of financial integration, a prudent fiscal policy fosters the perception of public-sector solvency, thus reducing sovereign risk and avoiding having the public sector itself become a source of shocks. It also helps to generate a fiscal cushion that allows for some fiscal flexibility, providing an independent stabilization instrument and reducing incentives to restrict flows through capital controls. Where monetary sterilization is feasible, a solvent fiscal position permits the public sector to bear any fiscal costs associated with that policy, again enhancing the flexibility of an instrument of stabilization policy. Finally, a reputation for fiscal prudence allows the government to borrow long-term in domestic currency, thereby reducing its vulnerability to liquidity crises.

Turning to the conduct of monetary policy, we have seen that sterilized intervention still appeared to be possible among emerging economies in the early 1990s, though it may have become more difficult to implement it over time. This stabilization tool has the strong advantage of flexibility in its use. However, many caveats are associated with relying heavily on sterilization to resist the overheating pressures triggered by capital inflows. First, sterilization may not be effective in insulating the domestic economy from external shocks. Second, its costs may weaken the fiscal accounts. Third, the creation of a large stock of domestic public debt as a result of sterilization may undermine fiscal credibility. Finally, as we have seen, an excessive reliance on sterilization to resist the macroeconomic overheating threatened by the arrival of capital inflows seems to have done less well in its effects on domestic macroeconomic performance than a mix of policies that put more weight on tighter fiscal policy.

APPENDIX 24.1. EVIDENCE ON FACTORS DRIVING CAPITAL INFLOWS IN THE 1990S

No general consensus has emerged concerning the relative roles that various factors may have played in bringing foreign capital back to emerging economies in the early 1990s. This appendix provides an overview of the main studies that attempted to address this issue.

1. Calvo et al. (1993)

Calvo et al. (1993; hereinafter referred to as CLR) argued that though domestic factors, such as a changed policy regime, may have been important in attracting inflows, such factors cannot explain why inflows occurred in countries that had not undertaken reforms or why when reforms were started earlier, inflows did not begin to emerge in Latin America until 1990. They thus emphasized the role of external factors. Their formal analysis took the following form:

1. They used *principal component analysis* (an empirical analysis that extracts the common components driving multiple time series) to establish the presence of a significant degree of co-movement among foreign reserves and real exchange rates for 10 Latin American countries during 1990–1991. It turned out that the first principal component explained a larger share of the variation in the 10 reserve and real exchange rate series during 1990–1991 than in 1988–1989, suggesting that common factors exerted a stronger influence on these variables during that period.
2. They found that the first principal components of both the reserve and real exchange rate series in the Latin American countries in their sample displayed a large bivariate correlation with several U.S. financial variables that CLR used as indicators of foreign rates of return.
3. They conducted tests of *Granger causality* (based on the ability of one time series to predict future values of the other, after controlling for the past history of the series for which the future values are being predicted). In individual countries,

Granger causality tests most frequently had reserves "causing" (predicting) real exchange rates than the reverse. This pattern also held for the first principal components of the two sets of series.

4. Structural *vector autoregressions* (VARs, a series of regressions of current values of a set of variables on past values of all the variables in the set) involving reserves, real exchange rates, and the first two principal components of the U.S. financial variables suggested that the foreign factors exerted causal influences over the domestic variables (in the Granger sense), and both *variance decompositions* (measures of the variance of one series that can be explained by stochastic shocks to other series) and *impulse response functions* (estimated dynamic responses of one series to shocks in the others) indicated that the foreign factors played a large role in accounting for reserve and real exchange rate movements.

2. Chuhan et al. (1993)

Using monthly bond and equity flows from the United States to nine Latin American and nine Asian countries over the period January 1988 to July 1992, Chuhan et al. (1993; hereinafter referred to as CCM) estimated separate panel regressions explaining bond and equity flows as functions of country-specific variables (country credit rating, price of debt on the secondary market, price earnings ratio in the domestic stock market, and the black market premium) as well as external variables (U.S. interest rates and U.S. industrial activity). They found that bond flows (but not equity flows) responded strongly to the country credit rating, whereas price-earnings ratios were uniformly important. However, U.S. interest rates also entered significantly with the theoretically expected negative sign in all the regressions.

To assess the relative importance of domestic and foreign variables, they computed the sum of *standardized coefficients* (coefficients divided by their standard errors) for the two sets of variables, finding that domestic and external variables have been about equally important in Latin America but that domestic variables had sums of standardized coefficients that were three to four times larger than those of external variables in Asia for both bond and equity flows.

3. Fernandez-Arias (1994)

Fernandez-Arias (1994) argued that the attribution of variation in country-specific financial variables to domestic shocks in CCM was improper and, in particular, that country creditworthiness, as indicated by the price of debt on secondary markets, is itself heavily dependent on external factors.

To assess the role of such factors, he regressed deviations in portfolio (bond and equity) inflows from their 1989 values for 13 developing countries on corresponding deviations in the external interest rate and in the price of debt on the secondary market (based on a simple burden-sharing model that linked creditworthiness to this variable), using fixed-effect panel estimates for which the intercept term was interpreted as the change in the domestic investment climate.

For the average developing country in the sample, changes in international interest rates proved to be the dominant force in explaining surges in capital inflows, accounting for over 60 percent of the deviation in such flows from the 1989 level. An extra 25 percent was due to changes in creditworthiness, leaving only about 12 percent to be explained by improvements in the domestic investment climate. Moreover, when account was taken of the role of external interest rates in determining the secondary-market debt price used as the creditworthiness indicator, thereby decomposing the latter into domestic and foreign components, fully 86 percent of the surge in inflows was attributed to movements in external interest rates.

4. Dooley et al. (1994)

Dooley et al. (1994) argued that the price of emerging-economy debt to commercial banks in the secondary market is a sensitive proxy for capital inflows because shifts in the demand for claims on developing countries, whether emanating from changes in domestic or external factors, should be reflected in these prices. Thus, rather than explaining capital inflows directly, they attempted to account for the behavior of secondary-market prices on debt since 1989, which, consistent with their interpretation of the relationship between such prices and capital flows, rose markedly at that time.

They found that essentially all of the increase in secondary-market prices of debt could be accounted for by reductions in the face value of debt and international interest rates, leaving almost nothing to be explained by improvements in the domestic environment.

5. Schadler et al. (1993)

These findings concerning the role of foreign factors have not gone unchallenged, however. Schadler et al. (1993), for example, argued that though foreign phenomena may have been important, such influences cannot be regarded as dominant, for several reasons:

1. They maintained that the timing of the relevant changes in external factors did not coincide with that of the inflows.
2. They noted that the timing, persistence, and intensity of inflows varied considerably across countries that received inflows, suggesting that investors responded to changes in country-specific factors over time.
3. They pointed out that surges in capital inflows were not universal within regions of emerging economies so that external creditors clearly exercised some cross-country discrimination in the allocation of funds.

6. Hernandez and Rudolf (1994)

More systematic evidence supporting a role for domestic factors in attracting capital inflows was provided by Hernandez and Rudolf (1994). Noting that previous work tended not to provide a careful specification of domestic factors, Hernandez and Rudolf examined the extent to which standard creditworthiness indicators could explain long-term capital inflows for a sample of 22 developing countries over the period 1986–1993. They used two methodologies:

1. First, they split their sample of countries into groups of high-capital-inflow recipients (HCIR) and low-capital-inflow recipients (LCIR). They found that the former had domestic saving rates twice as large as the latter, invested a much larger proportion of gross national product, exhibited significantly lower fiscal deficits and inflation rates, and had lower stocks of debt as well as larger stocks of foreign exchange reserves and faster rates of export growth. The HCIR countries were also more stable, in the sense that they exhibited lower variability of inflation and real exchange rates and scored lower on a political risk index.
2. Second, arranging their data into a panel of annual observations, they estimated capital-flow equations for a broad category of long-term flows as a function of lagged domestic consumption and investment rates, external interest rates and the ratio of net external debt (gross debt minus foreign exchange reserves) to *GNP*, the variability of the real exchange rate, and the presence of a Brady bond deal (see Chapter 25). They found statistically significant (albeit not very precisely estimated) roles for domestic creditworthiness indicators but no role for the external interest rate.

7. Taylor and Sarno (1997)

Taylor and Sarno (1997) tried to explain monthly portfolio flows from the United States to nine Asian and nine Latin American countries from January 1988 to September 1992. They used *Institutional Investor* credit ratings and the parallel market premium as country-specific factors and the U.S. Treasury bill rate and a U.S. long-term government bond rate, as well as U.S. industrial production, as global factors.

Taylor and Sarno (1997) found cointegrating relationships between capital inflows and the explanatory variables in all 36 cases and based conclusions about the relative importance of the two types of variables on the relative frequency of significant coefficients in country-specific error-correction models for inflows, specified using a Hendry general-to-specific methodology. They found that push and pull factors were equally important in explaining equity flows in both Asia and Latin America but that global factors (especially U.S. interest rates) were more important in explaining bond flows in both regions.

8. World Bank (1997)

All the evidence cited earlier pertains to the early years of the capital inflow episode of the 1990s, that is, 1989–1993. Additional evidence provided by the World Bank (1997) suggested that the factors driving inflows may have been changing over time and, in particular, that domestic factors may have played a more prominent role during 1994–1995 than they had done earlier. Adopting the CLR methodology, the World Bank found that quarterly portfolio flows from the United States to 12 emerging markets in East Asia and Latin America were characterized by a substantial amount of co-movement (measured by the proportion of the variation captured by the first principal component) during 1990–1993 and that the first principal component of these series was highly negatively correlated with the first principal component of a set of representative U.S. asset returns. Both these findings are consistent with the findings of CLR for this period, as described previously.

However, over the years 1993–1995, co-movements among portfolio flows became much weaker (the contribution of the first principal component dropped to 45% from 75% of the variance), and the correlation with U.S. asset returns reversed signs and became much weaker. The implication is that idiosyncratic country factors may have played a much larger role in the last part of the episode than they did in the early years.

What can we learn from all this research? There are three things we might want to know:

1. What factors contributed to the transition from a constrained to an unconstrained capital-flow regime?
2. How sensitive have capital flows to emerging economies been to each of their potential determining factors?
3. Which factors have actually been driving flows during any particular sample period?

Studies that have estimated capital-flow equations (such as those by Chuhan et al. (1993), Hernandez and Rudolf (1994), and Taylor and Sarno (1997)) tell us about the reduced-form coefficients of capital-flow equations and thus address the second issue raised. The evidence suggests that flows were responsive to both domestic and external factors.

The strongest evidence against the push view during the early years of the inflow episode is that provided by Hernandez and Rudolf (1994). However, their evidence is not necessarily inconsistent with the push view, despite the poor performance of the U.S. interest rate in their capital-flow regressions. Specifically, their focus on long-term capital flows and the weight given to the 1990–1986 period in their data suggest that their results may primarily apply to FDI flows and are not necessarily applicable to other types of capital flows such as portfolio or short-term flows.

Finally, the co-movement papers (CLR and World Bank (1997)) tell us something about the third issue, suggesting that external factors may have accounted for most of the variation in capital inflows early in the inflow episode (until about 1993) but that country-specific factors may have played a more important role later on.

REVIEW QUESTIONS

1. What types of factors may potentially account for the emergence of large capital inflows in emerging and developing economies? How does the geographic diversity in the magnitude of such flows help us assess the relative roles of such factors?
2. Why have large inflows of capital often been perceived as a mixed blessing in emerging and developing economies?
3. Explain the mechanism through which the arrival of large capital inflows may trigger macroeconomic overheating in fixed–exchange rate, open economies.
4. In formulating a policy response to the arrival of capital inflows, is there a trade-off between avoiding overheating and avoiding real exchange rate appreciation? Explain.
5. Describe the pros and cons of policy strategies to resist macroeconomic overheating that rely primarily on tight monetary policy (sterilization) or tight fiscal policy.

EXERCISES

As you have learned in this chapter, many developing countries periodically experience a capital-inflow problem, in which the country is swamped by large amounts of capital inflows that are driven by exogenous factors. Sudden surges of capital inflows may be driven by sharp decreases in foreign interest rates or by an exogenous increase in demand for the bonds issued by the domestic economy (in terms of the model of Chapter 7, think of this as a positive shock to the fraction b^* in the portfolios of foreign residents). In the first four of the following problems, think of these inflows as indeed being triggered by an unanticipated positive shock to the fraction b^*. Assume that the country that is affected by the capital-inflow problem is characterized by imperfect capital mobility, that its central bank maintains a fixed exchange rate, and that it initially operates a monetary policy regime in which it targets the stock of domestic credit.

1. Explain what effect you would expect the shock to b^* to have on the domestic interest rate, the price level, the real exchange rate, exports, and net exports if the central bank keeps the stock of domestic credit constant.
2. How would these effects differ if the central bank was targeting the money supply instead of domestic credit? If it was targeting the domestic interest rate? How

would the effects of the shock on the central bank's stock of foreign exchange reserves differ under these two alternative monetary policy regimes?

3. In the domestic credit-targeting case, if the central bank seeks to prevent the shock from affecting the domestic price level, how could it use exchange rate policy to do so (assume that the central bank does not monetize exchange rate changes)? What would the implications of this use of exchange rate policy be for net exports?
4. If the central bank does not respond to the sudden inflow of capital, is there anything the finance ministry could do to stabilize the domestic price level in response to the inflow? Explain.
5. Consider a small, open economy with a fixed exchange rate that is imperfectly integrated with international capital markets. Suppose that this economy experiences a capital inflow driven by an exogenous increase in foreigners' demand for domestic bonds. Assume that the central bank responds to the inflow by increasing the required reserve ratios of the economy's commercial banks.
   a. How would such a measure affect the causal link between a capital inflow and aggregate demand in the recipient economy?
   b. Assume that demand deposits in the recipient economy are only very weakly sensitive to the deposit interest rate. Use GM-BB analysis to show how the central bank's action would affect the economy (i.e., contrast the macro equilibrium in the presence of this measure with what would have happened if the central bank had done nothing).
   c. What is the argument against the central bank responding to the inflow in this way?

REFERENCES AND FURTHER READING

Bartolini, Leonardo and Alan Drazen (1997), "Capital Account Liberalization as a Signal," *American Economic Review*, vol. 87, pp. 138–154.

Calvo, Guillermo, Leonardo Leiderman, and Carmen Reinhart (1993), "Capital Inflows to Latin America: The Role of External Factors," *IMF Staff Papers*, vol. 40, pp. 108–151.

Chuhan, Punam, Stijn Claessens, and Nlandu Mamingi (1993), "Equity and Bond Flows to Latin America and Asia: The Role of Global and Country Factors," IECIF Policy Research Working Paper 1160, World Bank.

Dooley, Michael P. (1997), "A Model of Crises in Emerging Markets," working paper 6300, National Bureau of Economic Research.

Dooley, Michael, Eduardo Fernandez-Arias, and Kenneth Kletzer (1994), "Is the Debt Crisis History? Recent Private Capital Inflows to Developing Countries," working paper 4792, National Bureau of Economic Research.

Fernandez-Arias, Eduardo (1994), "The New Wave of Private Capital Inflows: Push or Pull?" IECIF Policy Research Working Paper 1312, World Bank.

Gooptu, Sudarshan (1994), "Are Portfolio Flows to Emerging Markets Complementary or Competitive?" policy research paper 1360, World Bank.

Hernandez, Leonardo, and Heinz Rudolf (1994), "Domestic Factors, Sustainability, and Soft Landing in the New Wave of Private Capital Inflows," unpublished manuscript, World Bank.

International Monetary Fund (2007), *World Economic Outlook* (October).

Laban, Raul, and Felipe Larrain (1993), "Can a Liberalization of Capital Outflows Increase Net Capital Inflows?" unpublished manuscript, Department of Economics, Pontifica Universidad Catolica de Chile.

Mathieson, Donald J., and Liliana Rojas-Suarez (1993), "Liberalization of the Capital Account: Experiences and Issues," occasional paper 103, International Monetary Fund.

Montiel, Peter J. (1996), "Managing Economic Policy in the Face of Large Capital Inflows: What Have We Learned?" in G. Calvo and M. Goldstein, eds., *Private Capital Flows to Emerging Markets after the Mexican Crisis*, Washington, DC: Institute for International Economics, pp. 189–218.

Mussa, Michael (1986), "The Effects of Commercial, Fiscal, Monetary, and Exchange Rate Policies on the Real Exchange Rate," in S. Edwards and L. Ahmed, eds., *Economic Adjustment and Exchange Rates in Developing Countries*, Chicago: University of Chicago Press, pp. 43–88.

Ostry, Jonathan (1991), "Trade Liberalization in Developing Countries, Initial Trade Distortions, and Imported Intermediate Inputs," *IMF Staff Papers*, vol. 38, pp. 447–480.

Schadler, Susan, Maria Carkovic, Adam Bennett, and Robert Kahn (1993), "Recent Experiences with Surges in Capital Inflows," occasional paper 108, International Monetary Fund.

Taylor, Mark P., and Lucio Sarno (1997), "Capital Flows to Developing Countries: Long- and Short-Run Determinants," *World Bank Economic Review*, vol. 11, pp. 451–470.

Thomas, Vinod, John Nash, and Associates (1991), *Best Practices in Trade Policy Reform*, London: Oxford University Press.

World Bank (1997), *Private Capital Flows to Developing Countries: The Road to Financial Integration*, New York: Oxford University Press.

PART 7

Varieties of Emerging-Market Crises

25

Sovereign Debt Crises

The process of *emergence* – of becoming integrated with international financial markets – has not been a smooth one for developing countries. As we have seen over the last several chapters, capital flows to these economies have been episodic, and inflows of foreign capital have often been disruptive, posing significant policy challenges. But even more serious have been the consequences of capital-flow reversals, which have become known as *sudden stops*.[1] These have often been associated with severe financial crises, which have frequently created substantial economic hardship for emerging-market economies in the form of sharp output contractions.

A *financial crisis* can loosely be defined in general terms as a situation in which a macroeconomically important economic agent fails to fulfill its financial obligations. Macroeconomically important agents include the government, the central bank, or a significant portion of the banking system. Such agents undertake financial obligations in the form of commitments to exchange financial assets for each other on some preannounced terms, that is, at par. For example, governments commit to redeeming their debt for currency at par, central banks to exchanging domestic currency for foreign currency at par, and banks to exchanging their deposits for currency at par. But it can happen that any of these agents may become unable or unwilling to meet its obligations. This creates a crisis because credible commitments to exchange assets at par sustain the relative values of the assets being exchanged, and the abrogation of such commitments thus gives rise to the prospect of discrete changes in asset prices and large redistributions of financial wealth, creating the incentives for massive portfolio reallocations by individuals seeking to avoid large losses or make large gains.

[1] The term is from Calvo (1998).

In this part of the book, we will study the emergence of three types of financial crises in emerging economies:

- *sovereign debt crises*, which arise when a government's ability or willingness to retire its debt (exchange it for currency) according to originally contracted terms comes into question
- *banking crises*, associated with an inability by commercial banks to redeem their deposits for currency at par
- *currency crises*, which emerge when central banks' ability to redeem their notes for foreign exchange at par comes into question

Because all these crises are macroeconomic in character, they may interact with each other. For example, the emergence of a sovereign debt crisis impairs not just the value of the government's debt but potentially also that of all assets within the government's political jurisdiction because the government may tax such assets (or in the limit, confiscate them) in seeking to meet its own financial obligations (that is why the government's risk premium typically sets a floor on the risk premium paid by private borrowers). This being so, the emergence of a sovereign debt crisis would trigger portfolio reallocations out of domestic assets and into assets outside the government's political jurisdiction, that is, into foreign assets (a phenomenon known as *capital flight*). Because this means a run both on banks and on the central bank's foreign exchange reserves (under fixed exchange rates), it is likely to trigger both banking crises and currency crises. A banking crisis may also be triggered if banks hold a large amount of government debt. The likelihood that a debt crisis will spill over in this way depends in principle on the country's stock of reserves, on its degree of capital mobility, on the soundness of its banks, and of course, on the government's policy response to the crisis.

Similarly, a banking crisis can trigger both debt and currency crises. The emergence of a banking crisis induces portfolio reallocations into all types of nonbank assets, including foreign assets. This directly creates pressure on the central bank's reserves, increasing the likelihood of a currency crisis. The banking crisis may impair the solvency of the entire banking system (whether or not the system was solvent when the crisis broke out), which may trigger a debt crisis if the government has to recapitalize the system and, in doing so, impairs its own ability to meet its financial obligations. In turn, this would give rise to capital flight, which would magnify the banking crisis and increase the probability of a currency crisis. The likelihood of spillovers in this case depends on the government's solvency as well as on the size of the banking crisis.

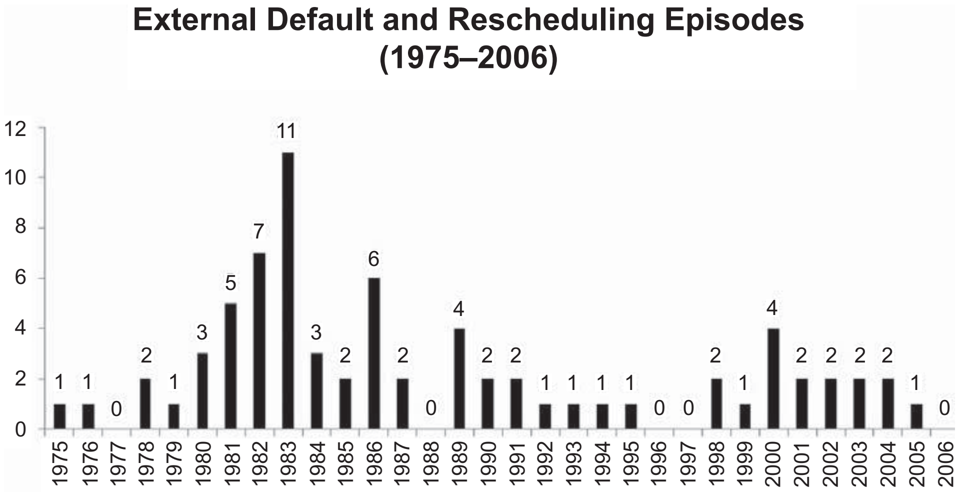
Finally, the emergence of a currency crisis triggers portfolio reallocations into foreign currency–denominated assets. Because agents would sell domestic-currency claims on domestic banks, this makes a banking crisis more likely. Moreover, changes in currency values may impair the financial health of the government, of banks, or of domestic firms if their foreign currency–denominated liabilities exceed their assets (i.e., if they suffer from *currency mismatches*). If firms are affected, then effects on the government and banks will be magnified through falling tax revenues and nonperforming loans, increasing the likelihood of both debt crises and banking crises. In this case, the emergence of spillovers depends on factors such as the severity of currency mismatches, the financial health of banks and firms, and the composition of the government's balance sheet.

It follows that the macroeconomic effects of any of these types of crises depend on the economy's vulnerability to each of the other crises. In particular, as we will see in this chapter and the next, severe real output effects are associated with debt and banking crises, so the consequences of currency crises for the performance of the real economy depend on the extent to which they trigger banking or debt crises as well as on how severe these induced crises turn out to be. An important distinction between currency crises in industrial countries and those in emerging and developing countries is precisely in the nature of these links, as we shall see in this part of the book.

We begin our study of emerging-market crises in this chapter by examining sovereign debt crises. Recall from Chapter 10 that credit risk has to do with the ex ante probability that the government may not meet its debt-servicing obligations. But what happens to the economy if the government indeed finds itself ex post in a situation in which it is unable to service its debt on the terms it had agreed on with its creditors? In that case, the government becomes insolvent, and a sovereign debt crisis erupts. In the context of a sovereign debt crisis, the excess of the government's outstanding debt over the present value of its debt-servicing capacity is referred to as a *debt overhang*. In this chapter, we will explore the macroeconomic implications of the emergence of sovereign debt crises and debt overhangs. Section I defines the term *debt overhang* more precisely and considers the likely macroeconomic effects of debt crises and debt overhangs. Section II turns to the evidence concerning the effects of debt crises on real output, and Section III conducts a case study on the developing-country debt crisis of the 1980s.

I. SOVEREIGN DEBT CRISES AND DEBT OVERHANGS

Our analysis in Chapter 9 indicated that there may exist a revenue-maximizing rate of inflation, that is, a rate of inflation that maximizes the real revenue that the government can collect from the inflation tax. This maximum revenue imposes a natural limit to the contribution that seignorage can make to debt service. However, the adverse real effects of high inflation that we examined in Chapter 13 give governments good reason for stopping short of collecting the maximum feasible inflation tax. The implication is that if governments want to reduce their reliance on the inflation tax while retaining solvency, they have no choice but to undertake a fiscal adjustment – a reduction in spending or increase in revenue collection.



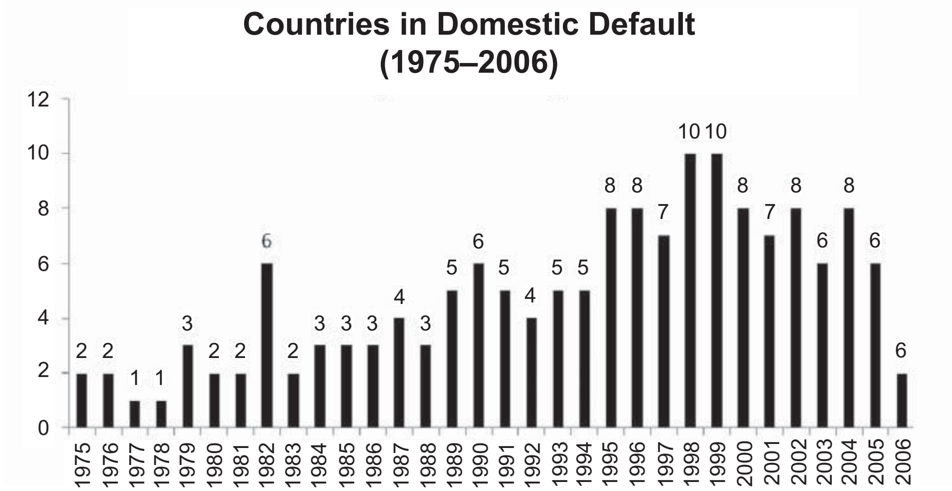


Figure 25.1. External and domestic default among emerging and developing economies, 1975–2006

The problem is, of course, that fiscal adjustment may be politically unpopular and thus difficult to achieve. This raises the possibility of opting for a third alternative – that of sacrificing the public sector's solvency through the nonpayment of public-sector debt. The attractiveness of this option, of course, depends on the costs associated with nonpayment compared with those of fiscal adjustment and/or reliance on the inflation tax. However, it is important to note that debt defaults have not been rare among emerging and developing economies. Figure 25.1 shows the incidence of external (Figure 25.1, top) and domestic (Figure 25.1, bottom) debt default episodes (debt crises) among emerging and developing economies from 1975 to 2006, based on data from Reinhart and Rogoff (2008a, 2008b).[2] As is evident from the chart, on average, some 8–10 default episodes on the external domestic debt of these countries have tended to occur per year.

1. The Debt Overhang

As a first step in exploring the macroeconomic implications of these events, we can develop the concept of debt overhang more precisely. As already mentioned, this term refers to the excess of the face value of the public sector's debt over the present value of the debt service payments that the public sector is prepared to make over the indefinite future. Using the notation of Chapter 9, recall that the sustainable value of the adjusted primary surplus of the consolidated public sector can be written as follows:

$$ps' = (r - n)d(0) - (\pi + n)m(r + \pi)$$

where ps' denotes the adjusted primary surplus of the consolidated public sector, r is the (real) risk-free interest rate, n is the economy's sustainable growth rate, π is the desired rate of inflation, and m denotes the base of the inflation tax. Now suppose that the macroeconomic costs of inflation impose a maximum value π_{max} on the long-run rate of domestic inflation that the government is willing to tolerate and that the social costs and/or political difficulty of fiscal adjustment place an upper limit ps'_{max} on the share of the adjusted primary surplus of the public sector in gross domestic product (GDP). Then, as in Chapter 9, we can derive the maximum sustainable value of the public sector's net debt (relative to GDP) as

$$d_{max} = [ps'_{max} + (\pi_{max} + n)m(r + \pi_{max})]/(r - n)$$

A debt overhang exists when $d(0) > d_{max}$, that is, when the existing debt is larger than the stock of debt that the public sector is willing and able to service on market terms.

How could such a situation emerge? Clearly, for reasons described in Chapter 9, the public sector's creditors would not willingly allow an accumulation of debt in excess of d_{max}, if they could anticipate it. Thus a debt overhang is likely to emerge as the result of an unanticipated change in circumstances. These could involve, for example, changes in economic circumstances that reduce the maximum primary surplus that the government is able to generate or in the inflation tax revenue that can be produced with the maximum tolerable rate of inflation π_{max}, a change in the time path of expected real interest rates or the country's growth rate, or even a change in political regime that alters the government's perceived willingness to produce ps'_{max} or π_{max}, even under unchanged economic circumstances.

[2] The specific countries are identified in Appendix 25.1.

2. Macroeconomic Effects of the Debt Overhang

What happens when a debt overhang emerges? One answer is that the public sector becomes insolvent, but this, of course, is tautological. By *definition*, the existence of a debt overhang is equivalent to insolvency. The interesting question is how the emergence of a debt overhang in the accounts of its consolidated public sector affects a country's macroeconomic performance.

One possibility is that the public sector's creditors simply write off the excess of $d(0)$ over d_{max}, in effect eliminating the debt overhang by adjusting the face value of the debt to more accurately reflect the public sector's ability to pay.[3] In that case, the only likely lasting consequence of the emergence of the debt overhang concerns creditors' view of the country's creditworthiness. As indicated earlier, creditors presumably allowed the outstanding stock of debt to exceed d_{max} in the first place because they had a more optimistic view of $ps'_{max} + (\pi_{max} + n)m(\pi_{max})$ than turned out to be justified after the fact, and the question then becomes what lessons creditors draw about the public sector's future reliability as a debtor from the need to revise their expectations. If these lessons are adverse, then the consolidated public sector may find itself restricted to a very limited supply of financing in the future.

However, there are several difficulties with arranging debt write-offs of this type:

- Creditors may be reluctant to write off debt because they are afraid to set a precedent for other debtors if they are too willing to arrange a write-off with any single debtor.
- Creditors may believe that some third party may be willing to bail out the debtor if they take a hard line and insist on full repayment.
- Writing off debt may be plagued by free-rider problems. This is so because it is in the interest of each individual creditor to have others write off his or her claims, thus leaving more of the public sector's limited resources available to service his or her own claims.

For all these reasons, individual creditors may be unlikely to voluntarily agree to undertake debt write-offs in the absence of some external coordinating mechanism. In a domestic economy, when the insolvent debtor is a firm, this coordinating mechanism is usually provided by a *bankruptcy* proceeding – a judicial proceeding in which an insolvent firm's assets are allocated to its creditors. However, when the insolvent debtor is a sovereign, as in the case of a country's consolidated public sector, matters are more complicated because there is no overriding legal authority

[3] It is important to distinguish this situation, in which part of the face value of the debt is written off permanently, from *debt rescheduling*, in which creditors essentially extend new loans to tide over a borrower that is temporarily having difficulties servicing debt. Rescheduling makes sense only if the borrower is perceived as solvent – i.e., as ultimately able to service the debt on market terms – and thus would not be willingly undertaken by creditors in the presence of a debt overhang.

that can impose an orderly workout of the competing claims on the sovereign's assets. The situation is especially problematic when the debt is denominated in foreign currency and creditors have legal recourse outside the country's own political jurisdiction, as in the case of external debt.

Suppose, then, that – perhaps because of the absence of such a coordinating mechanism – the debt overhang persists. This means that creditors do not give up their claims, but nonetheless, the public sector does not commit itself to servicing the debt as originally agreed. One immediate implication of this situation is that the public sector would no longer be able to borrow on market terms. The reason is that any creditors that extended funds to an insolvent government would immediately find themselves part of the pool of claimants – on par with the existing creditors – for the fixed amount of resources the government has available to service debt.[4] Because the share of the existing debt being serviced is $d_{max}/d(0)$, the value of the new creditor's claims would immediately collapse to a proportion $d_{max}/d(0)$ of their face value. Because the face value of these resources can be preserved at par by lending elsewhere at the riskless rate of interest, no creditor would have an incentive to willingly lend to an insolvent government. In this situation, therefore, overall deficits of the consolidated public sector could only be financed through seignorage.

But the existence of a debt overhang may also have direct effects on economic growth. To see how, note that we can measure the magnitude of the debt overhang in present-value (stock) terms as the ratio of the existing debt $d(0)$ to the present value of the resources that the consolidated public sector can make available for future debt servicing, given by d_{max}. Alternatively, we can measure the overhang in flow terms as the permanent flow of resources by which the government would need to augment its debt-servicing budget to eliminate the overhang, that is, to service the debt on market terms. Call this flow overhang measure τ. Then τ is given by

$$\tau = (r - n)(d(0) - d_{max})$$

where τ measures the additional fiscal effort (expenditure cuts or tax increases) that would be required each period to satisfy the public sector's creditors. For concreteness, we can just think of τ as indicating the additional tax revenue that the government would need to generate each period to satisfy its creditors.[5]

Notice that as long as creditors do not relinquish their claims on the domestic public sector, τ represents a potential additional future tax burden on domestic

[4] This assumes that new creditors are not granted *senior status*, i.e., are not paid before existing creditors. Legal covenants in syndicated loan contracts typically require existing creditors to share payments, and default clauses may be triggered if new creditors are paid before existing ones.

[5] For the subsequent argument to be made, it makes no difference whether the additional fiscal effort is conceived of in the form of additional tax revenue or reduced expenditures.

economy activity. In the presence of a debt overhang, any activity that is undertaken in the expectation of future returns (such as the accumulation of physical capital) thus faces a disincentive in the form of potentially higher future taxes. This disincentive can be expected to have a negative effect on domestic investment and thus on economic growth.

But that is not all. After all, if the government were to eliminate the debt overhang by announcing its intention to make a stronger fiscal effort – that is, by somehow increasing ps'_{max} – this disincentive would still be there. The fact that it appears in the form of a debt overhang (an unallocated resource shortfall rather than one that is eliminated by allocating it) may make it more onerous as a form of prospective taxation, for three reasons:

1. The process that creditors undertake to extract resources from an unwilling government may itself be costly (e.g., through disruption of trade), potentially making the permanent flow costs of transferring the resources to the creditors *exceed* τ. This potential excess burden would magnify the disincentive effect on investment of the public sector's debt obligations when they appear in the form of a debt overhang rather than just as debt to be serviced.
2. The size of the total resource costs that will eventually be incurred in making the transfer to creditors is unknown before the fact. Thus the existence of a debt overhang creates substantial *macroeconomic* uncertainty.
3. Even if the total resource cost were known, its allocation among domestic agents would tend to be unknown before the fact. Thus, in addition to macroeconomic uncertainty, the emergence of a debt overhang also creates *microeconomic* uncertainty. This, of course, means that the debt overhang is associated with a significant increase in uncertainty about future tax obligations at the microeconomic level.

As we have already seen, higher uncertainty tends to discourage investment by increasing the value of the option to wait until the uncertainty is resolved and has more general disincentive effects on any activity that involves incurring a sunk cost in the expectation of future returns such as the costly reallocation of productive resources. For all these reasons, the existence of a substantial debt overhang would tend to have negative effects on economic growth through adverse effects both on factor accumulation and on growth of total factor productivity.

It is worth emphasizing that these negative effects on growth can be magnified through a variety of macroeconomic channels. As mentioned earlier, the anticipation of potentially higher future taxes is a burden that falls on productive activities within the tax authority of the domestic government and can therefore be evaded by moving assets abroad. This means that a debt overhang creates incentives for capital flight. But massive amounts of capital leaving the country imperils both the stability of the domestic financial system and that of the exchange rate, creating the potential for banking and currency crises.[6] If capital flight results in a banking and/or currency crisis, then the instability in the domestic macroeconomic environment rises to a higher order of magnitude. As discussed in Chapter 3, such instability is very harmful for long-term growth.

II. DEBT CRISES AND REAL ECONOMIC ACTIVITY: EVIDENCE

How important have these debt-overhang effects been empirically in emerging economies? In the rest of this chapter, we will consider two types of evidence: cross-country econometric evidence and a case study of the international debt crisis of the 1980s. This section describes a recent cross-country study by Sturzenegger (2004), and the section that follows examines the crisis of the 1980s.

Sturzenegger (2004) examined a sample of 100 countries over the period 1974–1999. He implemented two empirical approaches to examine the growth effects of debt default in this sample, using cross-sectional and annual data. In the cross-sectional approach, he looked at the effect of default on the country's average growth performance after controlling for a large number of potential growth determinants, including initial *GDP* per capita, the ratio of investment to *GDP*, the size and rate of growth of each country's population, the rate of growth of government consumption, the initial level of education, an indicator of civil unrest, the change in the terms of trade, a measure of openness, the average rate of inflation, a measure of inflation volatility, and the incidence of banking crises. He used two indicators of default: a dummy variable indicating whether the country ever defaulted over the sample period and a dummy variable that took on the value of unity if the country defaulted in the 1980s or 1990s, the value of 2 if it defaulted in both periods, and the value of zero otherwise. He found that defaulters grow by 0.6 percent per year slower than nondefaulters, suggesting that the experience of default indeed penalized long-run growth.

However, one problem with the cross-sectional approach is that the default dummies may not capture the effects of default at all but rather may be proxying for omitted fixed country characteristics that simultaneously induce debt default and poor growth performance. Sturzenegger (2004) attempted to address this problem by using a fixed-effects estimator with annual data. Under this approach, the country dummy variables would pick up the growth effects of unchanging country characteristics that may induce a spurious correlation between the default experience and low growth. To address the fact that the growth effects of default may be prolonged over time, Sturzenegger introduced two new default dummies: the first

[6] Recall from Chapter 22 that even a healthy financial system is potentially vulnerable to failure if subjected to a sudden and large withdrawal of funds.

took on the value of unity in the default year and in the year that followed, whereas the second took on the value of unity in the default year and the next *five* years. He introduced separate dummies for the decades of the 1980s and the 1990s. The results were consistent with the cross-country evidence: the short-run dummy was associated with a reduction of the growth rate by 2 percent per year in the default year and the subsequent year, whereas the longer-run dummy was associated with a growth reduction of 0.8 percent per year on average over the six-year period.

III. THE DEBT CRISIS OF THE 1980S[7]

The international debt crisis of the 1980s provides an interesting case study with which to supplement this cross-country evidence. In this section, we review the experience of several of the major emerging economies that were most severely affected by the international debt crisis during that decade and interpret it in light of the mechanisms described in Section I, with a view to determining whether the events of that decade were consistent with what the results of the previous section would lead us to expect.

1. Run-up to the Debt Crisis

The decade of the 1970s was one during which international inflation was high and world nominal interest rates were relatively low. Thus the real interest rates faced by developing countries, calculated by deflating nominal interest rates either by the dollar export prices or import prices of these countries, were relatively low (and often negative), at a time when many such countries were growing rapidly. This implies that for many developing countries, perceived sustainable growth rates were in excess of the real interest rate at which these countries could borrow in world financial markets ($n > r$).

Under these conditions, solvency did not impose an effective constraint on the external borrowing of the public sectors in these countries because debt servicing could be financed with new borrowing – that is, without using the countries' own resources – without approaching the constraint $d(t) < d_{max}$. Not surprisingly, the governments of several large developing countries rapidly increased their stocks of debt relative to *GDP* over the course of the decade. Table 25.1 shows the evolution of the stock of total public debt relative to *GDP* for seven heavily indebted countries over the six-year period 1976–1981.[8] As Table 25.1 shows, ratios of public debt to *GDP* were particularly large in the major Latin American countries, but the phenomenon of rapid debt accumulation was not limited to countries in Latin

[7] This section draws on Montiel (1992).
[8] These are the seven largest economies in the International Monetary Fund's (IMF's) category of heavily indebted countries, used by the IMF for analytical purposes during the 1980s.

Table 25.1. *Seven Heavily Indebted Countries: Ratios of Public Debt to GDP, 1976–1981*

	1976	**1977**	**1978**	**1979**	**1980**	**1981**
Argentina	15.0	18.0	18.7	20.1	21.3	29.6
Brazil	19.8	20.9	21.3	22.2	28.5	30.1
Chile	36.1	26.5	30.1	26.8	21.5	18.8
Mexico	22.0	31.4	32.2	30.8	29.1	34.3
Nigeria	3.9	5.8	9.4	12.2	13.9	16.9
Philippines	31.0	33.8	36.8	35.5	34.6	36.3
Venezuela	11.7	17.2	28.6	35.2	36.3	40.0

Note: Values are in percentages.
Source: Guidotti and Kumar (1991).

America, as shown by the experiences of Nigeria and the Philippines.[9] Thus the combination of low real interest rates and high growth rates during the decade of the 1970s permitted many governments in emerging economies to accumulate a sizable stock of external debt by the beginning of the decade of the 1980s.

2. Triggering Events

The oil price shock of 1979–1980 fundamentally changed this situation. The combination of high oil prices and tight monetary policies implemented in industrial countries to combat their potentially inflationary impact resulted in two important changes in the economic environment in developing countries in 1981:

1. International real interest rates rose sharply.
2. Developing-country growth slowed substantially because of a variety of factors. These included a growth slowdown in industrial countries, adverse changes in the developing countries' terms of trade, and poor domestic policies in those countries.

Figure 25.2 shows the sharp reversal in 1981 of the relationship that had existed during the late 1970s between international real interest rates and the average growth rate among highly indebted developing countries. International real interest rates began to rise in 1980, just as growth among heavily indebted developing countries slowed markedly. The real interest rate first exceeded the average growth rate in 1981, and the gap widened thereafter. The implication of this development was that to retain solvency, the public sectors in highly indebted developing countries

[9] Notice that Chile is an exception within Latin America, having reduced its ratio of public debt to *GDP* substantially during this period. This was an outcome of the fiscal austerity imposed by the Pinochet government.

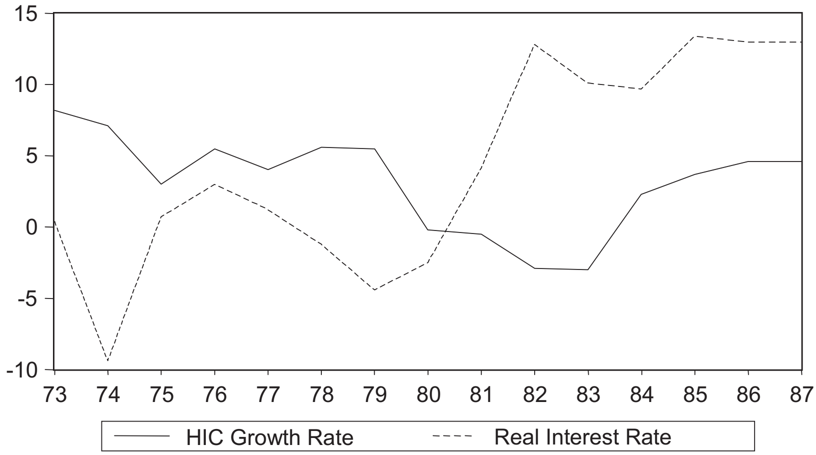


Figure 25.2. Real interest rate and growth rates in the heavily indebted countries, 1973–1987

found themselves with the need to generate large fiscal adjustments. An indication of the magnitude of the required adjustments is provided in Table 25.2. Using the formula

$$ps' = (r - n)d(1981) - (\pi + n)m$$

the second column of values of the table estimates the sustainable value of the primary surplus required to service the debt stocks from Table 25.1 for each of the countries in that table. This can be compared to the average values of the primary surplus generated by these countries during the late 1970s, given in the first column of values.[10] It is evident that substantial fiscal adjustments would have been required in all these cases for the heavily indebted public sectors to retain their solvency under the new circumstances.

3. Policy Responses

What most heavily indebted countries actually did initially was to implement relatively small fiscal adjustments, mainly through reduced public investment, and to rely on an increase in the inflation tax. These fiscal measures proved not to be enough to enable the public sectors in these countries to be able to continue servicing their debt on market terms. Thus the fiscal adjustments actually undertaken were insufficient to restore public-sector solvency in many of the heavily indebted countries.

[10] For Argentina, Chile, and Mexico, these are averages for 1974–1982. For the Philippines, the average is for 1981–1982 only. For the remaining counties, the average pertains to 1976–1982.

Table 25.2. *Actual and Sustainable Primary Surpluses for Seven Heavily Indebted Countries, 1982*

Primary Surpluses (in Percentage of *GDP*)	Actual	Sustainable
Argentina	–7.6	7.3
Brazil	–5.5	1.9
Chile	1.9	6.9
Mexico	–4.7	2.7
Nigeria	–	2.2
Philippines	–4.7	3.6
Venezuela	2.6	1.9

Source: Montiel (1992)

Figure 25.3 illustrates what happened, using the tools we developed in Chapter 9. Recall that the (flow) budget constraint of the consolidated public sector could be expressed as follows:

$$\Delta d = -(ps' + (\pi + n)m) + (r - n)d$$

As we saw in Chapter 9, taking ps', π, r, and n as permanent values of the relevant variables, the relationship between Δd and d can be plotted as a straight line in (d, Δd) space, with slope $(r - n)$ and vertical intercept $-\,(ps' - (\pi + n)m)$. Suppose

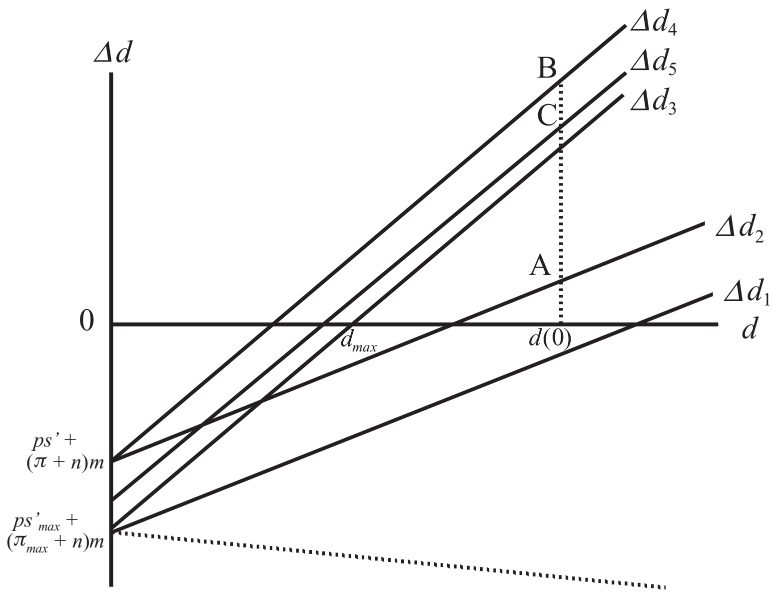


Figure 25.3. Fiscal aspects of the international debt crisis

that $ps'_{\max}$ and π_{max} are the maximum sustainable values of the adjusted primary surplus and domestic inflation rate, respectively, consistent with a maximum debt-*GDP* ratio of

$$d_{\max} = [ps'_{\max} + (\pi_{\max} + n)m(r + \pi_{\max})]/(r - n) \quad \text{for } r - n > 0$$

Graphically, $d_{\max}$ can be found as the horizontal intercept of a line with slope $(r - n)$ that cuts the vertical axis at $-[ps'_{\max} + (\pi_{\max} + n)m]$. During the 1970s, when $r - n < 0$, such a line would have had a negative slope (such as that of the dotted line in Figure 25.3) and thus would not have intersected the horizontal axis at positive values of d, implying that the public-sector solvency condition did not constrain the amount of debt that governments could accumulate, as we have seen.

When $(r - n)$ turned positive in 1980, the slope of the line from $-[ps'_{\max} + (\pi_{\max} + n)m]$ turned positive as well, as in the line marked Δd_1 in Figure 25.3. Up to this time, the public sectors in the heavily indebted countries were incurring new debt; that is, their economies were located at a point such as A in Figure 25.3 on the line marked Δd_2, with existing debt given by $d(0)$. As long as the stock of debt in existence at that time lay to the left of the intersection of Δd_1 with the horizontal axis, the highly indebted countries remained solvent. In 1982, however, the value of $(r - n)$ increased sharply. Thus the line from $-[ps'_{\max} + (\pi_{\max} + n)m]$ rotated further in a counterclockwise direction, to a position such as Δd_3 (with higher r and lower n, the intercept would also have moved up, but we can ignore this for now). At this point, $d(0) > d_{max}$, and the heavily indebted countries would have become insolvent in the absence of a fiscal adjustment. A fiscal adjustment intended to retain solvency would have shifted the budget line down, but such a fiscal adjustment proved not to be feasible; that is, it would have required $[ps' + (\pi + n)m] > [ps'_{\max} + (\pi_{\max} + n)m]$. Thus the actual fiscal adjustment turned out to be partial – only to, say, line Δd_5. Because $d(0)$ remained to the right of the horizontal intercept, the public sectors of these countries became technically insolvent. Not surprisingly, when secondary markets were created for the sovereign debt of developing countries in the mid-1980s, much of this debt sold at a substantial discount from its face value.

4. Macroeconomic Outcomes

The previous section suggested that the emergence of fiscal insolvency should have had negative effects on economic growth, through adverse effects on investment. What actually happened in these countries during the 1980s? First, as was to be expected, domestic investment contracted severely in the heavily indebted countries that faced fiscal insolvency. The data are presented for six countries in Figure 25.4. Using 1980 as a base year, Figure 25.4 shows the behavior of the investment-*GDP* ratio in six of these countries in index form. As shown in the figure,

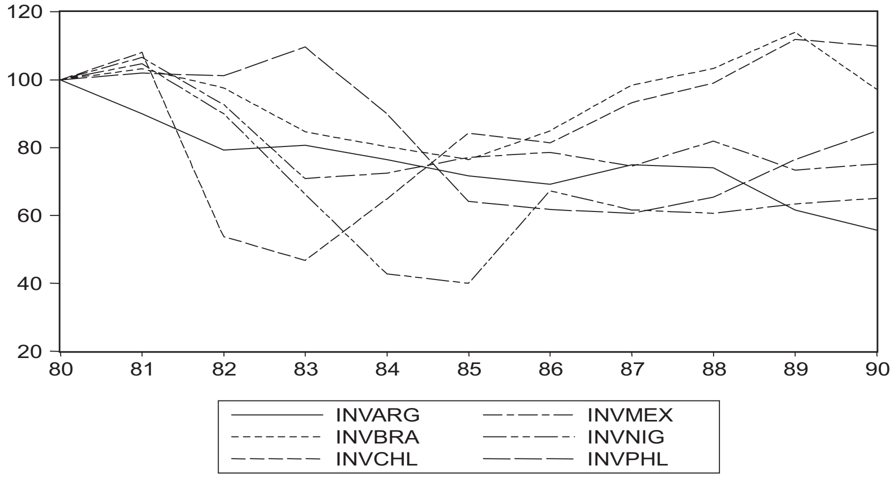


Figure 25.4. Investment ratios for six heavily indebted countries, 1980–1990

investment-*GDP* ratios fell in 1982 in all these countries, except the Philippines, where the decline did not materialize until 1984. More important, the ratio of investment to *GDP* remained below its 1980 value beyond the end of the decade in all the countries, except Brazil and Chile, where the ratio did not recover its 1980 value until 1989.

As one might expect, this investment collapse was associated with a significant reduction in the rate of growth of real *GDP*. The rate of growth of real *GDP* dropped so much in these countries that the growth rate of *GDP* per capita actually became *negative* for many of them during the 1980s, leaving them with lower incomes per capita at the end of the decade than at the beginning (Figure 25.5). This experience

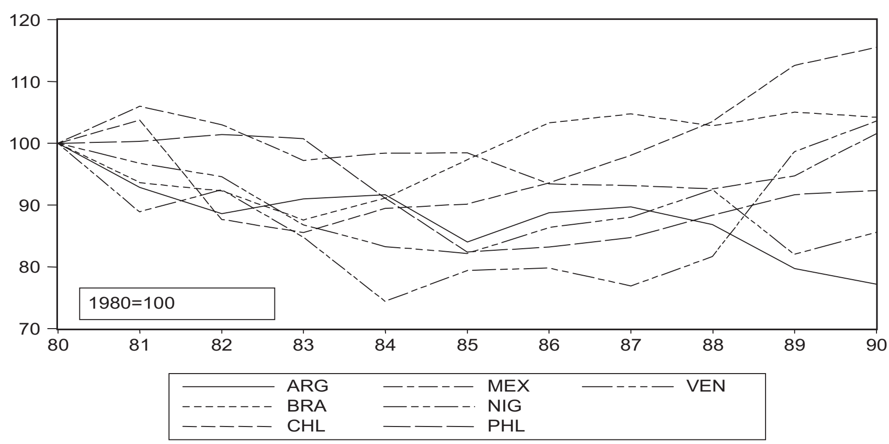


Figure 25.5. ***GDP*** **per capita in seven heavily indebted countries, 1980–1990**

caused many in Latin America to refer to the decade of the 1980s as the "lost decade."

The debt crisis was ultimately resolved – at least for the major heavily indebted market borrowers – through the Brady Plan, proposed by U.S. Treasury secretary Nicholas Brady and implemented beginning in 1989. This plan involved a partial write-down of the external debt of these governments, coordinated by the U.S. Treasury and the international financial institutions, in return for a series of policy undertakings by the indebted governments and a series of debt "enhancements" (guarantees) extended by the coordinating institutions.

This strategy for resolving the crisis can be readily interpreted in terms of the analysis in this chapter. Going back to Figure 25.3, note that the crisis arose because of the emergence of a debt overhang (an excess of $d(0)$ over d_{max}). Each of the three components of the Brady strategy would have contributed to eliminating this overhang by closing the gap between $d(0)$ and d_{max}. First, the enhancements provided by the industrial countries and the international financial organizations in effect increased the resources available to the heavily indebted countries for debt service. Thus they can be interpreted as having effectively supplemented the fiscal resources generated by those countries themselves (see Box 25.1); that is, they were equivalent to an increase in ps'_{max}. In terms of Figure 25.3, such an increase would cause the budget constraint line Δd_5 to shift downward. Second, the growth-enhancing measures that the countries participating in the Brady plan were encouraged to adopt had the intended effect of increasing n, thus reducing $r - n$ and causing the budget constraint line to rotate in a clockwise direction. The combination of these first two measures would tend to increase d_{max} for the participating countries. Finally, the third element – the write-down of existing debt – would close the gap by reducing $d(0)$. Thus the plan essentially attempted to close the gap between $d(0)$ and d_{max} by moving each of the two components in the direction of the other.

5. Interpreting the Evidence

The general outline of the experience of the heavily indebted countries during the 1980s thus turns out to have been in line with what the analysis of Section I would have led us to expect. The prospective fiscal insolvency that emerged in these countries after 1981 appears to have had a severe impact on the countries' subsequent growth performance, at least until the crisis was resolved at the end of the decade.

But though the evidence is strongly suggestive of such effects, it is not conclusive. The difficulty in interpreting the evidence is that the decade of the 1980s was a period of severe macroeconomic disruption in all these countries, featuring adverse behavior in their terms of trade and a host of poor domestic policy

Box 25.1. Postcrisis Fiscal Adjustment in Latin America

By the late 1980s, the heavily indebted countries in Latin America exhibited very large differences in their ratios of consolidated public-sector deficits to *GDP*. As described by Edwards (1996), many of them finally succeeded in making very large fiscal adjustments in the early 1990s (e.g., Argentina eliminated a fiscal deficit of 22% of *GDP* in 1989, and Nicaragua one of 58% of *GDP* in 1990, both within three years). But other countries found fiscal adjustment to be much more difficult (e.g., Brazil had achieved reasonably good fiscal outcomes in the mid-1980s, but it actually slipped back into annual inflation of more than 1,000% by 1992, until finally achieving a durable fiscal adjustment after 1994). Fiscal adjustment was sought in these countries through both revenue and spending measures.

The revenue measures included tax reforms designed not just to raise revenue but also to eliminate distortions and simplify the tax structure; improvements in tax administration; increases in public-sector prices; and sales of state-owned enterprises. In most countries, the top income tax rate was reduced, while the minimum rate was increased. At the same time, most countries raised the exemption level (from the equivalent to the average level of *GDP* per capita 1991 to twice that level), with the objective of improving efficiency while reducing regressivity. The marginal tax rate on corporate income tax was also reduced, and the number of corporate tax rates decreased. Many countries adopted a value-added tax (VAT). Others that already had a VAT in place reduced the number of rates and broadened the base of the tax. Finally, most countries adopted the taxation of capital gains as ordinary income and reduced withholding rates on foreign remittances. Edwards concluded that although tax revenue was higher in 1991 than it was in 1987–1988, overall, these reforms had weak effects on total revenues owing to problems with compliance and deficient tax administration.

Consequently, the substantial adjustments that were achieved in primary deficits in these countries were achieved through expenditure reductions. Edwards found that in all countries, except Colombia and Venezuela, public expenditures were much lower in the 1990s compared to their levels in 1987–1988. Significantly, capital expenditures were reduced in drastic fashion in almost every country.

decisions. Thus one can reasonably wonder whether the investment and real output collapses might not have been due to factors other than the debt overhang. Undoubtedly, some of it was. How can we separate the role of debt-overhang effects?

In principle, whether the presence of the debt overhang contributed to the investment collapse can be tested by estimating investment equations for this period that control for other determinants of investment in these countries. Unfortunately, in the case of the debt overhang, this procedure is not straightforward, for two reasons. First, it is difficult to obtain precise measures of the extent of the debt overhang, as opposed to the size of the stock of debt outstanding. Recall that it is prospective insolvency, rather than the debt itself, that creates uncertainty for prospective investors. Second, the effects of the debt overhang may be transmitted through a variety of domestic variables that would typically be included in empirical investment equations, for example, domestic real interest rates, changes in the stock of public capital, real exchange rate variability, and so on. Disentangling the independent effects of the debt overhang is thus a challenging task, and while the experience of these countries supports the notion that the types of mechanisms explored in this chapter can be very important empirically, the case remains to be proven.[11]

IV. SUMMARY

We began this chapter by pointing out that for a prospectively insolvent government, the alternative to making a fiscal adjustment or relying on seignorage revenue to service debt as scheduled is to stop servicing its debt entirely or service it incompletely. The difference between the present value of the debt service payments owed by the government and the payments that it actually intends to make is its debt overhang. If creditors do not forgive the excess debt obligations, the existence of the overhang creates an unresolved future obligation of the government that represents a potentially large and highly uncertain future tax liability for the private sector, thus discouraging costly economic activities that are justified by future payoffs such as the accumulation of physical capital. The upshot is that the existence of a debt overhang can have adverse effects on growth.

We explored the empirical support for this proposition in two ways: by examining some recent cross-country evidence on the growth effects of debt defaults and by using the debt crisis of the 1980s as a case study. The cross-country evidence was consistent with the view that defaults are harmful for growth, and this result was buttressed by the case study. We found that among several countries that encountered debt-servicing difficulties in the early 1980s, events unfolded in a manner quite consistent with what theory would predict, although the magnitude of the harmful effects of the debt overhang on physical investment and economic growth among these countries during the "lost decade" of the 1980s remains a matter of dispute.

[11] E.g., Warner (1992) finds that the behavior of the investment ratio in the heavily indebted countries can be explained entirely without debt-overhang effects, whereas Serven and Solimano (1992) conclude that debt-overhang effects are the most important single determinant of the investment-*GDP* ratio in these countries during the 1980s.

Appendix 25.1. Reinhart-Rogoff Default Episodes

External Default and Rescheduling Episodes	
1975	Angola
1976	Peru
1977	
1978	Peru, Turkey
1979	Nicaragua
1980	Bolivia, Peru, Sri Lanka
1981	Central African Republic, Costa Rica, Honduras, Poland, Romania
1982	Argentina, Dominican Republic, Ecuador, Mexico, Nigeria, Sri Lanka, Turkey
1983	Brazil, Central African Republic, Chile, Costa Rica, Côte d'Ivoire, Morocco, Panama, Philippines, Uruguay, Venezuela, Zambia
1984	Costa Rica, Egypt, Peru
1985	Angola, South Africa
1986	Bolivia, Guatemala, Morocco, Nigeria, Paraguay, Romania
1987	Panama, Uruguay
1988	
1989	Argentina, Bolivia, Guatemala, South Africa
1990	Uruguay, Venezuela
1991	Algeria, Russian Federation
1992	Nigeria
1993	South Africa
1994	Kazakhstan
1995	Venezuela
1996	
1997	
1998	Indonesia, Russian Federation
1999	Ecuador
2000	Côte d'Ivoire, Indonesia, Kazakhstan, Zimbabwe
2001	Argentina, Nigeria
2002	Indonesia, Myanmar
2003	Paraguay, Uruguay
2004	Nigeria, Venezuela
2005	Dominican Republic
2006	

Note: Sovereign default (*external debt crisis*) is defined as the failure to meet a principal or interest payment on the due date (or within the specified grace period). The episodes also include instances in which rescheduled debt is ultimately extinguished in terms less favorable than the original obligation.
Source: Reinhart and Rogoff (2008a).

Years in Domestic Default	
1975	Dominican Republic, Vietnam
1976	Angola, Dominican Republic
1977	Dominican Republic
1978	Dominican Republic
1979	Congo, Dominican Republic, Ghana
1980	Dominican Republic, Mozambique
1981	Dominican Republic, El Salvador
1982	Argentina, Bolivia, Dominican Republic, El Salvador, Ghana, Mexico
1983	Dominican Republic, El Salvador
1984	Dominican Republic, El Salvador, Myanmar
1985	Dominican Republic, El Salvador, Peru
1986	Brazil, Dominican Republic, El Salvador
1987	Brazil, Dominican Republic, El Salvador, Myanmar
1988	Dominican Republic, El Salvador, Panama
1989	Argentina, Dominican Republic, El Salvador, Liberia, Panama
1990	Argentina, Brazil, Dominican Republic, El Salvador, Kuwait, Liberia
1991	Dominican Republic, El Salvador, Kuwait, Liberia, Sudan
1992	Angola, Dominican Republic, El Salvador, Liberia
1993	Angola, Croatia, Dominican Republic, El Salvador, Liberia
1994	Angola, Croatia, Dominican Republic, El Salvador, Liberia
1995	Angola, Croatia, Dominican Republic, El Salvador, Liberia, Rwanda, Solomon Islands, Venezuela
1996	Angola, Croatia, Dominican Republic, El Salvador, Liberia, Solomon Islands, Sri Lanka, Venezuela
1997	Angola, Dominican Republic, Liberia, Mongolia, Sierra Leone, Solomon Islands, Venezuela
1998	Angola, Antigua and Barbuda, Dominican Republic, Liberia, Mongolia, Russian Federation, Sierra Leone, Solomon Islands, Ukraine, Venezuela
1999	Angola, Antigua and Barbuda, Dominican Republic, Ecuador, Gabon, Liberia, Mongolia, Russian Federation, Solomon Islands, Ukraine
2000	Angola, Antigua and Barbuda, Dominican Republic, Gabon, Liberia, Mongolia, Solomon Islands, Ukraine
2001	Angola, Antigua and Barbuda, Dominican Republic, Gabon, Liberia, Solomon Islands, Suriname
2002	Angola, Antigua and Barbuda, Argentina, Gabon, Liberia, Madagascar, Solomon Islands, Suriname
2003	Antigua and Barbuda, Argentina, Dominica, Gabon, Liberia, Solomon Islands
2004	Antigua and Barbuda, Argentina, Cameroon, Dominica, Gabon, Grenada, Liberia, Solomon Islands
2005	Antigua and Barbuda, Argentina, Dominica, Gabon, Grenada, Liberia
2006	Liberia, Zimbabwe

Note: The definition given for external debt (*domestic debt crisis*) in Reinhart and Rogoff (2008a) applies. In addition, domestic debt crises have involved the freezing of bank deposits and/or forcible conversions of such deposits from dollars to local currency.

Source: Reinhart and Rogoff (2008a); see also Reinhart and Rogoff (2008b).

REVIEW QUESTIONS

1. What common feature do financial crises of all types share? What distinguishes sovereign, banking, and currency crises? How might such crises interact with each other?
2. What is a debt overhang? How could its magnitude be measured? Because creditors would not willingly lend money to governments that they do not expect to be in a position to repay, how can debt overhangs emerge?
3. What are the mechanisms through which a debt overhang could affect real economic activity?
4. Explain why debt overhangs are unlikely to be resolved through voluntary actions by private creditors.
5. What role did private creditors, indebted countries, and the international community play in the resolution of the debt crisis of the 1980s?

EXERCISES

1. Explain in intuitive terms why a negative value of $r - n$ would make it possible for governments to become heavily indebted. What would creditors have to believe about the negative value of $r - n$ for your explanation to make sense?
2. The presence of a debt overhang has been said to harm growth not just by deterring investment but also by discouraging the adoption of pro-growth policies in the indebted countries. Can you explain why this might be true?
3. The resolution of the international debt crisis of the 1980s was a resolution only for relatively large emerging economies that had borrowed heavily from private banks. Many low-income countries that had large stocks of debt owed to bilateral and multilateral *public* lenders, typically at fixed and concessional (below-market) interest rates, did not benefit from the Brady Plan and did not have their debt problems addressed by the international community until the much later implementation of the Heavily Indebted Poor Country initiative. Would you expect the debt overhang problems of these countries to have emerged in the same way as those of the emerging-economy market borrowers? Explain why or why not.
4. One method that several indebted countries used to attempt to address their debt-overhang problems during the 1980s was debt-for-equity swaps, in which a creditor would write off part of the debt owed by the country in exchange for an ownership stake in one of the country's state-owned enterprises. Under what conditions would such a swap have eased the government's intertemporal budget constraint?
5. The "lost decade" of the 1980s in Latin America was characterized by high inflation in many of the region's heavily indebted countries. Do you think that the indebtedness and inflation problems were linked? If so, how?

REFERENCES AND FURTHER READING

Calvo, Guillermo (1998), "Capital Flows and Capital-Market Crises: The Simple Economics of Sudden Stops," *Journal of Applied Economics*, vol. 1, pp. 35–54.

Easterly, William (1988), "Fiscal Adjustment in High-Debt Countries," unpublished manuscript, World Bank.

Edwards, Sebastian (1996), "Public Sector Deficits and Macroeconomic Stability in Developing Countries," working paper 5407, National Bureau of Economic Research.

Guidotti, Pablo, and Moammar Kumar (1991), "Domestic Public Debt of the Externally Indebted Countries," occasional paper 80, International Monetary Fund.

Montiel, Peter J. (1992), "Fiscal Aspects of Developing-Country Debt Problems and DDSR Operations," unpublished manuscript, World Bank.

Reinhart, C., and K. Rogoff (2008a), "The Forgotten History of Domestic Debt," working paper 13956, National Bureau of Economic Research.

Reinhart, C., and K. Rogoff (2008b), "This Time Is Different: A Panoramic View of Eight Centuries of Financial Crises," working paper 13882, National Bureau of Economic Research.

Serven, Luis, and Andres Solimano (1992), "Economic Adjustment and Investment Performance in Developing Countries: The Experience of the 1980s," in V. Corbo, S. Fischer, and S. Webb, eds., *Adjustment Lending Revisited*, Washington, DC: World Bank, pp. 117–135.

Sturzenegger, Federico (2004), "Tools for the Analysis of Debt Problems," *Journal of Reconstructing Finance*, vol. 1, pp. 1–23.

Warner, Andrew M. (1992), "Did the Debt Crisis Cause the Investment Crisis?" *Quarterly Journal of Economics*, vol. 107 pp. 1161–1186.

26

Banking Crises

The wide reach of the government's taxing powers and the productivity of public goods help to explain why sovereign debt crises can have dramatic impacts on the economy through anticipations of future fiscal measures. Similarly, the critical role that financial intermediaries play in allocating resources and administering payment systems would suggest that breakdowns in financial intermediation associated with banking crises would tend to have dramatic effects on the functioning of the economy as well. In this chapter, we will see that this is indeed the case. Like sovereign debt crises, banking crises have been both common in emerging and developing economies and especially severe in their macroeconomic effects. In this chapter, we will explore why banking crises happen, from both theoretical and empirical perspectives. When examining the empirical determinants of banking crises, we will consider both cross-country and case study evidence, as we did for sovereign debt crises in Chapter 25.

Figure 26.1 provides a recent compilation of the frequency of banking crises by Kroszner et al. (2007).[1] There are two observations to take away from this figure. First, like sovereign crises, banking crises have not been infrequent. Second, the frequency of banking crises appears to have increased rather markedly after the late 1980s, when financial liberalization took hold in many emerging and developing economies. In the rest of this chapter, we will consider why this may have been so.

The chapter is organized as follows. Section I presents an overview of the analytics of banking crises, focusing on why they happen and why they have real effects on the economy. Section II looks at some cross-country evidence on the determinants of banking crises, whereas section III turns to the case study evidence. Section IV summarizes. In discussing currency crises in Chapter 27, we will round out our discussion of banking crises by considering in detail the interaction between

[1] The specific countries involved are identified in the appendix.

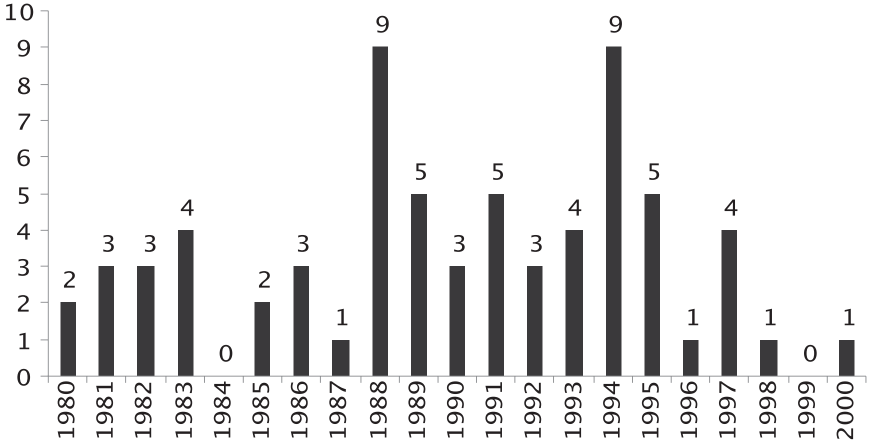


Figure 26.1. Banking crises per year, 1980–2000. *Source*: Kroszner et al. (2007)

banking and currency crises, a phenomenon that has come to be known as the *twin crises.*

I. BANKING CRISES: THEORY

We saw in Chapter 22 that because of the opaque character of their assets and the liquid nature of their liabilities, coupled with the sequential servicing constraint, banks are inherently vulnerable to panics (liquidity crises). While public policy, in the form of a lender of last resort and deposit insurance, can reduce such vulnerability, it can increase moral hazard because policies of this type can diminish the incentives for depositors to monitor their banks. The solution is for the government itself to act as a *delegated monitor*, through the implementation of an appropriate regulatory and supervisory system for banks.

1. Causes of Bank Crises

Unfortunately, however, implementing an appropriate regulatory and supervisory framework is not a trivial matter. Bank regulation may fail for many reasons. The most basic of these is that the legal framework for appropriate bank regulation may be inadequate. The legal basis of regulation must provide regulators the means with which to acquire the information they need to do their jobs (this refers to the adequacy of accounting and disclosure standards applied to banking) as well as the authority to compel compliance with their strictures, when necessary.

But even when the legal means are in place, inadequate knowledge or resources on the part of regulators can impair the regulatory process, particularly after a transition from financial repression or during a period of rapid financial innovation (when neither banks nor regulators are well prepared for loan evaluation and risk assessment).

Even with well-informed and well-funded regulation, regulation may fail for more insidious reasons. First, principal-agent problems may arise between regulators and taxpayers because the interests of the individual regulators may differ from those of the framers of regulatory policy. This may permit regulators to be captured by the banks and may encourage *regulatory forbearance* (a permissive approach to regulation) in the case of bank failure for *bureaucratic gambling* motives (hoping a problem of insolvency will go away through good luck so as not to be blamed for allowing it to happen). Second, regulation may be confounded by political pressures placed on regulatory agencies from politicians' constituencies as well as by corruption.

Inappropriate regulation arising from any of these sources means that banks may become vulnerable to shocks that could wipe out their capital, resulting in solvency crises. Indeed, we can think of bank vulnerability as arising when the shocks to which a bank's balance sheet may be subjected are large relative to the size of the bank's capital. Thus vulnerability can arise when banks have large risk exposures and/or limited capital, both of which, as we have seen, may reflect regulatory failures.

It is obvious from this discussion, however, that vulnerability is a matter of degree. Banks, like other firms, are in the business of taking risks, and they can be more or less vulnerable to shocks in their environment; that is, even a well-functioning banking system can always be thrown into insolvency by a sufficiently large shock (the *100-year flood phenomenon*). What matters is the degree of sensitivity of bank balance sheets to the underlying volatility of the economic environment, and one function of regulation is to adapt prudential standards for bank lending and capital adequacy requirements to the volatility of the environment in which banks operate.

These considerations suggest that regulatory failure is only one possible source of bank vulnerability. The origins of vulnerability can be classified into three types.

a. Bank Specific

This essentially refers to incompetence and/or fraud at the level of individual banks. This type of problem can emerge at individual banks even under an efficient regulatory system, but in the presence of effective mechanisms to prevent the spread of bank panics, problems of this type should not have systemic repercussions.

b. Industry Specific

Regulatory failure can be described as an industry-specific problem. Probably the most common and most important source of regulatory failure arises from the deregulation of bank activities without an adequate regulatory and supervisory mechanism in place, which allows the inappropriate incentives associated with moral hazard to have free rein, as we have seen. But in some cases, this general problem has been aggravated by the way that some aspects of the liberalization process were handled. For example, in some of the banking crises that we shall review in Section III (e.g., Chile in 1982 and Mexico in 1994), the privatization of previously nationalized banks that occurred before the crisis happened had been conducted in such a way as to allow unqualified buyers to acquire banks or to allow inappropriate links between banks and nonfinancial firms. In other cases, previously liberalized and still weak banks were subjected to a loss of franchise value from the allowing of new entry or as the result of competition arising from capital account liberalization.

As we shall see in Chapter 27, some of the major financial crises experienced by emerging economies in recent years may have had their origins in problems of this type. The Southern Cone, Mexican, and Asian crises were all preceded by lending booms, which are a prime symptom of premature domestic financial liberalization, that is, financial liberalization with an inadequate regulatory structure.

c. Macroeconomic

Vulnerability can also arise from macroeconomic sources. For example, it may be created by perceived guarantees (which turn out not to have been valid ex post) for the assets of bank owners, which in effect makes them behave as if their own resources are not at risk. This creates a moral hazard problem that encourages banks to take risks.[2] Somewhat paradoxically, macroeconomic booms may also make banks vulnerable, in the sense of reducing the average quality of the loans in their portfolios.[3] The reason is that maturity-based monitoring devices (i.e., short-term loans that have to be repaid frequently and thus reveal information about the financial health of the borrowers) may fail to provide much information during good times, when most debtors (even bad ones) will be able to make payments on schedule.

Once vulnerability has arisen from any of these sources, a crisis can be triggered by a variety of factors, either microeconomic or macroeconomic in nature. Microeconomic factors include prospective insolvency at one or more large banks, which, as we have seen, may spread panic elsewhere in the banking system when panic-prevention policies are not effective. But even when such policies are effective, macroeconomic shocks can nevertheless trigger banking crises because such shocks can undermine the solvency of vulnerable banks. Macroeconomic sources of banking crises may come in a variety of forms, depending on the nature of banks' exposure.

i. High interest rates. Because they are engaged in maturity transformation, banks' liabilities are likely to be more sensitive to changes in interest rates than their

[2] This may be the most severe potential indictment of crony capitalism in the East Asian case.
[3] See Gavin and Hausmann (1996).

assets. This means that banks are typically exposed to interest rate risk. Moreover, because higher interest rates tend to worsen adverse selection problems among bank borrowers, banks may also restrict credit expansion when interest rates rise. For both reasons, higher interest rates may impair the value of bank capital.

ii. Exchange rate devaluation. If either banks themselves or their customers have significant currency mismatches in their balance sheets, a large devaluation may also impair the value of bank capital, either directly or indirectly.

iii. Asset price collapses. Typically, these involve stocks and/or real estate. The obvious channel through which such events could impair the value of bank capital is through bank ownership of such assets. But even if banks do not own stocks or real estate, a collapse in the price of these assets will lower the net worth of bank borrowers, which, by worsening agency problems, will increase the banks' cost of doing business (i.e., increase the external finance premium) and thus the profitability of banking activity. As we shall see, property booms and collapses featured prominently in some of the crises to be reviewed in Section III.

iv. Recession. The obvious channel through which a recession would hurt banks is through credit risk, that is, the effect of reduced earnings on the capacity for bank debtors to service their debt.

v. Adverse terms of trade shocks. For banks not diversified outside their home countries, adverse terms of trade shocks could trigger many of the negative events listed previously and could directly affect the banks through negative effects on the business prospects of many bank customers.

vi. Unanticipated deflation. A decrease in the average price level that was not anticipated at the time interest rates were set has the effect of increasing ex post real interest rates. This reduces the net worth of bank debtors who are highly leveraged.

vii. Government debt-servicing difficulties. When banks hold large amounts of government securities, the inability of the government to service these obligations puts the banks at risk in exactly the same way as when banks' private customers are unable to repay their loans. As mentioned in the preceding chapter, this is one of the ways that sovereign debt crises can trigger banking crises.

This list suggests that there are many macroeconomic events that can arise to impair the value of bank capital. When any of these events are likely to happen, and when banks lack sufficient capital to allow them to withstand the losses that these events would entail, a generalized banking crisis is likely to emerge.

2. Macroeconomic Effects of Banking Crises

But why worry about this possibility? The key reason, of course, is that banking crises tend to have significant effects on the real economy. Banking crises invariably have fiscal effects because when the government guarantees the value of deposits, the difference between banks' deposit obligations and the value of their assets that emerges when banks are insolvent has to be made up by taxpayers. These are often described as the costs of the crisis. However, this fiscal cost represents a transfer from taxpayers to depositors. Though it does reflect a social cost (that in the absence of the transfer would have been borne by the depositors), it does not reflect a cost of the crisis itself but rather of the resource misallocations that may have helped cause the crisis.

How may the crisis itself affect the real economy? There are various reasons to worry about generalized bank failures. First, generalized bank failures disrupt the economy's payment mechanisms. Banking crises encourage a switch from deposits to cash that, in the absence of countervailing policies, would cause the money supply to contract. Because it economizes on transaction costs, money is socially productive, so a drastic reduction in the money supply has the effect of curtailing economic efficiency.

Second, by reducing the money supply, the switch from deposits to cash that occurs in a banking crisis raises domestic interest rates, which reduces aggregate demand through the traditional channels of monetary transmission. This was represented in our bank-based macro model of Chapter 23 by an increase in the parameter γ. Banking crises are thus likely to trigger recessions or even more severe macroeconomic downturns. The traditional remedy would be expansionary monetary policy, which would not tend to be inflationary because it would simply match the supply of monetary base to the increased demand for it, but as we noted in Chapter 22, fine-tuning this injection of reserves may pose a difficult challenge for the central bank.

Third, monetary policy may also work through a *credit channel*; that is, even in economies with well-developed securities markets, there may be a class of borrowers with agency costs sufficiently high that they have no access to such markets and can only borrow from banks. If so, this will tend to magnify the negative effects on the economy operating through the traditional mechanism of monetary transmission because the disruption of bank credit associated with generalized banking crises will curtail spending among such bank-dependent firms.[4]

[4] Even contractions of bank capital, without outright insolvency, may trigger a decline in bank lending because prudential regulations in the form of capital adequacy ratios link the amount of lending banks do to the amount of capital they hold. If there is a credit channel to monetary policy, this could slow real economic activity. A credit crunch of this sort, because of inadequate bank capital, has been assigned part of the blame for the 1990 recession in the United States.

Finally, to the extent that banks are liquidated as the outcome of a generalized banking crisis, the economy loses the knowledge capital that the liquidated banks had accumulated, which reduces the efficiency of the financial intermediation mechanism.

The cumulative impact of all these effects can be substantial. Caprio and Honohan (2009), for example, reported the typical output loss over the course of a crisis episode (based on a sample of 39 such episodes) at about 14.6 percent of the gross domestic product (*GDP*) of the affected economies.

II. BANKING CRISES: CROSS-COUNTRY EVIDENCE

Thus the potential stakes involved in the adoption of an appropriate policy framework for the domestic financial system are high indeed. We saw in Chapter 21 that both theory and evidence suggest that the real costs of financial repression are paid in the form of slower growth of the economy's productive capacity as resources are misallocated and dissipated and as the incentives for accumulation are reduced. We have just seen, on the other hand, that an inappropriate liberalization of the financial system may also be costly, both by inducing misallocation of resources as a result of incentive problems and by making the banking system vulnerable to crises with potentially severe macroeconomic impacts. But so far, the case for the harmful effects of premature liberalization has been based on a priori reasoning. Next, we will review some of the international evidence on the causes of banking crises. We examine some evidence from cross-country studies in this section and then turn to several case studies in the section that follows.[5]

1. Demigurc-Kunt and Detragiache (1998)

An early cross-country study investigating the empirical determinants of banking crises was by Demigurc-Kunt and Detragiache (1998). The authors relied on annual data drawn from samples of 45–65 countries (depending on data availability for specific equation specifications) over the period 1980–1994. They determined whether each country in their sample experienced a generalized banking crisis during each year in the sample by examining whether any one of four conditions were met. These included whether the ratio of nonperforming loans to total assets in the banking system exceeded 10 percent, whether there was a bank rescue operation during the year with fiscal costs in excess of 2 percent of GDP, whether banking-sector problems resulted in widespread bank nationalizations, and whether extensive bank runs, deposit freezes, or prolonged bank holidays took place in that year or generalized deposit insurance was enacted in response to banking difficulties. If any of these things happened, the year in question was classified as characterized by a

[5] For a recent comprehensive survey of the cross-country evidence, see Eichengreen and Arteta (2002).

generalized banking crisis. They found 31 crisis episodes out of the 546 observations in their sample.

Next, they used a logit regression procedure designed to estimate the factors that influenced the probability that any observation in the sample would be classified as a crisis or noncrisis situation. Among the crisis determinants they considered were indicators of institutional development intended to capture the adequacy of the regulatory environment as well as measures of macroeconomic volatility. Their empirical model performed well in explaining whether observations would be classified as crisis or noncrisis.

They found that a low value of the law-and-order index that they used to indicate the adequacy of the institutional environment helped predict crises and that the presence of deposit insurance *increased* the probability of a crisis in their sample. There was weak evidence that rapid growth of credit to the private sector – such as would emerge in a liberalized environment with inappropriate regulation of banks – predicts subsequent crises. A high ratio of M2 to the central bank's stock of foreign exchange reserves increased the likelihood of a crisis. Among macroeconomic variables, slow domestic *GDP* growth, poor performance of the country's terms of trade, and high domestic nominal and real interest rates were associated with crisis, but exchange rate depreciation and fiscal variables were not.

2. Hutchinson and McDill (1999)

Another well-known cross-country study was by Hutchinson and McDill (1999). They supplemented a statistical study of the determinants of banking crises in a large sample of countries with an examination of the "typical" time series behavior of macroeconomic variables and financial variables in crisis countries. They found that in crisis countries, compared to the others, the precrisis period was characterized by a faster rate of currency depreciation, a higher rate of inflation, and a higher ratio of M2 to foreign exchange reserves. Stock prices were somewhat higher in crisis countries, but fiscal performance showed no noticeable differences. In economies that had crises, real output growth experienced a boom before a crisis but slowed gradually prior to the crisis. It dropped sharply at the onset of the crisis and gradually recovered. Credit growth was strong prior to the crisis, contracted during the first year, and then rebounded slowly. Exchange rate depreciation jumped significantly at the onset of the crisis, and stock prices dropped markedly.

In their examination of these time series patterns, Hutchinson and McDill thus found stronger evidence for the role of macroeconomic booms triggered by periods of rapid credit expansion in triggering subsequent banking crises than had Demigurc-Kunt and Detragiache (1998) in their formal statistical analysis. The formal statistical analysis conducted by Hutchinson and McDill (1999) confirmed these results. Among institutional variables, they found that financial liberalization, deposit insurance, and their interaction all increased the likelihood of a subsequent crisis, whereas greater central-bank independence reduced it. Among macroeconomic variables, slower growth of real *GDP* and lower stock prices helped to explain the incidence of crises.

3. Upshot

The cross-country evidence in these two studies is thus consistent with the perspective adopted earlier in this chapter: financial liberalization in the presence of a weak domestic institutional environment for regulating and supervising banks gives free rein to moral hazard problems in the banking system, resulting in excessive expansion of risky lending. While this may trigger a macroeconomic boom in the short run, it makes banks vulnerable to an adverse turn of macroeconomic events, resulting in an increased likelihood of generalized banking crises.

III. BANKING CRISES: CASE STUDY EVIDENCE

We now turn to some case study evidence on the same issues drawn from a wide range of country experiences.

1. Historical Case Study Evidence for the United States

Because of the prevalence in the United States of *unit banks* (banks with only one branch), which tend to be undiversified, banking failures have been relatively common in American economic history. According to Mishkin (1989), the typical sequence of events in American bank panics has been as follows:

1. Most of these crises started with a sharp rise in interest rates, a stock market crash, and an increase in uncertainty arising after the start of a recession.
2. This led to increased adverse selection (more perceived risk led to higher interest rates, which squeezed out the more creditworthy borrowers) and moral hazard problems (lower firm net worth increased agency costs).
3. Weakened bank balance sheets and higher costs of intermediation gave rise to bank panics, which in turn aggravated adverse selection and moral hazard problems because of the implied reduction in intermediation activity by informed lenders.
4. In the most severe panics, unanticipated declines in the average price level, leading to debt deflation, aggravated this process by further reducing borrower net worth. This happened in 1873 and 1929.

The most severe banking crisis in the United States was during 1930–1933. It was preceded by a credit boom during the decade of the 1920s, by a recession that began in mid-1929, and by the stock market crash in October of that year. The percentage of banks failing in each year from 1930 to 1933 was 5.6, 10.5, 7.8, and 12.9 percent. By the end of 1933, just above half the number of banks that existed in 1929 were still in business. It ended with a bank holiday in March 1933.

2. The Nordic Crises of the Early 1990s[6]

Prior to the early 1980s, financial repression prevailed in Finland, Norway, and Sweden (interest rates were controlled, directed credit was used, capital flows were restricted, and the terms and quantities of bond issuance were restricted). Deregulation, in the form of the rapid removal of these restrictions, was carried out during the early to mid-1980s. At the time, bank capitalization was very low in all three countries, and formal deposit insurance existed in Finland and Norway, as did implicit guarantees in Sweden. No consideration was given to adapting regulatory standards to the new environment. Bank-lending and risk-management practices received no in-depth review during the immediate post-liberalization period. No exposure limits were in place, so many banks had concentrated exposure to small groups of borrowers.

Liberalization took place in an environment of rapid economic growth. Following liberalization, nominal interest rates rose and credit boomed from the middle to the late 1980s, resulting in consumption and investment booms. The household saving rate fell sharply in all three countries. The expansion of bank lending was accompanied by more risk taking, in the form of lending to more volatile sectors such as real estate, construction, and services. Credit expansion was in part financed by large capital inflows at relatively low foreign interest rates as well as through money markets. Foreign exchange–denominated lending increased sharply. The decline in interest rate spreads at this time suggested increased competition at the time of liberalization, accompanied by reduced franchise values.

The lending boom was ended in part by tighter monetary policy as well as by reductions in the tax deductibility of interest payments. Combined with a collapse in oil prices (an important export for Norway), in paper and pulp prices (important exports for both Sweden and Norway), and in trade with the Council for Mutual Economic Assistance countries of the former Soviet bloc (important for Finland), the result was recession and asset price deflation as well as exchange rate depreciation in all three countries.

Finance companies showed the effects of these developments first, as property prices collapsed. The difficulties of the finance companies spilled over to banks, initially through their involvement in lending to finance companies. Direct bank losses were initially in the real estate market, but as the recession deepened, nonperforming loans grew in activities not connected with real estate. Failed institutions tended to have lower capital ratios and to rely more extensively on funding from the money market and foreign loans than those that survived. Only a few small banks were actually liquidated. The authorities mostly assumed ownership of banks or provided funds to the banks that continued to operate. In Norway, the direct fiscal impact of the crisis amounted to about 3.5 percent of GDP. By the end of 1991, the

[6] This description is based on Drees and Pazarbasioglu (1995).

government had become the sole owner or majority shareholder of the three largest banks.

3. Seven Developing Countries

Balino and Sundararajan (1991) compiled a set of stylized facts about the experience of banking crises in seven developing countries. These crises include the Southern Cone crises in Argentina during 1980–1982, in Chile during 1981–1983, and in Uruguay during 1982–1985 as well as those in the Philippines during 1983–1986, in Thailand during 1984–1986, in Spain during 1978–1983, and in Malaysia during 1985–1986. In all countries, generalized insolvency of banks was a problem during the crises. In all these cases, nonperforming loans grew sharply just prior to and during the crisis (so none of these crises could be interpreted as a self-fulfilling bank panic).

All these episodes were preceded by periods of deregulation, though the extent of deregulation and the lead time before the crisis varied widely. Lending to related parties figured prominently in Chile (where the related parties were conglomerates known as *grupos*) and Spain.[7] The crises in these countries happened both with and without deposit insurance. Full deposit insurance was abandoned in Argentina in November 1979, and in the Philippines, the deposit insurance agency had insufficient funds, causing the settlement of claims to be delayed. Consistent with the experience we have reviewed so far, in each country, the crisis occurred after a period of rapid economic growth (a boom) with substantial variations in relative performance among sectors.

Movements in asset prices (real estate and stocks) were important elements in some cases (e.g., in the Southern Cone crises as well as in Malaysia). As in the Nordic countries, the outbreak of these crises was associated with major external shocks and balance of payments problems. Sharp adjustments in exchange rates and interest rates occurred around the time of the crises, though in some cases, the balance of payments crisis came before, and in others, after, the banking crisis. In most countries, however, external imbalances were severe just before the crisis.

The effects of the crises were those that our previous discussion would lead us to expect: there was a significant shift from deposits into currency and/or a decrease in the interest elasticity of currency demand following the crises in all these countries. There were sharp reductions in money multipliers in Argentina, the Philippines, Spain, and Uruguay, but not in Thailand and Chile (in Thailand, this was because the crisis had affected mainly finance companies rather than banks). The crises were associated with strong reductions in *GDP* growth and initial deceleration of inflation, though the latter was subsequently reversed in some cases. Confidence was restored by last-resort lending (which was used in all cases), intervention of some of the troubled institutions, and reimposition of deposit insurance.

[7] Connected lending was also important later in Mexico, during the banking crisis that began in 1994.

Last-resort lending, however, soon gave way to long-term lending at concessional rates because of the generalized insolvency of financial institutions. Thus governments wound up subsidizing failed institutions. To minimize the moral hazard implications of these actions, nonpecuniary penalties were imposed such as replacing management, requiring the surrender of shares, or preventing dividend distribution. The losses to depositors were minimal in all cases (though Argentina and Thailand permitted some losses), and the fates of troubled banks differed across countries (they were liquidated in some cases; merged or restructured in others; and nationalized, or subsidized, in still others). Bank borrowers were assisted with financial support, technical assistance, and debt-equity conversions.

4. Chile (1982)

Chilean banks had been nationalized under the socialist government of Salvador Allende in the early 1970s. The government of Augusto Pinochet, which deposed the Allende government in 1973, adopted a laissez-faire approach to the financial sector. First, banks were privatized by auctioning them off or returning them to their original owners without paying much attention to the character of the new bank owners or their connections with nonfinancial enterprises. Second, interest rate controls were lifted. Third, minimal regulations and supervisory requirements were imposed, but the government explicitly announced that deposits were not guaranteed, counting on depositors to police the banks. Fourth, reserve requirements were lowered. By 1980, they had been reduced to less than 10 percent. Finally, capital account restrictions were relaxed in 1979, at which time, the exchange rate was fixed against the U.S. dollar.

The result of the policies just described toward the financial sector included an economic boom during the late 1970s, a very rapid rate of credit expansion in 1978–1981, and a severe financial crash in the first half of 1982, involving the widespread failure of banks as well as an exchange rate crisis.

An interesting aspect of the Chilean experience is that it demonstrates why assertions by governments that deposits are not guaranteed may not tend to be credible and therefore why depositors may not tend to police banks adequately, counting on the government to bail them out later. The Chilean government, which was well known for its free-market orientation, expressly declared before the collapse of the domestic financial system that deposits were not guaranteed by the government, but after the financial crisis, depositors were nevertheless bailed out.

5. Mexico (1994)

We will return to the Mexican case in much more detail in the next chapter, but it is useful to examine it briefly in the current context for the sake of comparison with the international experience reviewed in this section.

Mexican banks had been nationalized in September 1982, after the Mexican government's announcement in August of its inability to service its external debts as scheduled led to the outbreak of the international debt crisis that we studied in Chapter 25. After that time, half of the lending business of Mexican banks went to the government. Consequently, when these banks were rapidly reprivatized in the early 1990s, they had little expertise in evaluating private credit risk. Moreover, the agency in charge of supervising banks, the National Banking Commission, did not have the capability to adequately monitor banks' portfolios. As we have already seen, connected lending was one symptom of poor bank regulation and supervision in Mexico during the early 1990s.

An enormous bank-lending boom was driven by capital inflows in the early 1990s, following Mexico's signing of a Brady Plan agreement to restructure its debt in 1989. This boom in lending not only depressed asset quality in banks but also stressed the regulators. It was associated with a consumption boom and a deterioration in Mexico's current account performance. All these phenomena are already familiar from our discussion of similar events elsewhere.

A sharp increase in U.S. interest rates beginning in February 1994, together with pressures on the Mexican peso arising from domestic political unrest, raised interest rates in Mexico in the first half of 1994 and increased adverse selection and moral hazard problems in the economy, partly because short debt maturities immediately transferred the effects of higher interest rates to the net worth of domestic firms through a financial accelerator mechanism such as that discussed in Chapter 20. Political events later in the year further increased uncertainty and, together with high interest rates and a 20 percent stock market decline in the fall, weakened the domestic banking system.

For reasons that we will examine in the next chapter, Mexico's eventual devaluation and subsequent floating of the exchange rate in early 1995 resulted in a liquidity crisis, which raised domestic interest rates to very high levels (in excess of 100%), and the stock market crashed. Together with the effects of the devaluation on the balance sheets of firms, these high interest rates would have been enough to make most of the financial system insolvent and trigger a run on banks, but as in Chile, the government stepped in and guaranteed deposits. However, banks' loss of capital restricted their ability to lend, and this credit channel factor contributed to a sharp contraction in the real economy in the last three quarters of 1995, when real *GDP* fell by nearly 10 percent.

6. Overview

There is, then, a fairly consistent story in all these different types of evidence about what causes banking crises. Moral hazard arising from financial liberalization with poor regulation and supervision results in lending booms, often accompanied by macroeconomic booms. The poor asset quality acquired by banks in this environment makes them vulnerable to macroeconomic shocks. These macroeconomic shocks have come in a variety of forms, including the onset of recessions, high interest rates, asset price collapses, and problems in the external sector. Shocks such as these provide the immediate trigger for the crisis. The crisis results in a decline in bank lending and a sharp contraction in real economic activity. Because the government often bails out depositors, whether or not a formal deposit insurance system is in place, and because the contraction in economic activity reduces fiscal revenues, these crises tend to have significant fiscal implications.

IV. SUMMARY

Banking crises involve generalized insolvency of the banking system. Such crises ultimately arise out of an interaction between the inherent vulnerability of a liberalized banking system and the moral hazard problems that plague banking. Vulnerability arises because bank assets are information-intensive, and therefore opaque and illiquid, whereas bank liabilities are highly liquid and subject to a sequential servicing constraint. This means that banks are susceptible to panics in which depositors withdraw their funds and cause banks to have to liquidate assets at fire-sale values, rendering otherwise solvent banks effectively insolvent. Deposit insurance addresses this vulnerability to bank panics, but at the cost of aggravating moral hazard problems, because the availability of deposit insurance significantly weakens the already-weak incentives for depositors to monitor their banks. The solution to this dilemma is for the government to act as delegated monitor on behalf of depositors by implementing an appropriate regulatory and supervisory framework for the banking industry.

The problem, unfortunately, is that the implementation of such a framework is subject to a variety of pitfalls. When it fails, moral hazard lending typically leads to rapid credit expansion and to an allocation of bank lending to highly risky assets. In this situation, bank portfolios become vulnerable to macroeconomic shocks, in the sense that the poor quality of bank assets can cause their value to decline sharply in the face of adverse macroeconomic shocks, often more than wiping out the value of bank capital and therefore rendering banks insolvent. When this situation arises, the disease tends to spread quickly around the banking system, and its macroeconomic effects are often severe, as the payment system is disrupted, interest rates rise, and many firms are completely denied access to credit. The result is a severe output contraction. This experience has been repeated many times around the world, and the pattern described shows up in both cross-country and case study evidence. Susceptibility to banking crises appears to have increased with financial liberalization, and it highlights the importance that a well-designed regulatory and supervisory framework has in the process of financial reform.[8]

[8] How to achieve such a framework is analyzed in detail in Barth et al. (2006).

APPENDIX 26.1. BANKING CRISES, 1980–2000

1980	Argentina, Morocco
1981	Chile, Mexico, Uruguay
1982	Colombia, Ghana, Turkey
1983	Equatorial Guinea, Niger, Philippines, Thailand
1984	
1985	Kenya, Malaysia
1986	Algeria, Bolivia, Kuwait
1987	Cameroon
1988	Benin, Burkina Faso, Central African Republic, Côte d'Ivoire, Madagascar, Nepal, Panama, Senegal, Tanzania
1989	Argentina, Jordan, Nicaragua, South Africa, Sri Lanka
1990	Brazil, Norway, Sierra Leone
1991	Egypt, Finland, Nigeria, Sweden, Tunisia
1992	Congo, Indonesia, Japan
1993	Cape Verde, India, Kenya, Togo
1994	Bolivia, Brazil, Burundi, Costa Rica, Jamaica, Mexico, Turkey, Uganda, Venezuela
1995	Argentina, Cameroon, Swaziland, Zambia, Zimbabwe
1996	Ecuador
1997	Indonesia, Korea, Malaysia, Thailand
1998	Philippines
1999	
2000	Turkey

Note: Banking crisis dates are the first year of the crisis episode reported in Caprio and Klingebiel (2002). A crisis occurs when much or all bank capital is exhausted.
Source: Kroszner et al. (2007).

REVIEW QUESTIONS

1. Why are bank regulation and supervision often unable to prevent the emergence of banking crises?
2. Do banking crises tend to be caused by "bad policy, bad banking, or bad luck" (the phrase is from Caprio and Klingebiel (1997, page 79))? Explain.
3. What are some possible macroeconomic roots of banking crises?
4. Describe the channels of transmission through which banking crises can affect the real economy.
5. Why do banking crises tend to result in large increases in public-sector debt?

EXERCISES

1. Why are the fiscal costs of banking crises (the cost to the government of making bank depositors whole) not an appropriate measure of their true social costs?
2. Explain why empirical studies of the determinants of banking crises have sometimes found that the likelihood of such crises is *increased* by the presence of deposit insurance.

3. Some commentators have suggested that to assist regulators in doing their jobs, banks should be required to hold some fraction of their portfolios in the form of the *subordinated debt* of other banks (these are securities that, in the event of bank liquidation, are paid after other bank creditors are paid, but before bank owners). How would this requirement help bank regulators?
4. Assume that the residents of a country hold no debt denominated in foreign exchange. How would the real effects of a banking crisis in that country be affected by whether the country maintains a fixed or a floating exchange rate?
5. What would you expect to happen to the balance sheet of the central bank of a country that is experiencing a banking crisis? Explain why.

REFERENCES AND FURTHER READING

Balino, Tomas, and V. Sundararajan (1991), *Banking Crises: Cases and Issues*, Washington, DC: International Monetary Fund.

Barth, James, Gerard Caprio, and Ross Levine (2006), *Rethinking Regulation: Till Angels Govern*, Cambridge: Cambridge University Press.

Caprio, Gerard, and Patrick Honohan (2009), "Banking Crises," unpublished manuscript, Williams College.

Caprio, Gerard, and Daniela Klingebiel (1997), "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?" in M. Bruno and B. Plescovic, eds., *Annual World Bank Conference on Development Economics 1996*, Washington, DC: World Bank, pp. 79–104.

Caprio, Gerard, and Daniela Klingebiel (2002), "Episodes of Systemic and Borderline Banking Crises" in D. Klingebiel and L. Laeven, eds., *Managing the Real and Fiscal Effects of Banking Crises*, World Bank Discussion Paper No. 428, pp. 31–49.

Caprio, Gerard, William C. Hunter, George C. Kaufman, and Danny M. Leipziger (1998), *Preventing Bank Crises: Lessons from Recent Global Bank Failures*, Washington, DC: World Bank.

Demigurc-Kunt, Asli, and Enrica Detragiache (1998), "The Determinants of Banking Crises in Developing and Developed Countries," *IMF Staff Papers*, vol. 45, pp. 81–109.

Drees, Burkhard, and Ceyla Pazarbasioglu (1995), "The Nordic Banking Crisis: Pitfalls in Financial Liberalization?" working paper WP/95/61, International Monetary Fund.

Eichengreen, Barry, and Carlos Arteta (2002), "Banking Crises in Emerging Markets," in M. Blejer and M. Skreb, eds., *Financial Policies in Emerging Markets*, Cambridge, MA: MIT Press, pp. 47–94.

Gavin, Michael, and Ricardo Hausmann (1996), "The Roots of Banking Crises: The Macroeconomic Context," in R. Hausmann and L. Rojas-Suarez, eds., Banking Crises in Latin America, Baltimore: Johns Hopkins University Press, pp. 27–63.

Hutchinson, Michael, and Kathleen McDill (1999), "Are All Banking Crises Alike? The Japanese Experience in International Comparison," working paper 7253, National Bureau of Economic Research.

Kroszner, R., L. Laeven, and D. Klingebiel (2007), "Banking Crises, Financial Dependence, and Growth," *Journal of Financial Economics*, vol. 84, pp. 187–228.

Mishkin, Frederic S. (1989), "Asymmetric Information and Financial Crises: A Historical Perspective," in G. Hubbard, ed., *Financial Markets and Financial Crises*, Chicago: University of Chicago Press, pp. 69–108.

27

Currency Crises and Crisis Interactions

This chapter attempts to bring together much of the material covered in the rest of the book by analyzing the links among currency crises, sovereign debt crises, and banking crises in emerging economies, both analytically and by conducting case studies of both the 1994 Mexican and 1997 Asian crises. These more recent crises are contrasted with the fiscally driven international debt crisis of the 1980s that we discussed in Chapter 25 and are compared to the 1982 Chilean financial crisis that has been mentioned at various places in the book. The objective of this chapter is to illustrate how the three general topics treated in this book – fiscal and monetary management, management of the exchange rate, and management of the financial system – lie at the heart of the most severe macroeconomic crises that developing countries have faced over the past two decades.

The structure of the chapter is as follows. We begin in Section I with a brief review of the traditional analytics of currency crises, a type of financial crisis that we have discussed at various points in the book but have not treated separately up to now. In Section II, we broaden the analytical discussion to include liquidity crises, or "sudden stops" of capital inflows. These have figured prominently in recent emerging-market crises, but they were not featured in the traditional currency-crisis literature until recently. As we will see, these liquidity crises involve interactions among debt, banking, and currency problems. Sections III and IV apply these analytical tools to seek an understanding of the factors that drove two of the most important emerging-market currency crises in the 1990s: the 1994 crisis in Mexico and the 1997 crisis in Thailand. Section V summarizes and concludes.

I. TRADITIONAL MODELS OF CURRENCY CRISES

As discussed in Chapter 25, a currency crisis arises when the credibility of a central bank's commitment to fix the exchange rate comes into question. Faced with the prospect of a sharp depreciation of the exchange rate, either through a discrete

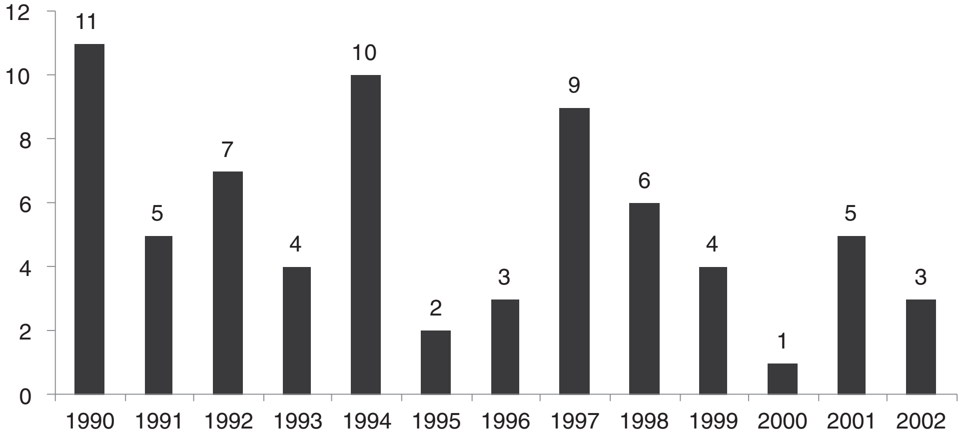


Figure 27.1. Currency crisis frequency, 1990–2002. ***Source*****: Frankel and Wei (2004)**

devaluation or through a depreciation as the result of a switch to a floating exchange rate, individuals holding the domestic currency in the crisis country have a strong incentive to switch out of domestic currency–denominated assets. This gives rise to capital outflows and increases in domestic interest rates, which, if sustained, often result in a devaluation of the official exchange rate or its complete abandonment in favor of a floating rate. Like sovereign debt crises and banking crises, currency crises have been quite common in recent years. Figure 27.1 provides the evidence for the period 1990–2002, based on Frankel and Wei (2004).

Models of currency crises have progressed through several "generations" since the seminal work of Krugman (1979) on the topic. The various generations of currency crisis models have several characteristics in common. In each case, the central bank has to decide whether to fulfill a previous commitment to sustain a par value for the domestic currency, and in each case, the central bank's decision whether to fulfill or abrogate its commitment depends on the state of the economy. The various generations of crisis models differ from each other according to the specific characteristics of the domestic economy that are taken to influence the central bank's decision (the fundamentals), how the evolution of the relevant fundamentals over time is described, and how the central bank's decision to sustain or abandon the parity is modeled. That decision may be modeled as an arbitrary decision rule, as the outcome of an optimization problem solved by the central bank, or as an outcome that is forced on the central bank. The first two are emphasized in first- and second-generation crisis models, respectively. We will consider them in this section. The third features prominently in the more recent third-generation models, which we will consider in the section that follows. Because first- and second-generation models are relatively familiar, the discussion of those classes of model will be brief, and most of our attention will be devoted to third-generation models in Section II.

1. First-Generation Models

First-generation crisis models are models of fiscal dominance. In these models, the central bank maintains a fixed exchange rate but is obligated to finance an ongoing fiscal deficit by expanding the stock of domestic credit. As we saw in Chapter 8, continued credit expansion in excess of growth in the domestic demand for money under a fixed exchange rate must inevitably lead to an exhaustion of the central bank's foreign exchange reserves, forcing the central bank to abandon the fixed exchange rate. Anticipation of the resulting loss of value for the domestic currency leads to speculative capital outflows that actually cause the central bank's foreign exchange reserves to be exhausted all at once in a massive speculative attack rather than through gradual reserve depletion (Krugman 1979). First-generation currency crisis models have been applied to the exchange rate crises associated with the fiscally driven international debt crisis of the early 1980s.

2. Second-Generation Models

In first-generation crisis models, the central bank behaves in a mechanical fashion: it is obligated to expand the stock of credit continuously by the government's financing needs, and it pegs the exchange rate as long as it is able to do so, abandoning the peg only when it runs out of reserves or its reserves reach a minimum threshold. By contrast, in second-generation models, central banks play a much more active role. When the expectation of a devaluation arises for whatever reason, the central bank defends the exchange rate with high domestic interest rates as long as it is not too costly for the domestic economy for it to do so. When the interest rate cost of maintaining the peg is too high, however, the central bank *chooses* to abandon it, whether by devaluing or abandoning the peg.

The interest rate cost of defending the peg can be high under a variety of circumstances, for example, when the economy is in recession, when the exchange rate is overvalued, or when the government's fiscal position or the solvency of the banking system are precarious. Recall that a second-generation perspective on the sustainability of fixed exchange rates underpinned the bipolar view of exchange rate regimes that we examined in Chapter 18. Second-generation models were developed to explain currency crises in settings where fiscal deficits did not seem to play a role in driving the abandonment of the exchange rate peg, especially in the context of the crisis of the European Exchange Rate Mechanism (ERM) of 1992.

More recent (third generation) crisis models emphasize a variety of different mechanisms, especially sudden stops of capital inflows associated with *balance sheet effects* (effects having to do with the composition of assets and liabilities in the portfolios of important economic agents). Because these models have been explicitly designed to address recent emerging-market crises, they are discussed separately in the next section.

II. MACROECONOMIC POLICIES AND VULNERABILITY: BALANCE SHEET EFFECTS AND SUDDEN STOPS

In Chapter 24, we discussed the capital-inflow problem created by the capital account openness that accompanied financial reform in emerging economies. One reason that capital inflows were perceived as a problem was the possibility of capital-flow reversals (sudden stops); that is, the arrival of capital inflows was a symptom of an enhanced degree of financial integration that left countries vulnerable to capital-flow reversals and to the financial (banking and currency) crises that would accompany them.

Vulnerability to crisis has two dimensions. These involve the probability that a country will experience an abrupt reduction (or reversal) in capital inflows as well as the costs to the economy if such an event were to come to pass. As we have seen, the probability of a reversal depends in turn on the likelihood that creditors will come to believe that the value of their claims on domestic economic agents has become impaired. This situation could arise when debtors become insolvent, as in the international debt crisis of the 1980s. However, as we saw in our analysis of banking crises in the last chapter, it could also materialize under an entirely different set of circumstances: when the debtor is merely *illiquid*, that is, when the claims that can be presented to that debtor for payment within a given period of time exceed the debtor's capacity to make such payments during that period, even though the debtor could do so (in present value terms) if given a longer period of time to repay.

The next section will take the third-generation perspective that in contrast to the international debt crisis of the 1980s, and the ERM crisis of 1992, several prominent recent financial crises in emerging economies have been liquidity crises. The 1994 Mexican crisis and the 1997–1998 Asian crisis that originated in Thailand will be used as examples. The basic argument is that policies undertaken during the capital-inflow period in these countries, especially toward the domestic financial system, but also with regard to the exchange rate and other policies, resulted in a currency crisis, the (mis)-handling of which left the economy vulnerable to a much more costly liquidity crisis after the exchange rate had been floated. To make this argument, we will first review some theory on liquidity crises in this section, before turning to the specific cases of Mexico and Thailand in the next two sections. Because the forms taken by the liquidity crises in Mexico and Thailand were superficially different – in the case of Mexico, the liquid assets held by foreigners were liabilities of the public sector, whereas in Thailand, they were liabilities of the private sector – the theoretical discussion in this section is divided into two parts discussing public- and private-sector "debt runs" respectively.

1. Public-Sector Debt Runs

Calvo (1988) argued that the existence of a large stock of nominal long-term government debt makes the public sector vulnerable to "confidence crises." Long-term, nominal (domestic currency), fixed-interest debt is vulnerable to taxation by the government through inflation (or devaluation, in the case of debt held by foreign creditors).[1] If the government's creditors believe that inflation is a possibility, they will demand a premium in the yield on long-term nominal debt over that on short-term or indexed debt, to compensate them for expected inflation and for bearing the attendant risk.

Suppose that the public comes to believe that in the event of a loss of confidence in the government's inflation performance (and thus the emergence of high nominal interest rates), the government will be unwilling to raise taxes to service the debt and will instead print money. In that case, a confidence crisis on government debt may actually be self-fulfilling, in the sense that the high nominal interest rates caused by a lack of confidence may indeed induce the government to inflate the debt away. Thus two equilibria could arise: a good equilibrium without a crisis, high taxes, or inflation and a bad equilibrium with high nominal interest rates and high inflation.

Alesina et al. (1990) pointed out that vulnerability to a self-fulfilling confidence crisis could also arise in the context of short-term debt. What creates the possibility of a bad equilibrium in the Calvo framework is the government's reluctance to make the fiscal adjustment required to meet a crisis-driven increase in its debt service requirements. But according to Alesina et al., a short maturity structure of the public debt may actually increase the likelihood of a confidence crisis on the debt. Indeed, the shorter and more concentrated are the maturities, the more likely a confidence crisis may be. The reason for this in their model has to do with convexity of tax collection costs (which implies that the marginal cost of raising revenue increases with the amount of revenue to be raised). This makes a given present value of tax revenue more costly to raise the less the government can smooth its collection over time, that is, the more front-loaded its collection. Because this increases the (utility) cost of servicing a given stock of debt the more bunched are maturities, it makes the government more likely to default if creditors become unwilling to roll over short-term debt. Thus the unwillingness of creditors to roll over short-term debt in fear of default may actually become self-fulfilling. In this case, the crisis takes the form of a debt run, comparable to the bank runs that we discussed in Chapter 26.

Giavazzi and Pagano (1990) summarize these results by noting that the likelihood of a Calvo bad equilibrium depends on three things: the size of the public debt, its maturity structure, and the time pattern of maturing debt. The logic is that when a substantial amount of debt has to be serviced at a point in time, if a confidence

[1] It is also, of course, vulnerable to default (in the extreme case, repudiation), but no more so than short-term or indexed debt.

crisis breaks out at that moment, the government would have to refinance a large portion of its debt on unfavorable terms.[2] The utility cost of doing so would be high, and thus the likelihood that the government will repudiate would be greater. This makes the confidence crisis more likely to happen. Under these circumstances, debt management, in the form of the issuance of long-term indexed debt, could push the economy to the good equilibrium because such debt cannot be monetized away and does not create large short-run amortization obligations.

2. Private-Sector Debt Runs

In the literature just reviewed, self-fulfilling confidence crises could afflict the public debt because of a choice made by the government in response to a loss of confidence; that is, the costs of generating the fiscal resources to service its debt would lead the government to ratify the fears of its creditors. But what if the debt is issued by the private sector? The point of this subsection is that this difference is inessential. What matters for vulnerability is not the identity of the debtor but the emergence of a situation in which, in the event of a run, creditors stand to lose part of the value of their claims. In the case of public-sector debt, this would happen because of the government's reluctance to incur the fiscal costs of continuing to service the debt on market terms in the event of a run. In what follows, we will consider a very simple and stylized model in which a run on *private* debt has similar consequences.

Consider an economy that maintains a fixed exchange rate, with the monetary base M backed by foreign exchange reserves F_C^* and central-bank credit B_C:

$$M = SF_C^* + B_C \tag{27.1}$$

where S is the nominal exchange rate. The demand for the monetary base depends on the domestic nominal interest rate R, which in turn is given by uncovered interest parity:

$$M = P(\underset{+}{S})L(R) \tag{27.2}$$

$$R = R^* + \hat{S} \tag{27.3}$$

where P is the domestic price level, taken to be an increasing function of the nominal exchange rate; R^* is the foreign interest rate; and $\hat{S}$ is the expected (and actual) rate of depreciation of the currency. This describes financial-market equilibrium in the model. At each moment in time, given the exogenous world interest rate, the policy-determined rate of depreciation determines the domestic nominal interest rate, which, through equation (27.2), determines the money supply M, as in the

[2] The term *crisis* here could refer to fears either of a repudiation or a devaluation because either would affect the return on domestic-currency debt.

fixed exchange rate model with perfect capital mobility of Part 2. Equation (27.1) then determines the stock of foreign exchange reserves, given domestic monetary policy B_C, through capital inflows and outflows.

The balance of payments in this economy (the change in the stock of foreign exchange reserves) is given by

$$\dot{F}_C^* = B[\underset{+}{S/P(S)}, \underset{+}{\theta}] + R^*(F_C - D) + (F - \gamma D) \tag{27.4}$$

where $\dot{F}_C^*$ is the change in the central bank's stock of foreign exchange reserves, $B(\ )$ denotes the trade balance, which depends on the real exchange rate $S/P(S)$ (the foreign price level is set equal to unity) and a shift factor θ; D denotes the stock of net external debt; F is the magnitude of gross capital inflows; and γ is the ratio of amortization to the stock of debt, an index of the maturity of the outstanding stock of debt: the larger γ, the shorter the maturity of existing debt. If $\hat{S}$ is constant (for simplicity, suppose $\hat{S}$ is zero), M must be constant, and if B_C is constant as well, the balance of payments must be zero. In that case, equation (27.4) determines the volume of gross capital inflows F. These flows determine the evolution of the stock of debt over time according to

$$\dot{D} = F - \gamma D \tag{27.5}$$

which completes the model. Notice that creditors' willingness to supply the level of capital inflows F must reflect their view that the economy is solvent, that is, that the projected path of the trade balance (and thus implicitly of θ) is such that the existing stock of debt D will be serviced on market terms.

How can a liquidity crisis arise in this simple world? Suppose there is a threshold value of the nominal exchange rate, say, S^*, above which domestic debtors are unable to service their foreign debt on market terms. This could happen, for example, because of currency mismatches in the balance sheets of domestic agents, in which their assets are predominantly denominated in domestic currency, while their liabilities are in foreign currency. In this case, the economy could be vulnerable to a self-fulfilling liquidity crisis, even if all external debt is private.

To see how, assume that in the event of a speculative attack on the currency that eliminates the central bank's foreign exchange reserves, the country would adopt a floating exchange rate. Now suppose that a liquidity crisis materializes in which foreign creditors cease extending new loans to domestic residents, withdrawing from the country as existing debt matures. In this case, $F = 0$. If F was positive initially, the liquidity crisis could trigger a self-fulfilling speculative attack on the currency because the floating exchange rate that would emerge after such an attack would be more depreciated than the initial fixed rate. This can be shown as follows: if the liquidity crisis triggers a successful speculative attack, the post-attack economy would be described by equations (27.1), (27.2), (27.4), and (27.5), with $B_C = 0$ in equation (27.1), $\dot{F}_C^* = F = 0$ in equation (27.4), and S determined endogenously.

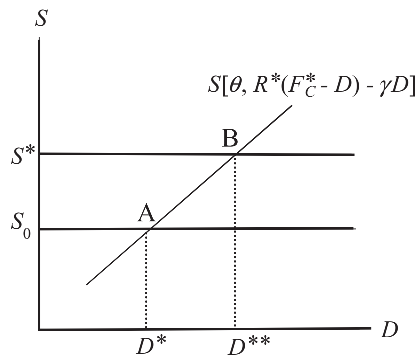


Figure 27.2. Determination of the range of multiple equilibria

The cutoff of foreign lending means that the country is placed in a position of financial autarky, so equation (27.3) no longer holds. Under these conditions, S is determined by flow equilibrium in the balance of payments, given by[3]

$$B[S/P(S), \theta] + R^*(F_C^* - D) - \gamma D \qquad (27.6)$$

Thus the post-attack exchange rate must satisfy

$$S = S[\underset{-}{\theta}, R^*(\underset{-}{F_C^*} - D) - \gamma D] \qquad (27.7)$$

This means that the floating exchange rate that would prevail after a successful attack, which is conventionally referred to as the "shadow" exchange rate, will be more appreciated the larger are the shift factors that increase the trade surplus and the *smaller* are scheduled debt service payments.

The stock of external debt D is the predetermined variable in this model, so equation (27.7) is depicted in (D, S) space in Figure 27.2. Its positive slope reflects the fact that the larger is the stock of external debt, the larger is the value of debt service payments in the post-attack situation, and thus the more depreciated the floating exchange rate must be. In the figure, D^* is the value of D that would be consistent with balance of payments equilibrium with $F = 0$ at the original fixed exchange rate. For values of D larger than D^*, but smaller than D^{**}, the shadow exchange rate is more depreciated than the initial fixed rate, so given a liquidity crisis, a rational speculative attack on the currency is possible. However, there is no reason for a liquidity crisis to emerge in the range between D^* and D^{**} because the floating rate that would emerge after a successful speculative attack would be below the critical value S^*, so foreigners would have no reason to cut off lending within this range.

[3] It is implicitly assumed here that F cannot be negative, which would be the case, e.g., if controls on capital outflows by domestic residents are adopted in the event of a liquidity crisis.

Consider, however, the range $D > D^{**}$. In this range, dual equilibria can arise because the shadow value of S is above S^*. In the absence of a liquidity crisis, debt would be serviced on schedule, and the exchange rate would remain below S^* at S_0, but if a crisis-cum-speculative attack happens, individual creditors would have acted rationally ex post because the exchange rate would immediately depreciate above S^*.

As is evident from equation (27.7), vulnerability to a confidence crisis (in the sense of a shadow exchange rate that is above its threshold value S^*) is increased in this model by a larger current account deficit, a larger stock of external debt, a larger share of short-term debt, and higher world interest rates.[4] All of these tend to shift the $S[\ ]$ curve upward in Figure 27.2, which increases the range of values of D for which $D > D^{**}$.[5] Because vulnerability arises at the moment when the shadow exchange rate equals the critical rate S^*, a larger current account deficit also brings vulnerability forward in time by increasing the stock of external debt.[6]

How can policy avoid vulnerability in this case? One option is for the central bank to maintain a sufficiently large stock of reserves (and/or official credit lines) to insulate itself from an attack (see Footnote 6). A second option would be to adopt policies that would have the effect of increasing S^*, for example, regulatory policies that would restrict the scope for currency mismatches in the financial system (either in the balance sheets of banks or those of their clients). Third, exchange rate and aggregate demand policies could be targeted to the achievement of a relatively low value of the trade deficit B, even if a larger trade deficit would be compatible with the country's solvency constraint. Fourth, capital account policies could be adopted to lengthen the maturity structure of external debt (i.e., to reduce γ). Finally, if none of these options are feasible, because vulnerability will be approached gradually over time when $F > \gamma D$, a country would be well advised to abandon its exchange rate peg *before* $S = S^*$.

We will now see that the recent dramatic instances of capital-flow reversals – the crises in Mexico and Thailand – followed a common pattern that culminated in the two types of liquidity crises described earlier. In both cases, policies followed during the inflow period resulted in currency crisis and an eventual floating of the currency. But the policy response to the currency crisis left both countries vulnerable to a subsequent liquidity crisis – centered on public liabilities in the case of Mexico and private ones in the case of Thailand. The liquidity crises greatly increased the real

4 On the other hand, it would be *decreased* by a reliable flow of foreign direct investment, which would represent an additional (exogenous) additive term in equation (27.4).

5 If F can be negative – i.e., if domestic residents are able to pull capital out of the country freely in the event of a liquidity crisis – the range of values of D for which dual equilibria can exist is larger still because negative values of F would also have the effect of shifting $S[\ ]$ upward.

6 Notice that in this setup, the size of the stock of foreign exchange reserves (which is determined by $\hat{S}$ and B_C) does not affect vulnerability as long as $F_C^*/S < M$ (i.e., as long as B_C is positive). However, the government can insulate itself from an attack by ensuring that $F_C^*/S > M$.

costs of the exchange rate adjustments in both countries through their effects on fiscal policies, domestic interest rates, and the exchange rate. These real effects were in turn magnified by financial-sector insolvency.

III. VULNERABILITY IN PRACTICE: MEXICO

The story of Mexico's descent into vulnerability is by now well known, but it bears reviewing in comparison with what happened in Asia two and a half years later. Mexico, in fact, had two crises at the end of 1994 and the beginning of 1995: a garden-variety balance of payments crisis and a public-sector debt run, respectively. The two phenomena were, of course, closely linked.

Like Thailand in 1997, and unlike Mexico itself in 1982, Mexico did not enter 1994 with a public-sector debt problem. By international standards, the stock of government debt outstanding was relatively small. Mexican public debt had been reduced from 67 percent of gross domestic product (*GDP*) in 1989 to a little over 30 percent in 1993. The reduction in the debt-*GDP* ratio was achieved in part through the use of privatization revenues to retire debt and in part through operational surpluses on the fiscal accounts.[7] Of the total public debt, about two-thirds was external (mostly long term, as a result of a Brady Plan debt restructuring in 1989) and one-third domestic, with an average maturity of about 200 days.

However, as noted by several observers at the time, Mexico did enter 1994 with an exchange rate problem. The problem consisted of a real exchange rate overvaluation caused by an exchange rate–based stabilization undertaken in 1988 (called the Solidarity Pact).[8] The symptoms of overvaluation included the large cumulative real exchange rate appreciation that the peso had undergone during the capital-inflow period, a substantial current account deficit, with a reduction in private saving as its counterpart, and slow economic growth. All these made devaluation a possibility. The government had, in fact, depreciated the peso substantially in the context of a widened exchange rate band that had been adopted in November 1991. Nonetheless, ex post estimation of the long-run equilibrium real exchange rate by several researchers, using the methods described in Chapter 16, concur that the peso had been substantially overvalued in 1994.[9]

[7] Of US$20.2 billion in privatization income obtained in 1991–1992, no less than 60% was used to retire debt (Steiner 1995), and operational surpluses of 2%–3% of GDP were achieved from 1990 to 1993.

[8] Dornbusch and Werner (1994) called attention to the overvaluation of the peso in spring 1994.

[9] E.g., Loayza and Lopez (1997) estimated a cointegrating equation for the Mexican real exchange rate using the country's stock of net foreign assets, and the relative productivity level in its traded and nontraded goods sectors, and applied the results to a calculation of real exchange rate misalignment in 1994, finding an overvaluation of 27%. Similarly, Warner (1997) estimated a cointegrating equation using the terms of trade and the ratio of capital inflows to GDP as fundamentals and estimated an overvaluation of about 25%.

The role of the exchange rate as a nominal anchor in the context of the Solidarity Pact, and doubts about whether the observed real appreciation may have been an equilibrium phenomenon, however, made the authorities reluctant to undertake a discrete exchange rate adjustment. Three exogenous events in the first quarter of 1994 magnified the pressures in the foreign exchange market: an uprising in Chiapas Province in January, the announcement of a tighter monetary policy by the U.S. Federal Reserve Board in February, and the assassination of presidential candidate Luis Donaldo Colosio in March. These events intensified expectations of devaluation, essentially for second-generation reasons. The exchange rate moved to the top of its band in the first quarter of 1994, the central bank suffered a large reserve loss, and from March on, a large premium emerged on the domestic currency–denominated government bonds (CETES) over dollar-denominated ones (TESOBONOS).

An exchange rate crisis materialized in the Mexican case through the interaction of these expectations of devaluation with the perception that the government would be unwilling to mount a high–interest rate defense of the currency, that is, precisely through the mechanisms we discussed in Section I. A sustained tight money defense did not seem likely because of a combination of ongoing recession (likely due to the perceived overvaluation itself), the poor state of the domestic financial system, and upcoming elections in fall 1994 that appeared to pose a significant challenge to the ruling party. As we have seen, these circumstances are precisely those in which speculative attacks tend to happen because markets do not believe the government will bear the political cost of mounting the high–interest rate defense called for by the rules of the game (see, e.g., Eichengreen et al., 1995).

An important point of similarity with the Asian case (as well as with the case of Chile in 1982) was the fragility of the financial system. As is now well known, this fragility emerged as the result of rapid financial liberalization in Mexico in the late 1980s and early 1990s, in a context in which the government's capacity to regulate and supervise the domestic financial system was not well developed (see Chapter 26). Poorly supervised banks with weak capital positions, operating under the assumption of government backing of their deposit liabilities and with unconstrained access to external funds, undertook a rapid expansion of credit in the early 1990s that fueled a consumption boom and left banks with a large stock of questionable assets and thus in a perilous financial situation. Given the unavoidable maturity mismatch involved in banking, and the weak position of many bank clients in the midst of the recession in Mexico, a high–interest rate defense of the currency would have imperiled the financial system and potentially severely aggravated the recession.

In the event, the government reacted to the emerging balance of payments crisis in two ways. First, it undertook sterilized intervention during the first quarter of the year, expanding the stock of credit to keep the monetary base relatively constant as foreign exchange reserves declined. Second, it shifted the composition of its debt from CETES to TESOBONOS. The share of TESOBONOS in total debt rose from less than 5 percent at the beginning of 1994 to over 55 percent by the end of the year. The share of privately held CETES, on the other hand, went from 60 percent of total debt in February to 20 percent by November.[10] The benefits to the government of this debt transformation consisted, of course, of lower debt-servicing costs, thus protecting the public sector's operational balance, but there may also have been an element of projecting credibility to counteract devaluation expectations.[11]

The net effect of the two policies, however, was to create a large increase in the government's net short-term (liquid) dollar liabilities and a reduction in its liquid dollar assets. As we have seen, in the presence of deposit insurance, M2 is backed by base money. Under officially determined exchange rates, on the other hand, base money is backed by reserves. With the two together, as in Mexico, M2 is backed by reserves. Thus the loss of reserves with a fairly stable value of M2 implied a worsening government liquidity position. This vulnerability to a confidence crisis was clearly magnified by debt management. By converting longer-term domestic-currency liabilities into short-term dollar liabilities, the central bank was essentially using (net) foreign exchange reserves to pay off those liabilities. By the time of the crisis in December 1994, M2 amounted to US$110 billion (Calvo and Mendoza 1996), and TESOBONOS maturing in early 1995 were another US$17.8 billion (Sachs et al. 1996). Foreign exchange reserves at the end of October were US$17 billion, compared to a minimum level of US$10 billion targeted by the central bank. Moreover, no extraordinary arrangements were in place on the part of external public institutions to lend to the Mexican government in the event of a private attack. At this point, the government was clearly vulnerable to a liquidity crisis. An attack merely awaited a coordinating mechanism.

The first attack came in mid-November 1994, when reserves fell by US$5 billion. The coordinating mechanism was clear: the new administration took office on December 1, and markets anticipated that the outgoing administration would devalue as a present to the incoming one (by removing the onus for doing so from them). The attack was halted by a public statement by president elect Zedillo supporting the Solidarity Pact. The new administration took office with reserves of $12.5 billion. There obviously was not much room to maneuver. Clearly the exchange rate was on the new administration's agenda, and apparently a leak did them in.[12] By December 20, after only three weeks in office, the new administration

10 Werner (1995) presents evidence that the government altered its mix of instruments in response to interest rate differentials.

11 The idea here is similar to that of establishing central-bank independence to establish fiscal credibility in Chapter 10. Only a government that firmly intended not to devalue would presumably denominate its debt in foreign exchange because if it did devalue under such circumstances, it would substantially increase the domestic-currency value of its outstanding debt. Thus switching to foreign-currency debt would be intended as a signal that the government intended to resist devaluation.

12 As reported in the *New York Times*, March 2, 1995.

effectively devalued by raising the ceiling on the exchange rate band by 15 percent. This was not perceived as enough, and the crisis was on, resulting in a float on December 21.

The interesting subsequent development for present purposes is that the immediate aftermath of the float was an increase in sovereign risk. Weekly TESOBONOS–U.S. Treasury bill interest rate differentials began to climb steeply in the third week of December, rising from 1.7 percent in the second week of December to nearly 20 percent by the end of January 1995. On the surface, this is puzzling, for at least two reasons. First, given the Mexican government's low stock of debt and its fiscal track record, why should sovereign risk have been a problem? Second, why should sovereign risk *increase* after a devaluation? If an overvalued currency was the main obstacle to growth in Mexico, as Dornbusch and Werner (1994) had argued before the crisis, one would have thought sovereign risk would have fallen as growth prospects improved, especially because a renewal of growth and a reduction of domestic interest rates would have improved the prospects of the domestic financial system.

One interpretation is that the premium reflected not a *solvency* problem but *liquidity* risk, and that the latter must be related to the balance of payments crisis. What were the links? The balance of payments crisis made the public sector more vulnerable to a debt run by lengthening the maturity of its assets and shortening that of its liabilities. The asset structure was made less liquid through the loss of liquid foreign exchange reserves, which were replaced on the asset side of the central-bank balance sheet by illiquid claims on development banks as a result of sterilization. On the liability side, the endogenous change in the maturity structure of government debt also made its liabilities more liquid. A bunching of maturing short-term debt in the first half of 1995 made a run more likely. Moreover, the change in the currency composition of debt magnified the size of the peso repayment obligations after the devaluation.

This leaves the question, however, of why creditors may have converged on the view that in the event of a public-sector debt run, default was a possibility. Faced with large short-term payment obligations, there are four things the government can do in the event of a debt run:

1. generate a fiscal surplus of sufficient size to service the debt
2. borrow from nonmarket sources (i.e., draw down reserves and/or borrow from official creditors)
3. print money[13]
4. default (e.g., by stretching out payments)

[13] Notice that, under fixed exchange rates, second and third options are identical. Thus the third is only feasible if the second is, and the four choices collapse to three. But by January 1995, Mexico was operating a floating exchange rate system. This means that it would indeed have had four choices in the event of a debt run.

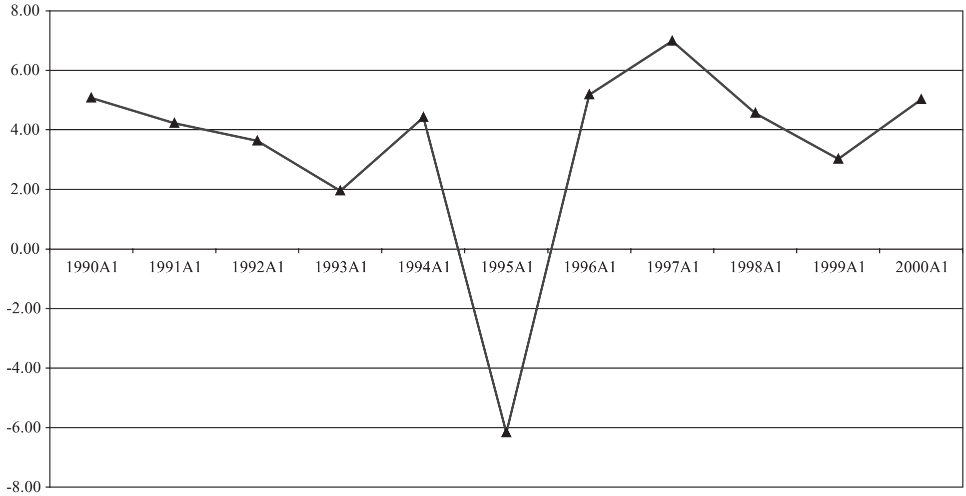


Figure 27.3. Real GDP growth in Mexico, 1990–2000

To analyze what the government would have been likely to do, creditors would have had to look at both the feasibility and the consequences of each of these options. Because reserves were essentially depleted and official borrowing in sufficient magnitude had not been arranged as of January, the second option was problematic. On the other hand, because the payments due were large, the distortions that would have been associated with the first option (e.g., tax distortions and an aggravated recession) as well as with the third (in the form of increased inflation) would have been large. This left the last as a realistic possibility, triggering the debt run.[14] The implications of the run were that the macroeconomic costs of the exchange rate adjustment were greatly magnified.

As shown in Figure 27.3, the combination of draconian fiscal adjustments under financial duress and very high domestic interest rates, combined with exchange rate overshooting as a result of the loss of confidence, resulted in a very severe recession in Mexico during 1995, despite the eventual resolution of the crisis through the provision of sufficient liquid official funds in March to pay off the government's liquid debt.

IV. VULNERABILITY IN PRACTICE: THAILAND

The central theme of this section is that the emergence of macroeconomic vulnerability in Thailand had much in common with the Mexican experience. Three

[14] Some observers have emphasized alternative coordinating mechanisms, including the way the initial currency crisis was handled (with repeated assurances that no devaluation was imminent, the emergence of leaks, a devaluation that was perceived as inadequate, and a forced float), and Mexico's financial history during the previous decade, which included debt default, bank nationalizations, and forced conversions of peso into dollar deposits.

areas of similarity stand out. First, as in the case of Mexico in 1994, fiscal solvency was not an issue in Thailand in 1997. The Thai government had undertaken a sustained effort at fiscal adjustment since the mid-1980s and, by 1997, had succeeded in sharply reducing the stock of government debt relative to GDP. Second, as in Mexico, there were strong reasons to believe that the Thai real effective exchange rate had become overvalued by 1997. While the real appreciation of the baht at the outbreak of the crisis relative to the early 1990s was much smaller than that of the Mexican peso, several factors suggested that the *equilibrium* real exchange rate may have depreciated – perhaps substantially – in Thailand by that date, implying the existence of a gap between the actual and equilibrium real exchange rates.

Indeed, the size of Thailand's current account deficit relative to GDP by 1997 exceeded that of Mexico in 1994. Third, as in the case of Mexico, Thailand underwent a significant financial liberalization several years before the crisis, which was followed by a domestic lending boom and a weakened financial system. Finally, Thailand's crisis, like Mexico's, played itself out over the space of about a year, during which time, the authorities failed to adjust in any of the ways suggested in Part 2 of Section II. Like Mexico, they responded to reserve outflows by incurring short-term dollar-denominated debt, thus impairing the country's liquidity position.

This being said, there were also obvious differences between Thailand's macroeconomic situation in 1997 and Mexico's in 1994. Most important, Thailand was growing very rapidly before its financial difficulties became apparent, in contrast to Mexico, which was experiencing a recession in 1994. In addition, the increase in Thailand's current account deficit reflected a sharp increase in domestic investment rather than a reduction in domestic saving. Finally, Thailand's outstanding stock of short-term debt was owed by the private sector rather than the government. An important feature of this debt was that, unlike in the case of Mexico, it was quite large relative to the feasible scale of resources that could be made available to Thailand by the official international financial community. The question is, of course, whether these differences make the Thai crisis sui generis or whether the Thai crisis reinforced the policy lessons of the earlier Mexican crisis. To address this issue, it is worth reviewing the development of events in Thailand.

1. Precrisis Developments

Southeast Asian countries faced two potentially important changes in their external economic environments during the early to mid-1990s. One was of a long-run nature and the other was cyclical. The long-run change was the emergence of China as a major exporter of labor-intensive manufactured goods in the early 1990s. China's share of world exports, which had grown at an average annual rate of about 6 percent during the 1980s, increased at a 10 percent annual rate from 1990 to 1995. Given that Chinese exports were directly competitive with the exports of several

Southeast Asian countries, this development implied that to remain competitive, Southeast Asian countries would have had to export at lower prices. Thus the emergence of China as a major exporter is equivalent in its effect on the long-run equilibrium real exchange rate of countries producing similar products to a terms of trade deterioration. The implication is that this development in itself would cause the long-run equilibrium real exchange rate of such countries to depreciate.

The cyclical development concerned monetary policy in Japan. The explicit adoption of an expansionary monetary policy in early 1995 to combat that country's long recession led to a rapid expansion in money growth after the second quarter of the year, which was accompanied by a sharp reduction in domestic interest rates and a strong depreciation of the yen relative to the U.S. dollar. The implications for Southeast Asia were two. First, the steep reductions in Japanese interest rates acted as a push factor driving capital to the region. As Figure 27.4 illustrates, the five Asian countries depicted all registered a sharp acceleration of reserve growth in 1995–1996.[15] Despite countervailing domestic measures (see later), this acceleration of capital flows into the region resulted in macroeconomic overheating in the form of accelerating inflation. Second, the depreciation of the yen relative to the U.S. dollar caused the nominal effective exchange rates of the Southeast Asian countries – for which Japan was the dominant trading partner, but which tended to stabilize their exchange rates against the U.S. dollar – to appreciate after mid-1995.[16]

2. Real Exchange Rate Misalignment in Thailand

These two developments had important implications for real exchange rates in Southeast Asia. Consider specifically the case of Thailand. From 1987 to 1997, the Thai baht fluctuated in a narrow range of ±3 percent around a U.S. dollar parity of 25.4 baht per dollar. This bilateral nominal exchange rate stability against the U.S. dollar proved to be consistent with a trend in real effective depreciation of the baht during the second half of the decade of the 1980s, because of the secular depreciation of the dollar in nominal effective terms, and with a fairly stable real effective exchange rate from 1990 to 1994. During the latter period, the appreciation of the dollar was offset by favorable inflation performance in Thailand relative to that of the country's trading partners, making a stable real effective exchange rate consistent with a stable bilateral exchange rate against the U.S. dollar.

The monetary developments in Japan intervened to change this situation. First, the Thai economy began to overheat in response to the new surge in capital inflows. The inflation rate, which had gradually come down in the early 1990s to the

15 Kaminsky and Reinhart (1998) document the importance of interregional financial links in East and Southeast Asia, while Montiel and Reinhart (1999) provide empirical evidence of the independent role of Japanese interest rates as a push factor for Asian economies.

16 See Frankel and Wei (1994).

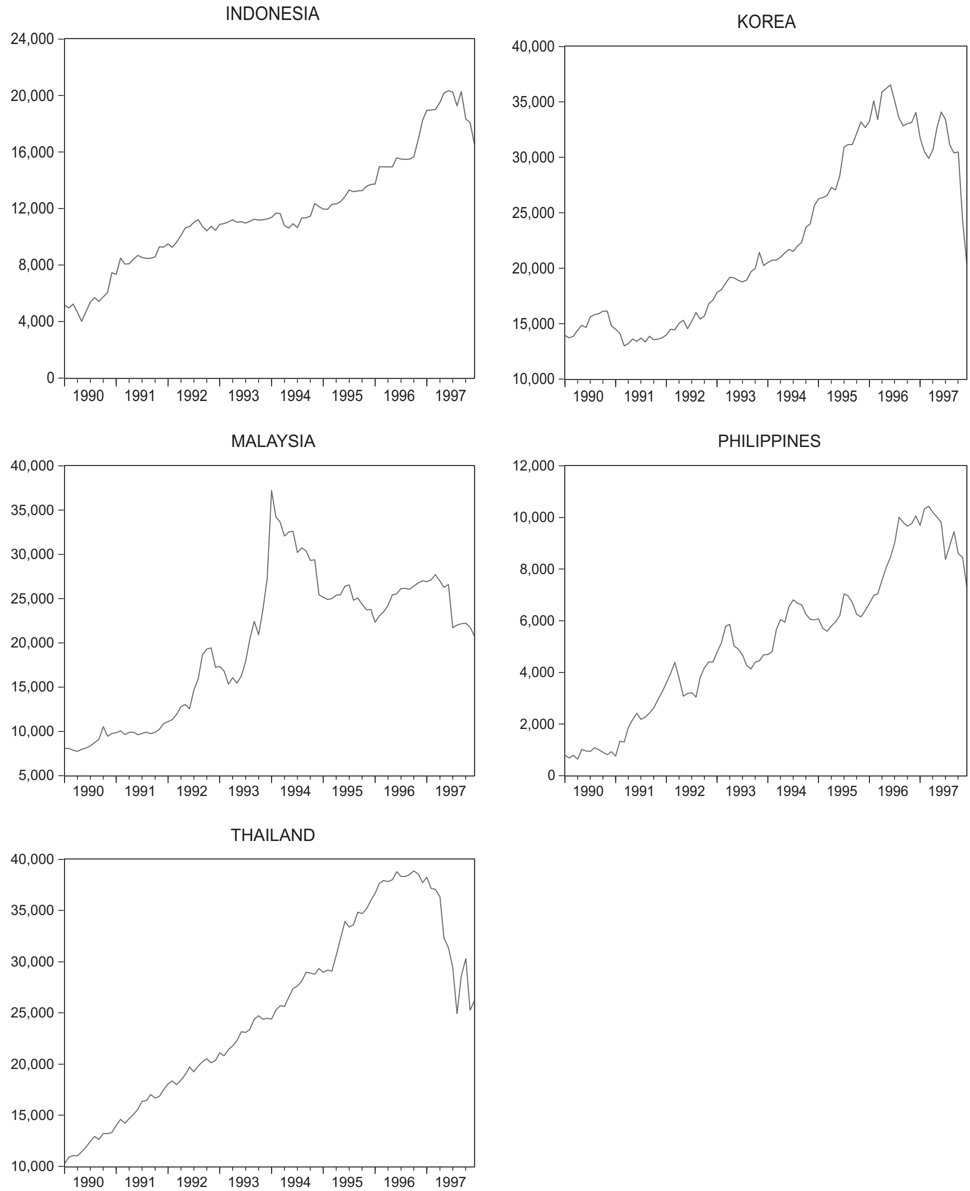


Figure 27.4. Reserve accumulation in five Asian crisis countries, 1990–1998

2.5 percent annual range – substantially below that of the country's trading partners – began to accelerate in mid-1993, gradually ascending to the 7 percent range by early 1996. This deterioration in relative price performance contributed to an appreciation of the real effective exchange rate. Second, the depreciation of the Japanese yen in the second quarter of 1995 implied that after depreciating sharply during the first quarter of the year, the U.S. dollar began to appreciate quite rapidly against the currencies of Thailand's other trading partners. After mid-1995, the baht depreciated in nominal terms from the bottom to the top of the range within which it had traded since 1987, but this proved to be far short of sufficient to offset the combined effects of accelerating domestic inflation and nominal effective appreciation of the U.S. dollar. As a consequence, the baht appreciated by nearly 20 percent in real effective terms over the two-year period from mid-1995 to mid-1997, a performance that represented a sharp break with experience since the mid-1980s. Even measured relative to its average value during 1990–1994, the baht had appreciated in real terms by about 12 percent by mid-1997. The significance of this change depends on what had been happening to the equilibrium real exchange rate in the meantime. As already mentioned, an important medium-term development affecting the equilibrium real exchange rate was that the countries in the region may have been losing competitiveness during the first half of the decade of the 1990s owing to the emergence of China as a major exporter. However, this effect has proven to be controversial. Fernald et al. (1998) have argued, for example, that China's export growth during the early 1990s came at the expense of Hong Kong, Korea, Singapore, and Taiwan rather than at the expense of the Southeast Asian countries.[17]

In any case, as we saw in Chapter 16, the equilibrium real exchange rate depends on a variety of other fundamental factors as well. To assess the extent of real misalignment in Thailand on the eve of its balance of payments crisis, recall the cointegrating equation reported in Chapter 16 for the estimation of the long-run equilibrium real exchange rate in Thailand (based on annual data from 1970 to 1996):

$$\log(REER) = 6.113 - \underset{(0.028)}{0.746}\log(CGR) - \underset{(0.016)}{0.581}\log(TOT) + \underset{(0.002)}{0.012}\ INFL$$

where *REER* is an index of Thailand's real effective exchange rate (with an increase representing a depreciation), *CGR* is the share of government consumption in GDP, *TOT* is an index of Thailand's terms of trade, and *INFL* is the world rate of inflation. Standard errors are in parentheses.[18] The theoretical rationale for the roles of these variables can be derived from the model of Chapter 16. As shown

[17] On the other hand, in a simple bivariate regression, China's share of world exports explains about 80% of the variation in Thailand's terms of trade during 1970–1994, suggesting that assessing China's impact on the equilibrium real exchange rates of the Southeast Asian countries may require going beyond an examination of export shares.

[18] This equation was estimated using the Johansen maximum-likelihood method. All the variables in the cointegrating equation possessed a single unit root, according to the results of augmented Dickey-Fuller tests, and there was no evidence of more than a single cointegrating vector among them. Other potential fundamentals, including the share of public investment in GDP, the openness ratio (exports plus imports over GDP), the world real interest rate, and a time trend to capture secular Balassa-Samuelson effects, entered the cointegrating equation with theoretically inappropriate signs and/or were statistically insignificant.

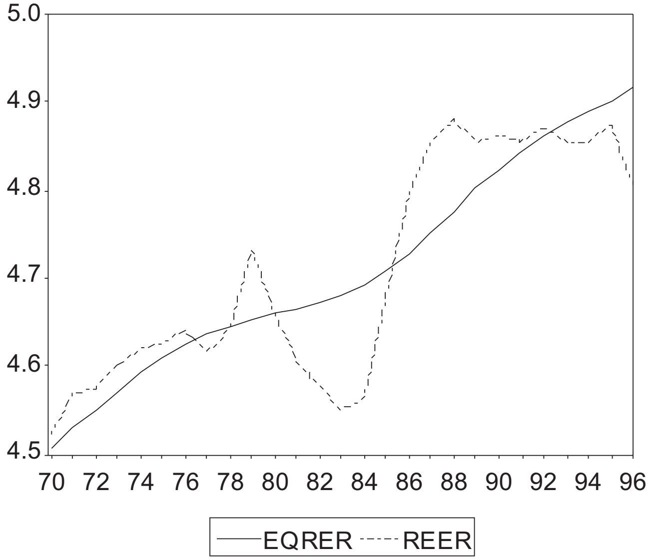


Figure 27.5. Thailand: Actual and estimated long-run equilibrium real exchange rates

in Chapter 16, an estimate of the long-run equilibrium real exchange rate for Thailand can be generated by calculating the fitted values of this equation using estimated permanent values of the fundamentals. The actual and estimated long-run equilibrium real exchange rates for Thailand were plotted in Figure 16.3 and are reproduced in Figure 27.5.

Notice that a gap between the actual and the estimated equilibrium real exchange rates began to emerge after 1992. By 1996, the estimated gap amounted to 11 percent of the estimated equilibrium real exchange rate.[19] How large would this gap have had to be before misalignment became a perceived problem? Recall that by comparison, Loayza and Lopez (1997) estimated misalignment of 27 percent for Mexico in 1994. A more useful general benchmark was provided by Goldfajn and Valdes (1999), who found, based on a large sample of countries, that the probability of a smooth return to the equilibrium real exchange rate from a short-lived overvaluation of 15 percent was about 16 percent in their sample for episodes that started after 1980. An even more relevant comparison might be with Thailand's own prior history. According to the methodology just described, the estimated degree of misalignment in Thailand in 1983–1984, just prior to a 15 percent devaluation of the baht in November 1984, was nearly 13 percent. The implication, then, is that the prospect of a nominal devaluation had to be taken seriously in Thailand by 1996.

[19] The range of available estimates of misalignment in Thailand before the crisis is large. Haque and Montiel (1999), using a general equilibrium simulation methodology, estimate misalignment of 17%. By contrast, Chinn (1998) finds the baht to have been approximately in equilibrium, while Tanboon (1998) estimates a misalignment of 30%.

3. Financial-Sector Vulnerability in Southeast Asia

Unlike exchange rate misalignment, which emerged rather quickly in Southeast Asia during the mid-1990s, financial fragility accumulated gradually during the inflow period. As is now well known, the late 1980s and early 1990s saw significant moves toward financial liberalization in several of these countries, but as in Mexico, these measures were implemented before the institutional apparatus was put in place to properly regulate and supervise these institutions. As in the case of Mexico, Southeast Asian countries, including Thailand, maintained very open capital accounts. Indeed, the establishment of the Bangkok International Banking Facility (BIBF) in Thailand in 1993 went further by actively promoting capital inflows.

The results were as outlined in Chapter 26: domestic lending booms, fueled in large part by external borrowing, emerged in the early 1990s in Indonesia, and later in Malaysia, the Philippines, and Thailand. In each of these countries, real bank and nonbank credit growth far exceeded output growth during those years. In addition, the exposure of banks to the property sector was far greater by 1997 in all four Southeast Asian countries than in Asian countries that did not experience similar lending booms such as India and Korea. That this lending proved to be excessive is suggested by the fact that vacancy rates were much higher in the capital cities of these countries than elsewhere.[20] The upshot is that, as in Mexico earlier, exchange rate overvaluation and financial-sector fragility had emerged among the Southeast Asian countries during the two to two and a half years prior to the crisis, and particularly so in Thailand.

4. Macroeconomic Policies in 1995–1996

This situation was exacerbated by the policy response to an upsurge in capital inflows during 1995–1996. Capital inflows to Southeast Asia increased sharply during these years, with short-term inflows playing a much larger role than they had earlier. As mentioned earlier, rising inflation and increasing current account deficits increased concerns about macroeconomic overheating throughout the region during this time.

The policy response essentially took the form of tight monetary policy to sustain domestic interest rates at relatively high levels. Because, as had Mexico before, countries in the region continued to pursue nominal exchange rate targets (despite occasional official statements to the contrary), tight money essentially meant the sterilization of balance of payments surpluses, conducted through a variety of means. Indonesia, Malaysia, the Philippines, and Thailand all tended to sterilize consistently in response to large inflows. Thailand increased the intensity of

[20] Evidence on property-sector exposure and vacancy rates is provided in Alba et al. (1998).

sterilization in 1993, in response to an upsurge in private capital inflows associated with the establishment of the BIBF. As capital inflows continued to increase in size, domestic interest rates continued to be increased.

Though Southeast Asian countries had achieved a substantial fiscal adjustment over the course of the inflow episode (which began in the region roughly in 1987–1988), this fiscal adjustment was in part the product of a medium-term fiscal strategy rather than consisting solely of a short-run countercyclical policy response to an expansionary external financial environment. Indeed, according to Alba et al. (1998), in Indonesia, Malaysia, and the Philippines, the *fiscal impulse* (the change in the full-employment surplus, a measure of the stance of fiscal policy) moved *pro-cyclically* throughout the 1991–1996 period, whereas in Thailand, fiscal policy was countercyclical in 1992–1993 but had also become pro-cyclical by 1995–1996. The short-run inflexibility of fiscal policy as a stabilization instrument placed the short-run burden of stabilization on monetary policy. Consistent with the inflow experience described in Chapter 24, one implication of this policy mix was an intensification of short-term inflows intermediated through the domestic financial system.[21]

Though it is easy to understand why funds would flow into the domestic financial system under these conditions, it is less obvious how the loan demand on which financial institutions depend to service their liabilities could have been sustained under relatively high domestic real interest rates. The answer is the emergence of asset price inflation associated with the rapid financial expansion in the first half of the 1990s described earlier. The legacy of this situation, however, was a financial system with borrowers whose creditworthiness and the value of whose collateral were heavily dependent on inflated asset values. This made the net worth of these institutions vulnerable to a downward correction of domestic asset prices. Such a downward correction could come about in two ways: through a negative reassessment of the earning streams associated with these assets or through an increase in the discount rates applied to those earning streams. As shown earlier, the second of these began to play a role by late 1995, as the sharp monetary tightening continued to raise domestic interest rates throughout the region. Indeed, owing to high interest rates, stock market performance in the region turned poor in 1994–1995.

In addition to premature financial liberalization and mismanagement of the exchange rate, the heavy reliance on tight money to combat overheating was a key mistake in short-run macroeconomic policy in Southeast Asia during the inflow episode prior to the events of 1997. The excessive weight placed on monetary policy put strains on domestic asset values and increased the stock of short-term

[21] For cross-country evidence on the links between sterilization and the magnitude of short-term inflows, see Montiel and Reinhart (1999).

external liabilities. The combination of real exchange rate appreciation, financial-sector fragility, fiscal rigidity, and the accumulation of large short-term external liabilities recalls the ingredients of the Mexican crisis at the end of 1994. As in Mexico, these policy mistakes did not imply the *necessity* of a crisis (i.e., they did not constitute sufficient conditions for a crisis), but they did create a state of heightened vulnerability centered on the financial sector. Nonetheless, in the absence of negative shocks, a crisis may have been avoidable, or at least postponable.

5. Outbreak of the Crisis in Thailand

Sufficient conditions for the crisis materialized in the form of adverse external conditions during 1996. The directly observable shock that preceded the Southeast Asian currency crisis was a collapse in export growth. Poor export performance materialized throughout the region in 1996. Asian export growth slowed markedly in the first half of 1996 (in dollar terms, 7% growth, compared to 20% in 1995) and, despite expectations to the contrary, failed to recover as the year wore on. The implications were the following:

- Poor export growth implied reduced *GDP* growth through the standard aggregate demand channel captured in the model of Part 2. This reduced the income streams expected to be associated with domestic assets and dampened domestic asset values through this channel, which reinforced the negative impact of high domestic interest rates.
- Poor export growth introduced an element of noise into the medium-term competitiveness calculation. To the extent that the export slowdown reflected a loss of competitiveness, of course, it implied that the gap between the actual and equilibrium real exchange rates may have been larger than had earlier been imagined.

Why did export growth slow? Possible sources of loss of competitiveness include those already mentioned, that is, the effects of real exchange rate appreciation and growth in Chinese export capacity. Other factors included a slowdown in economic growth in Japan and a collapse in semiconductor prices (blamed by the press on worldwide overcapacity and a weak personal computer market due to poor growth performance in both Japan and Western Europe). These may have been transitory real shocks, which do not in themselves affect the equilibrium real exchange rate. However, as we have seen (Chapter 18), the classic textbook response to such a shock for the purpose of stabilizing aggregate demand is a temporary exchange rate depreciation. The contribution of this shock to the crisis may therefore have come in the form of increasing the perceived likelihood of a devaluation.

The upshot is that negative shocks during 1996 had two effects:

1. They may have increased the perceived likelihood of a nominal exchange rate adjustment.
2. They increased the vulnerability of financial sectors in the region by depressing asset values and weakening the balance sheets of financial institutions.

In other words, these shocks reinforced the two sources of an emerging balance of payments crisis analyzed previously. The emergence of apparent misalignments implied the perception in financial markets that nominal exchange rate adjustments might be forthcoming, while fragility in the financial sector suggested that the costs of resisting such a misalignment, under pressure from speculation, would likely be perceived by the authorities in the region as prohibitively high. Because high interest rates would have impaired both the balance sheets and cash flows of domestic financial institutions, market participants would have perceived that fighting off a speculative attack through this traditional method would have been judged too costly by the authorities, particularly in light of the region's traditional commitment to the competitiveness objective. These two factors interacted, as in the case of Mexico, to make Thailand vulnerable to a crisis of confidence.

The first hints of currency problems in Thailand emerged during late July and early August 1996, triggered by worries over export competitiveness. Following a bleak report on economic prospects issued by the Bank of Thailand, the central bank was forced to spend US$1 billion to support the baht.[22,23] By the end of the month, though, the bank still possessed US$39.4 billion in foreign exchange reserves. Nevertheless, in September, Moody's downgraded Thailand's short-term foreign debt, noting financial-sector problems and the rapid accumulation of foreign debt during 1995 (amounting to a 40.7% increase in the stock of debt outstanding over the course of the year, to a year-end value of US$41.1 billion).

Whereas Mexico's currency crisis played itself out during a 10-month period in 1994, the crisis in Thailand took almost a full year (August 1996 to July 1997) before it culminated in the abandonment of the exchange rate parity. Over the course of that year, news on export performance, economic growth, and financial-sector problems grew progressively worse, and outflow episodes became progressively more severe. As growth slowed and domestic interest rates were maintained at relatively high levels to defend the currency, the stock market fell, losing 35 percent

[22] The Bank of Thailand revised 1996 projected real GDP growth from the 8.3% earlier forecast to 7.8% and revised projected export growth to 10.2%, compared to 17.4% projected earlier and to 23.6% in 1995.

[23] Earlier blips had affected the Indonesian rupiah (but for idiosyncratic reasons, i.e., when a medical trip abroad by President Suharto fueled concerns over political instability) and the Malaysian ringgit in early January 1996, when information became available about the size of Malaysia's current account deficit.

of its value during 1996. The government's handling of the crisis left much to be desired from its outbreak, until the final abandonment of the peg on July 2, 1997, and beyond. Key mistakes were made both with respect to exchange rate policy and policies toward the financial sector.

In light of the analysis in Section II, the government's biggest mistake was to attempt to hold the nominal value of the baht for almost a year. Its adherence to the exchange rate peg in the context of devaluation expectations resulted in large capital outflows and very high domestic interest rates, which, through their effects on the government's liquidity position and the balance sheet of the financial system, together magnified the uncertainty and instability that resulted after the baht was eventually floated. Despite the pressure on the currency in mid-1996 and again in February 1997, at the end of March 1997, the Bank of Thailand continued to report foreign exchange reserves of US$38 billion, almost unchanged from the July 1996 level. By the time the baht was floated in July 1997, reserves had officially fallen only to US$33 billion. It later became apparent, however, that reserves had been maintained through large *swap* transactions (currency trades that would be reversed at an agreed future price and future date), leaving the bank with a stock of future dollar liabilities in excess of US$23 billion after the flotation of the currency.

As suggested by the model of Section II, the loss of liquidity created uncertainty in the market after the baht was floated as to whether the country would be able to meet its large short-term external obligations in the event that private agents did not roll over a significant portion of them, and this perceived vulnerability to a liquidity crisis undermined the value of the currency after the float. As the model of that section would suggest, the perception that the exchange rate would have to move substantially in the event of a confidence crisis may itself have accelerated the onset of the crisis.

Given the exchange rate policy, the second key mistake was to postpone the resolution of the problems of the financial system. The government initially (during most of 1996) denied the vulnerability of the country's 15 commercial banks and 90 finance companies. However, as mentioned earlier, maintaining the fixed exchange rate required very high domestic interest rates, which in turn continued to undermine the value of assets held by the financial system that the government was otherwise trying to support. In late February 1997, for example, depositors engaged in a run on Thai finance companies, transferring deposits to the relatively safer banks and moving money abroad.

Not coincidentally, renewed pressure on the baht emerged in February. The Bank of Thailand responded by tightening monetary policy. High interest rates resulted in a succession of financial crises triggered by falling property values. The central bank, engaged in its lender-of-last-resort function, extended credit to these institutions, which was later estimated to total US$15.7 billion. The effect of these policies was to jeopardize the solvency both of finance houses and banks, which added to the stock of nonperforming assets in the financial sector and to the eventual cost of resolving the sector's difficulties.[24] The increase in this cumulative cost impaired the government's fiscal position, and this unresolved liability overhang was a second source of uncertainty that increased instability in the period after the currency was ultimately allowed to float. It also magnified the ultimate fiscal burden associated with the crisis.

The government finally began to signal a tougher stance toward the financial sector in late June. Merger laws to accommodate the takeover of companies in difficulties were eased, and 16 finance companies were ordered to suspend operations and seek merger partners within 30 days. The government indicated that troubled finance companies that failed to merge would be allowed to go under. This suggested for the first time that though the government's implicit guarantee of bank deposits continued to hold, shareholders in the finance companies would not be bailed out.

Unfortunately, the measures adopted toward the financial sector in late June proved to be too little too late. The central bank announced in June that its gross foreign-currency reserves had fallen to US$33.3 billion in the course of five straight months of balance of payments deficits, not including short-term foreign exchange liabilities in the form of swaps undertaken to support the baht. The baht was finally floated on July 2. In the first day, it fell by over 16 percent, and in the space of a month, it depreciated by over 25 percent. Despite the float, however, the country's problems were not over. The shortage of liquidity and the accumulation of nonperforming loans in the financial sector left a significant overhang of uncertainty in the economy, and the Bank of Thailand intervened to seek to avoid excessive depreciation of the baht. Rather than moving toward lower domestic interest rates, on the occasion of the float, the central bank raised its discount rate by two percentage points. Despite the government's insistence as late as July 7 that it required no International Monetary Fund (IMF) assistance, the uncertainty surrounding the liquidity and fiscal problems led it to begin negotiations with the IMF on July 28, and a program was agreed on one week later. As in the case of Mexico, Thailand's financial crisis resulted in a severe short-run contraction in real *GDP*.

V. SUMMARY

The experiences reviewed in this chapter have taught us that whatever the macroeconomic costs associated with the overheating potential created by the arrival of large capital inflows to financially open and solvent emerging economies, they pale beside those created by the abrupt cessation and reversal of capital flows that

[24] In March 1997, Standard & Poor's rating service estimated the cost to the government of bailing out insolvent institutions at 6% of GDP, while by August, that figure had increased to 12%–15% of GDP.

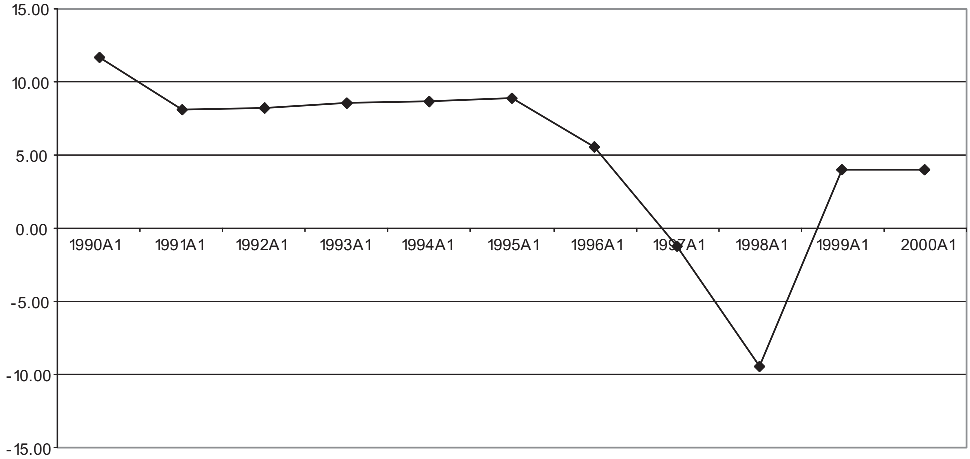


Figure 27.6. Real GDP growth in Thailand, 1990–2000

are associated with liquidity crises in such countries. The two issues are related, however: vulnerability to such crises is the result both of policies followed during the inflow episodes and of the ways that countries responded to the implications of these policies.

In both Mexico and Thailand, perceived exchange rate misalignment emerged during the inflow period as the result of policies pursued during that period – specifically, an open capital account, an inflexible exchange rate policy, and excessive reliance on monetary rather than fiscal policy to combat overheating after a premature domestic financial liberalization.

Overvaluation interacted with financial-sector fragility to produce the currency crisis. Thailand's higher growth rate and higher rate of investment could not protect it from this fate. These features of Thailand's economy would have suggested a superior long-term debt-servicing capacity but would not have been expected to insulate it from a sharp short-run downturn as a result of a financial-sector collapse brought about by a high–interest rate defense of the currency. In both countries, postponing the exchange rate adjustment absorbed valuable liquid assets and left the economy in a vulnerable liquidity position. The failure of external creditors to roll over short-term debt subsequent to the flotation of the currency created a situation in which the debt was unlikely to be serviced on market terms in Mexico as well as in Thailand – albeit for different reasons, arising from the fact the short-term external debt was public in Mexico's case and private in the case of Thailand.

What can we learn from the experiences of these two crises? There are two main lessons. The first is that in the presence of high capital mobility – that is, once countries have become reintegrated with world financial markets – exchange rate overvaluation in a context in which the domestic economic costs of a high interest rate defense make the sustainability of such a defense problematic constitutes a recipe for a currency crisis.

Crisis avoidance can thus take several forms. The most obvious, of course, is to avoid reintegration with world financial markets in the first place. But as we saw in Chapter 22, once countries have removed long-standing capital account restrictions, there is weak evidence at best that the implementation of new market-based instruments can be effective in slowing capital movements. This does not mean, however, that countries have no choice about the pace of their reintegration with world capital markets. If they do, it may be best to postpone reintegration until the sources of vulnerability to currency crises are under control.

Where reintegration is a fact, on the other hand, escaping currency crises means avoiding overvaluation as well as the conditions that make a high–interest rate defense of the currency less than credible. One way of achieving the first objective, of course, is allowing the exchange rate to float (and accepting the associated volatility). If the exchange rate is to be officially determined, however, it should be managed actively to avoid misalignment, perhaps with some assistance from the adjustment of policy-based fundamentals (e.g., fiscal policy), if feasible. Both Mexico and Thailand fell short on this score. A corollary of this point is that exchange rate–based stabilization is a dangerous strategy with an open capital account. Given that stabilizing from a high (but not hyperinflationary) rate of inflation typically includes an initial period of real appreciation, the challenge is how to permit an adjustment of the real exchange rate without going through a balance of payments crisis. Chile in 1982, Mexico in 1994, and a subsequent crisis that we have not discussed in this book – that of Brazil in 1999 – suggest that moving to the flexibilization stage without trauma is a challenging task. On the other hand, Israel (1987) and Poland (1990) suggest that it is not a hopeless one.

The other component is enhancement of the credibility of a high–interest rate defense of the currency, were that to become necessary. The experience of the European Monetary Union in 1992–1993 suggests that the achievement of complete credibility is likely to prove an elusive goal. As we have seen, credibility can be enhanced through institutional innovations such as increased central-bank independence or the adoption of a currency board, but the experiences of Argentina in 1995 and Hong Kong in 1997 suggest that even under relatively favorable circumstances (arising from Argentina's inflationary history and Hong Kong's ample foreign exchange reserves), this remains a challenge. At the very least, though, governments can avoid measures that undermine the credibility of their resolve. In both the Mexican and Thai cases, loss of credibility was greatly magnified by the fragility of the financial sector due to inappropriate liberalization during the early 1990s. Given that maturity mismatches are unavoidable in banking, policies that ensure the health of the financial system can increase the likelihood that a defense of the currency will be sustained.

The second main lesson is that once vulnerability has arisen, in the form of perceived overvaluation and inability to mount a high–interest rate defense, postponing the inevitable merely serves to aggravate its consequences, primarily by increasing vulnerability to a subsequent liquidity crisis, which involves a much more dramatic loss of confidence and larger exchange rate movements than would otherwise happen. After all, a puzzling feature of both the Mexican and Thai experiences is that, unlike the United Kingdom and Italy in 1992, the floating of the currency was followed by a severe real-sector crisis rather than an acceleration of growth. One difference, of course, was the advanced stage of financial fragility in both Mexico and Thailand, and the banking crises that accompanied the exchange rate crises. But one might have expected this to have been alleviated by the elimination of the misalignment, reinforcing a tendency for an acceleration of growth. The liquidity crisis that followed the float in both cases undermined this possibility.

In the Mexican case, postponement was financed by reserve losses and short-term foreign-currency financing of the public debt, reducing the government's liquid dollar assets and increasing its liquid dollar liabilities. The prohibitive fiscal costs of servicing these liabilities in the event of a run made it rational for individual creditors not to roll over their claims, an event that seems to have greatly magnified the short-run costs of Mexico's real exchange rate adjustment through the requirement of draconian fiscal adjustment and loss of investor confidence, which was associated with high interest rates and an overshooting nominal exchange rate. In the case of Thailand, postponement was financed by reserve losses, and illiquidity resulted from the combination of reserve depletion and a large stock of previously existing liquid private-sector external liabilities that arose in part from policies designed to combat overheating in 1995–1996. The value of these claims on the private sector was impaired by the adverse effects on the solvency of private debtors of a large nominal exchange rate depreciation, again making it rational for individual creditors not to roll over their claims. The nominal exchange rate indeed moved very sharply after the currency was floated, in the context of a liquidity crisis.

The upshot is that the similarities between Mexico and Thailand mattered much more than the differences, and the policy message from the two experiences is the same: perceived real exchange rate misalignment in the context of an open capital account with a fragile domestic financial sector is a reliable recipe for a currency crisis. Postponing the adjustment invites a liquidity crisis that, in the context of a vulnerable domestic financial system, creates very severe short-run real economic dislocations.

Unfortunately, Mexico and Thailand were not the only emerging economies to suffer through financial crises in the 1990s. The next chapter examines lessons that we have (and have not) learned from the broader worldwide crisis experience during that first decade in which several middle-income developing countries became emerging economies.

APPENDIX 27.1. CURRENCY CRISES FROM 1990 TO 2002

1990	Algeria, Congo, Côte d'Ivoire, Croatia, Ghana, Lebanon, Pakistan, Rwanda, Sudan, Uganda, Zimbabwe
1991	India, Jamaica, Jordan, Tunisia, Zambia
1992	China, Ecuador, Ethiopia, Malawi, Nicaragua, Nigeria, Trinidad and Tobago
1993	Cameroon, Côte d'Ivoire, Liberia, Senegal
1994	Algeria, Congo, Croatia, Kenya, Mexico, Niger, South Africa, Togo, Turkey, Venezuela
1995	Azerbaijan, United Arab Emirates
1996	Liberia, Myanmar, Pakistan
1997	Cameroon, Congo, Gabon, Guinea, Korea, Panama, Romania, Tanzania, Zimbabwe
1998	Indonesia, Moldova, Niger, Russian Federation, Sri Lanka, Ukraine
1999	Angola, Belize, Brazil, Nigeria
2000	Ghana
2001	Congo, Gabon, Turkey, Uruguay, Zambia
2002	Angola, Libya, Venezuela

Source: Frankel and Wei (2004)

REVIEW QUESTIONS

1. How do first- and second-generation crisis models explain the emergence of a currency crisis?
2. What distinguishes third-generation currency crisis models from their predecessors?
3. Explain the conditions under which a rational speculative attack could emerge in the third-generation crisis model that we developed in this chapter. Is such an attack ever inevitable?
4. How did vulnerability to speculative attacks arise in Mexico and Thailand? How were the causes of vulnerability in these countries related to the preceding capital-inflow episodes?
5. What are the key policy lessons from the 1994 Mexican and 1997 Thai currency crises?

EXERCISES

1. Use the third-generation currency crisis model of this chapter to explain the types of policies that countries with intermediate exchange rate regimes would have to implement to avoid vulnerability to currency crises.
2. How does the third-generation crisis model described in this chapter fit the empirical evidence on the determinants of currency crises described in Chapter 19?
3. What is the link between financial-sector regulation and supervision as analyzed in Chapter 22 and the vulnerability of a soft exchange rate peg to a third-generation currency crisis?

4. In what sense do first-generation crisis models help to explain the currency crises that emerged in some of the Latin American countries that defaulted on their external debts during the 1980s?
5. Explain how a fragile banking system (one that is near insolvency) increases the vulnerability of a country that maintains a soft peg to a second-generation currency crisis.

REFERENCES AND FURTHER READING

Alba, Pedro, Amar Bhattacharya, Stijn Claessens, Swati Ghosh, and Leonardo Hernandez (1998), "Volatility and Contagion in a Financially-Integrated World: Lessons from East Asia's Recent Experience," unpublished manuscript, World Bank.

Alesina, Alberto, Alessandro Prati, and Guido Tabellini (1990), "Public Confidence and Debt Management: A Model and a Case Study of Italy," in R. Dornbusch and M. Draghi, eds., *Public Debt Management: Theory and History*, Cambridge: Cambridge University Press, pp. 94–117.

Calvo, Guillermo (1988), "Servicing the Public Debt: The Role of Expectations," *American Economic Review*, vol. 78, pp. 647–661.

Calvo, Guillermo, and Enrique G. Mendoza (1996), "Reflections on Mexico's Balance of Payments Crisis: A Chronicle of a Death Foretold," *Journal of International Economics*, vol. 41, pp. 235–264.

Chinn, Menzie (1998), "Before the Fall: Were East Asian Currencies Overvalued?" working paper 6491, National Bureau of Economic Research.

Dornbusch, Rudiger, and Alejandro Werner (1994), "Mexico: Stabilization, Reform, and No Growth," *Brookings Papers on Economic Activity*, vol. 1, pp. 253–297.

Eichengreen, Barry, Andrew K. Rose, and Charles Wyplosz (1995), "Exchange Market Mayhem: The Antecedents and Aftermath of Speculative Attacks," *Economic Policy*, vol. 21, pp. 249–312.

Fernald, John, Hali Edison, and Prakash Loungani (1998), "Was China the First Domino? Assessing Links between China and the Rest of Emerging Asia," International Finance Discussion Paper 604, Board of Governors of the Federal Reserve System.

Frankel, Jeffrey A., and Shang-Jin Wei (1994), "Yen Bloc or Dollar Bloc? Exchange Rate Policies of the East Asian Economies," in T. Ito and A. Krueger, eds., *Macroeconomic Linkages: Savings, Exchange Rates, and Capital Flows*, Chicago: University of Chicago Press, pp. 295–329.

Frankel, Jeffrey, and S.-J. Wei (2004), "Managing Macroeconomic Crises: Policy Lessons," in J. Aizenman and B. Pinto, eds., *Managing Volatility and Crises: A Practitioner's Guide* (Cambridge: Cambridge University Press), pp. 315–405.

Giavazzi, Francesco, and Marco Pagano (1990), "Confidence Crises and Public Debt Management," in R. Dornbusch and M. Draghi, eds., *Public Debt Management: Theory and History*, Cambridge: Cambridge University Press, pp. 125–142.

Goldfajn, Ilan, and Rodrigo Valdes (1999), "The Aftermath of Appreciations," *Quarterly Journal of Economics*, vol. 114, pp. 229–262.

Haque, Nadeem, and Peter J. Montiel (1999), "Long-Run Real Exchange Rate Changes in Developing Countries: Simulations from an Econometric Model," in L. Hinkle and P. Montiel, eds., *Exchange Rate Misalignment: Concepts and Measurement for Developing Countries*, New York: Oxford University Press, pp. 381–404.

Kaminsky, Graciela L., and Carmen M. Reinhart (1998), "On Crises, Contagion, and Confusion," unpublished manuscript.

Krugman, Paul (1979), "A Model of Balance of Payments Crises," *Journal of Money, Credit, and Banking*, vol. 11, pp. 311–325.

Loayza, Norman, and J. Humberto Lopez (1997), "Misalignment and Fundamentals: Equilibrium Exchange Rates in Seven Latin American Countries," unpublished manuscript, World Bank.

Montiel, Peter J., and Carmen M. Reinhart (1999), "Do Capital Controls and Macroeconomic Policies Influence the Volume and Composition of Capital Flows? Evidence from the 1990s," *Journal of International Money and Finance*, vol. 18, pp. 619–635.

Sachs, Jeffrey, Aaron Tornell, and Andres Velasco (1996), "Financial Crises in Emerging Markets: The Lessons from 1995," unpublished manuscript.

Steiner, Roberto (1995), "The Mexican Crisis: Why Did It Happen and What Can We Learn?" unpublished manuscript, World Bank.

Tanboon, Surach (1998), "An Analysis of the Thai Currency in Retrospect," unpublished manuscript, Williams College.

Warner, Andrew M. (1997), "Mexico's 1994 Exchange Rate Crisis Interpreted in Light of the Nontraded Model," working paper 6165, National Bureau of Economic Research.

Werner, Alejandro M. (1995), "The Currency Risk Premia in Mexico: A Closer Look at Interest Rate Differentials," unpublished manuscript, International Monetary Fund.

28

Lessons from the Emerging-Market Crises of the 1990s and 2000s

In the preceding chapter, we examined the 1994 Mexican and 1997 Thai financial crises in some detail because these were two of the most important crises among emerging economies that entered the decade of the 1990s with a policy commitment to reform domestic financial systems and integrate themselves more fully with international financial markets. But the decade of the 1990s was characterized by a spate of financial crises, afflicting many other emerging economies as well. Partly, this may have reflected contagion (spillover effects) from the crises we have studied, for example, those of Argentina in 1995 or of several East and Southeast Asian economies in 1997. But partly, these emerging-economy crises have also been homegrown, for example, the Brazilian crisis of 1999 and the Argentine crisis of 2001–2002. Moreover, the incidence of financial crises has not been restricted to emerging economies. Crises have also occurred among industrial countries (e.g., the Nordic banking crisis that we reviewed in Chapter 26 and the European Exchange Rate Mechanism (ERM) crisis of 1992) as well as transition economies (the Russian crisis of 1998).

What lessons can we draw from the recent spate of financial crises in emerging economies for the central topic of this book: the conduct of domestic macroeconomic policies in emerging economies? To bring together the various themes that we have explored throughout the book, this chapter will attempt to summarize what we may and may not have learned about this issue from the financial turbulence of the 1990s and early 2000s.

It is worth noting at the outset that economists are still debating the lessons to be drawn from the macroeconomic experiences of earlier decades, that is, those of the 1930s, 1960s, and 1980s, among other decades. Thus it would be very surprising indeed if the profession had reached a consensus about the lessons to be learned from the financial crises that have characterized the two most recent decades. Many of the lessons from recent crises are still in dispute. In that sense, it is still a bit premature to guess what the enduring lessons of the recent spate of financial crises around the world will turn out to be. To acknowledge this unsettled state of affairs, this chapter will classify the lessons to be discussed into two types: those on which there is widespread agreement and those that are more problematic, with an emphasis on the first type. However, it is worth emphasizing that even with regard to those in the first category, there are important dissidents, and what you will read here is just one economist's views on where things stand.

We begin by discussing lessons on which there appears to be some broad agreement. These can be divided into several areas: the benefits and costs of financial integration, causes of currency crises, exchange rate policy, monetary policy, and fiscal policy. Afterward, we turn to issues that are important for macroeconomic management under current international circumstances and on which (sometimes strong) opinions have been expressed by knowledgeable observers, but on which substantial disagreement still exists. The chapter concludes with a discussion of what all this might mean for macroeconomic management in emerging economies.

I. LESSONS THAT WE HAVE LEARNED

The term *globalization* came into increased use during the decade of the 1990s. It refers to increased mobility of goods and services across international boundaries as well as the increased mobility of financial capital. In many countries, it has also come to mean increased international labor mobility. But globalization is not just an inevitable product of technological change. In many ways, the world economy was more globalized, that is, more integrated, from 1880 to 1914 than it has become during the last two decades. Focusing just on the financial aspects of globalization, not all countries have chosen to participate equally in the integration of world financial markets. China and India are well-known examples of countries that decided to go slowly in the financial area, and some countries have decided to withdraw, at least partially, from a state of high integration. In September 1998, for example, Malaysia, which had traditionally been a financially open economy, imposed severe capital controls in response to the Asian financial crisis that had broken out during the previous year.[1] Thus integration reflects in part a set of domestic policy decisions, and the decision to pursue more intense integration involves trading off the benefits and costs of doing so. The first set of lessons concerns what we have learned in recent years about these benefits and costs.

1. Benefits and Costs of Financial Integration

a. Importance of External Financial Shocks

Economists believe that everything has benefits and costs, of course, and as we saw in Chapter 22, financial integration is no exception to this rule. In particular,

[1] These controls were subsequently phased out gradually.

economists have always been aware that enhanced financial integration meant exposure to external financial shocks. At the same time, increased financial integration offers, among its other benefits, to dampen the effects of certain domestic shocks. What may not have been sufficiently appreciated prior to the experience of recent crises was how large the external financial shocks were likely to be. The first lesson, then, is the following:

> Increased financial integration, while offering significant benefits to developing countries, also comes with some important costs. The key costs involve exposure to external financial instability that, it turns out, can be quite substantial.

This is not an altogether new experience. A key lesson learned from the post Bretton-Woods era of floating exchange rates turned out to be that floating rates were much more volatile than economists had anticipated. It now appears that we also underestimated the volatility of the external financial environment facing emerging economies that become more integrated with world capital markets. In other words, the financial shocks faced by these countries have turned out to be larger and more common than was previously thought they would be. Such shocks can arise from at least three sources:

1. monetary policy in capital-exporting countries
2. heightened sensitivity to domestic macroeconomic developments
3. contagion from shocks afflicting other capital-importing countries

The first of these was already familiar before the crises of the 1990s. Indeed, as we saw in Chapter 24, some of the early capital-inflow literature (see Calvo et al. 1993) worried quite a bit about this problem before any of the recent crises materialized. What the crisis experience of the 1990s has suggested, however, is the following:

- In the face of domestic macroeconomic problems, the severity of the punishment inflicted by international financial markets may be far out of proportion with the magnitude of the apparent crime committed by the domestic macroeconomic authorities. The crisis in Mexico first made this clear (see especially Calvo 1996).
- Countries have to worry not just about financial developments in the capital-exporting countries and financial-market reactions to macroeconomic developments at home but also about the effects on the supply of funds facing the domestic economy of macroeconomic developments in other borrowing countries. This refers to the phenomenon of contagion, which afflicted Argentina, Brazil, and the Philippines in 1995, after the Mexican crisis in December 1994; several East and Southeast Asian countries in 1997, after the Thai crisis in July; and many countries around the world, after the Russian crisis of August 1998.

The implication of these two developments is that the external financial environment for countries that pursue increased integration with world financial markets may be much less stable than we previously thought, with attendant challenges for domestic stabilization policies.

b. Domestic Distortions and Financial Integration

A second lesson concerning financial integration has to do not with the external financial environment but rather with domestic circumstances that may affect the benefits and costs of financial integration:

> Financial opening is particularly dangerous when the domestic financial system is heavily distorted. Not only may this result in immiserizing external borrowing, but it will also increase the probability and likely severity of a domestic financial crisis.

As we saw in Chapter 22, a previously existing literature on the optimal sequencing of macroeconomic reform put financial opening at the end of the process of liberalization, after macroeconomic stabilization was achieved, commercial policy was liberalized, and domestic financial repression was removed. The fear was that the presence of financial repression would result in resource misallocation if the economy was opened up financially, largely through capital outflows.

The experience of the Southern Cone countries in the late 1970s and early 1980s alerted economists to a different kind of risk: that the combination of an open capital account and an improperly liberalized financial system could lead to a combination of domestic banking crisis and balance of payments crisis. As we saw in Chapter 26, the basic problem is that the combination of explicit or implicit deposit insurance, low net worth of banks, and poor prudential regulations and bank supervision creates incentives for bank managers to invest in highly risky ventures, which will cause many banks to fail, triggering a financial crisis.[2] Such crises have large fiscal costs and costs to real economic activity operating through the so-called credit channel of monetary policy. Because the resources absorbed by the banks can be much larger when the capital account is open, an open capital account magnifies the scale of the crisis. Moreover, under an open capital account, the fiscal costs and macroeconomic dislocations caused by a banking crisis are also likely to trigger a balance of payments crisis as both domestic residents and foreign creditors switch to foreign assets.

As we saw in the preceding chapter, with an open capital account and a weak financial system, not only can a domestic financial crisis trigger a balance of payments crisis, but the reverse is also true; that is, a balance of payments crisis can trigger a domestic financial crisis. The two phenomena thus tend to be closely associated. Indeed, a paper written at the end of the turbulent decade of the 1990s by Kaminsky and Reinhart (1999) documented the close association that has existed historically between crises of both types, so much so that they refer to them as

[2] Recall the analysis of the Chilean experience by Diaz-Alejandro (1985). See also Velasco (1987).

"twin" crises. In the preceding chapter, we argued that the Mexican and Asian crises in many ways replicated the Southern Cone experience and confirmed the dangers of combining an open capital account with a weak domestic financial system. Indeed, the major difference between the macroeconomically destructive Mexican and Asian balance of payments crises of the 1990s and the macroeconomically harmless 1992 crisis in the ERM was precisely the fragility of the domestic financial systems in Mexico and the Asian countries, a fragility that was not present in the European countries affected by the 1992 ERM crisis.

Because of the vulnerability it creates to "twin" crises, the combination of an open capital account and a weak financial system also has important implications for the conduct of exchange rate and monetary policies. We will discuss these implications later.

c. Factors Driving Capital Flows

As we saw in Chapter 24, during the first half of the 1990s, increased financial integration was accompanied by large capital inflows into several emerging economies. During the early years of this episode, many observers worried whether this new capital-inflow episode would culminate in a new debt crisis, just as the previous episode during 1974–1981 had done. One point of view, associated with former U.K. chancellor of the exchequer Nigel Lawson and former International Monetary Fund (IMF) official Walter Robichek, was that the new capital inflows were unlikely to prove problematic because in contrast with the inflows of the 1970s, which largely reflected external borrowing by the public sectors of the indebted countries, the new inflows were directed to private agents in the borrowing countries so that each individual transaction was subject to a market test. However, an important lesson emanating from the recent crises has been the following:

> The Lawson-Robichek doctrine that capital inflows reflecting market transactions between private agents do not pose macroeconomic risks has not held up well. The sectoral identity of the agents on the borrowing side of the market is not definitive in determining whether capital inflows are likely to be followed by capital flow reversals.

An important stylized fact is that all the major emerging-economy financial crises during the decade of the 1990s were preceded by periods of substantial capital inflows, and in each case, the domestic borrowers were predominantly private agents. Each of these crises was accompanied by persistent capital-flow reversals in the form of a cessation of private capital inflows or actual outflows, not unlike those that characterized the outbreak of the international debt crisis during the decade of the 1980s. The interpretation of this fact is a subject of dispute, with views ranging from those who emphasize the role of public-sector guarantees in driving both the capital inflows and subsequent reversals that characterized the crises of the 1990s (e.g., Dooley 1997) to those who blame the crises on characteristics of international capital markets (e.g., Radelet and Sachs 1998; Calvo and Mendoza 1996). What is common to both views, however (as well as to those in between), is that, contrary to simple versions of the Lawson-Robichek doctrine, the identity of the borrower does not constitute grounds for complacency concerning the macroeconomic risks associated with large capital inflows.

2. Causes of Currency Crises

The first three lessons listed previously suggest that financial integration is not a panacea and that indeed, in some circumstances, it can be downright dangerous, even if the associated capital flows are strictly market-driven transactions between private agents. A natural follow-up question is what can be done to minimize these dangers – more specifically, what can countries do to avoid homegrown currency crises as well as the adverse implications of crises elsewhere? To address this question, the next set of lessons from recent experience concerns the causes of currency crises as well as their implications for a variety of macroeconomic policies.

a. Multiplicity of Causes

A first lesson from the recent crises in this regard is the following:

> There is no single cause of currency crises. Thus there is no single cure or single policy that it is necessary to get right.

As described in the preceding chapter, in trying to understand the causes of currency crises, the analytical literature in macroeconomics has proceeded through a succession of crisis models. First-generation crisis models interpreted crises as resulting from an inconsistency between exchange rate and monetary policies – specifically, as the result of a monetary policy that was excessively expansionary given the officially determined path of the nominal exchange rate. In terms of the model of Part 2, this family of crisis models involved continuous and excessive credit expansion in a situation of high capital mobility, typically presumed to be driven by the need to finance a fiscal deficit. As we saw in Part 2, the eventual result of such a policy is to deplete the country's stock of foreign exchange reserves. These models thus blamed crisis on monetary and fiscal fundamentals, in the form of overly expansionary policies. Such models, originating with Krugman (1979), predated the international debt crisis of 1982, but that experience was interpreted as consistent with the first-generation models because of their emphasis on unsustainable fiscal policies.

Second-generation models originated with the ERM crisis of 1982. As described in the preceding chapter, in these models, the government does not follow a mechanical rule of holding on to a fixed exchange rate until it runs out of reserves but instead makes an optimal choice among competing macroeconomic objectives. A fixed exchange rate may be abandoned when the constraints that it imposes on monetary policy run counter to the government's desire to run a more expansionary monetary policy for domestic stabilization (or fiscal) reasons. Notice that, like the first-generation models, the basic conflict is between expansionary monetary policy and a fixed exchange rate. But in second-generation models, it is not *actual* but rather *prospective* monetary expansion that creates the basic inconsistency. This has two implications:

1. The fundamentals are appropriate, in the sense that the exchange rate would be sustainable in principle.
2. As in the examples we saw in the previous chapter, currency crises can be self-fulfilling because the fact of an attack can alter the costs the government perceives in adhering to the fixed rate.

Both classes of models emphasize the choices facing the monetary authorities. They note that monetary policy is caught between conflicting objectives: an external one (the exchange rate) and an internal one. The models differ with regard to the internal objective. In the case of the first-generation models, it is the need to generate seignorage revenues to finance a fiscal deficit, whereas in the second-generation models, based on the experience of Western Europe, it was the desire to combat recession or to lower public-sector borrowing costs, both of which might be achieved with a more expansionary monetary policy than would be consistent with a fixed exchange rate.

The Mexican and Asian experiences reviewed in the preceding chapter emphasize a different set of objectives for monetary policy that could conflict with maintaining a fixed exchange rate – that is, to maintain low domestic interest rates to protect a fragile domestic financial system – and they have given rise to third-generation crisis models that emphasize factors such as moral hazard, currency mismatches, liquidity problems, and associated sudden stops of capital flows. The main point is that the potential causes of crises have gradually been expanded with experience, from macroeconomic policy inconsistencies to the existence of a recession or a large stock of public debt to a weak domestic financial system. The crises of the 1990s have changed our understanding of the causes of currency crises by expanding the range of circumstances under which crises can reasonably be expected to emerge.

b. Traditional Recipes

This has an important implication for the rewards that countries can expect to receive from what has traditionally been understood to constitute macroeconomic good behavior:

> Neither a strong (stock and flow) fiscal position, a relatively small current account deficit, nor a larger current account deficit accounted for by a large share of investment in gross domestic product (*GDP*) and accompanied by rapid growth can necessarily protect a country against a severe *homegrown* currency crisis.

As we saw in the preceding chapter, both Mexico and Thailand (as well as other Asian countries that were affected by the crisis of 1997) had strong fiscal positions, both in the sense of small overall deficits or actual surpluses in their overall fiscal accounts (a flow) as well as low public-sector debt (a stock). Thus fiscal prudence was not enough to insulate these countries from the outbreak of financial crisis. One possible explanation is that, though these countries were not characterized by large *public-sector* imbalances, they did have large *private-sector* imbalances. As we have seen, this was true for Mexico and Thailand, which had large private sector–led current account deficits (in the neighborhood of 8% of *GDP*) in 1994 and 1997, respectively. However, among other countries that underwent crises in the 1990s, both Indonesia and Korea had relatively small current account deficits at the time of their crises, suggesting that the size of the current account deficit is not the critical variable either; that is, minimizing both public- and private-sector flow imbalances is not enough to protect an economy from a currency crisis. Finally, one lesson that at one time had been thought to have been learned from Mexico – that what mattered was the behavior of domestic saving and investment underlying the current account deficit – was proven wrong by Thailand, which registered very large rates of domestic saving and investment, together with rapid growth at the time of its crisis.

Some people have drawn from this observation the conclusions that currency attacks are whimsical (not based on fundamentals) and thus driven by pure panic. Others have concluded that they are driven by conspiracies or by the actions of a single large speculator (Malaysian prime minister Mahathir expressed this view at the time of the Asian financial crisis). Finally, as we saw in Chapter 18, based partly on this experience, some economists have come to the conclusion that there is no combination of domestic policies that can insulate countries with officially determined rates from succumbing to speculative attacks and thus that fixed exchange rates for national currencies are no longer feasible. We will come back to this issue later. The point for now is that the traditional rules of good behavior have clearly not been enough to protect countries from the onset of currency crises.

c. The Role of Debt Composition

Note that the traditional measures of good behavior are all solvency-based macroeconomic fundamentals. If these are not sufficient to explain the incidence of currency crises, there are several possibilities. The most obvious one, of course, is that the value of claims on residents of a country can become impaired through means other than general macroeconomic or fiscal insolvency. A crisis may arise not as the result of actual or prospective insolvency but as the result of anticipated policy choices, driven by changes in circumstances, that would have the effect of shifting the burden of taxation toward agents who hold such claims. As we saw in the preceding chapter, a special case would be that of a liquidity crisis, in which the change in circumstances is triggered by the holders of the claims themselves in the form of a panic, as in the case of runs on solvent banks.

Models of self-fulfilling speculative attacks triggering currency crises do not exonerate the domestic authorities from responsibility for such attacks, however, because they suggest that these attacks can be successful only if the values of certain macroeconomic fundamentals (the identity of which is model specific) fall within a specific range. One such liquidity-based fundamental is the stock of short-term external liabilities. Short-term external liabilities indeed appear to have played an important role in both the Mexican and Asian crises. They did so by creating a liquidity crunch when creditors refused to roll over these debts. Thus a third lesson concerning causes of crises is the following:

> Contrary to some views expressed early in the capital inflow experience, the maturity composition of a country's external liabilities seems to matter in determining a country's vulnerability to capital flow reversals.

However, the consensus on this particular lesson from recent crises is not as strong as for the prior ones. For one thing, liquidity-based stories are not the only way to rationalize the inability of the traditional solvency-based fundamentals to explain the crises of the 1990s. For any single crisis unexplained by the traditional fundamentals, it remains possible to adhere to the view that the crisis was ultimately driven by a solvency problem. This could happen if the insolvency existed before the crisis, but would not tend to be revealed by the traditional indicators, or if insolvency arises only in the event of a crisis.

The role of actual or prospective insolvency in explaining the crises of the 1990s is actually an unresolved issue in some instances. As already mentioned, the leading interpretations of the 1992 ERM crises relied on domestic recession and a high stock of domestic public debt as factors undermining the fixed exchange rate by weakening the authorities' resolve to keep interest rates high to deter speculation against the domestic currency. Floating the exchange rate to pursue a more expansionary monetary policy amounts to a change in policy that has the effect of imposing a capital levy on holders of assets denominated in the depreciating currency. The flight from the domestic currency is driven by this prospective tax, not by doubts about the solvency of the government implementing the policy change. On the other hand, whereas one interpretation of the Mexican and Asian crises is that a fragile domestic financial system can produce the same result through a similar mechanism, some observers have offered a different interpretation. In this alternative view, these balance sheet difficulties are perceived as having undermined the fixed exchange rate by creating a large unfunded liability for the public sector that impaired its solvency. The traditional indicators would simply miss this source of insolvency.

Despite this ongoing disagreement about the roles of solvency versus liquidity in the Mexican and Asian crises, the weight of the broader econometric evidence suggests an important role for liquidity indicators such as the ratio of M2 to reserves and the maturity composition of debt in explaining the incidence of currency crises more generally (see, e.g., Sachs et al. 1996). The weight of the evidence thus creates a strong presumption for the view that liquidity matters, and if liquidity matters, so does the maturity composition of external debt. As we shall see next, this particular lesson has important implications for a variety of policies, including monetary policy.

3. Policy Lessons

We now turn to lessons from recent crises for domestic macroeconomic policies. We will consider in turn capital account restrictions, exchange rate policies, monetary policy, and fiscal policy.

a. Capital Controls

As discussed in Chapter 22, capital controls have always been controversial among economists because they reflect direct interference with market allocations and may thus prevent economies from reaping the gains that would otherwise be expected from intertemporal trade. However, economists have also long recognized that there may exist valid second-best reasons for the imposition (or retention) of capital account restrictions. In Chapter 22, for example, we reviewed the sequencing literature that advocated the retention of capital account restrictions until several other domestic problems had been addressed. On the other hand, there has always been a substantial amount of skepticism among economists that capital account restrictions could be effective, in the sense that private agents may find a way to circumvent them. The lessons from the recent crises have concerned both the motivations for adopting capital account restrictions and their likely effectiveness. Concerning the first, the recent crises have strengthened the case for capital account restrictions for at least five reasons:

1. as protection against instability in the international financial environment
2. as a second-best measure to minimize the costs of financial-sector distortions
3. as a means of minimizing the risks of speculative attack by increasing the costs of speculating against the currency
4. as a way of attempting to preserve some monetary autonomy
5. as a way of attempting to lengthen the maturity of the country's external liabilities

We have already noted that two lessons from the crises of the 1990s are that the external financial environment is likely to be less stable than we may previously have thought and that capital account liberalization may greatly increase the costs associated with problems in the domestic financial sector. We have also noted that recent experience suggests that currency crises can be caused by a wide variety of factors and that avoiding vulnerability to such crises is not a simple matter of following the traditional rules of good behavior. We will also suggest that monetary policy may have very little room to maneuver under fixed exchange rates when capital mobility is high, especially when the domestic financial sector is weak. Finally, we have noted that the maturity of a country's external liabilities has come to be seen as an important factor in determining its vulnerability to currency crises. All these reasons have tended to strengthen arguments for capital account restrictions.

However, we should be clear about what this does and does not mean. These factors tend to strengthen the case for restrictions, relative to what may have been thought before experiencing the crises of the 1990s. They do not clinch the case that countries should necessarily intervene in their capital accounts. They suggest that the benefits of intervening may be higher than we may have thought, but there remain costs (i.e., benefits of liberalizing), and the issue of effectiveness remains important. Regarding the latter, the lessons from recent crises can be interpreted as follows:

> Though the evidence that capital account restrictions can affect the total quantity of capital flows remains weak, there is substantial evidence that they can affect the composition of inflows.

It may make a difference for the effectiveness of controls whether a country has previously liberalized its capital account, whether controls are intended to apply to inflows or outflows, and whether they are applied in normal times or in the midst of a crisis. Several developing countries that had previously liberalized their capital accounts imposed controls on capital inflows before the Mexican crisis. As we saw in Chapter 22, these episodes, particularly the one in Chile, have received a substantial amount of attention. The evidence we reviewed there indicated that there is weak support for the view that these controls have succeeded in affecting the volume of total capital inflows, but there is stronger evidence that they have affected the composition of inflows. Whether the reimposition of controls in the midst of a crisis (as in Malaysia in 1998) can effectively restrict capital outflows, however, remains problematic.

b. Exchange Rate Policy

i. Misalignment. We have emphasized previously that the combination of an open capital account and a fragile financial system is a dangerous one. It is dangerous because in the event of a speculative attack, the government is unlikely to tolerate the high interest rates necessary to defend the currency, and thus the attack is more likely to be successful. But a weak financial sector does not *in itself* mean that the exchange rate peg is unsustainable. Thus the peg would collapse only if the attack happened; that is, the currency becomes vulnerable to a *self-fulfilling* attack. But why should an attack happen at all? For an attack to happen, something has to coordinate the expectations of speculators on the likelihood that the exchange rate will indeed change.

A likely candidate is overvaluation of the domestic currency. If the currency is overvalued, then a real depreciation must happen. The only question is whether it will be brought about through a nominal devaluation or through an adjustment in relative prices. But as we saw in Chapter 18, with sticky prices, an adjustment in relative prices typically requires a domestic recession. Because this would undermine the viability of the domestic financial sector, the authorities are unlikely to tolerate this mode of adjustment when the domestic financial sector is fragile. Thus currency overvaluation is likely to be corrected through a nominal devaluation under these circumstances, and the knowledge that this is so will help to trigger speculation against the currency. The implication is the following:

> The costs of exchange rate misalignment can be very high when the capital account is open and the domestic financial sector is fragile.

This was probably the central lesson of the Southern Cone policy experience, and the most recent crises in Mexico and Thailand have served to confirm it. As we saw, in both cases, the domestic financial sector was weak as a result of inappropriate previous liberalization in the context of an open capital account, and the currency was overvalued, though for different reasons. When speculation against the domestic currency materialized, the authorities found themselves unable to mount the traditional high-interest defense, and the currency collapsed.

ii. Credibility. Exchange rate misalignment and a fragile financial system increase the costs to the authorities of maintaining a fixed exchange rate. Thus they make it more likely that a fixed exchange rate will be abandoned if attacked, and the increased likelihood that an attack will be successful makes it more likely that an attack will indeed happen. One way that the monetary authorities can try to prevent attacks is to signal their commitment to the fixed exchange rate. If this signal is credible, it makes it less likely that an attack will be successful and thus less likely that an attack will happen. The authorities can signal commitment by willingly imposing costs on themselves for changing the exchange rate. Unfortunately, one lesson of the recent crises is that successful precommitment mechanisms are hard to come by. Two such precommitment mechanisms, the adoption of currency boards and the issuance of foreign currency–denominated debt, have been tried recently, and neither one managed to insulate the countries that tried them from speculative attacks. One lesson is the following:

> Exchange rate credibility cannot automatically be achieved through the adoption of a currency board.

As we have seen previously, Hong Kong implemented a currency board in 1984, and Argentina did so in 1991, the latter as part of an exchange rate–based stabilization plan. Because the adoption of a currency board precludes a country from orienting monetary policy to domestic objectives, it sends a strong signal that the authorities are prepared to sacrifice domestic for external objectives. Moreover, because the currency board arrangement, including the prevailing exchange rate, is typically enshrined in the constitution of the country, it is costly to change. Nonetheless, it is not *impossible* to change, and recent experience suggests that this seems to make all the difference.

Despite its currency board, the Argentine peso came under heavy speculative attack in the first quarter of 1995, in the wake of the Mexican crisis. This loss of credibility of the Argentine currency board shows up in the form of large increases in the peso-dollar interest rate differential in the first quarter of 1995. The collapse of the monetary base almost destroyed the Argentine banking system, which lost nearly one-third of its deposits during this time. As a result, domestic interest rates increased sharply, and the economy went into a severe recession, sending the unemployment rate to 18 percent. Another attack, driven by fiscal problems in Argentina, finally brought the currency board down at the end of 2001. Similarly, Hong Kong failed to escape the Asian "flu," having been hit by a very strong speculative attack in October 1997. As in the case of Argentina, domestic interest rates rose, asset values fell, and the real economy fell into recession.

It is worth noting that in both cases, there were forces at work that would have tended to increase the credibility of the currency board – specifically, the popularity of the inflation stabilization that Argentina achieved through the currency board arrangement and the large foreign exchange reserves and sound banking system in Hong Kong. These advantages were clearly not sufficient to convince speculators that these countries would necessarily adhere to their currency board arrangements.[3] It is notable, moreover, that the 1995 speculative attack on the Argentine peso and the 1997 attack on the Hong Kong dollar both failed, implying that in those instances the authorities' true resolve was underestimated by speculators. Thus, in these cases, the currency board arrangement proved inadequate as a means for communicating the authorities' true commitment.

A less drastic commitment device is for the government to issue indexed debt. This is a poison-pill strategy for achieving credibility in which the authorities would actually impose a large fiscal cost on themselves by devaluing. This helps address the time-consistency problem that may underlie expectations of devaluation because it reduces the net benefits that the government can reap from devaluation. As we saw in the preceding chapter, this was the strategy pursued by Mexico in response to reserve outflows and expectations of devaluation triggered by the assassination of

[3] To be fair, there were also disadvantages that may have made both currency boards appear fragile, specifically, the weaknesses of Argentina's banking system and the fairly recent (summer 1997) change in Hong Kong's political status.

presidential candidate Luis Donaldo Colosio in March 1994.[4] Mexico's subsequent currency crisis suggests the following lesson:

> Exchange rate credibility cannot be bought by converting domestic-currency government debt into foreign-currency debt.

The problem with this strategy is that, though issuing indexed debt may reduce the net benefits of devaluation to the government, this may not be enough to tip the scales in favor of retaining the exchange rate peg, either in actuality or in the eyes of speculators. Moreover, if the strategy does not succeed in sustaining the peg, then the costs of devaluation could be greatly magnified. It is at least arguable, for example, that the costs of the Mexican crisis arose largely because of the liquidity crisis (the refusal of creditors to roll over maturing TESOBONOS) that followed the December 20 devaluation and that this crisis in turn was brought on by the fiscal implications of the devaluation in the presence of a large stock of indexed debt.

c. Monetary Policy

The most obvious lesson from the recent crises for monetary policy is a familiar one:

> Maintaining an officially determined exchange rate, without following the gold standard rules of the game (i.e., sterilizing), has always been and continues to be a dangerous strategy.

This is indeed an old lesson, learned under the Bretton-Woods system, which was brought down by the refusal of the United States to devote its monetary policy to the maintenance of the dollar gold parity and by that of the European countries to accept the monetary growth rates that would have been implied by the gold standard rules of the game, given the expansionary monetary policy in the United States. It has only been confirmed by recent events:

1. In 1992, the ERM crisis was triggered by the conflict between, on one hand, the high interest rates required in Italy and the United Kingdom to maintain their parity to the Deutschmark and, on the other hand, the low interest rates required to combat recession in both countries.
2. The December 1994 Mexican crisis was the product of the decision to resist the domestic deflationary effects of an increase in foreign interest rates and in the country risk premium by sterilizing reserve losses and borrowing short term in foreign exchange. The result was to undermine the exchange rate and, as already indicated, to weaken the government's fiscal position when the inevitable devaluation eventually came.

[4] Recall that the share of indexed TESOBONOS in total public debt increased from less than 5% at the beginning of 1994 to over 55% by the end of the year, while the share of privately held (nonindexed) CETES fell from 60% of total public debt in February to 20% by November 1994.

3. The Asian countries did the same thing under the opposite circumstances in 1995–1996. They sterilized capital inflows, increasing their burden of short-term external debt. Thailand also sterilized when capital began to flow out in 1996–1997.

In all these cases, countries were unwilling to give up domestic objectives for external ones, and the result was the sudden collapse of the currency.

Because it will always be impossible for monetary policy, which is only one instrument, to achieve two objectives, this implies the need to abandon an objective or complement monetary policy with another instrument. We will come back to this later.

The next lesson for monetary policy has to do with the effectiveness of the gold standard rules of the game in achieving the objective of protecting the currency:

> It may generally be possible to create a crisis with excessively expansionary monetary policy, but under certain circumstances, it may not be possible to prevent one with restrictive monetary policy.

The circumstance in question is one in which the financial sector is extremely fragile. When the domestic financial system is weak, restrictive monetary policy may be unable to save the currency. The reason is that the rules of the game call for the central bank to adopt a restrictive monetary policy in response to a loss of reserves triggered by a capital outflow. When this happens, the government has two choices:

1. If it is unwilling to sacrifice the domestic financial system by allowing domestic interest rates to rise, then it will be unable to defend the currency. That just means that it is unwilling to play by the rules of the game because the costs to the domestic economy of doing so when the rules call for a monetary contraction are very steep.
2. But even if it is willing to let domestic interest rates rise, that may not save the currency. High domestic interest rates may cause the domestic financial system to fail. Depositors may switch into foreign currency in the expectation that the government may have to inflate to pay the costs of a bailout.

In other words, the attempt to protect the currency may bring about a financial crisis that leads to a balance of payments crisis, which in turn may cause the fixed exchange rate to be abandoned anyway. Under these circumstances, the currency peg is doomed.

In the previous case, the central bank may be unable to achieve its external objective, even if it is willing to abandon any concern it may have for domestic objectives to do so. Though this is bad enough, a third lesson from recent crises is that things can actually get worse: there is an alternative set of circumstances under which the central bank may be able to achieve neither domestic nor external objectives:

> When currency mismatches are important and the financial sector is fragile, monetary policy may be left with no desirable options.

This appears to be the central lesson of the Asian financial crises for monetary policy. When the financial sector is weak and currency mismatches are absent, the central bank may not be able to safeguard the currency, but it can at least protect the domestic economy by allowing the exchange rate to float. A depreciation of the currency in this case would imperil the financial system much less than would an increase in domestic interest rates, so abandoning the currency peg and allowing a depreciation would at least prevent a domestic economic contraction. But when currency mismatches are substantial, either in the balance sheets of their banks or in those of their customers, allowing the currency to depreciate would undermine the net worth of the banks, either directly or indirectly through that of their customers, and thus bring on a financial crisis and economic contraction.

In this case, monetary policy is simply left with no desirable options. There are, in fact, three ways to respond to an attack on the currency under these circumstances:

1. The central bank can attempt to defend the currency with tight monetary policy. This policy would reflect the view that the currency mismatch makes currency depreciation a greater danger to the solvency of banks and their customers than high domestic interest rates. This was the controversial policy that the International Monetary Fund initially pursued in the Asian countries, until currency markets appeared to stabilize. As mentioned earlier, it is a dangerous policy if the credit channel for monetary policy is strong (so many domestic firms do not have access to borrowing outside the domestic banking system) and banks are weak because a credit crunch may cause both firms and banks to fail, thereby driving capital out of the country and perhaps even failing to sustain the currency.
2. It can abandon the currency peg and expand the money supply. This policy would be indicated to safeguard the domestic economy if currency mismatches do not pose a danger to domestic firms and banks. It is the policy that was advocated for Brazil in the midst of its 1999 crisis by Professor Jeffrey Sachs of Harvard, and it does appear that floating the currency in Brazil did not have the disastrous effects that it had in the Asian countries or in Mexico (where the currency mismatch was in the public sector's balance sheet).
3. It can expand the money supply but impose controls on capital outflows as a temporary expedient to permit the central bank to continue to defend the value of the currency. This is an option that some observers have argued needs to be taken seriously, and it is the one controversially adopted in Malaysia in September 1998. The obvious doubts about it concern the possible distortions

created, the potential loss of credibility, and the question of whether the controls can be made to stick. The jury is still out on the Malaysian experience.

The three previous lessons have one common theme: under high capital mobility, monetary policy is a single policy instrument that at best can achieve either one of two (domestic and external) objectives; that in the intermediate case can achieve one objective; and that in the worst of all cases can achieve neither objective. This has a clear policy message, important enough to enshrine as a separate lesson:

> Financial-sector difficulties can severely constrain what monetary policy can achieve. Thus policies to ensure a sound financial sector (i.e., one with limited vulnerability to high domestic interest rates and real exchange rate depreciation) are important not only because they help to promote efficient intermediation and to avoid financial crises but also because they increase the effectiveness of monetary policy.

But even under the best of circumstances – that is, a healthy and well-functioning financial system – the impossible trinity of open-economy macroeconomics suggests that monetary policy will find itself having to choose among competing external (the exchange rate) and domestic (economic activity) objectives. One instrument cannot be used to achieve two objectives, so this means abandoning an objective or finding additional instruments. One way to abandon an objective is to make the need to achieve it less pressing. The constraint imposed by the external objective can be softened somewhat by holding large stocks of net international reserves or by maintaining large international credit lines. But how large do reserve stocks have to be? Another lesson from recent crises is the following:

> With high capital mobility, international reserve adequacy should not be measured with respect to traditional flow variables (e.g., months of import coverage) but rather by comparing M2 plus short-term government debt to reserves.

The reason is that the sustainability of the exchange rate depends on a comparison between the central bank's stock of liquid foreign exchange–denominated assets and its liquid liabilities. The latter consist not just of the monetary base but also of assets that can readily be converted into base, which in turn includes all the liabilities of financial institutions explicitly or implicitly insured by the government. When capital mobility is high, all these claims can be presented to the central bank in short order (rather than gradually through the purchase of imports). Thus reserve adequacy must be assessed by a comparison between the outstanding stocks of such claims and the means to satisfy them.

Holding large stocks of reserves can prolong the time during which an economy can pursue a domestic objective without having to give up an exchange rate peg. This allows more time for automatic adjustment mechanisms to take effect (e.g., for transitory shocks to play themselves out or for domestic wage-price adjustments to take place) or for other adjustment policies to be deployed. Thus the existence of large stocks of reserves makes it less likely that the exchange rate will eventually succumb to a speculative attack, which in turn makes it less likely that the currency will be attacked.

An alternative to easing the constraints imposed by an objective is to deploy additional instruments. Economic theory tells us that under perfect capital mobility, monetary and exchange rate policy are not independent instruments because either monetary policy must be at the service of the exchange rate (i.e., there is no monetary autonomy under a fixed exchange rate) or the exchange rate must adjust automatically to be consistent with the prevailing monetary policy, making the exchange rate a function of monetary policy (and thus preventing the two instruments from being deployed independently). Thus we have the following lesson:

> The use of the exchange rate and monetary policy as independent instruments (in the short run) requires the presence of capital account restrictions. As natural barriers to capital movements decline, monetary autonomy tends to decrease.

One interpretation of the recent crises, as mentioned earlier, is that central banks in the emerging economies that experienced crises (specifically, those in Mexico and Thailand) ignored this lesson and continued to pursue domestic objectives with monetary policy that were inconsistent with the fixed exchange rate. One way to explain their doing so is that they overestimated the degree of monetary autonomy they retained given the degree of integration their economies had achieved with world capital markets. A policy implication of this lesson is that to continue to exercise a degree of monetary autonomy comparable to what these economies had enjoyed in the past, it would have been necessary to replace the natural barriers that were weakened by technological and institutional developments with policy-induced ones, though the feasibility of this option remains problematic, as indicated previously.

It is worth noting that barriers to capital movements, if effective, by definition increase the costs of moving funds in and out of individual countries and as such make these countries less vulnerable to speculative attack. This suggests a final lesson for monetary policy, which combines elements of the two previous lessons:

> A combination of large reserve stocks and barriers on capital movements appears to have been effective in resisting financial contagion, even when financial-sector fragility and possible exchange rate overvaluation would have suggested vulnerability to speculative attacks.

This appears to have been a key lesson from the experience of China, which survived the Asian financial crisis in much better condition than many of its neighbors.

d. Fiscal Policy

Another way to match instruments to targets would be to use fiscal policy in a countercyclical fashion. Observers such as Guillermo Calvo have pointed out that countercyclical (i.e., tight) fiscal policy would have had desirable effects during the capital-inflow period because it would have reduced pressures for macroeconomic overheating and real exchange rate appreciation, while permitting domestic interest rates to fall and building up the net worth of the public sector as a cushion to facilitate a less restrictive fiscal stance during future downturns. A key lesson from both the capital-inflow period and the recent crises, however, has been the following:

> Except for the effects of automatic stabilizers, countercyclical fiscal policies have not been easy to implement in capital-importing developing countries. During the good times of the capital-inflow period, fiscal tightening appears to be politically difficult, whereas during crisis times, tight fiscal policies are often advocated to facilitate the adjustment of the current account and promote capital inflows through market confidence.

The phrase "are often advocated" is used advisedly in the preceding. Though the desirability of fiscal tightening during the inflow period has not been controversial (though its feasibility has been questioned), the appropriate role of fiscal policy during crises has been. Particularly when the cause of the crisis is not actual or prospective fiscal insolvency, it is hard to defend discretionary fiscal tightening, especially in the midst of a severe crisis-induced recession. Two possibilities exist. The first arises when the fiscal implications of banking system restructuring imperil the solvency of the public sector. In that case, a fiscal adjustment (in present-value terms) will be required to avoid triggering wholesale capital flight caused by prospective taxation or confiscation of assets. The second is when currency mismatches threaten the viability of firms and banks. In that case, fiscal tightening may be indicated to minimize the overshooting of the real exchange rate required to adjust balance of payments flows.

The second lesson regarding fiscal policy is less dramatic:

> High-interest, long-term debt may ultimately be cheaper than low-interest, short-term debt.

This issue arose for the public sector in the case of Mexico and with reference to private debt in the case of the Asian countries. The problem, of course, is that short-maturity debt increases the liquidity of the liability side of the balance sheet, increasing the vulnerability of the public sector – or of the economy – to liquidity crises.

II. LESSONS THAT WE HAVE NOT LEARNED

As mentioned, the lessons listed previously would probably command reasonably wide, though certainly not universal, agreement. Other lessons have also been drawn from the recent crises, which, if valid, would have important implications for domestic macroeconomic management. These lessons, however, are much more controversial than the previous set. Because there are many controversies surrounding the recent crises, there is a large number of such lessons from which to draw. The following listing is selective, emphasizing only a few among the lessons that we have *not* learned that appear to be the most important.

1. Structural Reforms

An important component of the IMF-supported adjustment programs in response to crises in Indonesia, Korea, and Thailand was a broad package of structural reforms, intended to address problems not just in the financial sector but also in the areas of corporate governance and in the real sector more broadly. These real reforms have proven to be quite controversial. The question is whether they should have been pursued in the midst of the crises. Thus we can express our first controversial lesson as follows:

> Crises present an opportunity to undertake needed (i.e., productive) structural changes.

The argument in favor of this proposition is that crises may tend to disrupt political equilibria that sustain harmful structural policies in place. Thus the occasion of a crisis may provide an opportunity to implement needed and long-lasting changes that would have been politically impossible to carry out during normal times. The controversy around this "lesson" concerns the fact that even if they are needed and appropriate, structural reforms are usually politically sensitive and difficult to implement. Conditioning balance of payments assistance (i.e., liquidity support) on the implementation of such reforms therefore creates uncertainty about whether that assistance will be forthcoming to crisis countries. The result may be to undermine confidence in the resources underpinning the macroeconomic adjustment program designed to extract the country from the crisis, thereby weakening the program and possibly aggravating the crisis.

2. Exchange Rate Regimes

As we discussed in Chapters 18 and 19, many observers have drawn sweeping conclusions from the crises of the 1990s about what types of exchange rate regimes remain feasible for emerging economies that maintain open capital accounts. Specifically, they have concluded the following:

> Given high capital mobility, developing countries must choose between extreme exchange rate regimes: either floating exchange rates or dollarization/currency boards.

Many economists have taken this view. As we saw in Chapter 18, it depends on the observation that in a situation of high capital mobility, any fixed exchange rate is subject to speculative attack, and the suspicion that in such circumstances, most countries will not find it desirable to incur the costs associated with a sustained high–interest rate defense. Because we have already examined this argument in that chapter, there is no need to explore it further here.

3. Sources of Currency Mismatches

As we saw in Chapter 27, currency mismatches in the balance sheets of domestic firms and financial institutions can trigger speculative attack. We also discussed earlier in this chapter how currency mismatches can complicate monetary management during crises. The question is why such currency mismatches arise. One "lesson" that has been drawn about the source of such mismatches from recent experience is the following:

> Officially determined exchange rates encourage currency mismatches in the balance sheets of domestic agents.

The argument in support of this view is that when the exchange rate is fixed, domestic agents tend to discount exchange rate risk. Consequently, when domestic interest rates exceed foreign ones, domestic agents tend to borrow in foreign currency, even if their assets are denominated in domestic currency, thus incurring exchange rate risk.

However, there are several problems with this story. First, higher interest rates on domestic-currency loans than on foreign-currency loans presumably reflect a perception of exchange rate risk among domestic lenders. It is unclear why there should be an asymmetry in perceived exchange rate risk between domestic lenders and borrowers. Second, if a fixed exchange rate really lowers the perception of exchange rate risk among all market participants, it is not clear why loans extended by foreign lenders to domestic borrowers would tend to be denominated systematically in foreign exchange. Presumably, a credibly fixed exchange rate would lower exchange rate risk for both parties in the transaction, leaving indeterminate the currency of denomination of the loan. Finally, it is not at all clear that the mere announcement of a fixed exchange rate indeed reduces perceived exchange rate risk. As we saw earlier in this chapter, even countries that have buttressed their fixed exchange rates with the institutional apparatus associated with hard pegs (e.g., currency boards) have not been immune from the perception of currency risk. A more credible explanation for the emergence of currency mismatches probably has to rely on a moral hazard argument; that is, domestic borrowers are willing to assume exchange rate risk because they believe that in the event of a large depreciation of the domestic currency, they will receive financial support from the government. But the factors giving rise to currency mismatches remain to be clearly identified.

4. Monetary and Fiscal Responses to Crises

A final controversial "lesson" concerns the appropriate monetary and fiscal response to the types of crises that some emerging economies experienced in the 1990s – specifically, to crises featuring potential financial-sector insolvency associated with currency mismatches in balance sheets.

> Responses to crises should feature tighter monetary and fiscal policies. The argument for fiscal tightness can take two forms:

- If the financial sector is prospectively insolvent, the public sector will have large contingent liabilities associated with the cost of recapitalizing domestic financial institutions. Because such liabilities may threaten the solvency of the public sector itself, tighter fiscal policies (in the form of a larger adjusted primary surplus, as in Chapter 9) are required to safeguard the solvency of the public sector.
- Even if the public sector's solvency is not prospectively at risk, tighter fiscal policies will facilitate the adjustment in the current account of the balance of payments required by the drying up of capital inflows, thus reducing the required adjustment in the exchange rate (as in the liquidity-crisis model of Chapter 27).

The counterargument is that tightening fiscal policy in the context of a crisis, when the government's solvency is not at risk, is to cause fiscal policy to behave pro-cyclically, thereby aggravating the severity of the recession associated with the crisis. The jury is still out on this issue. It is easy to see that the outcome should depend on a host of characteristics that may vary from one country to another (e.g., on the composition of the fiscal contraction, the impact of fiscal policy on aggregate demand, the sensitivity of the current account to the exchange rate, or the extent of currency mismatches in domestic balance sheets).

With respect to monetary policy, the argument for tight money is that it is required to defend the exchange rate and thus minimize the damage to balance sheets from currency mismatches. The counterargument is that tight money will be ineffective in protecting the currency because the damage that high domestic interest rates can do to the real economy under circumstances of financial fragility may actually repel capital flows rather than attract them. This remains an unsettled issue and is the subject of current research.

III. SUMMARY AND CONCLUSIONS

On the basis of the lessons that we have reviewed in this chapter, what can we say about the likely components of desirable domestic macroeconomic management in emerging economies in the context of an increasingly integrated world? A successful strategy would seem to entail the following:

1. First and foremost, the crises of recent years in emerging and developing countries (and the Southern Cone crises of the early 1980s) suggest strongly that the financial system can be the Achilles' heel of domestic macroeconomic management. Thus structural policies to maintain a healthy financial system without severe currency mismatches and with substantial net worth, minimizing its vulnerability to exchange rate depreciation and high interest rates, should be high on the policy agenda. This is good for its own sake but also helps to preserve the flexibility of monetary policy.
2. Structural policies that would tend to promote wage-price flexibility would be a useful complement. But what such policies are, and how successful they can be, are difficult questions to answer. We do not know enough about the causes of nominal wage-price rigidities at present to say much about these issues.
3. For countries that choose to maintain an officially determined exchange rate, the maintenance of a large stock of net international reserves – or an equivalent arrangement to provide liquid foreign-currency assets in a pinch – would seem to be desirable. This reduces the urgency of having to continuously achieve external balance objectives. Side benefits are likely to be to reduce currency risk premia and the incidence of speculative attacks. Many emerging and developing countries indeed undertook the accumulation of very large reserve stocks in the wake of the Asian crisis.
4. Even for such countries, it would seem desirable for exchange rate policy to allow some flexibility within a band, with a central parity that tracks the equilibrium real exchange rate. The difficulty, of course, is whether we can know what this equilibrium rate is. Specifically, can we know when fundamentals change permanently and how large their empirical effect on the equilibrium real exchange rate should be? Chapter 16 suggested that this remains at present more of an art than a science.[5]
5. Fiscal flexibility is a highly desirable goal in a financially integrated world, to provide an additional stabilization instrument to achieve internal balance. But the amount of fiscal flexibility that countries can realistically achieve should probably not be overestimated.

[5] The importance of this issue is indicated by the fact that the IMF has recently undertaken to estimate equilibrium real exchange rates as part of its regular consultations with many of its member countries, using several of the methodologies described in Chapter 16 (see Lee et al. 2008).

Finally, to end this chapter on a positive note, it is worth noting that the lessons reviewed in this chapter can indeed be learned. The case of Chile provides an example. The crises in Mexico and Thailand during the 1990s had as a precursor the Chilean crisis of 1982, and though Mexico and Thailand do not seem to have learned the lessons of the earlier Chilean crisis, Chile itself seems to have done so.

The management of capital inflows in Chile during the 1990s provides an interesting contrast with that of Mexico and Thailand. Like Mexico and Thailand, Chile entered that inflow episode in a solid fiscal position. Because of lessons learned from its earlier financial crisis, moreover, Chile also did so with a liberalized but secure and well-regulated financial system. Its exchange rate policy was explicitly geared to the maintenance of a competitive real exchange rate, and the authorities managed the exchange rate flexibly to avoid misalignment, including revaluing the rate when necessary, a measure that signaled their intention to track the equilibrium rate.[6] These policies together precluded the emergence of the exchange market pressures that confronted Mexico and Thailand.

Perhaps more important, Chile also took measures to protect itself from vulnerability to a liquidity crisis, a situation which, as the model in Chapter 27 indicated, could have emerged independently of a prior exchange rate crisis and could itself have triggered such a crisis (of the self-fulfilling variety). Chile did so by explicitly targeting both its current account deficit through the use of exchange rate and aggregate demand policies as well as the composition of its external debt through policies such as exchange rate variability within a wide band and the imposition of capital account restrictions designed to affect the structure of external debt.

No matter how appropriate its policies, no small country that is as highly integrated with international goods and financial markets as Chile has become over the last two decades can insulate itself from international economic upheavals, and Chile has not been an exception to this rule. Overall, however, the country has weathered the volatile international financial environment that has characterized this period remarkably well. The crisis lessons reviewed in this chapter suggest that Chile's successful performance has not been accidental.

REFERENCES AND FURTHER READING

Calvo, Guillermo (1996), "Capital Flows and Macroeconomic Management: Tequila Lessons," unpublished manuscript, International Monetary Fund.

Calvo, Guillermo, and Enrique Mendoza (1996), "Petty Crime and Cruel Punishment: Lessons from the Mexican Debacle," *American Economic Review Papers and Proceedings*, vol. 86, pp. 170–175.

[6] In addition, not only was Chile among those countries (as Thailand had been earlier) that explicitly attempted to use fiscal policy in a discretionary fashion as a tool to combat the overheating pressures emanating from the inflows of the early 1990s, but as described in Chapter 11, it also appears to have found a way to implement flexible countercyclical fiscal management through its Structural Surplus Rule.

Calvo, Guillermo, Leonardo Leiderman, and Carmen Reinhart (1993), "Capital Inflows to Latin America: The Role of External Factors," *IMF Staff Papers*, vol. 40, pp. 108–151.

Diaz-Alejandro, Carlos (1985), "Good-Bye Financial Repression, Hello Financial Crash," *Journal of Development Economics*, vol. 19, pp. 1–24.

Dooley, Michael P. (1997), "A Model of Crises In Emerging Markets," National Bureau of Economic Research *Working Paper* # 6300 (December).

Kaminsky, Graciela, and Carmen Reinhart (1999), "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems," *American Economic Review*, vol. 89, pp. 473–500.

Krugman, Paul (1979), "A Model of Balance of Payments Crises," *Journal of Money, Credit, and Banking*, vol. 11, pp. 311–325.

Lee, Jaewoo, Gian Maria Milesi-Ferretti, Jonathan Ostry, Alessandro Pratti, and Lucca Ricci (2008), "Exchange Rate Assessments: CGER Methodologies," occasional paper 261, International Monetary Fund.

Radelet, Steven, and Jeffrey Sachs (1998), "The Onset of the East Asian Financial Crisis," working paper 6680, National Bureau of Economic Research.

Sachs, Jeffrey, Aaron Tornell, and Andres Velasco (1996), "The Collapse of the Mexican Peso: What Have We Learned?" *Economic Policy*, vol. 22, pp. 13–64.

Velasco, Andres (1987), "Financial Crises and Balance of Payments Crises," *Journal of Development Economics*, vol. 27, pp. 263–283.

29

Lessons from the Great Recession

As indicated in the preceding chapter, the decade of the 1990s was punctuated by a series of financial crises in emerging-market economies. We saw in Chapter 3 that, partly as a result of these crises, significant structural and macroeconomic reforms were undertaken in a large number of emerging and developing economies during the decade of the 1990s. Many of these reforms involved the key areas that we have explored in this book: the institutional underpinnings for the formulation of fiscal policy, the governance of the central bank, the monetary and exchange rate regimes implemented by the bank, policies toward the domestic financial sector, and policies toward the capital account of the balance of payments. These reforms continued to be implemented in many emerging and developing economies during the decade of the 2000s.

After the Argentine crisis of 2001–2002, the world economy entered a five-year period of rapid growth, with low inflation and substantial macroeconomic stability, that proved to be a prosperous time for most emerging and developing countries. However, this time came to an end after 2007 with the outbreak of the Great Recession, a significant slowdown in economic activity among the advanced economies that became especially acute after the failure of the investment bank Lehman Brothers in the United States in September 2008. That event was followed by an international financial crisis more severe than any the world had seen since the Great Depression. What distinguished this crisis from the ones that we have reviewed in the preceding four chapters is that, unlike the crises we studied there, this one originated in the industrial countries – specifically, in the United States – rather than in emerging-market economies. However, the crisis spread internationally very quickly. From the perspective of the emerging and developing economies, therefore, the Great Recession represented an enormous external shock.

This shock was a severe test for the reformed macroeconomic policy environment that many countries had implemented over the previous two decades. This chapter attempts to discern lessons about how the reforms that these countries undertook in the 1990s and before the outbreak of the crisis affected their ability to cope with this external shock. As we will see, the news is relatively good. The world economy's emergence from the crisis has in fact been strongly assisted by renewed growth in emerging economies, even though the crisis initially affected them very severely. This growth resumption in turn was partly made possible by the ability of these countries, in the context of their improved institutional and macroeconomic policy environments, not just to avoid having to implement pro-cyclical macroeconomic policies but actually to be able to adopt strongly countercyclical fiscal and monetary policies.

The structure of the chapter is as follows: in the next section, we examine macroeconomic policies and performance in five emerging- and developing-country regions during the decade of the 2000s, up to the time of the crisis. This regional perspective is important because though macroeconomic reforms and policies have shared many common features across emerging and developing countries over the past decade, there have been important regional differences, and these differences serve to give us a cross-sectional perspective on the extent to which the reforms to the macroeconomic policy environment that emerging economies implemented in the precrisis period have affected their responses to the crisis. The outbreak of the crisis in the United States and its transmission to other industrial countries is described in Section II. This is followed by an examination of the somewhat different channels of crisis transmission to the five emerging- and developing-country regions in Section III. Section IV considers the state of vulnerability in these countries at the time the crisis broke out. Section V describes how emerging and developing economies responded to the crisis, and how this response differed from the response to previous crises, whereas Section VI examines what the macroeconomic impacts of the crisis have been in those economies as of the time of writing in summer 2010. Finally, the policy lessons from this experience are summarized in Section VII.

I. BACKGROUND: POLICIES AND OUTCOMES IN EMERGING ECONOMIES DURING THE DECADE OF THE 2000S

Both the reforms that were undertaken in the years before the outbreak of the Great Recession and the transmission of the crisis tended to differ across regions of the world, in part because different emerging- and developing-country regions had very different previous growth and crisis experiences and in part because there are significant regional differences in economic links between emerging and developing economies, on one hand, and advanced economies, on the other. Accordingly, we will consider the lessons of the Great Recession from a regional perspective, following the geographic classification of countries used by the International

Monetary Fund.[1] This section begins by considering the experience of emerging and developing countries in each region during the decade or so preceding the outbreak of the Great Recession.

1. Asia and the Pacific

Prior to the 1997–1998 Asian crisis (see Chapter 27), emerging economies in Asia had pursued an outward-oriented development strategy based on manufactured exports to industrial economies. The Asian financial crisis turned out to be a defining moment for those economies, and in many ways, the situation in which these countries found themselves at the outbreak of the Great Recession was the outcome of the policy measures that they had implemented in response to the experience of their own earlier crises.

These countries did not close themselves off from the international economy after the Asian crisis. Instead, their real and financial integration with the rest of the world actually *increased* during the subsequent decade. During the decade of the 2000s, the emerging economies in this region remained highly open to trade, and they continued to be heavily dependent on exports to the Organization for Economic Co-operation and Development (OECD) countries, primarily of manufactured goods, and especially of electronics, which accounted for about one-third of the region's exports during this time. Intraregional trade within Asia also increased strongly over the decade of the 2000s, but for much of the decade, it consisted primarily of exports of capital-intensive intermediate goods from other countries in the region to China, where they were assembled for export to OECD countries. This intraindustry trade existed to exploit China's comparative advantage in labor-intensive activities, and China essentially became an export platform from East and Southeast Asia to the rest of the world. As in the precrisis period, countries in the region continued to maintain very high saving rates, but the crisis had the effect of lowering investment rates in the region, with the implication that the region as a whole ran a persistent current account surplus with the rest of the world during the 2000s.

As a result of the capital account liberalization that they undertook after 2000, Asian emerging economies also became more financially integrated with the rest of the world. For example, firms in the region significantly increased their issuance of foreign exchange–denominated bonds in international financial markets after 2003, and private external debt almost doubled from 1997 to 2005 (International Monetary Fund (IMF) 2006a). U.S. holdings of Asian securities increased from

[1] Unless otherwise noted, the empirical information in this chapter is drawn from various issues of the IMF's *Regional Economic Outlook* reports, which are available on a semiannual basis after 2005 on the IMF's Web site (http://www.imf.org/).

4.6 percent of the gross domestic product (GDP) of those economies in 1994 to 13.1 percent in mid-2006, whereas emerging Asian holdings of U.S. securities increased from 10 percent of the Asian countries' GDP to 28 percent during the same period. The correlations in stock market returns between Asian emerging economies and the United States increased sharply from 1990–1996 to 2000–2007. According to the IMF (2006b), Asian-focused hedge funds more than quintupled in number (to approximately 1,150) from 2002 to 2006.

One consequence of this increased financial integration was that, as described in Chapter 24, Asian countries began to receive large gross private capital inflows after 2003, including portfolio flows, and the IMF estimates that gross capital inflows, even excluding the regional financial centers of Hong Kong and Singapore, were comparable in size to their mid-1990s levels, at 5–6 percent of *GDP* (IMF 2007a). Despite these large inflows, however, foreign liabilities relative to *GDP* remain smaller in Asia than in Latin America and are particularly small in China and India. As documented in Chapter 24, relative to the early-1990s capital-inflow episode, the share of foreign direct investment (FDI) in capital inflows during this more recent episode was much larger, whereas that of portfolio inflows was much smaller, and larger inflows have been accompanied by larger gross outflows, largely as the result of capital account liberalization. Consequently, *net* capital inflows in Asia were on average smaller than in the mid-1990s. Though the volatility of both gross inflows as well as gross outflows increased relative to that associated with the earlier inflow episode, that of net inflows did not appreciably change. As one would expect, large inflows were associated with lower sovereign risk premia for Asian emerging-market economies during this period. Such premia were especially low for countries such as China, Malaysia, and Thailand, and they declined after early 2004 for Indonesia and the Philippines as well.

Aside from increased real and financial openness, another legacy of the Asian crisis was a change in the way that exchange rate policy was conducted in many (but not all) Asian countries. Exchange rate regimes in the region were on the whole (with the important exceptions of China and Malaysia) substantially more flexible in the decade of the 2000s than they had been before the Asian crisis. However, these regimes remained heavily managed, as central banks engaged in heavy sterilized intervention to avoid real exchange rate appreciation.[2] As a result, by 2005, real effective exchange rates among emerging-market economies in Asia remained more depreciated than prior to the crisis. However, rapidly growing exports and capital inflows during the boom years between 2003 and 2007 made real appreciation difficult to resist. Consequently, Korea's and Indonesia's real effective exchange rates appreciated by some 30 percent from 2002 to 2007, Thailand's by more than 20 percent, and that of the Philippines by about 15 percent (IMF 2007b). Even China, which had not wavered from a fixed exchange rate policy, found itself

[2] For the evolution of exchange rate regimes in postcrisis Asia, see Hernandez and Montiel (2003).

pressured to allow its exchange rate to appreciate, and on July 21, 2005, it announced a move toward a more flexible exchange rate arrangement, revaluing the renmimbi by 2 percent and announcing that its value would henceforth be set against a basket of currencies rather than against the U.S. dollar (it had previously been fixed at RMB 8.3 per dollar). On the same day, Malaysia abandoned the fixed dollar peg that it had implemented in response to the crisis in September 1998 in favor of a tightly managed float. Thus most exchange rate regimes in the region had effectively become managed floats by 2005. Nonetheless, heavy central-bank intervention in foreign exchange markets resulted in Asian economies building up truly massive stocks of foreign exchange reserves by 2007.

A third major policy area in which the crisis had an important impact was in the structure of regional financial systems and the institutional framework for corporate governance. Financial systems in Asia have traditionally been bank dominated, with firms tending to rely for financing on their own funds and bank loans. As noted in Chapter 27, banking-sector weaknesses were an important ingredient in the 1997–1998 crisis mix, and the crisis severely undermined the solvency of many Asian banks. Consequently, substantial reforms of domestic banking systems were undertaken in several Asian countries after the crisis. Banking systems were recapitalized, privatized, and opened to foreign investment; directed lending was sharply curtailed; bank supervision was improved; and nonperforming loan (NPL) ratios were drastically reduced.

By contrast, bond markets in Asia were relatively very small at the time of the crisis. After the crisis, however, the Asian Bond Market Initiative was launched as a joint effort by countries in the region to promote the development of local and regional bond markets by improving market infrastructure and the regulatory environment. The initiative created several funds to invest in bonds issued by Asian governments. Originally, those funds were supported by the foreign exchange reserves of member countries, but they were later opened to investment from the private sector as well.

Though some emerging economies in Asia (Hong Kong, Singapore, India, Malaysia, and Taiwan, in particular) have large equity markets, and though equity issue tends to provide much larger shares of corporate finance in Asia than in other emerging-market regions, equity markets in the region are also on average smaller than those in industrial countries. However, partly because of improvements in market infrastructure and corporate governance (the legal framework for corporate governance was improved by strengthening accounting standards, limiting cross holdings, and increasing shareholder rights), these markets grew rapidly in the postcrisis period. They also increased in liquidity (turnover) and breadth (measured as the percentage of market capitalization and turnover accounted for outside of the 10 largest firms). Portfolio equity inflows to these markets grew rapidly, and foreign participation in Asian equity markets was about one-third of the total at the end of 2005. As mentioned earlier, movements in these markets became more synchronized with those in global equity markets after the early 1990s.[3] Other nonbank financial intermediaries, such as mutual funds, pension funds, and insurance companies, have also grown rapidly in the region. Finally, in 1997, many Asian firms were highly leveraged, with large unhedged foreign-currency exposure, large short-term debt, and low profitability. After the crisis, firms in emerging Asia decreased their vulnerability to financial shocks by reducing their debt-equity ratios (*deleveraging*) from peaks of over 300 percent in 1996 to about 75 percent in 2007 (IMF 2009a), strongly increasing their profitability and holding more liquid assets.

The crisis also had implications for the conduct of fiscal and monetary policies in Asia. Partly as a legacy of the costs of financial restructuring and fiscal stimulus after the Asian crisis, the ratio of public debt to *GDP* for the region increased from 26 to 36 percent from 1996 to 1999. By 2008, government debt still stood at about 33 percent of *GDP* in emerging Asia. Though higher than its precrisis value, this remains relatively low by international standards, especially for a fast-growing region. With respect to monetary policy, the move away from fixed exchange rates required the adoption of an alternative nominal anchor, and many central banks in the region moved to inflation-targeting frameworks after the crisis. As suggested by Taylor rules, policy rates have been sensitive to the behavior of inflation, increasing when inflation has accelerated and decreasing when it has slowed. An indication that monetary policy credibility has become well established in Asia is that inflation expectations increased much less than the increase in headline inflation when higher world food and energy prices drove headline inflation higher in 2007–2008.

In short, countries in Asia became even more highly integrated with the world economy during the postcrisis period, along both real and financial dimensions. Though this has increased their exposure to external shocks, a better composition of external financing, stronger domestic financial systems with improved corporate governance, better monetary policy frameworks, strong fiscal positions, more flexible exchange rate regimes, and large reserve accumulation have all reduced vulnerability and placed these economies in a more favorable position to respond to such shocks with expansionary policies.

2. Emerging Europe

Most emerging economies in Europe are formerly centrally planned economies that undertook the transition to the market during the 1990s. Whereas the Asian crisis was the defining experience for Asian emerging and developing economies, for emerging economies in central and eastern Europe, the corresponding event

[3] However, markets in China and India remain illiquid, those in Hong Kong and the Philippines remain dominated by a few large firms, and throughout the region, equities remain a small portion of household wealth as well as of financial-sector assets, so domestic economies appear to remain somewhat insulated from fluctuations in global equity prices.

was the transition away from central planning that began – rather traumatically – in the early 1990s. These economies consist of the Baltic countries (Estonia, Latvia, and Lithuania), countries in central and eastern Europe (such as the Czech Republic, Hungary, Poland, and Russia, among others), and countries in southeastern Europe, including the Balkans and Turkey (which is not a formerly centrally planned economy). All the countries in emerging Europe – including Turkey – have actively pursued real and financial integration with industrial Europe over the past two decades, and several countries in the region achieved very rapid growth during the decade of the 2000s. Growth was particularly strong in the period before the crisis in the Baltics and in Russia, a major oil exporter.

In many ways, the experiences of these countries over the past two decades resemble those of other emerging economies in *previous* decades. Outside Russia, for example, many emerging European countries had substantial fiscal deficits during the 2000s, and discretionary fiscal policy was pro-cyclical, with structural deficits either increasing or improving only very slightly during those years, which turned out to be boom years for those countries. Even in Russia, which had fiscal surpluses, those surpluses were driven by oil revenues, and the Russian nonoil fiscal deficit behaved pro-cyclically, increasing from 2004 to 2008, a time when the Russian economy was growing rapidly. Despite this, public-sector balance sheets remained relatively strong in the sense that public-sector debt-*GDP* ratios tended to be lower than those in other emerging-economy regions. Similarly, again with the exception of Russia, which benefited from high oil prices, almost all emerging European economies had large current account deficits in the period leading up to the crisis. Although FDI inflows contributed substantially to financing these deficits, many European emerging economies funded them with substantial foreign bank borrowing, including borrowing from the domestic branches of foreign (western and northern European) banks that was in turn funded by loans from their parent banks. This situation was particularly prevalent in the Baltics (where the parent banks tended to be Scandinavian) and in southeastern Europe (where many parent banks were Austrian) but less so in central and eastern Europe.

As in Asian countries, the financial sectors of emerging European economies are largely bank based. Outside the Czech Republic, Hungary, Poland, and Turkey, bond and equity markets are not well developed. Corporate bond markets are both very small and illiquid. Equity markets are illiquid and not very broad, even in the countries where those markets are most highly developed. Nonbank financial intermediaries, such as pension funds, mutual funds, and insurance companies, are just beginning to develop.

Except in Belarus, which remains financially repressed, banking sectors in the region have been liberalized. However, the institutional infrastructure for the financial sector is deficient in many countries, and in particular, the local subsidiaries of foreign banks have not always been well capitalized (IMF 2007b). As mentioned earlier, many of these banks rely heavily on nondeposit funding, especially from parent banks in western Europe. Though their domestic loans were generally denominated in foreign currency, they tended to be extended to unhedged domestic borrowers, supporting activities in the nontraded goods sector such as real estate investments. This type of financial intermediation created substantial currency mismatches for bank borrowers and therefore credit risk for the banks themselves. Banks in emerging Europe were also exposed to interest rate risk indirectly because, even though they often issued mortgages and other loans on variable rates, the credit risk on those mortgages made the banks vulnerable to interest rate increases. Because much of their debt was in foreign exchange and at variable rates, domestic firms and households in emerging Europe were exposed both to exchange rate risk and to interest rate risk in the precrisis period.

In seeking to integrate with western Europe, many of the countries in emerging Europe maintained fixed exchange rates. In countries with fixed exchange rates, especially in the Baltics, high financial integration rendered monetary policy ineffective in the run-up to the crisis, and monetary policy was overwhelmed by large precrisis capital inflows. As a result, many emerging European countries experienced very rapid domestic credit growth, and countries such as Bulgaria, Croatia, Estonia, Hungary, and Latvia all faced the crisis with high levels of private-sector debt. Moreover, external debt-*GDP* ratios were very high – significantly higher than in other emerging economies – and left those countries with a large share of loans denominated in foreign currency.

Finally, though emerging European economies received large capital inflows in the precrisis period, those inflows were just enough to finance the large current account deficits mentioned earlier. Therefore, unlike countries in Asia and Latin America (see later), emerging economies in eastern Europe entered the Great Recession with low ratios of reserves to short-term external debt. For half the countries in the region, this ratio was well below 100 percent, and for another quarter, it was close to or somewhat above 100 percent. Only in the Baltics and Belarus was this ratio well above 100 percent (IMF 2008a).

3. Latin America and the Caribbean

The extensive macroeconomic reforms that Latin American countries began to implement in the 1990s had the effect of making the region increasingly integrated with the world economy. The region's increased openness has made it more vulnerable to external shocks, both financial and real. Indeed, Osterholm and Zettelmeyer (2007) estimate that external factors account for more than half of the medium-term variance of Latin American growth, with external financial conditions accounting for about 35 percent, foreign growth for 10–15 percent, and commodity price fluctuations for 5–6 percent. Another indicator of increased financial openness is that in the past, sovereign spreads for most emerging-market economies in Latin America have tended to move with those of *high yield* (risky) U.S. corporate bonds, which are relatively volatile. But more recently, countries (such as Brazil, Colombia, Mexico, and Peru) that have made significant institutional and policy improvements have seen their spreads move more closely with those of *investment grade* (relatively safe) corporations, which tend to be much more stable (IMF 2008b). Moreover, over the course of the decade of the 2000s, firms in Latin America gained increased access to international capital markets. According to the IMF (2008b), by 2007, corporations in Latin America derived some 25 percent of their net new financing from foreign sources. In addition, foreign banks account for an increasing share of the region's deposits. In principle, pressure on the capital bases or the liquidity of their home offices could cause them to curtail lending in the region, increasing the strength of financial transmission channels.

As in other regions of emerging and developing countries, the period from 2003 to 2007 was a prosperous one for Latin America. A very favorable external environment was an important contributing factor. High commodity prices after 2002, substantial capital inflows dominated by FDI, and sharply increased flows of workers' remittances toward some countries, made foreign resources plentiful. These flows put upward pressure on real effective exchange rates, and as in Asia, many countries in the region responded by intervening in foreign exchange markets, accumulating very large stocks of foreign exchange reserves. During this time, Latin American exports became more diversified geographically (relying less on the United States and more on Asia), spreads on sovereign bonds narrowed sharply, and the credit ratings of countries in the region improved.

The result of all this was that growth was rapid in the period through 2007 – indeed, from 2003 to 2007, the Latin American and Caribbean region registered its highest growth rate since the 1970s – and regional stock markets boomed. Growth averaged over 5 percent from 2003 to 2007, compared to an average of 3.5 percent during 1970–2000. Rapid growth was driven by booming domestic consumption and investment. Yet, in contrast with previous booms in Latin America, current accounts were largely in surplus, implying that though domestic absorption grew rapidly, it did not outstrip growth in productive capacity. As a result, external debt decreased relative to GDP, and inflation remained low.

A large part of the explanation for the behavior of absorption relative to income reflects a change in the policy environment, with both fiscal and monetary policies remaining relatively restrained during the boom of the 2000s, partly reflecting institutional reforms but partly also the result of more policy discipline within the existing institutional framework.

On the fiscal side, the primary fiscal surplus averaged over 3 percent of GDP on a GDP-weighted average basis in Latin America and the Caribbean during 2003–2007. At the time the crisis broke out, fiscal responsibility laws were in place in Argentina, Brazil, Chile, Colombia, Ecuador, Panama, Peru, and Venezuela (Corbacho and Schwartz 2007). Other countries in the region that are heavily dependent on commodity-based revenues benefited from high commodity prices.[4] Though public-sector spending increased during this period in many countries, the increase in spending proved to be restrained relative to previous boom times. Overall, the IMF (2007d) estimated that most countries in the region (with the exception of Venezuela) were running *structural*, or *cyclically adjusted*, primary surpluses in 2007 (primary surpluses estimated at full-employment levels of *GDP*). This helped reduce public-sector debt-*GDP* ratios from a regional average of 77 percent of *GDP* in 2002 to 50 percent in 2007. At the same time, governments in the region made a conscious effort to improve the structure of public debt in the sense that they replaced a large share of their foreign-currency debt with domestic-currency debt, and several countries (Colombia, Mexico, Peru, and Brazil) began to be able to issue long-maturity domestic-currency debt, a sign of confidence in long-run price stability.

On the monetary side, formal inflation targeting (IT) was implemented in Brazil, Chile, Colombia, Mexico, and Peru. Average inflation rates fell to about 3.5 percent in these countries, and the gap between targeted and actual inflation closed from 2.5 percent to about 0.5 percent (IMF 2006c). Costa Rica, Uruguay, and Paraguay also moved toward inflation targeting during this period. As would be expected under a well-implemented IT regime, central banks in the region, especially in Brazil and Mexico in 2005, and in Chile, Colombia, Peru, and Mexico in 2007, responded to the boom by raising their real policy interest rates. Policy interest rates were further increased in the first half of 2008 as the increases in international food and fuel prices of 2007 continued to affect headline inflation in the region. As in Asia, there is evidence that inflation targeting has made inflation expectations in these countries more firmly anchored in the sense that these changes in headline inflation had very small effects on expected future inflation (see IMF 2008b). Aside from this exogenous external shock late in the decade, inflation remained subdued in most countries of the region during the 2003–2007 boom.

As in Asia, exchange rate regimes also evolved during this period. Argentina and Brazil, which had pegs in 1996–1997, were operating managed floats by 2005, and Chile, Mexico, and Peru all were allowing more exchange rate variability in their managed floats by 2005 than they had in the mid-1990s. Importantly, firms in the region appear to have substantially reduced currency mismatches in their balance sheets, making them less vulnerable to currency risk.[5]

Finally, financial reform continued in the region during the boom in the precrisis period. Credit market infrastructure improved in countries such as Brazil and Mexico, including the strengthening of credit information systems and loan recovery

[4] These include Bolivia, Chile, Colombia, Ecuador, Mexico, Peru, and Venezuela.

[5] A survey of 1,200 firms in Argentina, Brazil, Chile, Colombia, Mexico, and Peru by the IMF (2008c) found that the share of foreign-currency liabilities on these firms' balance sheets had been reduced from 37% in 1998 to 17% in 2007.

frameworks. Indicators of banking system health had also been improving. The IMF (2007d) cites several indicators of such improvement, including high levels of bank profitability and capitalization, reduced overhead costs in banking, reduced ratios of nonperforming loans, and reduced foreign-currency exposure.[6] Importantly, unlike in emerging Europe, foreign banks that have affiliates in Latin America tend to fund their operations largely from domestic sources rather than by borrowing from their parent banks. Indeed, most banking systems in the region – whether domestically or foreign owned – relied overwhelmingly on domestic rather than foreign financing, and they maintained positive net foreign asset positions.[7] Financial dollarization has been decreasing on both sides of banks' balance sheets over time, a sign of increased confidence in the domestic currency. Though the banking system is dominant in financial intermediation, countries such as Chile, Colombia, Brazil, Mexico, Panama, Peru, and Trinidad and Tobago have well-established domestic stock and bond markets. As in Asia, these markets have become larger and more liquid during recent years, and domestic firms have been raising an increasing amount of capital in those markets.

4. Middle East and Central Asia

Emerging and developing economies in the Middle Eastern and central Asian region have undergone neither the traumatic crises that afflicted countries in Asia and Latin America nor the experience of transition in central and eastern Europe. For these countries, the dominant source of macroeconomic instability has been fluctuations in the price of oil. The region consists of countries that are major oil exporters and whose macroeconomic performance tends to be dominated by the price of oil (Algeria, Azerbaijan, Bahrain, Iran, Iraq, Kazakhstan, Kuwait, Libya, Oman, Qatar, Saudi Arabia, Syria, and the UAR), economies that are lesser oil exporters or are tied through trade or remittance flows to the major oil exporters (Egypt, Jordan, Lebanon, Morocco, Pakistan, and Tunisia), emerging economies that are tied to Europe through exports and remittances (Morocco and Tunisia), and a large group of low-income oil-importing countries such as Afghanistan and Pakistan.

These economies are very open and are becoming more so over time. For the region as a whole, foreign trade increased from 50 percent of GDP in 1990 to 90 percent in 2007 (IMF 2007b). Oil-exporting countries in the region tend to have roughly a quarter of their economies accounted for by the oil sector on average, and oil revenues tend to account for something like three-quarters of both exports and

[6] According to the IMF (2009e), banks in the region had median capital-asset ratios of about 15%, compared to a mandated ratio under the international Basel II guidelines of 8%.

[7] The IMF (2008a) reports that outside Panama, regional banking systems had less than 10% of their liabilities funded by nonresidents.

government revenue. Consequently, the high oil prices that were associated with the precrisis global boom were a major boon for them. The expansionary effects of the increase in oil prices were transmitted from the oil exporters to other countries in the region through intraregional remittances and FDI flows. The oil-importing countries also benefited from strong growth in remittance inflows from outside the region during the precrisis years, especially from Europe to North Africa and from Russia to central Asia. Growth in partner countries, primarily in Europe and Russia, also fed increases in foreign exchange inflows to these countries in the form of export receipts. Finally, like emerging and developing economies elsewhere, countries in the Middle East and central Asia experienced a large increase in FDI flows from outside the region after 2003.

Foreign exchange regimes in these countries and the structure of domestic financial markets caused the combined effects of all these inflows to be expansionary. About one-third of the countries in the region maintained fixed exchange rates against the U.S. dollar, whereas those with more flexible currencies, like emerging economies elsewhere, managed those rates to avoid real appreciation of their currencies. Consequently, both countries with fixed regimes and those with managed floats accumulated large stocks of foreign exchange reserves prior to 2007. Though this was especially true of the oil exporters, it was true for the other country groups as well.

The structure of the financial system in this region made the induced increase in the money supply difficult to sterilize. Financial intermediation is even more strongly dominated by banks in this region than elsewhere, and many of those banks are still government owned. In the Middle East and northern Africa, for example, 85 percent of all outstanding financial assets consisted of bank assets, with 12 percent equities and only 3 percent bonds in 2005 (IMF 2005). The small size of domestic bond markets has made it difficult for governments to sterilize large inflows of foreign exchange. Consequently, bank credit to the private sector expanded rapidly throughout the region during the pre-2007 international boom, and banks in some countries (especially Azerbaijan and Kazakhstan) actually added to the inflow pressure by borrowing abroad to fund nontraded activities at home, as in eastern Europe. The rapid expansion in private credit caused stock and real estate markets to boom, and the precrisis years proved to be years of widespread and strong growth for emerging economies in the Middle East and central Asia.

In the oil-exporting countries, the effects of the boom were dampened by fiscal policies. Though most oil-exporting countries have no formal fiscal rules, during the precrisis period they tended to base their budgets on conservative oil price projections and to save excess oil receipts in stabilization funds. Consequently, except for Kazakhstan, these countries tended to save most of the additional revenues associated with higher oil prices. Though higher oil prices resulted in rapid growth, therefore, they did so in the context of large current account surpluses in the oil-exporting countries.

Fiscal policies were less restrained in the other countries in the region. The low-income countries tend to run small fiscal deficits on average, and the emerging-market economies tend to run substantially larger ones. However, fiscal deficits were sufficiently moderate that ratios of public-sector debt to *GDP* declined over time in the majority of these countries. Whereas the oil-exporting countries tended to have positive (and in some cases, very large) public-sector assets, the other countries had relatively large stocks of government debt at the beginning of the decade of the 2000s. Debt-*GDP* ratios were in excess of 70 percent among the emerging-market economies and averaged in the neighborhood of 80 percent in the low-income countries, but these ratios fell sharply over the course of the decade in low-income countries and declined as well in the emerging-market economies. Debt-*GDP* ratios remained high, however (in excess of 60%), in Egypt, Jordan, Lebanon, and Mauritania.

The state of the banking systems in the region was mixed before the Great Recession. In the Middle East, they were basically sound, in part because of conservative lending policies. Prudential indicators were strong, and capital adequacy ratios were high. However, the banking sectors in the formerly centrally planned economies of the Caucasus and central Asia remained weak and fragmented, with the government playing a very large role and state banks remaining inefficient.

5. Sub-Saharan Africa

For the purposes of this chapter, the distinguishing features of the 44 countries in the sub-Saharan African region are their low incomes, weak institutional environments, susceptibility to real shocks, and limited degree of integration with international financial markets. Countries in this region had very poor growth performance from the mid-1970s to the mid-1990s. Indeed, per capita incomes in the region were at the same level in the mid-2000s as in the mid-1970s, even after a significant growth acceleration in the last decade of the period. As a group, these countries perform poorly on international governance indicators: according to the World Bank, 16 of the 20 countries in the world with the most difficult business conditions are found in sub-Saharan Africa. They have also suffered disproportionately from civil conflict.

As in the Middle East and central Asia, this region contains a significant group of oil exporters. Angola, Cameroon, Chad, Republic of Congo, Equatorial Guinea, Gabon, and Nigeria are all oil exporters. Many other countries in the region are heavily specialized in exporting primary commodities. Those countries benefited from improvements in nonoil commodity prices after 2003. Sub-Saharan African countries have large agricultural sectors, and fluctuations in commodity prices, together with droughts and natural disasters, have caused them to have large variances in their growth rates. Poor infrastructure and high costs of doing business have impeded the growth of manufactured exports, although the share of such exports has been increasing over time, to over a quarter of total exports by 2007. Although most manufacturing is linked to processing of resources, nine countries have significant shares of their exports in textiles and clothing, primarily to the United States and the European Union. Several countries (Cape Verde, Guinea-Bissau, Lesotho, Senegal, and Togo) are large remittance recipients. Barriers to trade both inside and outside the region remain important. Though traditionally, Europe and the United States have been the dominant destinations for African exports, Asia has become a larger consumer of African primary commodities in recent years, with about one-quarter of African exports going to that region.

Except for South Africa, the countries in this region have traditionally had very weak links with international capital markets. Many of them are heavily dependent on official development assistance. The IMF (2009c) classifies sub-Saharan African countries as emerging markets (South Africa), frontier markets with less integrated domestic markets and fewer links with international financial markets (12 countries, including Botswana, Cape Verde, Ghana, Kenya, Mauritius, Mozambique, Namibia, Nigeria, Seychelles, Tanzania, Uganda, and Zambia), and financially developing countries (31 other countries). Private capital inflows have traditionally been dominated by FDI, especially to extractive industries. Although portfolio capital inflows increased over the course of the decade of the 2000s, they were overwhelmingly directed to South Africa. However, this situation had begun to change at the time that the Great Recession struck. International rating agencies have begun to rate sovereigns in sub-Saharan Africa (14 countries were rated by Standard & Poor's and 12 by Fitch by 2006), and although these ratings remain far below investment grade, some countries in the region (Botswana, Ghana, Kenya, Nigeria, Uganda, and Zambia) began to receive private portfolio flows in small amounts, despite that in addition to trade restrictions, de jure restrictions on capital flows remain extensive.

Many countries in sub-Saharan Africa continue to maintain fixed exchange rates. There are 14 countries in the CFA (the French acronym for the central African franc) zone, and Lesotho, Namibia, and Swaziland are pegged to the South African rand. Both Cape Verde and the Comoros are pegged to the euro, and Botswana is pegged against a basket of currencies. Three countries are pegged to the U.S. dollar. Moreover, as elsewhere, nominal exchange rate stability is common even in countries with managed floats.

Financial sectors are poorly developed. As in other regions, banks dominate, but banking sectors are small. Nineteen countries in sub-Saharan Africa experienced large banking crises in the late 1980s and early 1990s, and banking systems have been reformed in many of those countries over the two decades since. Banks tend to be adequately capitalized, but they are hampered by weak legal and regulatory frameworks. As a result, many are highly liquid. Moreover, reserve ratios are high. Bank assets are highly concentrated, and rates of return on assets are low. Many banks have high ratios of NPLs, and prudential measures, such as those limiting concentrated exposures to individual borrowers or sectors, are often violated. Given their market power, banks have high margins and are profitable. However, concentration ratios have been declining, and foreign ownership has been rising. The majority of banking assets are foreign owned in about 20 sub-Saharan African countries. Securities markets are poorly developed even by the standards of other regions. Outside South Africa, there are no meaningful secondary markets for government bonds. Of 44 countries in the region, only 22 have equity markets, and of those, only 9 have more than 20 listings (IMF 2008d).

Countries in the region (except for some oil exporters) have tended to sustain fiscal deficits in the vicinity of 3–4 percent of *GDP*, about one-third of which are financed by external grants, and public-sector debt-*GDP* ratios have traditionally been very high. However, in the precrisis period, oil exporters reacted to high oil prices with sound fiscal policies, saving most of the increased revenues, and in the region as a whole, public-sector debt-*GDP* ratios were substantially reduced from 2000 to 2007 as the result of debt relief under the heavily-indebted poor country (HIPC) initiative, rapid growth, and conservative fiscal policies. The median debt-*GDP* ratio in 2007 was 40 percent. As a result of their conservative fiscal policies, oil exporters and South Africa have accumulated substantial reserves, and other countries have kept reserves roughly stable, with FDI flows and concessional financing sufficient to offset current account deficits that tend to run at 3–4 percent of *GDP* on average.

On the monetary front, only three countries (South Africa, Ghana, and Mauritius) maintain an inflation-targeting regime. Other countries that do not target the exchange rate use some type of monetary anchor.

The decade of the 2000s saw democratic institutions develop through most of the region. This was accompanied by substantial improvements in macroeconomic performance. In contrast with the past, favorable external conditions in the form of improved terms of trade, larger FDI flows, and increased flows of official development assistance, together with debt relief, did not result in a loss of fiscal discipline. Consequently, the precrisis period was characterized by increasing growth and declining inflation, with more rapid growth becoming geographically widespread and the vast majority of countries registering single-digit inflation. The growth acceleration dates from the mid-1990s, and growth rates averaged 5 percent per year through 2007. Growth has been persistent, and its volatility has decreased to a 30-year low (IMF 2007f). Though it was higher in the oil-exporting countries, it was historically high in others as well. Though this improved growth performance was undoubtedly helped by exogenous external factors such as high prices of minerals and food, as well as rising import demand in OECD countries, reduced public-sector debt-*GDP* ratios and sharply lower inflation suggest that improved macroeconomic policies played a role as well. Indeed, the lessons drawn by the IMF (2008e) were that growth in the region was promoted by engagement with the global economy, macroeconomic stability, and the use of prices to allocate resources.

For present purposes the key observation is that this improved policy environment, with the attendant smaller current account and fiscal deficits, lower inflation, reduced debt, and larger foreign exchange reserves, are likely to have reduced the vulnerability of sub-Saharan African economies when external conditions deteriorated after 2007.

This, then, was the situation in emerging and developing economies when the Great Recession broke out. Next we will consider the origins of the crisis and how it became transmitted internationally.

II. ORIGINS OF THE CRISIS[8]

The Great Recession originated in a collapse of housing prices in the United States in mid-2006, triggered by systematic tightening of monetary policy that caused the Federal Funds rate – the policy interest rate targeted by the Federal Reserve System – to rise from 1 percent in June 2004 to 5.25 percent by June 2006. This rate increase punctured a housing price bubble that had been under way in the United States for some time, driven at least partially by financial innovation coupled with poor financial supervision, a pattern that we saw repeatedly in the banking crises that we reviewed in Chapter 26.

Both the supply and demand sides of the market for housing finance in the United States were affected by the combination of financial innovation and regulatory failure: on the demand side, *subprime mortgage loans* – typically adjustable-rate mortgage (ARM) loans with low initial teaser rates that later adjusted to reflect market rates – enabled many individuals to finance housing purchases who were in fact poor credit risks. The regulatory authorities failed to appropriately police predatory lending by banks to such individuals. On the supply side, financial innovation in the form of the *securitization* of mortgage loans (the packaging of many such loans into securities that could be sold by banks in the United States' highly liquid bond markets) attracted substantial nonbank funds into the housing loan market in that country, and financial regulation failed to ensure that these securities were properly evaluated by credit-rating agencies. This situation was sustainable – and was sustained by the expectation of rising housing prices – until rising market interest rates triggered changes in the terms of ARMs that made these loans unaffordable for many individuals, leading to a wave of foreclosures and a collapse of housing prices.

It is important to emphasize that though financial innovation and regulatory deficiencies undoubtedly contributed to the housing bubble in the United States, it cannot be the whole story, because housing price bubbles also emerged in countries (such as the United Kingdom, Spain, and Ireland) that did not share the

[8] For very readable accounts of the origins of the Great Recession by two Nobel Prize–winning economists, see Krugman (2009) and Stiglitz (2009).

U.S. experience with financial innovation and that administered independent regulatory systems. Low international interest rates, driven by some combination of the international savings glut and expansionary pre-2004 Federal Reserve policies, certainly share some part of the blame.[9] The story of how this low–interest rate environment interacted with domestic housing markets to result in dramatic housing price appreciation in some countries and not in others remains to be told.

The financial implications of the housing-price collapse in the United States were magnified by two factors. First, the housing bubble did not literally pop but rather deflated – in other words, housing prices did not immediately collapse to observable new sustainable values. Instead, they began a decline that appears not to have bottomed out until the second quarter of 2009. Moreover, because the process of foreclosure on delinquent home owners creates its own negative dynamics for house prices, the adjustment in housing prices need not be monotonic. This means that the ultimate size of the required housing price adjustment in the United States – and therefore of the losses to be suffered by lenders – remains unknown. Second, the mortgage-backed securities that were issued by the loan originators to finance housing loans are inherently opaque, and thus the ultimate allocation of these losses among financial institutions is also unknown. The combination of these factors resulted in an enormous increase in *counterparty risk* (the risk that the other party in a financial contract will fail to carry out their obligations) that caused credit markets to freeze in the United States, particularly after Lehman Brothers was allowed to fail in September 2008, strongly suggesting that the liabilities even of major U.S. financial institutions would not be backed by the full faith and credit of the U.S. government. The breakdown of the credit system in the United States transmitted the crisis to the American real economy in dramatic fashion in the last four months of 2008, and the U.S. economy contracted at an annualized rate of over 6 percent in the last quarter of that year.

III. CRISIS TRANSMISSION

Although it originated in the United States, the crisis spread internationally very quickly. An initial channel of transmission was financial and primarily affected western European countries whose private capital markets are tightly integrated with that of the United States. One link was direct: the solvency of European financial institutions that had purchased U.S.-issued mortgage-backed securities was called into question by the perception that those assets had undergone a dramatic loss in value and had high levels of uncertainty about what their ultimate value would be.

A second financial link was indirect and had more worldwide effect. The implications for the real economy of the wealth destruction caused by the collapse of the

[9] For a recent empirical attempt to estimate the contribution of the savings glut to lowering interest rates in the United States, see Craine and Martin (2009).

housing bubble had begun to adversely affect the U.S. stock market by the third quarter of 2007 (the U.S. stock market peaked in October 2007), when subprime problems began to surface, and aggravation of the credit crisis in late summer and early fall 2008 accelerated the decline of the market. All projections for the world economy became more uncertain as the crisis deepened and confidence in the competence of policy makers in the major industrial countries waned. The increase in worldwide economic and political uncertainty acted like a monsoon effect that sharply reduced productive asset values throughout the world. Stock markets elsewhere around the world moved in sympathy, and the crisis spread through a worldwide decline in equity prices.

A third financial channel of transmission was a reduced international appetite for risk. This led to a reallocation of international financial portfolios from risky assets to those assets perceived as safest: U.S. government obligations and gold. This had two implications for emerging and developing countries. First, increased sovereign borrowing costs placed strains on fiscal accounts in countries whose governments had incurred substantial market debt. As we saw in Chapter 10, this tended to put upward pressure on the entire spectrum of domestic interest rates in such countries. Second, countries with currencies closely tied to the U.S. dollar faced an additional negative shock in the form of real effective exchange rate appreciation, pulled along by the appreciation of the dollar, whereas those with floating rates faced pressure for their currencies to depreciate as capital flowed out of their economies. Whether this resulted in an additional negative effect on aggregate demand in these countries or a positive one depended on their domestic vulnerability to exchange rate depreciation, an issue to which we will return later.

The fourth channel of transmission arose from the consequences of the crises for real activity in the North Atlantic economies. The breakdown of credit markets and collapse in asset values in those economies caused economic activity to contract dramatically (including by over 12% at an annual rate in the fourth quarter of 2008 in Japan). This manifested itself in two "real" channels of transmission: a decline in the demand for the exports of developing countries and a sharp contraction in flows of worker remittances. The dramatic decline in industrial-country import demand after mid-2008 is evident in Figure 29.1. This demand contraction had two effects: it directly affected export volumes in developing countries, and as shown in Figure 29.2, it resulted in a sharp reduction in the prices of primary commodities after summer 2008, most prominently that of crude oil. The combination of these factors complicated the impact of this real channel of transmission: though almost all countries underwent significant reductions in their export volumes, the behavior of their terms of trade was varied, depending on whether they were primary commodity exporters or importers. Commodity importers had some of their export decline compensated for by improvements in their terms of trade; for commodity exporters, however, no such compensation emerged. Although the impact of declines in workers' remittances was uniformly negative, the importance of this

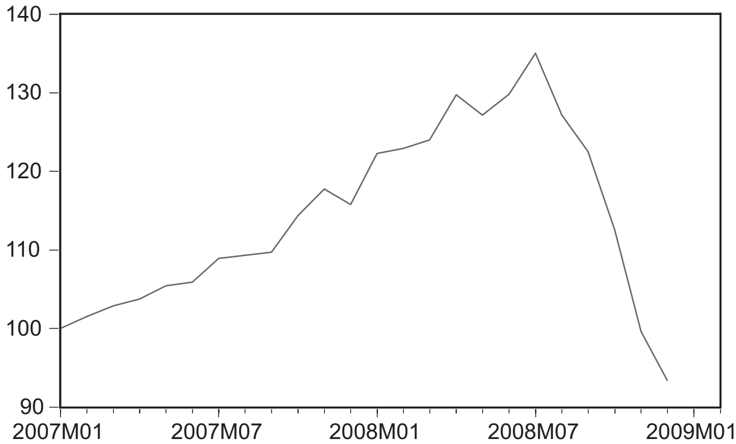


Figure 29.1. Import volume, all OECD countries (January 2007 = 100)

channel also varied widely among emerging and developing countries because not all such countries are remittance recipients.

From the perspective of emerging and developing economies, the preceding description of the origins of the crisis and its mechanisms of international transmission lead to three observations: (1) the crisis was of external origin, (2) its mode of transmission has been both financial and real, and (3) because most emerging and developing economies have become more open both commercially and financially in recent years, the impact of the shock is likely to be larger than it would have been before the market-oriented reforms that these countries undertook in the 1990s. These observations have implications for the magnitude of the shock confronting these countries and its distribution across countries. Specifically, as we will see, the shocks caused by the crisis have been large for all countries, but they have been worse for more financially integrated economies, for economies with more liquid

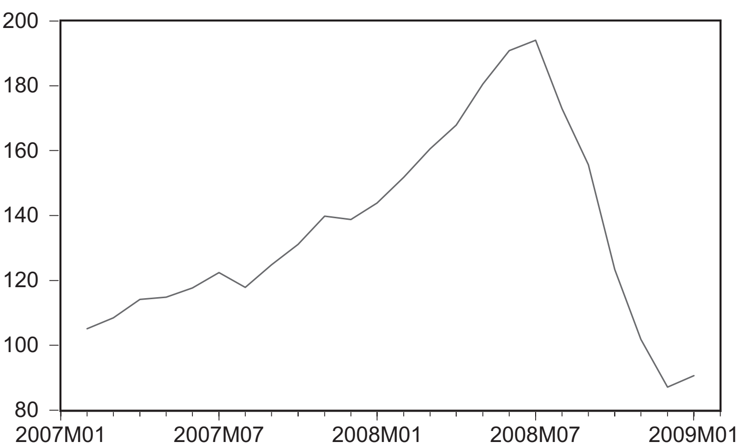


Figure 29.2. All-Commodities Price Index (January 2007 = 100)

financial markets, for economies that are more open commercially, for those with a larger share of manufacturing exports, for those with stronger ties to the United States (e.g., Mexico), and for those more dependent on commodity exports. In the rest of this section, we will examine the severity of the shock and the mechanisms of transmission to each of the major emerging- and developing-country regions.

1. Asia and the Pacific

As indicated in the preceding section, emerging economies in Asia became even more integrated with international financial markets after the 1997–1998 Asian crisis. Accordingly, the initial impact of the crisis on Asian economies was financial. As in the United States, stock markets peaked in many Asian countries in October 2007 and fell by nearly half from the end of 2007 to October 2008. By February 2009, they had fallen by 60 percent. Net portfolio equity inflows declined after October 2007 as well. In the last quarter of 2008, after the Lehman Brothers collapse, bank-lending flows to Asia turned heavily negative, and access to external bond financing became much harder. In fact, international bond markets were completely closed to Asian firms from mid-2008 to March 2009 (IMF 2009a).

Increased real integration also caused the crisis to be transmitted to Asia through markets for goods and services. According to the IMF (2008f), direct and indirect exposure through exports to the United States and the European Union in emerging Asia had increased from about 22 percent of emerging Asian GDP in 1994 to about 32 percent by 2006. Asia's concentration on electronic goods, motor vehicles, and capital machinery made it particularly vulnerable to the crisis because the Great Recession hit the manufacturing sector especially hard (because the sector produces lumpy goods that rely heavily on financing). Exports to these traditional markets from emerging Asia indeed fell sharply in 2007. Electronics exports peaked in the third quarter of 2007, and prices of semiconductors (which represent about one-third of electronics exports from emerging Asian countries) fell. However, exports from emerging Asia to nontraditional markets in eastern Europe, Latin America, Africa, and the Middle East, which, as we saw in the preceding section, have become a nonnegligible share of the region's exports, held up much better. Figure 29.3 shows the dramatic fall in real goods exports from the ASEAN-5 countries (Indonesia, Malaysia, the Philippines, Singapore, and Thailand) in 2008–2009.

As a result of these strengthened real and financial links, growth in Asia had become much more highly correlated with that of the United States in 2000–2007 than it had been in 1990–1996, making the seven Asian countries with the highest trade exposure to the United States, for example, as highly synchronized with the U.S. economy as Canada and Mexico. Indeed, fixed-effect panel regressions of growth in individual Asian countries on growth in the United States, European Union, and Asia as a whole; changes in terms of trade; and three crisis controls show that a 1 percent decline in U.S. growth leads to a decline of 0.6 percent in growth in

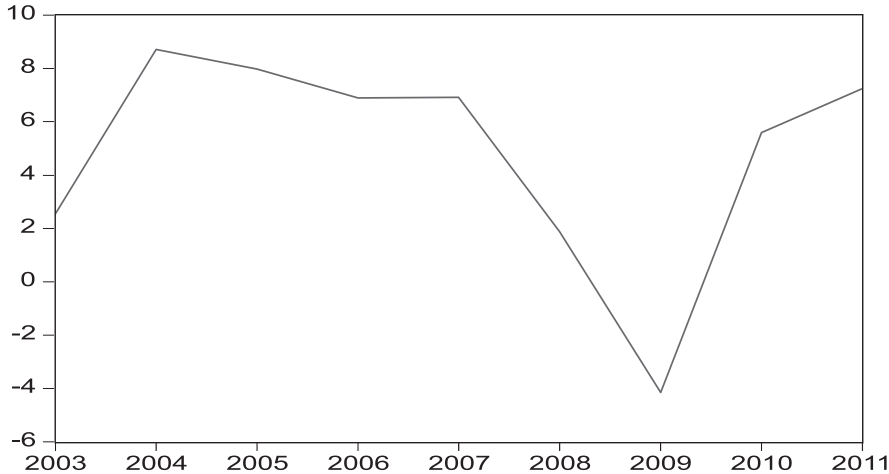


Figure 29.3. Growth rate of real goods exports, ASEAN-5

Asia, a greater impact than the United States has on western hemisphere countries (IMF 2009a). The impact of U.S. growth on China and India also appears to have increased sharply between these two periods. However, real *GDP* in emerging Asia, excluding China and India, contracted even more than would have been expected from these results – by 15 percent on an annualized basis in the fourth quarter of 2008, compared to a 6 percent contraction in the United States.

2. Emerging Europe

Emerging European economies are highly open. They export primarily to the rest of Europe, with a very small share of trade with Asia and oil-producing countries. They rely much more on external bank financing than do other emerging economies, and this type of financing has been notoriously volatile in previous crises. As in Asia, sovereign risk premia for economies in emerging Europe began to turn up in mid-2007 and increased continuously thereafter. Private external bond issues also contracted sharply in mid-2007, and stock prices turned down at the same time. Countries in the region that had large fiscal and current account deficits, and that were more dependent on bank capital flows, suffered larger increases in sovereign risk premia and larger stock price collapses (IMF 2009d). In part, this was because pressure to reduce leverage in the parent banks of eastern European subsidiaries led to a contraction of credit to the emerging economies in Europe (Figure 29.4), which caused a significant slowdown in credit growth there even before the collapse of Lehman Brothers. In addition, crisis management measures in advanced European economies increased funding costs for European emerging economies. For example, the extension of deposit guarantees in some European countries caused funds to flow out of banks in emerging European economies, which did not have the fiscal

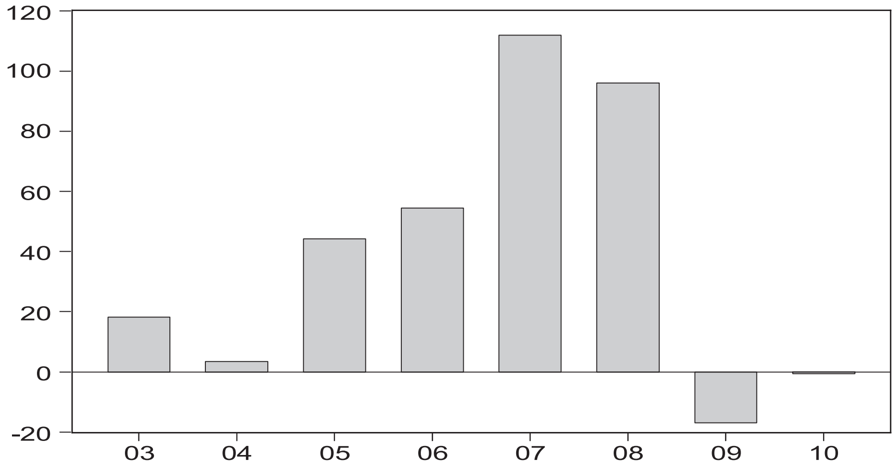


Figure 29.4. **"Other" private capital flows to emerging Europe, 2003–2010 (US$ billion)**

resources to extend similar guarantees (IMF 2009d). This sharp contraction in bank credit caused the collapse of the real estate boom that had emerged in the Baltic economies prior to 2007, which contributed to dramatic contractions in economic activity in these countries.

3. Latin America and the Caribbean

As mentioned in the preceding section, both financial and real openness have been increasing in Latin America. Consistent with strong integration between equity markets in Latin America and those of the United States, asset values collapsed in the seven largest Latin American countries during the second half of 2008, as shown in Figure 29.5. Stock market indexes in these countries held up well in the first half of the year but began to decline in sympathy with asset prices in the United States in June and shared the sharp contraction of the Dow Jones industrial average in October. In addition, Latin American governments faced a sharp increase in their costs of access to international financial markets, despite the deep reductions in policy interest rates that were implemented throughout the industrial world. The reason, of course, is that country risk premia rose very sharply, as shown in Figure 29.6. Another manifestation of heightened perception of financial risk was a substantial shortening of maturities of rolled-over debt (what Izquierdo and Talvi (2009) call *financial precarization*). The increase in the perceived risk of assets issued by Latin American governments and firms also resulted in capital outflows from the region.

The collapse in real economic activity in the region's major trading partners resulted in a commensurate collapse in demand for the exports of all the countries in the region because Latin American exports remain heavily oriented toward

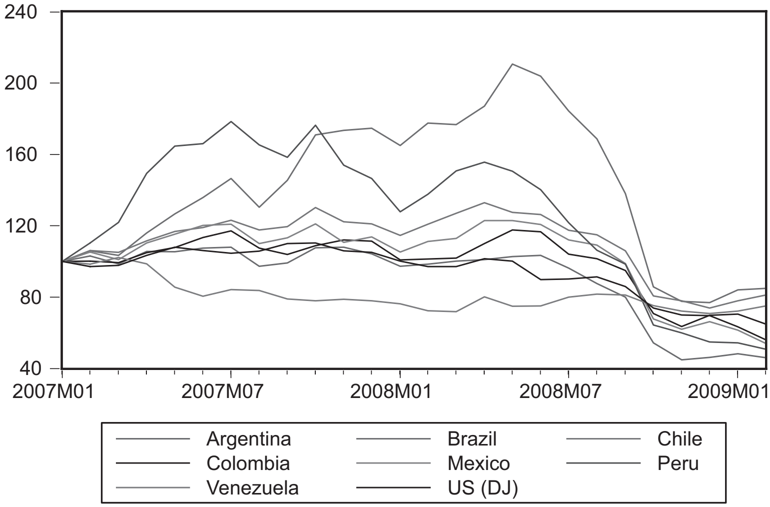


Figure 29.5. Stock market indexes for eight countries (January 2007 = 100)

western Europe and the United States. As shown in Figure 29.7, export revenues collapsed after mid-2008 in the seven largest countries in the region (the LAC-7).

As shown previously, the prices of primary commodities fell dramatically worldwide in summer 2008. The price of crude oil, for example, fell by more than two-thirds, which was obviously very damaging for the region's main oil exporters, including Venezuela, Mexico, and Ecuador. At the same time, oil-importing countries, such as Chile, Uruguay, and the Central American economies, benefited from lower oil prices. However, the reduction in oil prices only ameliorated the deterioration in the terms of trade that many of those countries would otherwise

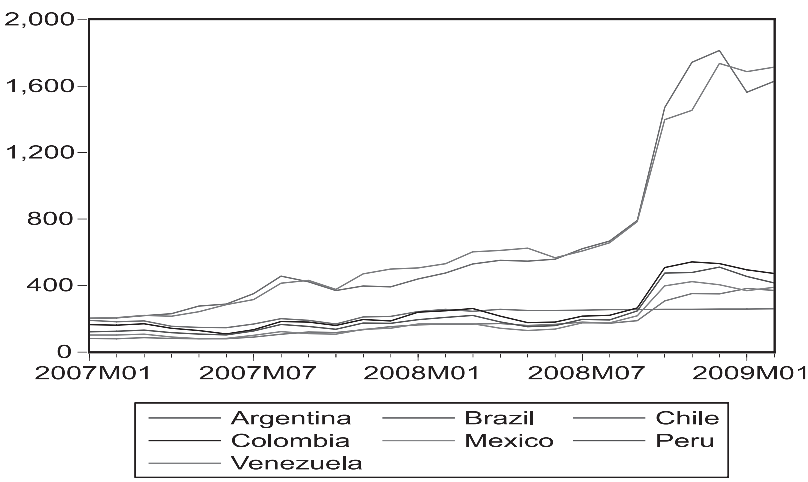


Figure 29.6. EMBI spreads, seven countries

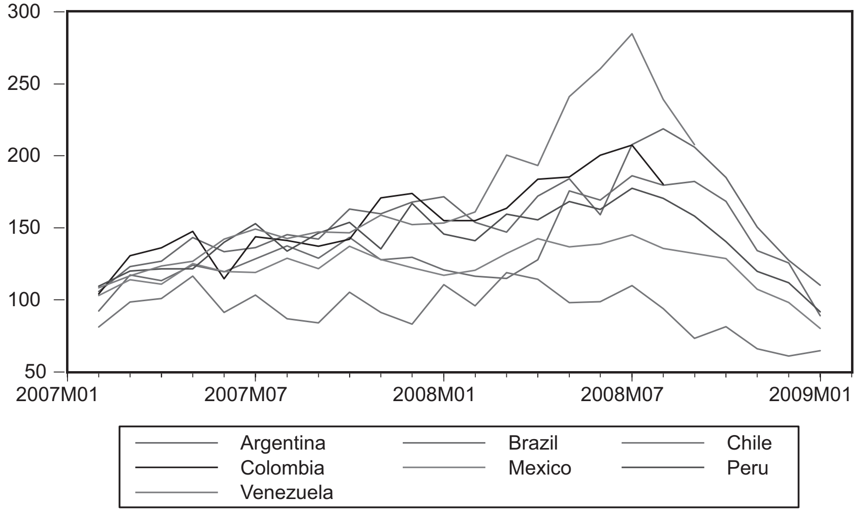


Figure 29.7. Index of U.S. dollar export values, seven countries (January 2007 = 100)

have suffered through declining prices of their own primary exports (e.g., copper in Chile, beef in Uruguay, coffee in Central America). Countries like Argentina, Brazil, and Colombia continued to enjoy more favorable terms of trade at the end of 2008 than they had at the beginning of 2007, whereas for Chile, Mexico, and Peru, the situation was decidedly worse. Nevertheless, for all these countries, the trend in their terms of trade was downward after early fall 2008.

A final channel of transmission that was important in Latin America, but with effects that were very different from country to country, was a significant reduction in remittance inflows as the result of reduced economic activity in the countries to which workers from the region had migrated, especially the United States. This factor was especially important in Mexico and various countries in Central America that are large remittance recipients (especially El Salvador, Guatemala, Haiti, and the Dominican Republic), but also for South American countries like Ecuador, which also have a large number of workers abroad.

4. Middle East and Central Asia

The transmission of the crisis to countries in the Middle East and central Asia was in many ways similar to that in other parts of the world, though the collapse in oil prices caused by the reduced level of economic activity in oil-importing countries played a more important role. Financial contagion emerged rapidly, manifesting itself in a contraction in regional stock markets and a widening of bond spreads for banks that borrowed heavily from abroad, especially in Kazakhstan. As in other regions, countries in the Middle East and central Asia also suffered a decline in FDI flows. In addition, the contraction of economic activity in Europe and Russia led to decreased demand for the region's exports, lower tourism receipts, and a sharp reduction in flows of worker remittances. A channel of transmission that proved to be particularly important in this region operated though the appreciation of the U.S. dollar as the result of a safe haven effect. Because many countries in the region were pegged to the dollar, their nominal effective exchange rates appreciated with the dollar, adding an additional contractionary shock.

5. Sub-Saharan Africa

Except for South Africa, countries in sub-Saharan Africa are only weakly integrated with international financial markets. Accordingly, crisis transmission through financial channels was more immediate in South Africa than in the rest of the continent. South Africa experienced portfolio flow reversals, depreciation of the rand, and stock market volatility. Financial transmission to the rest of the continent, while present, was much weaker. Though portfolio flows had been increasing rapidly in previous years, they remained very small when the crisis broke out. They indeed contracted sharply, and sub-Saharan African countries found themselves unable to issue foreign-currency bonds in the first half of 2008, but the small scale of these flows made this a less important channel of transmission in the rest of the continent than in South Africa. A more important financial channel of transmission for most countries in the region was reduced FDI inflows.

As in other regions, real transmission occurred through reduced demand for exports, lower commodity prices, and reduced flows of remittances. Other than South Africa, oil exporters were hit hardest, and first hit, in mid-2008.

IV. VULNERABILITY

The implications of the crisis-associated shocks just described for emerging and developing economies depend on the extent of vulnerability that characterizes those economies. Vulnerability determines the difference between a downturn and a crisis. In this respect, the news has been relatively good for many emerging and developing economies: though the reforms of the past two decades may have left them more exposed to external shocks, they have also at the same time made them more resilient in the face of such shocks. There are several reasons to believe that vulnerability has been substantially reduced for these countries by recent reforms.

First, a key source of macroeconomic vulnerability is the health of the financial system, as the United States and several other industrial countries have rediscovered to their dismay. As the result of financial reforms undertaken over the past decade and a half, including improvements in financial regulation and supervision, enhanced competition in the financial system, and in some cases, the recent resolution of banking crises, the financial systems of many emerging and developing countries were healthier at the outbreak of the crisis than they had been in the past.

The entry of foreign banks has also significantly contributed to the health of banking systems in many countries. Moreover, financial institutions in emerging and developing countries, unlike those in advanced Europe, did not tend to acquire the toxic assets (mortgage-backed securities) that caused so much trouble in advanced economies, so they did not experience the direct hit suffered by financial institutions in many industrial countries.

Second, central banks have been strengthened as macroeconomic institutions in many emerging and developing countries. Not only have they been accorded legal independence but they have taken responsibility for maintaining low and stable inflation rates, and to a significant extent, they have achieved that goal in recent years, enhancing their credibility and therefore their flexibility.

Third, many emerging economies have transitioned to more flexible exchange rate arrangements, reducing vulnerability to the disruptive discrete exchange rate depreciations that have been associated with currency crises and providing an automatic stabilizing effect in response to external financial shocks. As a result, bilateral exchange rates against the U.S. dollar indeed depreciated in many countries when the external financial environment turned adverse in fall 2008, as we will see later. Moreover, despite still being pronounced in some countries, financial dollarization has declined, reducing the impact of a factor that has weakened or even reversed the otherwise expansionary effect of exchange rate depreciation in the past.[10]

Fourth, fiscal reforms have enhanced the flexibility of fiscal systems in some cases, and many countries have recently demonstrated both the political will and economic ability to make significant fiscal adjustments. As a result, public debt stocks (as a proportion of *GDP*) have generally declined over the past several years, as shown for seven Latin American countries in Figure 29.8. Coupled with the reform of fiscal institutions in some countries (such as the Structural Balance Rule in Chile and the Fiscal Responsibility Law in Brazil that we examined in Chapter 11), these reforms should have enhanced fiscal credibility, while at the same time strengthening the effects of automatic fiscal stabilizers.

In addition to these institutional reforms in the financial sector, monetary, exchange rate, and fiscal policy areas, the accumulation of large stocks of international reserves has also contributed to reducing vulnerability to adverse external shocks among emerging and developing countries. Figure 29.9 shows the dramatic rate of accumulation of such reserves relative to the size of their economies for all emerging and developing countries after the 1997–1998 Asian crisis. These reserves have been accumulated both to serve as self-insurance against sudden stops of capital inflows (i.e., to serve as rainy day funds) and to prevent undesired appreciation

[10] To the extent that currency mismatches are induced by fixed exchange rate regimes combined with lax financial regulation, improved regulation and more flexible exchange rate management could be behind the reduction in the extent of such mismatches.

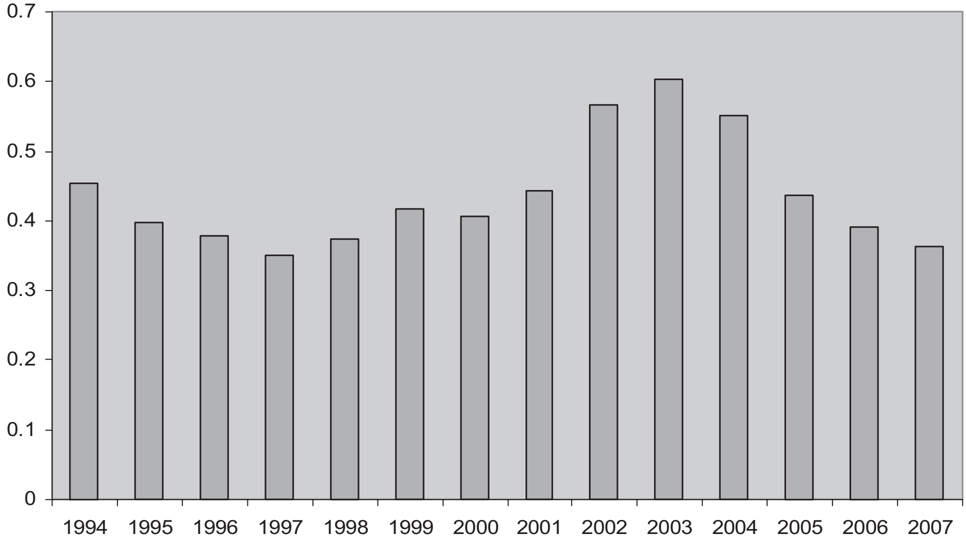


Figure 29.8. Total public-sector debt/GDP, LAC-7 countries ***Source.*** **Inter-American Development Bank**

of domestic currencies. At the outset of the crisis, they represented large stocks of liquid public-sector assets that could be deployed to prevent excessive exchange rate depreciation, if desired, or to finance temporary fiscal deficits or other fiscal outlays to support recovery, if necessary.

Finally, these reserves were strengthened by liquidity agreements with the U.S. Federal Reserve System (Brazil and Mexico received liquidity commitments for $30 billion each) and massive IMF resources pledged by G-20 countries to be used

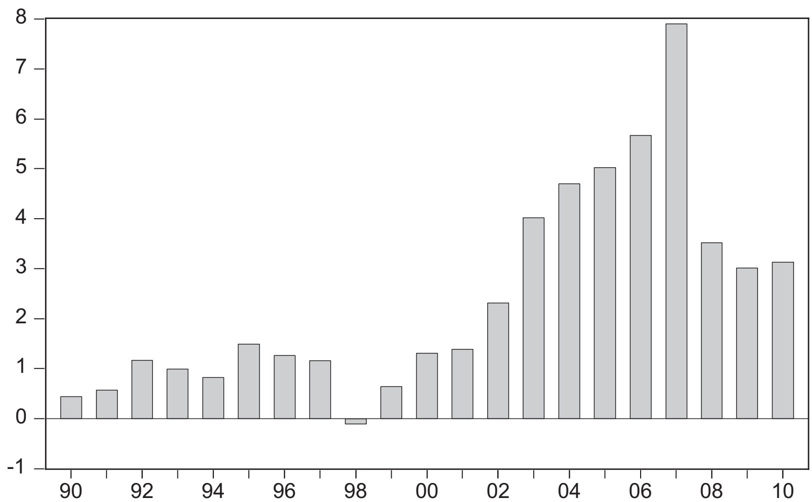


Figure 29.9. Ratio of reserve accumulation to GDP, emerging and developing countries

in new low-conditionality programs. For example, the IMF reached precautionary liquidity agreements in several Latin American countries for a total of over $60 billion and reached a similar agreement with Poland.

All these factors suggest that the large external shock that the Great Recession represented for emerging and developing countries arrived in a very different environment from that which characterized most of these countries in the previous decade. Consequently, despite their magnitudes, the external shocks associated with the Great Recession should have been expected to have been less disruptive to these economies than their macroeconomic history – and particularly that of the previous decade – might otherwise have led one to believe. Most important, perhaps, are that the sudden disruptions associated with banking and currency crises are less likely in the current context of reformed banking systems and more flexible exchange rate regimes; that financial and macroeconomic policy institutions have more credibility, thereby making short-run deviations from medium-term policy stances less disruptive to expectations; and that policy makers have the means at their disposal to counter shocks – in the form of large reserve stocks – that have not been available in the past.

V. POLICY RESPONSES

Did emerging-market economies actually respond to the shocks associated with the Great Recession in the ways that their improved circumstances have led us to expect? In this section, we address this question, once again taking a regional perspective.

1. Asia and the Pacific

The response to the crisis among Asian emerging economies is particularly instructive as it differed substantially between countries that had and had not implemented significant macroeconomic reforms by the time the crisis hit. Low inflation and monetary credibility among Asian emerging economies made it possible for these countries to respond to the crisis with monetary easing without fear of igniting inflation expectations, and countries in the region indeed took advantage of this opportunity. Monetary conditions began to be eased in spring 2007, and monetary expansion became more aggressive in the last quarter of 2008, when the recession became acute internationally. According to the IMF (2009b), the median decline in policy rates in the region from the third quarter of 2008 to the second quarter of 2009 exceeded 2.25 percent, five times as much as in past recessions. China, India, Korea, Taiwan, and the ASEAN-5 (including Vietnam) all lowered policy rates or decreased reserve requirements in the last quarter of 2008, and China, which continued to maintain a repressed financial system, also loosened credit quotas and mortgage lending conditions and provided greater support for lending to small and medium enterprises. In Korea, the policy rate was cut by more than 3 percentage points after November 2008, helping credit growth to remain strong. Policy rates in many countries were allowed to become negative in real terms, whether measured relative to headline or to core inflation.

Floating exchange rates also contributed to stabilizing aggregate demand in these countries. Exchange rates came under pressure to depreciate as the result of a safe haven effect that saw capital flow from all over the world into U.S. Treasury bills as well as the monetary expansion that these countries undertook to sustain demand. Emerging economies in Asia did not resist depreciation, though several countries accepted some reserve losses to smooth their exchange rate changes. The Korean won and Indonesian rupiah, in particular, depreciated about 20 and 10 percent in nominal effective terms from September 2008 to March 2009, respectively.

On the fiscal side, modest debt levels and large stocks of foreign exchange reserves allowed Asian economies to enter the crisis with room to implement countercyclical fiscal policies in the form of fiscal expansions without fear of perceived threats to fiscal solvency. Most Asian countries do not have automatic stabilizers such as unemployment insurance, so they had to rely on discretionary measures. They did so in an aggressive way. China announced a large fiscal stimulus package in November 2007, focused on a massive program of public investment. The four tiger economies of Hong Kong, Korea, Singapore, and Taiwan allowed an increase in their fiscal deficits by 2.25 percent of *GDP* in 2008. The IMF's (2009a) calculations showed a positive fiscal stimulus in China, India, the four tigers, and the ASEAN-4 economies (Indonesia, Malaysia, the Philippines, and Thailand) during 2009 that were comparable or larger in size than those in the G-20 countries on average (2.75% of *GDP*, compared to about 2% in the G-20 countries). Throughout the region, discretionary fiscal measures were heavily weighted toward spending, especially investment in infrastructure, and were implemented more quickly than in industrial countries. Importantly, the IMF (2009a) notes that the fiscal policy response to the current recession in Asia has been stronger than the response to past recessions, with a median increase in the fiscal deficit of over 3.5 percent, more than double the response after the Asian crisis.

With regard to the financial sector, financial institutions in Asia had limited exposure to subprime loans in the United States, and the functioning of Asian credit markets was not interrupted initially, though credit became less plentiful, and bank-lending rates increased. After the collapse of Lehman Brothers, however, confidence in some Asian banks suffered, and several governments (Hong Kong, Indonesia, Malaysia, Singapore, and Taiwan) created liquidity (lender of last resort) facilities and expanded deposit insurance. However, credit did not freeze up in Asia, as it did temporarily in the United States and western Europe.

By contrast with the experience of emerging economies in Asia, developing economies in the region were not in a position to mount a forceful policy response to the crisis. Those countries were hit by lower commodity prices, reduced demand for nontraditional exports such as garments, lower tourism receipts, and reduced

FDI flows. However, weak public solvency positions, reduced fiscal revenues because of lower commodity prices and lower revenue from import taxes, ineffective monetary transmission mechanisms, and inflexible exchange rates rendered countercyclical responses much more difficult to implement in those countries.

2. Emerging Europe

The contrast between the policy response of emerging economies in Asia and those in Europe is stark. As in past crises, and in contrast with emerging economies in Asia, many emerging economies in Europe negotiated adjustment programs with the IMF, in exchange for exceptional financing. As of April 14, 2009, IMF programs were in place in Belarus, Bosnia-Herzegovina, Hungary, Latvia, Romania, Serbia, and Ukraine, and as mentioned earlier, Poland availed itself of the IMF's unconditional new Financial Credit Line.

With respect to the monetary policy response to the crisis in emerging Europe, recall that many European emerging economies maintained fixed exchange rate regimes. Although the exchange rate was floated in the Ukraine, and Belarus devalued its fixed exchange rate by 20 percent against the U.S. dollar as well as switching its peg to a basket of currencies, exchange rate pegs were maintained in almost all these countries, with no change in parities. One consequence was that, in contrast with Asia, interest rates were actually *increased* in countries with fixed rates, driven by a need to defend the currency in the face of large capital outflows. However, some countries with floating rates (e.g., Hungary) also tightened monetary policy.

Turning to fiscal policy, the IMF (2009d) argued that fiscal stimulus was not an option in emerging European economies, essentially because these countries had no fiscal space when the external shock arrived. Instead, reminiscent of the Asian crisis, more fiscal adjustment was perceived as required for bank recapitalizations and to restore market confidence in public-sector solvency. Although the average central government balance relative to *GDP* went from a surplus of 0.3 percent in 2008 to a deficit of 5.2 percent in 2009 in emerging Europe, this outcome was largely an endogenous response to a very severe contraction in economic activity, and discretionary fiscal measures were intended to be *contractionary* in the IMF-supported programs in this region (IMF 2009d). Belarus, for example, committed itself to pursuing a balanced budget and undertaking public-sector wage restraint; Hungary undertook fiscal restraint to provide confidence and finance the enhancement of bank capital; the widening of the fiscal deficit was limited in Latvia; the fiscal stance was strengthened in Romania to reduce the government's financing needs and to assure the availability of resources to maintain adequate capitalization of banks as well as to help bring inflation within the central bank's target; the fiscal stance was tightened in Serbia; and Ukraine sought to achieve a balanced budget in 2009. Finally, regarding the financial sector, in addition to recapitalizing banks, all the emerging economies in Europe substantially increased deposit guarantees in response to the crisis.

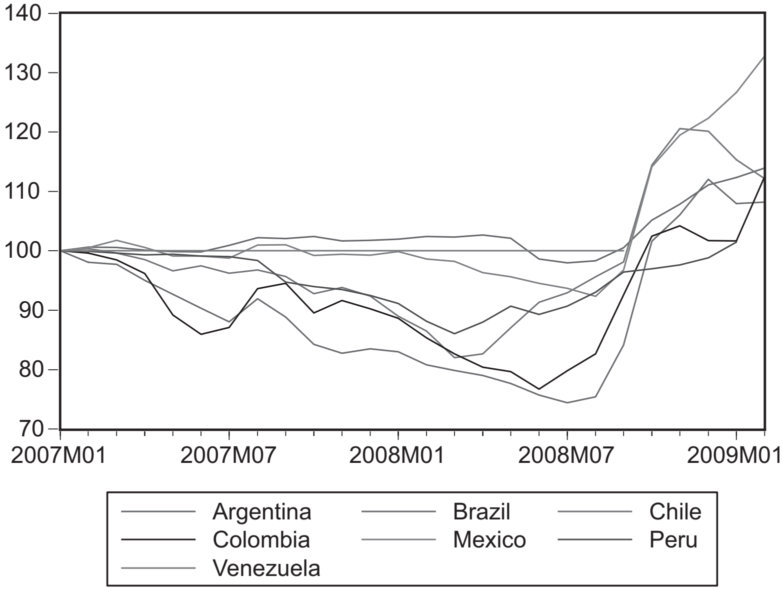


Figure 29.10. Bilateral U.S. dollar exchange rate indexes, LAC-7 (January 2007 = 100)

3. Latin America and the Caribbean

As already mentioned, the improved macroeconomic environment in Latin America caused sovereign risk premia for the emerging economies in that region to behave more like lower-risk corporate bonds in the United States than like junk bonds. Thus policy reforms may have benefited the region even before the implementation of policy responses by reducing the size of the financial shock to which Latin American emerging economies were subjected (IMF 2009e).

But the policy response also helped. Indeed, the policy response to the crisis in Latin American emerging economies was in many ways similar to that in Asia, reflecting the reforms that the region had implemented over the previous two decades. Monetary policy turned expansionary somewhat later than in Asia and did so especially after the Lehman Brothers collapse in September 2008, when the international crisis deepened dramatically. Policy rates were lowered in all the major inflation-targeting emerging economies in the region, beginning with Colombia in December 2008, as well as in non-inflation-targeting countries such as the Dominican Republic, Honduras, Paraguay, and Venezuela (IMF 2009e). The transmission from policy rates to bank-lending rates worked well in these cases. Unlike Europe, there was no major effort to defend exchange rates by raising interest rates, except in Jamaica. As a result, as in Asia, the crisis resulted in large exchange rate depreciations in Latin America (Figure 29.10). Again as in Asia, there was some drawing down of reserves to smooth the exchange rate adjustment.

The fiscal policy response reflected the extent to which countries had previously undertaken economic reforms and was therefore mixed within the region. Fiscal stimulus packages were adopted in Argentina, Brazil (0.6% of GDP), Chile (2.6% of *GDP*), Costa Rica, El Salvador, Mexico, Panama, Paraguay, and Peru. Bolivia, Costa Rica, and Paraguay also increased spending for countercyclical reasons. However, these packages were relatively modest compared to those in Asia. The fiscal authorities in commodity-exporting financially integrated countries – the IMF's identification of the countries in the region with the most developed institutions for policy formulation (Brazil, Chile, Colombia, Mexico, and Peru) – provided the most support, increasing their average domestic primary deficit by some 3.5 percent of *GDP* in 2009. The cyclically adjusted primary balance was loosened in 2009 among such countries, whereas it was tightened on average among the group of countries classified as "other" commodity exporters (Argentina, Bolivia, Ecuador, Paraguay, Suriname, Trinidad and Tobago, and Venezuela). Consequently, the former implemented a positive *fiscal impulse* (the change in the cyclically adjusted primary deficit), the latter a negative one. Fiscal policy was thus countercyclical in the former but pro-cyclical in the latter. The contrast between Chile, on one hand, and Ecuador and Venezuela, on the other, is instructive. Chile planned to increase real spending by 15 percent in 2009 by drawing down previously accumulated funds, while countries without such resources, such as Ecuador and Venezuela, planned to cut spending by 10 percent. Tourism-intensive commodity-importing countries, the IMF's term for a variety of small economies in the Caribbean, tended to have large stocks of public debt on average and maintained fixed exchange rates in the context of open capital accounts with a fairly high degree of financial integration. These countries consequently had little scope for either fiscal or monetary stimulus.

As in other regions, specific measures were adopted to safeguard the financial system. Liquidity support was provided to financial institutions in Argentina, Brazil, Chile, Costa Rica, Colombia, Guatemala, Mexico, and Peru. The scope of institutions with access to the central bank's discount window was widened, and reserve requirements on banks were lowered.

4. Middle East and Central Asia

The response to the crisis in the Middle East and central Asia was mixed. As mentioned earlier, many countries in the Middle East, both oil exporters and importers, maintain fixed or heavily managed exchange rates against the U.S. dollar. Such countries saw their nominal effective exchange rates appreciate, following the appreciation of the dollar. Because they also tended to have higher inflation rates than their trading partners, their real effective exchange rates appreciated as well. Those with more flexible exchange rate regimes generally allowed their currencies to depreciate, though countries in the Caucasus, and in central Asia, in particular, were constrained in the extent to which they could do so by currency mismatches in the balance sheets of their banks and corporations. As a result of reduced oil prices, oil exporters in the Middle East and North Africa, which largely maintained fixed exchange rates (except for Algeria and Sudan), registered much-reduced current account surpluses that were more than offset by capital account deficits, so they drew down their foreign exchange reserves to defend the values of their currencies.

An easing of monetary policy was nearly universal. On the fiscal side, however, the response was less uniform because of wide differences in the available fiscal space. Countries with substantial fiscal cushions, such as Saudi Arabia and the United Arab Emirates, in particular, but also Algeria and Libya, undertook substantial fiscal stimulus programs, as in Asia and Latin America. Where they were present, sovereign wealth funds responded to the contraction in capital inflows associated with the crisis by lending more actively in their domestic economies and funding public-sector projects. Fiscal stimulus was also widely undertaken by countries in the Caucasus and central Asia, especially in countries such as Azerbaijan, Turkmenistan, and Uzbekistan, which had saved during the boom. However, many countries in the region had no such fiscal cushions. This was true of oil exporters such as Iran, Sudan, and Yemen, which had to reduce their fiscal deficits even in the face of the crisis to preserve fiscal sustainability. It was also true on average of oil-importing countries in the Middle East, which averaged public-sector debt-*GDP* ratios of over 60 percent, four times as high as the oil-exporting countries or the countries in central Asia. According to the IMF (2009f), among Middle Eastern oil-importing countries, all but Morocco and Tunisia adopted pro-cyclical policies in response to the crisis, tightening fiscal policy to preserve fiscal sustainability in the face of declining revenues.

Finally, even though banking systems were relatively strong, their lending for real estate and stock purchases made them vulnerable to the regional collapse in asset prices. Moreover, in the Caucasus and central Asia, dollarization remained important in the banking sector, and exchange rate depreciation interacted with currency mismatches to create substantial losses for banks. Therefore governments in many of these countries provided support to their banking systems through deposit guarantees, liquidity support, capital injections, and equity purchases.

5. Sub-Saharan Africa

Though many countries in sub-Saharan Africa maintain fixed exchange rates, the high level of foreign exchange reserves in the region – the median value was 13 percent of *GDP* at the outset of the crisis, compared to 5 percent in the 1970s and less in the 1980s – has created more scope for financing balance of payments deficits and has made it less imperative for countries in the region to raise interest rates to defend their exchange rate parities. Accordingly, as in regions other than Europe, two-thirds of the countries in sub-Saharan Africa lowered one or more policy interest rates after the crisis began.

The IMF, using pairwise correlations between spending growth and *GDP* growth, as well as measures of *fiscal amplitude* (the difference between spending growth when *GDP* growth is above its median value and when it is below), found that fiscal policy was pro-cyclical in sub-Saharan Africa during 1980–2008 (IMF 2009g). This fiscal pro-cyclicality was more pronounced for countries with high debt ratios and weaker institutions and was more pronounced for capital than for current spending. But using the fiscal amplitude measure, pro-cyclicality declined from the 1980s to the 1990s and from the 1990s to the 2000s.

This experience reflects the gradual easing of constraints on countercyclical fiscal response in the form of high debt, high inflation, and large fiscal deficits during boom times. Consistent with this progression, though fiscal balances in the region had been strongly negative at the outbreak of previous crises in 1975, 1982, and 1991, they were slightly positive on average for the region as a whole when this crisis began. Seventy-two percent of all sub-Saharan African countries had surpluses in their primary fiscal balances during 2006–2008, compared to 28 percent in 1991–1995, and 71 percent had ratios of public-sector debt to GDP of less than 60 percent, compared to 33 percent in the earlier period. Many countries in the region therefore had room to implement countercyclical fiscal policy without jeopardizing their perceived fiscal solvency, and the IMF (2010b) notes that nearly two-thirds of the sub-Saharan African countries that experienced growth slowdowns as a result of the crisis were actually able to increase government spending to stabilize their economies.

Accordingly, fiscal policy has been eased in some three-quarters of sub-Saharan African countries in response to the Great Recession (IMF 2009g). Fiscal deficits increased as the result of automatic stabilizers as well as discretionary responses. The turnaround in fiscal balances amounted to some 6 percent of GDP on average in 2008, a sharp contrast with experience during past recessions. Countercyclical fiscal policies were implemented in particular in the middle-income countries of the region, where public-sector debt ratios were lowest. For those countries, the increase in deficits was due primarily to discretionary spending increases. In oil-exporting countries and low-income countries, by contrast, decreases in revenues dominated. Fiscal targets were loosened in about three-quarters of the countries in the region with IMF agreements (IMF 2009g).

VI. MACROECONOMIC OUTCOMES

How have these policy responses affected the impact of the Great Recession on emerging and developing economies? This section examines macroeconomic outcomes among such economies, again following a regional sequence.

1. Asia and the Pacific

The heavily export-dependent economies in the Asian region suffered the sharpest output contractions as the result of the crisis, but they also experienced the fastest recoveries. The international recession had a particularly severe effect on the demand for electronics and other manufactured goods all over the world, not only because the recession originated in high-income countries that are the primary market for goods such as consumer electronics and automobiles but also, as we have seen, because such goods tend to be lumpy, and their purchase therefore relies heavily on finance. As credit markets froze up in the industrial countries in the last quarter of 2008, final demand for such goods suffered accordingly. Production of manufactured goods was also affected by the behavior of inventories. As final demand for such goods contracted, firms ran down their inventories, causing production to decrease by more than final demand. Asian countries that specialized in the production for manufactured goods for high-income countries were therefore hit particularly hard by the Great Recession, as illustrated by the sharp export contraction among the ASEAN countries shown in Figure 29.3.

As financial conditions eased in early 2009, however, demand for manufactured goods began to recover in industrial countries, and the restocking of inventories caused total demand for manufactured goods to increase more than final demand. This led to a recovery in world trade that benefited Asian manufactured exporters. Accordingly, recovery started in Asia during the first quarter of 2009, led by exports. However – and this is the important point for our purposes – recovery in Asia was also supported by expansionary domestic policies. Simulations with the IMF's Global Integrated Monetary and Fiscal model reported in IMF (2010a) estimate that fiscal stimulus in those countries added 1.75 percent on average to growth in Asia during the first half of 2009. The region benefited in particular from an increase in regional exports to China, responding to the boost given to Chinese demand by the infrastructure investment associated with Chinese fiscal stimulus as well as to increased private investment in that country caused by the countercyclical relaxation of credit restrictions. These measures, together with the recovery in world trade, had the effect that by August 2009, alone among the major economies of the world, China was growing at rates above its long-run trend.

The quick and strong recovery in Asia caused risk premia for the region to return to their normal levels (though not to their unusually low precrisis levels), and foreign capital to begin to flow into Asian countries once again in the second quarter of 2009. Inflows of portfolio equity capital contributed to a sharp rebound in regional stock markets, which soon returned close to precrisis levels. Already in 2009, the combination of sustained current account surpluses and restored capital inflows caused Asian economies to begin to accumulate foreign exchange reserves once again.

2. Emerging Europe

As was true of other emerging-market regions, stock markets in emerging Europe rebounded during 2009, responding to reduced risk aversion on the part of international investors that caused stock markets to rebound all around the world. However, recovery in real economic activity was much more uneven within emerging Europe. Although the economic rebound among these countries was held back in general by the slow pace of recovery among industrial economies in western Europe, with which emerging European countries have strong commercial and financial ties, the pattern of recovery in emerging Europe was linked both to the policy response to the crisis and to precrisis vulnerabilities, especially in the financial sector. Countries such as Hungary, Latvia, and Romania had failed to build up fiscal space during the boom years to allow them to undertake countercyclical fiscal policies in response to the crisis, and their fiscal policies therefore responded procyclically, prolonging the output contraction in those countries. Financial-sector distress slowed the recovery process in Estonia and Lithuania. By contrast, Turkey, which was able to undertake countercyclical fiscal measures and avoid the financial-sector weaknesses of the Baltic countries, was projected by the IMF to recover to growth rates of over 5 percent in 2010 and over 3 percent in 2011, after contracting by 4.7 percent in 2009.

3. Latin America and the Caribbean

Contrary to past experience, there was no acceleration of inflation in Latin America, and growth did not drop more than in other parts of the world. Moreover, the region sustained a strong recovery in 2009.

Recovery in Latin America was driven by a combination of expansionary domestic policies, the recovery in world demand for manufactured goods mentioned earlier, and improved commodity prices, driven by the rapid recovery in the commodity-intensive Asian economies. Both these components of the external environment began to recover in 2009, though other components, such as remittances and tourism, did not. This meant that the large commodity-exporting countries in the region recovered more quickly than the smaller commodity-importing ones. For instance, real output began to recover in Colombia in the first two quarters of 2009 and in Brazil in the second quarter. Other commodity exporters also began to recover at about the same time, but small commodity importers lagged behind.

As in Asia, capital inflows also returned to the region rather quickly. The increase in sovereign risk spreads had been reversed by mid-2009 for the higher-rated countries of Latin America, and public-sector access to external financial markets had been restored by the second quarter of 2009, though access by private firms lagged somewhat behind. Large nonresident portfolio outflows turned out to be short-lived, and inflows returned to the larger countries in the first part of 2009. As in previous crises, flows of FDI turned out to be more stable than those of other types of external financing. Overall, foreign exchange reserves remained above end-2007 levels in the region.

The IMF (2009e) estimates that Latin America's improved policy environment cut the output cost of the external shocks associated with the Great Recession from the third quarter of 2008 to the second quarter of 2009 in half for Brazil, Chile, Colombia, Mexico, and Peru (the financially integrated commodity exporters), from a projected decline of 8 percent in real *GDP* to an actual one of about 4 percent. For the other subgroups of countries in the region, which were constrained from adopting equally stimulative fiscal and monetary measures, recovery from the Great Recession was much more dependent on the external environment. The other commodity exporters were assisted in recovery by increases in world commodity prices, whereas the commodity-importing countries in the region were hampered by the slow recovery of tourism and workers' remittances.

4. The Middle East and Central Asia

As in other regions, the trough of the external financial shock in Middle Eastern and central Asian countries was in the first quarter of 2009. Spreads on sovereign *credit default swaps* (financial instruments that reflect the cost of insurance against a sovereign default) fell continuously and substantially after their peak in the first quarter of 2009, and stock markets in the region began to recover after March 2009.

Oil-importing countries in the region were hit less hard by the crisis than oil exporters because of their limited financial links with the rest of the world as well as their limited manufactured exports. However, the crisis was acute for oil importers in the Caucasus and central Asia because of their close links with Russia, an oil-exporting country that suffered a sharp contraction in 2009. Though several countries in the region have implemented countercyclical monetary and fiscal policies, high levels of public debt have forced them to moderate the size of their fiscal stimulus and restrict its duration. Accordingly, these countries have been particularly affected by the crisis. On the other hand, countercyclical policies helped many oil exporters in the Middle East as well as in the Caucasus and central Asia to moderate their growth slowdowns.

Kazakhstan, however, is a special case. As indicated in Section III, banks in Kazakhstan relied heavily on borrowing from foreign banks and used those funds, often in the form of short-term liabilities, for lending to unhedged domestic borrowers in construction and real estate, very much as in the Baltics. The depreciation of the domestic currency and curtailed supply of external bank loans caused Kazakhstan to suffer a severe Asian-style banking crisis from which it has not yet emerged at the time of writing in summer 2010.

5. Sub-Saharan Africa

Except for South Africa, most countries in the sub-Saharan African region seem to have hit bottom in the first quarter of 2009. In contrast with past crises, recovery in the region was faster than in the rest of the world, both because more policy shock absorbers, such as lower interest rates, were at work and because countries in the region were able to resist harmful measures such as pro-cyclical fiscal policies and increased trade restrictions. Though growth rates averaged over 6 percent from 2003 to 2008, growth fell to 2 percent in 2009 but was already projected by the IMF to recover to 4.5 percent in 2010 and to over 5 percent in 2011 (IMF 2010b).

VII. SUMMARY

This book has highlighted a number of reforms, both in the institutional framework for policy formulation and in policy orientation more generally, that emerging and developing economies can implement to protect macroeconomic stability and avoid crises that often have harmful effects not just on short-run macroeconomic well-being but also on long-run capacity growth. The protection of macroeconomic stability has two dimensions: avoiding having policies themselves become a source of instability as well as permitting policies to be deployed countercyclically to offset the effects of instability with nonpolicy sources. Partly in response to the multitude of crises that emerging economies endured during the decades of the 1980s and 1990s, a large number of these economies implemented broad-based policy reforms, covering fiscal policy, central-bank governance, monetary and exchange rate policy regimes, and policies toward the domestic financial system and the capital accounts of the balance of payments. Notably, these reforms were intended not just to avoid policy-induced dislocations but also to improve the trade-off between credibility and flexibility in policy formulation so that macroeconomic policies could be deployed in countercyclical fashion without fear of triggering destabilizing expectations of fiscal insolvency, renewed inflation, exchange rate devaluation, or banking crisis.

The Great Recession put this new policy environment in emerging and developing economies to the test. After a worldwide boom during the first part of the decade of the 2000s, the recession began with the collapse of a housing bubble in the United States in mid-2007 that eventually culminated in a severe financial crisis in that country in fall 2008. That crisis spread rapidly to western Europe and Japan through financial links and to emerging and developing economies through a variety of financial and real channels, resulting in the most severe international macroeconomic crisis since the Great Depression.

Under previous policy regimes, emerging economies would have been forced to respond to the large contractionary external shocks associated with the crisis by tightening fiscal and monetary policies to safeguard their perceived fiscal solvency, defend their exchange rates, maintain anti-inflationary credibility, and repair their banking systems. Such a response would have magnified the crisis internationally by reducing demand from emerging economies, which have represented an ever-growing share of total world demand. The Latin American debt crisis of the 1980s, triggered by the second most severe international recession since the Great Depression, the 1982 recession, provides a case in point. As we saw in Chapter 25, the pronounced vulnerability of heavily indebted countries, and their weak domestic policy environments, created a so-called lost decade of growth for the affected Latin American countries during the 1980s.

This time turned out to be different. Most emerging economies that had previously implemented macroeconomic reforms and had followed conservative fiscal and monetary policies during the worldwide boom that preceded the crisis, especially in Asia and Latin America, but also in some emerging economies in the Middle East and central Asia, as well in sub-Saharan Africa, managed to avoid such destructive policies. They had sufficient fiscal credibility not just to allow automatic fiscal stabilizers to work but actually to implement discretionary fiscal expansions. They had sufficient anti-inflationary credibility to reduce policy interest rates and allow their exchange rates to depreciate. They had reformed their domestic financial sectors to prevent their banking systems from constituting a source of vulnerability. As a consequence, not only have emerging economies outperformed more advanced economies during this crisis, but for the first time, they are helping to lead the world out of recession. We can only hope, for the sake of international development, that the lessons that these countries have learned from the Great Recession will spread and endure.

REFERENCES AND FURTHER READING

Corbacho, Ana, and Gerd Schwartz (2007), "Fiscal Responsibility Laws," in M. Kumar and T. Ter-Minassan, eds., *Promoting Fiscal Discipline*, Washington, DC: International Monetary Fund, pp. 58–105.

Craine, Roger, and Vance L. Martin (2009), "The Interest Rate Conundrum," *B.E. Journal of Macroeconomics*, vol. 1(9), Article 8. http://www.bepress.com/cgi/viewcontent.cgi?article=1819&context=bejm.

Hernandez, Leonardo, and Peter J. Montiel (2003), "Post-crisis Exchange Rate Policy in Five Asian Countries: Filling in the 'Hollow Middle'?" *Journal of the Japanese and International Economies*, vol. 17, pp. 336–369.

International Monetary Fund (2005), "Regional Economic Outlook: Middle East and Central Asia (September)." http://www.imf.org/external/pubs/ft/reo/2005/eng/meca0905.pdf.

International Monetary Fund (2006a), "Regional Economic Outlook: Asia and Pacific (May)." http://www.imf.org/external/pubs/ft/reo/2006/eng/01/areo0506.htm.

International Monetary Fund (2006b), "Regional Economic Outlook: Asia and Pacific (September)." http://www.imf.org/external/pubs/ft/reo/2006/ENG/02/areo0906.htm.

International Monetary Fund (2006c), "Regional Economic Outlook: Western Hemisphere (November)." http://www.imf.org/external/pubs/ft/reo/2006/ENG/02/wreo.htm.

International Monetary Fund (2007a), "Regional Economic Outlook: Asia and Pacific (April)." http://www.imf.org/external/pubs/cat/longres.cfm?sk=20264.0.

International Monetary Fund (2007b), "Regional Economic Outlook: Asia and Pacific (October)." http://www.imf.org/external/pubs/ft/reo/2007/APD/ENG/areo1007.htm.

International Monetary Fund (2007c), "Regional Economic Outlook Europe: Strengthening Financial Systems (November)." http://www.imf.org/external/pubs/ft/reo/2007/EUR/ENG/ereo1107.htm.

International Monetary Fund (2007d), "Regional Economic Outlook: Western Hemisphere (April)." http://www.imf.org/external/pubs/ft/reo/2007/WHD/ENG/wreo0407.htm.

International Monetary Fund (2007e), "Regional Economic Outlook: Middle East and Central Asia (October)." http://www.imf.org/external/pubs/ft/reo/2007/MCD/ENG/mreo1007.htm.

International Monetary Fund (2007f), "Regional Economic Outlook: Sub-Saharan Africa (October)." http://www.imf.org/external/pubs/ft/reo/2007/AFR/ENG/sreo1007.htm.

International Monetary Fund (2008a), "Regional Economic Outlook Europe: Reassessing Risks (April)." http://www.imf.org/external/pubs/ft/reo/2008/EUR/eng/ereo0408.htm.

International Monetary Fund (2008b), "Regional Economic Outlook: Western Hemisphere (April)." http://www.imf.org/external/pubs/ft/reo/2008/whd/ENG/wreo0408.htm.

International Monetary Fund (2008c), "Regional Economic Outlook: Western Hemisphere (October)." http://www.imf.org/external/pubs/ft/reo/2008/WHD/ENG/wreo1008.htm.

International Monetary Fund (2008d), "Regional Economic Outlook: Sub-Saharan Africa (April)." http://www.imf.org/external/pubs/ft/reo/2008/AFR/eng/sreo0408.htm.

International Monetary Fund (2008e), "Regional Economic Outlook: Sub-Saharan Africa (October)." http://www.imf.org/external/pubs/ft/reo/2008/AFR/eng/sreo1008.htm.

International Monetary Fund (2008f), "Regional Economic Outlook: Asia and the Pacific (April)." http://www.imf.org/external/pubs/ft/reo/2008/APD/ENG/areo0408.htm.

International Monetary Fund (2009a), "Regional Economic Outlook: Asia and the Pacific (May)." http://www.imf.org/external/pubs/ft/reo/2009/APD/eng/areo0509.htm.

International Monetary Fund (2009b), "Regional Economic Outlook: Asia and the Pacific (October)." http://www.imf.org/external/pubs/ft/reo/2009/APD/eng/areo1009.htm.

International Monetary Fund (2009c), "Regional Economic Outlook: Sub-Saharan Africa (April)." http://www.imf.org/external/pubs/ft/reo/2009/afr/eng/sreo0409.htm.

International Monetary Fund (2009d), "Regional Economic Outlook Europe: Addressing the Crisis (May)." http://www.imf.org/external/pubs/ft/reo/2009/EUR/eng/ereo0509.htm.

International Monetary Fund (2009e), "Regional Economic Outlook: Western Hemisphere (May)." http://www.imf.org/external/pubs/ft/reo/2009/WHD/eng/wreo0509.htm.

International Monetary Fund (2009f), "Regional Economic Outlook: Middle East and Central Asia (October)." http://www.imf.org/external/pubs/ft/reo/2009/MCD/eng/mreo1009.htm.

International Monetary Fund (2009g), "Regional Economic Outlook: Sub-Saharan Africa (October)." http://www.imf.org/external/pubs/ft/reo/2009/AFR/eng/sreo1009.htm.

International Monetary Fund (2010a), "Regional Economic Outlook: Asia and the Pacific (April)." http://www.imf.org/external/pubs/ft/reo/2010/APD/eng/areo0410.htm.

International Monetary Fund (2010b), "Regional Economic Outlook: Sub-Saharan Africa (April)." http://www.imf.org/external/pubs/ft/reo/2010/AFR/eng/sreo0410.htm.

Izquierdo, A., and E. Talvi (2009), "Policy Trade-offs for Unprecedented Times: Confronting the Global Crisis in Latin America and the Caribbean," paper prepared for the annual meeting of the Board of Governors of the Inter-American Development Bank, March.

Krugman, Paul (2009), *The Return of Depression Economics and the Crisis of 2008*, New York: W. W. Norton.

Osterholm, Par, and Jeromin Zettelmeyer (2007), "The Effects of External Conditions on Growth in Latin America," working paper WP/07, International Monetary Fund.

Stiglitz, Joseph (2009), *Freefall: America, Free Markets, and the Sinking of the World Economy*, New York: W. W. Norton.

Index