Journal of Economic Growth, 8, 5–46, 2003

The United States as a Coastal Nation*

JORDAN RAPPAPORT†

Federal Reserve Bank of Kansas City, Kansas City, MO 64198

JEFFREY D. SACHS

The Earth Institute, Columbia University, New York, NY 10027

US economic activity is overwhelmingly concentrated at its ocean and Great Lakes coasts, reflecting a large contribution from coastal proximity to productivity and quality of life. Extensively controlling for correlated natural attributes and initial conditions decisively rejects that the coastal concentration of economic activity is spurious or just derives from historical forces long since dissipated. Measuring proximity based on coastal attributes that contribute to either productivity or quality of life, but not to both, suggests that the coastal concentration derives primarily from a productivity effect but also, increasingly, from a quality of life effect.

Keywords: economic growth, population density, productivity, quality of life

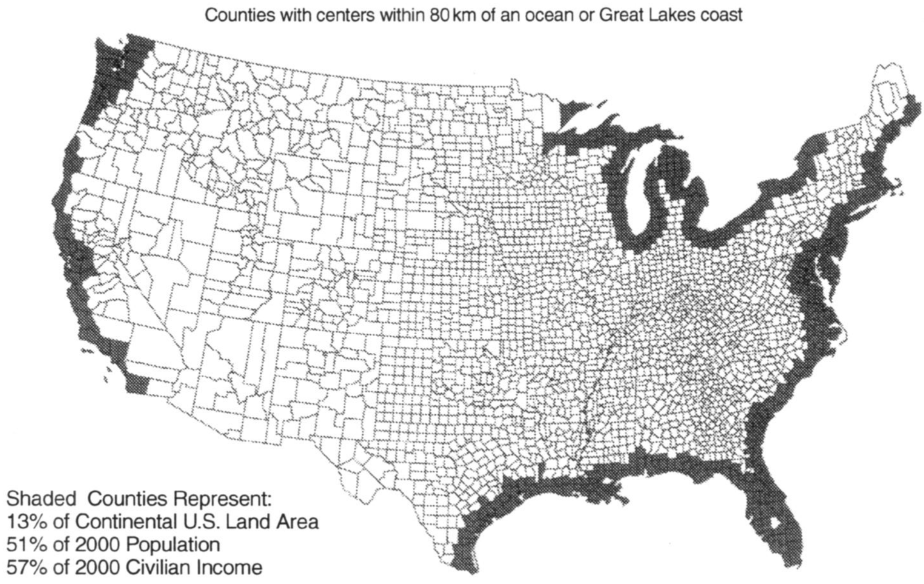
JEL classification: O40, O51, R11, R12

1. Introduction: Geography Matters

An abundance of rich, fertile land and an open frontier uniquely characterize US economic development. Less widely recognized is the extent to which the United States is and has always been a primarily coastal country. Consider Map 1: the shaded area represents the 559 counties with centers within 80 km of an ocean or Great Lakes coast. Collectively, these counties account for just 13 percent of the continental US land area but 51 percent of 2000 population and 57 percent of 2000 civilian income. Put differently, income per square kilometer of these coastal counties is more than eight times that of the remaining inland counties.

That the United States with its abundant land remains a primarily coastal nation underscores a basic economic fact: geography matters. In the search to understand the underlying determinants of growth and prosperity, economists have examined a myriad of

* Thank you for advice and suggestions to Patty Beeson, Jerry Carlino, David Cutler, Steven Durlauf, David DeJong, Claudia Goldin, Edward Glaeser, two anonymous referees, and seminar participants at Emory University, the University of Pittsburgh, the Federal Reserve Bank of Philadelphia, the 2000 Federal Reserve System Regional Conference, the 2001 Missouri Economics Conference, and the 2001 NBER Summer Workshop. Thank you to Michael Haines for sharing historical census data. A number of individuals have provided excellent research assistance including Nathaniel Baum Snow, Scott Benolkin, Anne Berry, Krista Jacobs, Jason Martinek, Peter Northup, Aarti Singh, Chris Yenkey, and Andrea Zanter. The views expressed herein are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of Kansas City or the Federal Reserve System.
† Corresponding author. Federal Reserve Bank of Kansas City, 925 Grand Blvd, Kansas City, MO 64111.



Map 1. Coastal counties.

country attributes ranging from the self evident (e.g., education) to the controversial (e.g., culture). But the role of geography, for the most part, has been neglected.

From a theoretical perspective, modern growth models focus on the accumulation of physical, human, and technological capital, which individually or together complement raw labor as the main factors of production. Land, when included, tends to serve as the intensive factor in a traditional sector away from which labor in shifting (Lewis, 1954; Jorgenson, 1961; Ranis and Fei, 1961; Harris and Todaro, 1970; Dixit, 1973; Drazen and Eckstein, 1988). More recently, theory has begun to grapple with the issue of space: increasing returns to scale in production, whether direct or via spillovers in technology and human capital, imply a spatial concentration of industry location (Henderson, 1988; Krugman, 1991; Fujita et al., 1999). While both approaches yield insights, neither addresses the constraints physical geography may place upon economic growth.

This was not always so. Adam Smith in Book 1 of *The Wealth of Nations* observes the importance of access to navigable water as an input to the development process:

> As by means of water carriage a more extensive market is opened to every sort of industry than what land carriage alone can afford it, so it is upon the sea-coast, and along the banks of navigable rivers that industry of every kind begins to sub-divide and improve itself, and it is frequently not till a long time after that those improvements extend themselves to the inland part of the country.

Thus, Smith laments the difficult preconditions for economic growth facing inland Africa and large parts of Russia, Siberia, and Central Asia.

> All the inland parts of Africa, and all that part of Asia which lies any considerable way north of the Black and Caspian Seas, the ancient Scythia, the modern Tartary and Siberia, seem in all ages of the world to have been in the same [economically undeveloped] state in which we find them at present... There are in Africa none of those great inlets ... to carry maritime trade into the interior parts of that great continent.

Smith's observation on the role of access to navigable water still holds in the late twentieth century. Recent cross-country empirical research affirms that the level and growth rate of per capita income continue to be strongly positively correlated with coastal proximity (Gallup and Sachs, 1998; Rappaport, 2000b). The observation also implicitly underscores the incredibly favorable economic geography enjoyed by the nations of Western Europe. Extensive ocean shorelines host a succession of natural harbors and numerous navigable rivers penetrate deep into the interior.[1]

The United States, however, stands out as a possible exception to the importance of access to navigable water in fostering growth. While the United States also enjoys long ocean shorelines and an extensive inland river network, its continental scale nevertheless implies that most of the United States land mass lies considerably far from navigable water. Rather than from coastal proximity, an argument can be made that the United States' prosperity derives from its natural resource abundance; indeed its land-based wealth is the stuff of American mythology. Consistent with such an argument, Wright (1990) shows that during the period when the United States moved into a position of world industrial preeminence, the factor content of its net exports was growing increasingly intensive in natural resources.

But such US exceptionalism is misleading. In fact, the US economic activity is overwhelmingly concentrated near its ocean and Great Lakes coasts. Moreover, this concentration has been increasing over the twentieth century.

Herein we argue that the concentration of US economic activity at its ocean and Great Lakes coasts reflects a large present-day contribution from coastal proximity to productivity and quality-of-life. Extensively controlling for correlated natural attributes and initial conditions, linear regressions decisively reject that the positive correlation of economic activity with coastal proximity is spurious or derives from historical forces long since dissipated. Measuring proximity based on coastal attributes that contribute to productivity or to quality of life but not to both suggests that the coastal concentration of economic activity derives primarily from a productivity effect but that the coastal contribution to quality of life is becoming increasingly important.

The paper proceeds as follows: Section 2 discusses the theoretical basis for using economic density as a measure of underlying productivity and quality of life and reviews related empirical literature. Section 3 presents some simple empirics illustrating the continual increase in the coastal concentration of US economic activity since the late nineteenth century. Section 4 lays out our econometric specification. Section 5 presents our results supporting coastal proximity's continuing contribution to productivity and increasing contribution to quality of life. A last section concludes.

2. Theory and Background

For comparing economic outcomes across countries, per capita income serves as a natural measure of welfare; but for comparing economic outcomes across local areas among which individuals and firms can easily move, it does not. A high level of per capita income may reflect high underlying productivity; but it may also reflect compensation for undesirable quality of life such as unpleasant weather or pollution. Hence it is not clear whether high per capita income represents good or bad underlying fundamentals.

Rather than per capita income, we use population density as a more natural measure for capturing underlying variations in local productivity and quality of life (Haurin, 1980; Glaeser et al., 1992, 1995; Ciccone and Hall, 1996). As will be shown below, alternatively using employment density as the dependent variable in our regressions effects identical results.

Define a "locality" as a geographic area where people both live and work. Consider a locality with a set of attributes that increase the productivity of resident firms. In addition to access to navigable water, some productivity-enhancing attributes might include abundant natural resources, temperate weather, and rule of law. Firms' high productivity increases the marginal revenue product of both labor and capital, in turn inducing an inflow of each; moreover, the complementarity between labor and capital implies these inflows are mutually reinforcing (Figure 1, panel (a)). In a long-run steady state, high productivity implies high population density, dL/dproductivity > 0.

Similarly, consider a locality with a set of attributes that directly increase the quality of

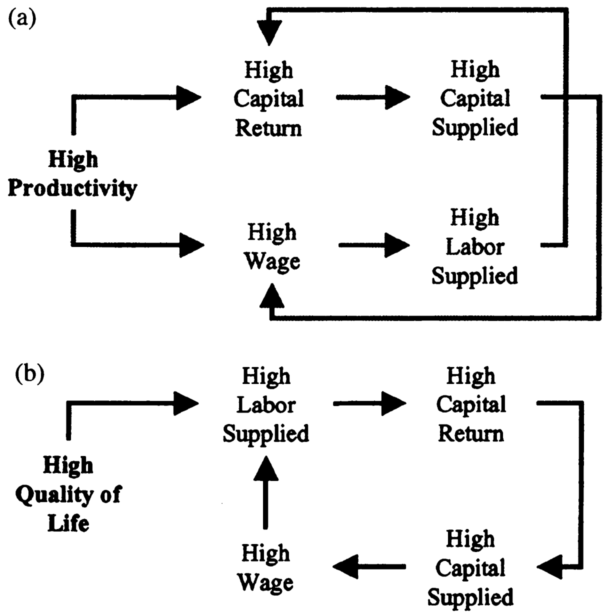


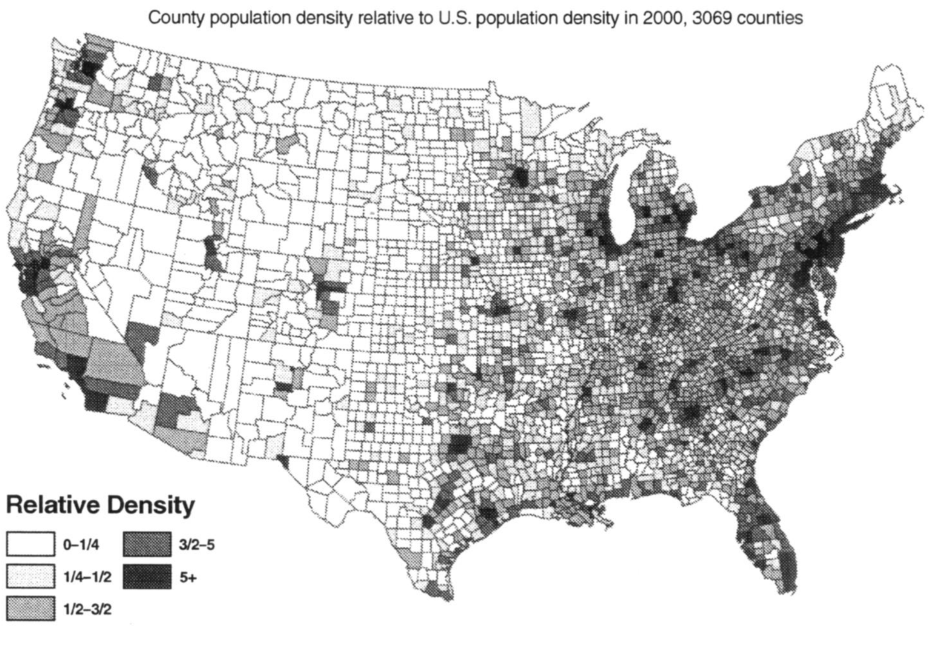
Figure 1. Population density as a measure of underlying productivity and quality of life. (a) High productivity increases population and employment. (b) High quality of life increases population and employment.

life of local residents. Some quality-of-life-enhancing attributes might include ocean vistas, pleasant weather, and low crime. The high quality of life induces an inflow of labor which in turn induces an inflow of capital (Figure 1, panel (b)). In a long-run steady state, dL/dquality-of-life > 0. Formal proofs of these are deferred to Appendix A. But the intuition should be straightforward.[2]

Consistent with the idea that people vote with their feet (Tiebout, 1956), population density reveals individuals' preferences over local areas by aggregating the indirect contribution to utility via productivity-driven higher wages with the direct contribution to utility via high quality of life. Map 2, which shows the relative population density of US counties in 2000, represents the result of such a vote: the higher the population density, the greater the productivity and quality-of-life benefits from underlying local attributes.[3]

Inherent in population density's aggregating over productivity and quality of life is that it cannot distinguish between the two. In a certain sense, the distinction does not matter. For a representative agent seeking to choose the ideal local attribute mix, population density already correctly weights productivity versus quality of life. But from the perspective of economic development, a positive income elasticity of demand implies that individuals living in the United States will tend to place a relatively high value on quality of life. The coastal concentration of US economic activity holds lessons for poorer nations only if it reflects—at least in part—high underlying productivity.

Of course, there is little doubt that living near a coast also may increase quality of life. In addition to overwhelming anecdotal evidence, the quality-of-life contribution from coastal proximity is persuasively established by the compensating differential empirical literature,



Map 2. Relative population density.

which values it by the sum of the lower wages individuals are willing to accept and the higher housing prices they are willing to pay to live in coastal areas (Rosen, 1979; Roback, 1982). Controlling for worker-specific and house-specific characteristics, Blomquist et al. (1988) estimate that location adjacent to an ocean or Great Lakes coast lowers annual incomes by $155 and raises the annual price of housing by $716; Gyourko and Tracy (1991) estimate these at $874 and $1,201 respectively (all values are in 1999 dollars). Hence the contribution from coastal location to an average working individual's quality of life would be valued in the range of $871 to $2,075 per year. Similarly, Stover and Leven (1992) value the coastal contribution to quality of life at $721 per year.

The compensating differential approach proves less useful in valuing coastal proximity's contribution to productivity. In theory, this contribution could be valued by the sum of the higher wages and nontradable input prices firms are willing to pay to operate near a coast. In practice, researchers have not had access to plant and office level data with sufficient detail and geographic scope to do so.

Equally problematic, anonymity requirements limit the geographic identification of microdata to local areas with high populations. The Blomquist et al. (1988) and Gyourko and Tracy (1991) results are based on respective cross-sections of 253 US urban counties and 130 US metropolitan areas. Because sparsely settled local areas represent non-random economic outcomes, such a sample selection strongly biases estimated valuations of geographic attributes.

In what follows, we primarily focus on the combined productivity and quality-of-life effects of coastal proximity as measured by population density. In our regression analysis, however, we also attempt to disentangle coastal proximity's contribution to productivity versus its contribution to quality of life. Specifically, we examine partial correlations with coastal proximity measures that we believe a priori influence productivity or quality of life but not both. Doing so suggests that the larger part of the coastal concentration of US population derives from a productivity effect; but increasingly, the coastal concentration is also being underpinned by a quality-of-life effect.

A second limitation of using population density to measure economic outcomes is the difficulty in distinguishing between present-day contributions to productivity and quality of life versus historical contributions to these. For example, coastal proximity may have greatly increased productivity and quality of life through the end of the nineteenth century, after which it affected neither. To the extent that adjustment towards the resulting new steady-state spatial distribution is slow, high population density near coasts circa 1900 would remain for a long time thereafter. Alternatively, historical contributions to productivity and quality of life could effect high present-day population density via economies of scale. As Fujita and Mori (1996) argue, "[port cities] should have disappeared a long time ago when the original advantage (of cheap water access) became unimportant. Clearly, their continued prosperity can be explained only when we consider the 'lock-in effect' of some self-reinforcing agglomeration forces." In other words, high historical population density near coasts subsequently contributed to high productivity and quality of life causing steady-state population density to be path dependent.

While delayed adjustment and steady-state path dependence are surely contributing forces, empirically they fall far short of being able to account for the coastal concentration of US economic activity. As documented below, such concentration has been increasing since the late nineteenth century. More definitively, quantitatively-large, statistically significant positive partial correlations between changes in population density and coastal proximity continue to hold even after extensively controlling for initial conditions. As a result, we are able to strongly reject that the present-day coastal concentration of US economic activity primarily reflects historical forces long since dissipated.

3. The Coastal Concentration of US Population, Historical and Present

Given the maritime nature of the European colonization of North America, the high coastal concentration of economic activity is hardly new. The establishment of settlements at locations affording easy access to ocean transport allowed for the communication and trade on which the Atlantic economy flourished. In the middle and deep South, tobacco and cotton plantations spread up from the ocean coast along the navigable rivers down which they could transport their goods. Indeed, the importance of access to navigable water is underscored by the huge public and private investments during the 1820s and 1830s for the construction of canals which in turn facilitated the westward spread of trade and industry (Tanner, 1840; Poor, 1860; Fogel, 1964). The resulting contribution to productivity is documented by Sokoloff (1988) who shows that US inventive activity over the period 1790 to 1846, as measured by patents per capita, is strongly positively correlated with proximity to navigable waterways.

Figure 2 picks up the story in 1880. At that time, the collective population density of all counties with centers within 80 km of an ocean coast is 2.4 times that of the contemporary continental United States (i.e., total population divided by total land for such counties relative to total population divided by total land for all continental US counties). The collective population density of all counties with centers within 80 km of a Great Lakes coast and those remaining counties with centers within 40 km of a river on which there was commercial navigation in 1968 are each 2.6 times that of the contemporary continental United States.[4]

Figure 2 (panel (a)) shows the subsequent rapid increase in relative population density for the ocean and Great Lakes coast counties. The ocean coast relative population density grows steadily from 1880, surpassing the Great Lakes coast relative population density in 1950, and rising to 4.2 in 1990 and 2000. The Great Lakes coast relative population density rises to 3.5 in 1930 where it approximately remains through 1970 before falling rapidly to 2.8 in 2000. On the other hand, navigable rivers relative population density steadily declines from 1880; by 2000 it has fallen to 1.5.

One explanation for the rising concentration of population near coasts is that it is picking up the effects of correlated attributes such as temperate weather and the shift in employment out of agriculture. As a first pass at addressing such potential explanations, Figure 2 (panel (b)) shows the partial correlation coefficients of population density regressed on categorical dummies for counties with centers within 80 km of an ocean or Great Lakes coast, or within 40 km of a river on which there was commercial navigation in 1968.[5] The regressions include 12 weather variables (January minimum temperature, July maximum heat index, mean annual precipitation, mean annual days with precipitation of at least 0.1 inch, mean annual days temperature falls below 32 °F, mean annual days

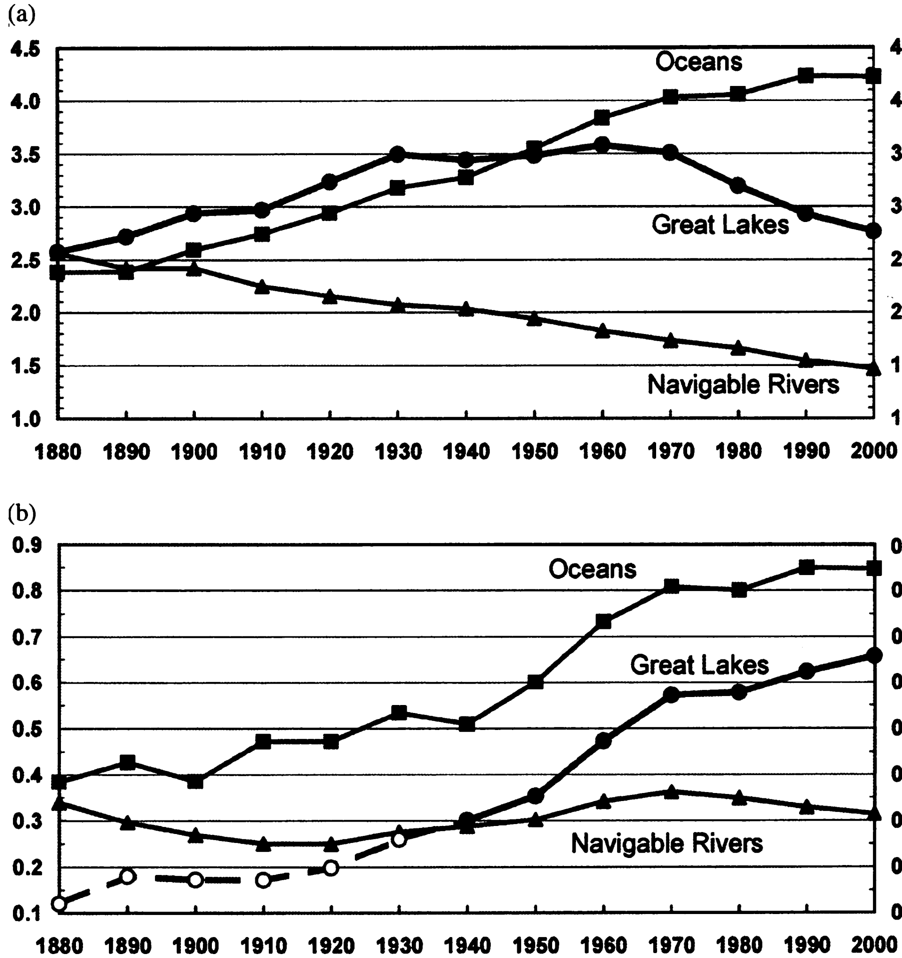


Figure 2. The ocean and Great Lakes categories are made up of counties with centers within 80 km of the respective coast; the navigable river category is made up of counties with centers within 40 km of a river on which there was commercial navigation in 1968. Panel (a) shows the aggregate population density of each of the categories relative to that of the continental United States in the same year. For Panel (a), counties that were included in the ocean or Great Lakes categories were excluded from the navigable river one. Panel (b) reports coefficients on category dummy variables from regressing log(1 + population density) on these along with weather and topography variables as enumerated in the text. Open points (connected by dashed lines) represent coefficients not significant at the 0.05 level (using standard errors robust to spatial correlation as described in the text).

temperature rises above 90 °F—each of which is entered linearly and quadratically) and two topography variables (the standard deviation of within-county elevation divided by county land area, entered linearly and quadratically).

Controlling for weather and topography greatly reduces the higher relative population density attributable to coastal proximity. In the 1880 regression, the positive, statistically significant coefficient on the ocean coast dummy implies that controlling for weather and topography, counties with centers within 80 km of an ocean coast have expected 1880 population density 1.5 times that of more inland counties. The coefficient on the ocean coast dummy remains approximately constant through 1900, rises steadily from 1900 to 1930, rises steadily again from 1940 to 1970, and thereafter rises slightly more from 1980 to 1990. Its 2000 value implies that the ocean coastal counties have expected population density 2.3 times that of more inland counties. Beeson, DeJong, and Troesken (2001) find a similar increasing partial correlation between ocean proximity and population density over the period 1840–1990.

The 1880 regression also admits a positive, statistically significant coefficient on the navigable river dummy, implying that counties with centers within 40 km of a navigable river have expected 1880 population density 1.4 times that of non-navigable river counties. The navigable river coefficient falls gradually through 1920, then rises gradually through 1970, and then again falls gradually through 2000. Its 2000 value implies that such counties have expected population density 1.4 times that of remaining counties, the same as in 1880.[6]

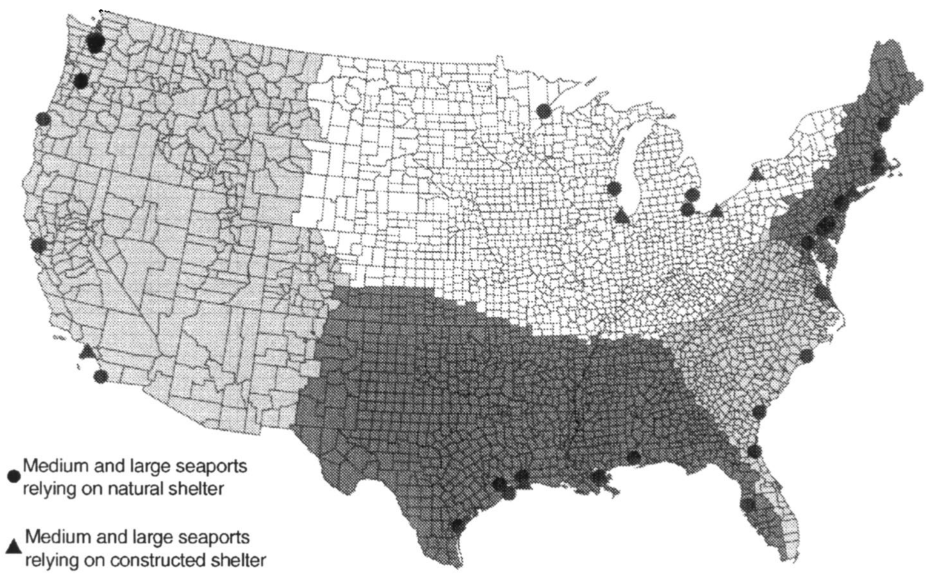
The coefficient on the Great Lakes dummy from the 1880 regression is close to zero and not statistically significant. But the Great Lakes coefficient steadily rises throughout the century, first statistically differing from zero in 1940 and attaining a value in 2000 implying the Great Lakes coastal counties have expected population density 1.9 times that of remaining counties.

The patterns of the partial correlation coefficients shown in Figure 2 (panel (b)) differ somewhat from those of the relative population densities shown in Figure 2 (panel (a)). For instance, despite having the highest relative population density in 1880, the corresponding Great Lakes coast partial correlation coefficient does not statistically differ from zero. In other words, favorable weather and topographical conditions largely account for the Great Lakes' high population density in 1880. Similarly, the 1920–1970 fall in relative population density near navigable rivers and the 1960–2000 fall in relative population density near the Great Lakes coasts contrast with rising coefficients on the corresponding dummies. Again, unfavorable weather and topographical conditions—that is, changing tastes and technology with respect to the navigable river and Great Lakes counties' weather and topography—largely account for the relative population density decline, which was actually slowed by proximity to navigable rivers and the Great Lakes. Most likely, the rising coefficients reflect delayed adjustment and agglomerative effects following historical productivity and quality-of-life contributions from navigable river and Great Lakes proximity. Such an interpretation emphasizes the need to extensively control for initial conditions in our regression analysis below (see also note 7).

Figure 3 breaks out the concentration of population density near oceans for each of four different coastal segments: the North Atlantic coast (Maryland north to Maine), the South Atlantic coast (Virginia south to Florida), the Gulf of Mexico Coast, and the Pacific Coast. Justifying a three-way split (Atlantic, Gulf, Pacific), for instance, are different trade opportunities offered by the varying locations. The North Atlantic versus South Atlantic split, on the other hand, follows from empirically observed differences between the two.

Figure 3 (panel (a)) emphasizes the extremely high population density of counties with centers within 80 km of the North Atlantic coast. The ratio of these counties' population density relative to continental US population density is 10.2 in 1880, rising to 12.1 in 1930, thereafter falling gradually to 11.7 in 1970, and then falling more steeply to 9.6 in 2000. The relative population density of counties with centers within 80 km of the South Atlantic coast is 1.3 in 1880, a level at which it roughly remains through 1940, thereafter rising rapidly to 2.8 in 2000. Relative population density of counties with centers within 80 km of Gulf of Mexico coast starts from 0.6 in 1880 slowly growing to 1.0 in 1940, thereafter growing more rapidly to 2.1 in 2000. Counties with centers within 80 km of the Pacific coast grow rapidly from relative population density 0.5 in 1880 to 4.4 in both 1990 and 2000.

Figure 3 (panel (b)) shows the partial correlation coefficients of population density with categorical dummy variables for location within 80 km of the four ocean coasts after controlling for Great Lakes and navigable river proximity, weather, and topography, along with region fixed effects for the ocean or Great Lakes coast to which a county is closest (i.e., the five regions shown in Map 3). Such controls dramatically reduce the higher relative population density attributable to North Atlantic coastal proximity. As with Figure 2, high relative population density along the North Atlantic coast is partly attributable to favorable weather and topography. In addition, coefficients on the region-specific dummies show that much of it also is attributable to the late-nineteenth-century high relative population density of the North Atlantic region as a whole. Compared to the actual 1880 North Atlantic coast relative population density of 10.2, a positive, statistically significant coefficient implies that the North Atlantic coastal counties have expected population density 1.4 times that of the remaining counties for which the North Atlantic is the closest coast. This coefficient rises steadily through 1970, remains constant from 1970



Map 3. Counties and harbors by nearest coast.

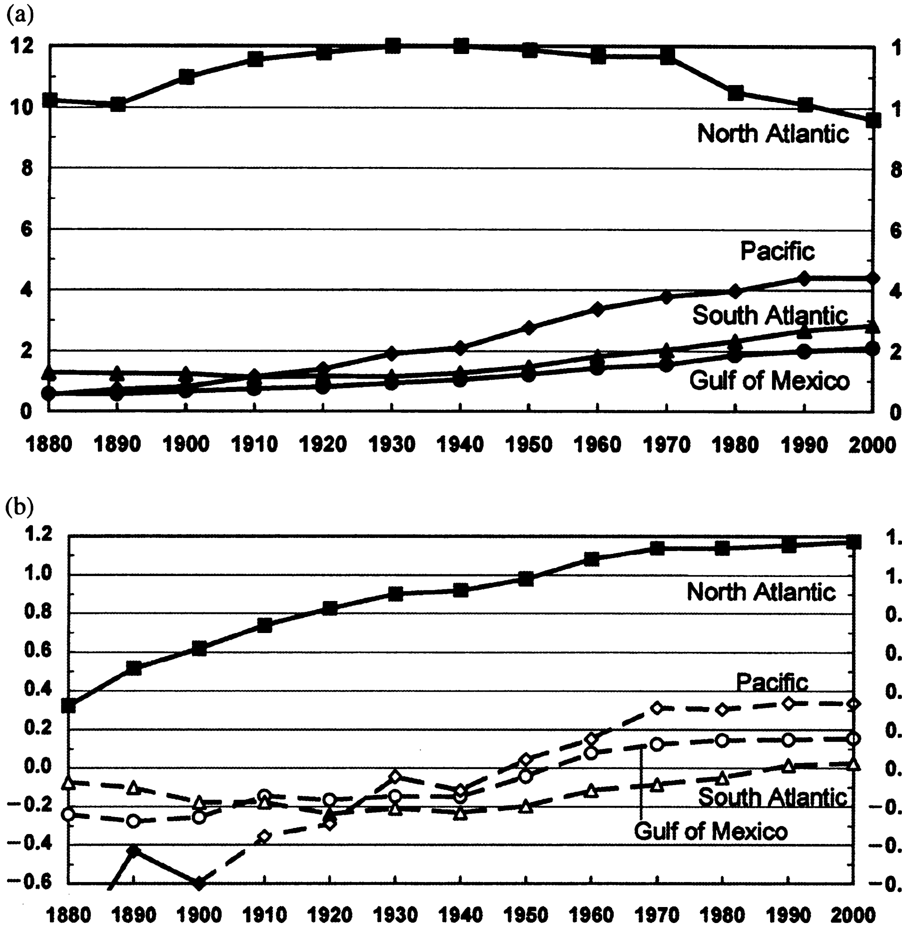


Figure 3. Categories are made up of those counties with centers within 80 km of the respective coast. South Atlantic is composed of Atlantic coastal counties closer to the Virginia coast extending south. North Atlantic is composed of Atlantic coastal counties closer to the Maryland coast extending north. Panel (a) shows the aggregate population density of each of the categories relative to that of the continental United States in the same year. Panel (b) reports coefficients from regressing log(1 + population density) on coast dummy variables controlling for Great Lakes and 1890 navigable river proximity, weather and topography as enumerated in the text, along with closest-coast-specific intercepts. Open points (connected by dashed lines) represent coefficients not significant at the 0.05 level (using standard errors robust to spatial correlation as described in the text).

to 1980, and then again rises slowly from 1980 to 2000. Its ending value implies that the North Atlantic coastal counties have expected 2000 population density 3.2 times that of the remaining counties for which the North Atlantic is the closest coast.

The coefficients on the South Atlantic, Gulf of Mexico, and Pacific coast dummies are all negative in the 1880 regression. The South Atlantic coefficient remains negative in all decades, but it statistically differs from zero only in 1920. The Gulf of Mexico steadily increases starting in 1940 and becomes positive in 1960; but it never statistically differs from zero. A large negative, statistically significant coefficient on the Pacific coast dummy in the 1880 regression implies that counties along the Pacific coast have expected 1880 population density 0.4 times that of other counties for which the Pacific is the closest coast. The Pacific coast coefficient rapidly increases starting in 1900; it no longer statistically differs from zero in 1910; and it turns positive in 1950. While not statistically significant, the Pacific coast coefficient in the 2000 regression implies that counties along the Pacific coast have expected population density 1.4 times that of other counties for which the Pacific is the closest coast.

The upward trends in the time series of coefficients on each of the ocean and Great Lakes coast dummies strongly suggest a contribution from coastal proximity to productivity and quality of life that is increasing over time. More definitively establishing such an increasing contribution is a main result of our regression analysis below. First, however, we briefly present some summary empirics characterizing differences between coastal and inland counties.

Table 1 recaps the very high coastal density of economic activity shown in Map 1 and Figures 2 and 3. Four alternative measures of economic density—population, employment, labor income, and capital income—all effect a similar ranking of the coastal categories: the North Atlantic counties followed by (in decreasing order) the Pacific counties, the Great Lakes or South Atlantic counties (depending on measure), and the Gulf of Mexico counties. Unsurprisingly, personal income levels are higher at the coast than inland. For all the coastal counties, per worker annual labor income in 2000 averaged \$41,070 versus \$31,494 for the remaining inland counties. Such higher income levels may have derived either from a coastal contribution to productivity or from the higher skills of coastal workers. In 2000, 27.7 percent of adults in the coastal counties had at least a Bachelor's degree and 10.5 percent had a graduate degree versus just 21.6 percent and 7.5 percent of adults in inland counties. Coastal counties also disproportionately attract immigrants: 15.9 percent of their 2000 population are not native US citizens (25.4 percent for the Pacific coastal counties) versus just 6.0 percent of the inland population. Finally, the industrial compositions of the North Atlantic, South Atlantic and Pacific coastal counties are disproportionately skewed toward services and finance; those of the Gulf of Mexico and inland counties are disproportionately skewed towards natural resources; and that of the Great Lakes coastal counties is disproportionately skewed towards manufacturing.

4. Econometric Specification

Based on the theory sketched in Section 2 and formalized in an appendix, we assume a data generating process for steady-state economic density in locality i,

$$\underset{\substack{\Uparrow\\ \text{log}\\ \text{steady-}\\ \text{state}\\ \text{density}}}{L^*_{i,t}} = \underset{\substack{\Uparrow\\ \text{time}\\ \text{invariant}\\ \text{included}\\ \text{attrib.}}}{\boldsymbol{\beta}_t(\mathbf{x}_i)} + \underset{\substack{\Uparrow\\ \text{time}\\ \text{intrcpt.}}}{\mu_t} + \underset{\substack{\Uparrow\\ \text{time}\\ \text{invariant}\\ \text{excluded}\\ \text{attrib.}}}{\nu_i} + \underset{\substack{\Uparrow\\ \text{idiosync.}\\ \text{attrib.}}}{\xi_{i,t}} \tag{1a}$$

Table 1. Coastal versus inland counties in 2000.

		Coastal Counties (Center within 80 km of Ocean or Great Lakes Coast)						Inland Counties (Non-Coastal Counties)		
	All Continental Counties	All Coastal Counties	North Atlantic	South Atlantic	Gulf of Mexico	Pacific	Great Lakes	All Inland Counties	Within 40 km of Navigable River	Remaining Inland Counties
Number of counties	3,069	559	101	122	111	55	170	2,510	443	2,067
Percent of continental US (%)										
Land area	100.0	13.3	1.6	2.2	3.0	2.6	4.0	86.7	7.8	78.9
Population	100.0	50.6	15.7	6.2	6.2	11.6	10.9	49.4	11.5	38.0
Civilian employment	100.0	50.7	15.9	6.0	5.9	11.9	11.0	49.5	12.0	37.4
Civilian labor income	100.0	57.3	20.4	5.6	5.5	14.3	11.4	42.7	11.0	31.8
Capital income	100.0	56.7	18.8	7.4	5.9	13.2	11.3	43.3	11.1	32.3
Density relative to continental US										
Population	1.00	3.79	9.62	2.85	2.10	4.41	2.76	0.57	1.47	0.48
Civilian employment	1.00	3.80	9.73	2.76	1.98	4.54	2.78	0.57	1.54	0.47
Civilian labor income	1.00	4.29	12.53	2.59	1.86	5.44	2.88	0.49	1.41	0.40
Capital income	1.00	4.25	11.56	3.36	2.00	5.06	2.86	0.50	1.42	0.41
Income										
Per worker labor income	$36,343	$41,070	$46,820	$34,046	$34,074	$43,567	$37,658	$31,494	$33,269	$30,923
Per person labor and capital income	$26,752	$30,216	$34,321	$25,809	$24,040	$32,531	$27,889	$23,209	$25,760	$22,439
Age, education, and immigrant status (%)										
Age 0–17	25.6	25.4	24.7	24.8	25.9	26.1	26.1	25.8	25.1	26.0
Age 18–64	62.0	62.3	62.6	62.8	60.6	63.2	61.4	61.7	61.9	61.7
Age 65 and older	12.4	12.3	12.7	12.4	13.5	10.6	12.5	12.5	13.0	12.3
Bachelors degree or higher (persons 25+)	24.8	27.7	30.7	29.1	21.5	29.1	24.0	21.6	22.8	21.3
Graduate or professional degree (persons 25+)	9.1	10.5	12.6	11.2	7.5	10.4	8.8	7.5	8.0	7.3
Not a native US citizen	11.1	15.9	16.7	15.1	10.8	25.4	8.1	6.0	3.5	6.7
Immigrated to US in previous 15 years	6.4	8.9	9.5	8.4	6.1	14.2	4.5	3.7	2.2	4.1
Civilian employed by industry (%)										
Natural resources	3.3	2.0	1.0	1.8	4.4	2.7	1.8	4.6	3.1	5.0
Manufacturing	11.5	10.7	9.2	6.5	7.4	11.2	16.6	12.4	12.3	12.4
Finance, insurance, real estate	8.1	8.9	10.2	8.3	7.7	8.9	7.8	7.3	7.5	7.2
Services	32.0	34.8	37.3	35.7	31.6	35.3	31.7	29.2	31.0	28.7

Notes: Navigable rivers are rivers on which there was commercial navigation in 1968. For further information, see data appendix.

The vector $\mathbf{x}_i$ includes measures of locality i's coastal proximity along with measures of correlated geography attributes such as weather and topography. The exogenous nature of these attributes eliminates a reverse-causal interpretation of partial correlations.

As the effects of coastal proximity may change with tastes and technology, steady-state economic density is assumed to be a time-varying function, $\boldsymbol{\beta}_t(\cdot)$, of the time-invariant $\mathbf{x}_i$.

Steady-state economic density is modeled as additionally depending on a time intercept term, μ_t; non-modeled time-invariant attributes, ν_i; and time-varrying idiosyncratic attributes, $\xi_{i,t}$.

In practice, steady-state economic density is not observable. To proxy for it, we use current economic density. Hence we estimate

$$\begin{aligned} L_{i,t} &= \mathbf{x}_i'\boldsymbol{\beta}_t + \mu_t + \underbrace{\nu_i + \xi_{i,t} + (L_{i,t} - L_{i,t}^*)}_{\text{error term}} \\ &= \mathbf{x}_i'\boldsymbol{\beta}_t + \mu_t + \varepsilon_{i,t}. \end{aligned} \tag{1b}$$

The difference between steady-state and current economic density is subsumed in the error term. This difference is likely to be non-trivial. Observed persistent US local population and employment flows suggest that economic density adjusts only very slowly towards its steady-state level (Rappaport, 2000a). With slow adjustment and the time-varying dependence of steady-state economic density on the exogenous attributes, $\mathbf{x}_i$, the latter will in general be correlated with $(L_{i,t} - L_{i,t}^*)$. Therefore, $\widehat{\boldsymbol{\beta}}_t$ estimated using (1b) will be biased. Intuitively, $\widehat{\boldsymbol{\beta}}_t$ captures a combination of the past and current dependence of economic density on $\mathbf{x}_i$.

So a possible interpretation of non-zero estimated coefficients on the $\mathbf{x}_i$ in (1b) is that rather than measuring a current functional relationship, $\partial L_{i,t}^*/\partial \mathrm{x}_i$, they instead capture a past functional dependence that no longer holds in the present. For instance, a particular attribute, x_i^k, may have had a positive impact on steady-state economic density at some time in the past, $\beta_{t-1}^k > 0$, but not in the present, $\beta_t^k = 0$. If so, the change regression,

$$dL_{i,t} = \mathbf{x}_i' d\boldsymbol{\beta}_t + d\mu_t + d\varepsilon_{i,t}, \tag{2}$$

might be expected to estimate $\widehat{d\beta_t^k} < 0$. Note, however, that the right-hand side variables in (2) may still be correlated with the error term, in which case $\widehat{d\beta_t^k}$ will be a biased estimator of the change in the structural parameter, $d\beta_t^k$. Such a correlation would arise, for instance, from the asymmetric movement of population density towards its steady-state level as documented in Glaeser and Gyourko (2001).[7]

Even more difficult is distinguishing between (1a) and an alternative data generating process characterized by steady-state path dependence,

$$\begin{array}{ccccccccccc} L_{i,t}^* & = \boldsymbol{\Gamma}_t(& L_{i,t-1}, & \mathbf{x}_i) & + & \mu_t & + & \nu_i & + & \xi_{i,t} \\ \Uparrow & & \Uparrow & \Uparrow & & \Uparrow & & \Uparrow & & \Uparrow \\ \text{log} & & \text{lagged} & \text{time} & & \text{time} & & \text{time} & & \text{idiosync.} \\ \text{steady-} & & \text{density} & \text{invar.} & & \text{intrcpt.} & & \text{invar.} & & \text{attrib.} \\ \text{state} & & & \text{incld.} & & & & \text{excld.} & & \\ \text{density} & & & \text{attrib.} & & & & \text{attrib.} & & \end{array} \tag{3a}$$

In practice, this is usually estimated by,

$$\begin{aligned} L_{i,t} &= \gamma_t L_{i,t-1} + \mathbf{x}_i'\boldsymbol{\delta}_t + \mu_t + \underbrace{\nu_i + \xi_{i,t} + (L_{i,t} - L_{i,t}^*)}_{\text{error term}} \\ &= \gamma_t L_{i,t-1} + \mathbf{x}_i'\boldsymbol{\delta}_t + \mu_t + \varepsilon_{i,t}. \end{aligned} \tag{3b}$$

The specification (3b) is structurally equivalent to the change regression (2) with the addition of initial density as a right-hand side variable. Put differently, (2) constrains the coefficient, γ, on lagged density in (3b) to be one.

The interpretation of the estimated $\widehat{\gamma}$ is problematic. Suppose that (1a) is the true data generating process. The slow adjustment of economic density towards its steady state implies that running (3b) should estimate $\widehat{\gamma} > 0$ despite that $\partial L^*_{i,t}/\partial L_{i,t-1} = 0$. Reinforcing this upward bias is that $L_{i,t-1}$ contains information on the time invariant excluded attributes proxied by ν_i. Indeed, the greater the variance of ν_i, the greater the tendency for (3b) to estimate $\widehat{\gamma} \approx 1$ (Islam, 1995; Caselli et al., 1996). Hence finding that $0 < \widehat{\gamma} \leq 1$ does not imply that steady-state economic density depends on lagged economic density.

On the other hand, under (1a) there is no reason to expect that (3b) should estimate $\widehat{\gamma} > 1$. Such a finding can be taken as sufficient evidence of history dependence, $\partial L^*_{i,t}/\partial L_{i,t-1} > 0$. Subtracting lagged density, $L_{i,t-1}$, from both sides of (3b) gives a conditional divergence interpretation of $\widehat{\gamma} > 1$: all else equal, places with higher economic density grow at a quicker rate. Conversely, $\widehat{\gamma} < 1$ is often interpreted as evidence of conditional convergence: all else equal, places with higher economic density grow at a slower rate.

Even more problematic is interpreting the $\widehat{\boldsymbol{\delta}}_t$ estimated from (3b). With the identifying assumptions that $\gamma < 1$ and that density grows at a rate linearly proportional to its gap from its steady state, $dL_{i,t} = -(1-\gamma) \cdot (L_{i,t} - L^*_{i,t})$, the $\widehat{\boldsymbol{\delta}}_t$ should estimate $(1-\gamma)\boldsymbol{\beta}_t$ and so have the same sign as $\boldsymbol{\beta}_t$. Hence the $\boldsymbol{\delta}_t$ are sometimes interpreted as measuring the signs of the structural relationships, $\partial L^*_{i,t}/\partial \mathbf{x}_i$.

In practice, our estimates of δ_t are almost always identical in sign and similar in magnitude to our estimates of $d\boldsymbol{\beta}_t \equiv \boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}$ from the change specification (2). This is true even after allowing for a richer, nonlinear dependence of current density on lagged density. Hence we interpret the $\widehat{\boldsymbol{\delta}}_t$ as capturing the change in effect of $\mathbf{x}_i$ on steady-state economic density, which corresponds exactly to the partial derivative of (3a) with respect to $\mathbf{x}_i$. To emphasize such an interpretation, we report results from change regressions that subtract lagged population density from both sides of (3b) rather than the identical coefficients from level regressions that control for lagged population density.

Note that assuming (1a) is the true data generating process with a particular $d\beta^k_t$ small, (3b) will tend to estimate $\widehat{\delta}^k_t \approx 0$. Such a finding in no way implies that $\beta^k_t = 0$. On the other hand, regardless of whether (1a) or (3a) is the true data generating process, identically signed estimates $\widehat{\beta}^k_t \gtrless 0$ from (1b) and $\widehat{\delta}^k_t \gtrless 0$ from (3b) can be taken as sufficient evidence that $\partial L^*_{i,t}/\partial x^k_i$ is also of this same sign.

Because any omitted geographic variables induce spatial correlations among the error terms, we use a generalization of the Huber–White heteroskedastic-consistent estimator based on Conley (1999) to report standard errors robust to such a spatial structure. For county pairs between which the Euclidean distance is beyond a certain cutoff $\bar{d}$, we impose that the covariance between error terms is zero. Within this distance, we impose a (weakly) declining weighting function, g(distance), on the covariance between errors. In essence, this amounts to allowing for a spatially-based random effect. Dropping time subscripts,

$$E(\varepsilon_i) = 0$$
$$E(\varepsilon_i \varepsilon_j) = g(\text{distance}_{ij})\rho_{ij} \tag{4}$$
$$\hat{\rho}_{ij} = e_i e_j,$$

$$\begin{aligned} g(\text{distance}_{ij}) &= 1 \quad \text{for distance}_{ij} = 0 \\ g(\text{distance}_{ij}) &= 0 \quad \text{for distance}_{ij} > \bar{d} \\ g'(\text{distance}_{ij}) &\leq 0 \quad \text{for distance}_{ij} \leq \bar{d}. \end{aligned} \tag{5}$$

Herein, we assume the weighting on the covariance between error terms falls off quadratically as the distance between county centers increases to 200 km, the cutoff beyond which we impose zero covariance. So $g(\cdot) = 1 - (\text{distance}_{ij}/200)^2$. Thus accounting for spatial correlation approximately doubles standard errors relative to the assumption of homoskedasticity.

Note that the error specification in (4) and (5) reduces to the Huber–White heteroskedastic-consistent estimator for standard errors when $\bar{d}$ equals zero; it reduces to a group-based random effect estimator for standard errors with a non-Euclidean distance measure and a one-zero step specification for $g(\cdot)$.

5. Empirical Results: The Coastal Determinants of Economic Density

To reject the hypothesis that coastal proximity does not influence economic density, Maps 1 and 2 and Figures 2 and 3 are sufficient. Our purpose in pursuing multivariate regression analysis, instead, is to better describe the magnitude and nature of coastal proximity's effect on economic density. In particular, we would like to distinguish whether this effect occurred only historically versus whether it continues up through the present and to what extent the underlying mechanism is a contribution to productivity versus a contribution to quality of life.

5.1. Current Economic Density and Coastal Proximity

We begin by describing the partial correlations between measures of current economic density and coastal proximity. As with Figure 2 (panel (b)), we measure coastal proximity with three dummy variables set equal to one for counties with centers within 80 km of an ocean coast, within 80 km of a Great Lakes coast, or within 40 km of a river on which there was commercial navigation in 1968. Summary statistics are shown in Table 2. Defining navigability based on usage in 1968 introduces an upward bias by excluding potentially navigable rivers near which there is insufficient economic density to support commercial navigation. Such criticism is anyway made moot by the essentially zero partial correlation between current economic density and navigable river proximity once we control for ''major'' river proximity.

Table 3 (column (1)) reports results from regressing log(1 + population density in 2000) on just the three dummies (and an intercept). Large positive coefficients on all three

Table 2. Summary statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Land area (sq. km)	3,069	2,497	3,390	59	51,936
2000 Population	3,069	91,099	296,317	67	9,519,338
2000 Civilian employment	3,069	53,519	188,142	123	5,492,154
1960 Population	3,063	58,262	205,986	208	6,038,771
1920 Population	3,014	35,073	108,959	37	3,053,017
2000 Population density	3,069	86.0	642.8	0.04	25,846
2000 Civilian employment density	3,069	59.8	883.1	0.05	47,271
1960 Population density	3,063	71.2	699.0	0.07	28,509
1920 Population density	3,014	53.8	804.5	0.02	40,086
Log(1 + 2000 population density)	3,069	2.93	1.43	0.04	10.16
Log(1 + 2000 civilian employment density)	3,063	2.34	1.39	0.05	10.76
Log(1 + 1960 population density)	3,063	2.65	1.29	0.06	10.26
Log(1 + 1920 population density)	3,014	2.50	1.12	0.02	10.60
1920-to-1960 change Log(1 + population density)*	3,014	0.40	1.25	−2.63	9.25
1960-to-2000 change Log(1 + population density)*	3,063	0.72	1.14	−2.33	8.10
Ocean coast dummy	389	1	0	1	1
Distance to ocean coast (km)	3,069	638.1	482.6	0.4	1875.0
Ocean natural harbor dummy	199	1	0	1	1
Distance to ocean natural harbor (km)	3,069	660.7	467.2	2.6	1,887.8
Ocean shoreline/sq.km	235	0.08	0.08	0.00	0.66
Log(1 + ocean shoreline/sq.km)	235	0.07	0.07	0.00	0.50
Great lakes dummy	170	1	0	1	1
Distance to great lakes coast (km)	3,069	841.2	594.7	0.3	2,686.1
Great lakes natural harbor dummy	30	1	0	1	1
Distance to great lakes natural harbor (km)	3,069	929.5	567.4	5.0	2,703.2
Great lakes shoreline/sq.km	84	0.04	0.03	0.00	0.18
Log(1 + great lakes shoreline/sq.km)	84	0.04	0.03	0.00	0.17
Navigable river dummy	508	1	0	1	1
Distance to navigable river (km)	3,069	248.6	245.0	1.2	1,246.0
Major river dummy	1,153	1	0	1	1
Distance to major river (km)	3,069	74.0	67.0	0.5	436.5
Small/Med/Lrg ocean harbor dummy	405	1	0	1	1
Dist. to Small/Med/Lrg ocean harbor (km)	3,069	412.9	313.5	2.6	1,354.0
Small/Med/Lrg Grt lakes harbor dummy	124	1	0	1	1
Dist. to Small/Med/Lrg Grt lakes harbor (km)	3,069	873.5	598.5	5.0	2,703.2
Very Sml/Sml/Med/Lrg ocean harbor dummy	514	1	0	1	1
Dist. to Very Sml/Sml/Med/Lrg ocean harbor (km)	3,069	403.8	317.0	2.0	1,340.3
Very Sml/Sml/Med/Lrg Grt lakes harbor dummy	147	1	0	1	1
Dist. to Very Sml/Med/Lrg Grt lakes harbor (km)	3,069	871.4	600.7	5.0	2,703.2
Major river-ocean junction dummy	121	1	0	1	1
Distance to major river-ocean junction (km)	3,069	696.1	478.2	6.1	1,947.8
Major river-grt lakes junction dummy	20	1	0	1	1
Distance to major river-grt lakes junction (km)	3,069	939.0	601.6	19.1	2,928.3

Notes: *Change in log population density shown on annual percentage basis.
Summary statistics are for all continental US counties. For dummy variables and shoreline measures, summary statistics are shown only for observations with values that do not equal zero.

Table 3. Economic density and coastal proximity.

Dependent Variable→	Log(1 + 2000 Population Density)				Log(1 + 2000 Employment Density)			
RHS Variables↓	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Weather/topography controls	No	Yes	No	Yes	No	Yes	No	Yes
State fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Ocean coast dummy	***1.608***	***0.847***	***0.969***	*0.283*	***1.482***	***0.890***	***0.913***	*0.276*
	(0.198)	***(0.233)***	***(0.177)***	*(0.163)*	***(0.205)***	***(0.239)***	***(0.179)***	*(0.172)*
Great Lakes coast dummy	***1.161***	***0.657***	*0.462*	***0.542***	***1.098***	***0.607***	*0.491*	***0.543***
	(0.211)	***(0.168)***	*(0.255)*	***(0.228)***	***(0.212)***	***(0.177)***	*(0.254)*	***(0.226)***
Navigable river dummy	***0.848***	***0.315***	***0.455***	***0.286***	***0.731***	***0.291***	***0.432***	***0.282***
	(0.120)	***(0.106)***	***(0.098)***	***(0.081)***	***(0.118)***	***(0.106)***	***(0.097)***	***(0.082)***
Observations	3,069	3,069	3,069	3,069	3,069	3,069	3,069	3,069
Number of indep. variables	3	17	51	65	3	17	51	65
Sum of squared residuals	5,060.6	3,070.5	3,478.1	2,623.6	4,854.5	3,335.9	3,641.9	2,882.8
R^2	0.199	0.514	0.449	0.584	0.177	0.434	0.383	0.511
Control variables R^2	—	0.481	0.408	0.576	—	0.398	0.343	0.502

Notes: Ocean coast and Great Lakes coast dummy variables are one if county center is within 80 km of the respective coast, zero otherwise. Navigable river dummy variable is one if county center is within 40 km of a river on which there was commercial navigation in 1968, zero otherwise. Standard errors in parenthesis are robust to spatial correlation using the Conley spatial estimator discussed in the text with a weighting that declines quadratically to zero for counties with centers 200 km apart. Bold type signifies coefficients statistically different from zero at the 0.05 level; italic type signifies coefficients statistically different from zero at the 0.10 level.

statistically differ from zero at well below the 0.05 level. The coefficient magnitudes imply that the 389 ocean coast, 170 Great Lakes coast, and 508 navigable river counties have respective expected 2000 population density 5.0, 3.2, and 2.3 times that of more inland counties. Approximately half of the higher expected coastal population density derives from favorable weather and topography conditions near such coastal locations. The regression reported in Table 3 (column (2)), augments the column (1) regression by including the twelve weather variables (January minimum temperature, July maximum heat index, mean annual precipitation, mean annual days with precipitation of at least 0.1 inch, mean annual days temperature falls below 32 °F, and mean annual days temperature rises above 90 °F—each entered linearly and quadratically) and two topography variables (the standard deviation of within-county elevation divided by county land area, entered linearly and quadratically) used in Figure 2 (panel (b)) and Figure 3 (panel (b)). Controlling for the weather and topography, the ocean coast, Great Lakes coast, and navigable river counties have respective expected 2000 population density 2.3, 1.9, and 1.4 times that of remaining inland counties.[8]

On their own, the ocean, Great Lakes, and navigable river dummies account for 20 percent of the variation in 2000 population density across counties. Additionally controlling for weather and topography, the augmented specification accounts for 51 percent of such variation. Additionally controlling for state fixed effects increases explanatory power to 58 percent. For comparison, controlling only for weather and topography (but not coastal proximity) accounts for 48 percent of the variation. Controlling only for state fixed effects accounts for 41 percent of the variation.

The statistically significant positive partial correlations between population density and coastal proximity reported in Table 3 (column (2)) are extremely robust. Similar results obtain from using state fixed effects rather than the weather and topography controls (Table 3, column (3)), from using log(1 + employment density in 2000) as the regression dependent variable (Table 3, columns (5)–(7)), and from varying the distances demarcating the coast dummies (Rappaport and Sachs, 2002). Only the combination of state fixed effects and the weather/topography controls causes a substantial drop in the coefficient on the ocean coast dummy (Table 3, columns (4) and (8)). Given that coastal proximity is among the most important characteristics describing states' location and the substantial correlation of weather with coastal proximity, we do not believe the latter result implies fragility.

5.2. *Controlling for History*

As discussed in the econometric specification section above, one interpretation of the positive partial correlation between current population density and coastal proximity is that it is picking up a past functional relationship that no longer holds. As a starting point towards addressing such concerns, Table 4 (column (1)) shows the results from regressing the change in population density, log(1 + 1960 population density) − log(1 + 1920

Table 4. Coastal proximity controlling for history.

Dependent Variable→	ΔPop Density (1920–1960)				ΔPop Density (1960–2000)			
RHS Variables↓	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Weather/topography controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Initial density/concentric pop.	No	No	Yes	Yes	No	No	Yes	Yes
Supplemental historical controls	No	No	No	Yes	No	No	No	Yes
Ocean coast dummy	***1.380***	***0.749***	***0.650***	***0.345***	***1.038***	***0.277***	***0.310***	0.174
	(0.143)	***(0.136)***	***(0.125)***	***(0.123)***	***(0.129)***	***(0.140)***	***(0.147)***	(0.112)
Great Lakes coast dummy	***0.800***	***0.712***	***0.647***	***0.538***	***0.263***	***0.472***	***0.489***	***0.367***
	(0.156)	***(0.151)***	***(0.119)***	***(0.115)***	***(0.120)***	***(0.129)***	***(0.136)***	***(0.101)***
Navigable river dummy	0.077	***0.232***	0.084	0.050	−0.075	−0.069	−0.118	−0.071
	(0.079)	***(0.076)***	(0.070)	(0.061)	(0.088)	(0.090)	(0.081)	(0.062)
Observations	3,014	3,014	3,014	3,013	3,063	3,063	3,063	3,063
Number of indep. variables	3	17	31	53	3	17	31	60
Sum of squared residuals	4,060.7	3,280.8	2,843.1	2,337.4	3,620.5	2,962.1	2,494.5	1,782.9
R^2	0.142	0.307	0.399	0.506	0.093	0.258	0.375	0.553
Control variable R^2	—	0.271	0.376	0.497	—	0.247	0.362	0.548

Notes: Ocean coast and Great Lakes coast dummy variables are one if county center is within 80 km of the respective coast, zero otherwise. Navigable river dummy variable is one if county center is within 40 km of a river on which there was commercial navigation in 1968, zero otherwise. Standard errors in parenthesis are robust to spatial correlation using the Conley spatial estimator discussed in the text with a weighting that declines quadratically to zero for counties with centers 200 km apart. Bold type signifies coefficients statistically different from zero at the 0.05 level; italic type signifies coefficients statistically different from zero at the 0.10 level.

population density), on the ocean coast, Great Lakes coast, and navigable river dummies (and an intercept) without any other controls.[9,10] The dependent variable has been normalized so that it can be interpreted as an annual growth rate. The regression admits large positive, statistically significant coefficients on the ocean coast and Great Lakes coast dummies and a coefficient that does not statistically differ from zero on the navigable river dummy. Table 4 (column (2)) shows results from the same 1920–1960 change regression after controlling for weather and topography.[11] The positive coefficients on the ocean coast and Great Lakes dummies continue to statistically differ from zero at the 0.05 level; but the magnitude of the ocean coast coefficient has been reduced by slightly less than one half and that of the Great Lakes coefficient has fallen by approximately fifteen percent. In other words, nearly half of the 1920–1960 increase in ocean coast relative population density and a smaller portion of the 1920–1960 increase in Great Lakes relative population density can be attributed to changing tastes and technologies with respect to the weather and topography. In contrast, controlling for the weather and topography causes an increase in the magnitude of the positive coefficient on the navigable river dummy which now statistically differs from zero at the 0.05 level. So for the navigable river counties have weather and topography conditions against which tastes and technology were shifting.

Absent adjustment frictions and agglomeration forces, the positive coefficients just reported would establish the continued, indeed increasing, contribution of coastal proximity to productivity and quality-of-life. More realistically, the positive coefficients may just be capturing delayed adjustment and steady-state path dependence. To rule out such interpretations, we include two additional sets of control variables measuring initial conditions in 1920.

A first set of variables measure a county's initial population density in 1920 entered as a seven-part spline. As discussed in the first working paper version of the present paper (Rappaport and Sachs, 2001), such a specification captures a nonlinear relationship between initial population density and subsequent growth characterized both by divergence (growth increasing with initial population density) and convergence (growth decreasing with initial population density).

A second set of variables measure total 1920 population within concentric rings emanating from a county's center. An innermost circle measures log(1 + total population of all counties with centers within 50 km from a county's own center) and, at a minimum, always includes the county's own population. A second ring measures log(1 + total population of all counties with centers 50–100 km from a county's own center); additional rings with outer circumference radii of 150, 200, 300, 400, and 500 km make for a total of seven rings. Together, these concentric population variables capture, for instance, the "market potential" available to local firms producing goods with nontrivial transportation costs (Krugman, 1991; Fujita et al., 1999; Black and Henderson, 2001; Hanson, 2001).

Table 4 (column (3)) shows the results from including these two additional sets of initial period variables as controls in the 1920–1960 change regression.[12] The positive coefficients on the ocean and Great Lakes coast dummies drop only slightly in magnitude and continue to statistically differ from zero at the 0.05 level. But the positive coefficient on the navigable river dummy drops toward zero from which it no longer statistically differs. The latter result suggests that the positive, statistically significant coefficient on the navigable river in Table 4 (column (2)) is indeed picking up some combination of delayed adjustment and agglomeration (see footnote 7). The seemingly successful ability of the initial population density and concentric total population variables to control for delayed adjustment and agglomeration suggests that the positive significant coefficients on ocean and Great Lakes coast dummies are indeed picking up increasing contributions from such locations to productivity and quality of life. Of course, it is still possible the positive coefficients are spurious due to an omitted variable bias.

Table 4 (column (4)) shows the results from the 1920–1960 change regression controlling for a supplemental set of 22 variables measuring initial conditions in 1920 including urbanization, age, education, agricultural intensity, and manufacturing intensity (enumerated in the left-hand panel of Table 5). The inclusion of such controls does cause a

Table 5. Supplemental historical controls.

1920 (22 Variables)	1960 (29 Variables)
Urbanization (4 variables)	*Urbanization* (4 variables)
percent population urban (linear, quadratic)	percent population urban (linear, quadratic)
percent population in cities of at least 25,000 (linear, quadratic)	percent population rural farm (linear, quadratic)
Age and Education (8 variables)	*Age and Education* (7 variables)
percent of population age 0–17	percent of population age 0–20
percent of population age 45 +	percent of population age 65 +
percent males 21 + literate (linear, quadratic)	median age
percent females 21 + literate (linear, quadratic)	percent of population 25 +, 0 to 4 years education
percent of population age 16–20 + literate (linear, quadratic)	percent of population 25 +, high school graduate or higher
	percent of population 25 +, median years of education
	college enrollment (percent of population)
Agriculture (5 variables)	*Industry and Occupation* (8 variables)
farmland percent of county land area (linear, quadratic)	percent of civilian employment in each of 7 industries
improved farmland percent of county area (linear, quadratic)	percent of civilian employment in "whitecollar" occupations
log(1 + farm output value density)	
Manufacturing (5 variables)	*Industrial Density* (8 variables)
log (1 + manufacturing employment density)	log (1 + manufacturing employment density)
log (1 + manufacturing wages density)	log (1 + manufacturing production employment density)
log (1 + manufacturing value added density)	log (1 + manufacturing wages density)
log (1 + manufacturing value of inputs density)	log (1 + manufacturing value added density)
log (1 + manufacturing horsepower of installed capital density)	log (1 + mineral and extraction employment density)
	log (1 + mineral and extraction wages density)
	log (1 + mineral and extraction sales density)
	log (1 + agriculture sales density)
	Bank Deposits (2 variables)
	log (1 + commercial and savings bank deposits density)
	log (1 + savings and loan associations deposits density)

substantial drop in the magnitude of the coefficient on the ocean coast dummy and a lesser drop in the magnitude of the coefficient on the Great Lakes coast dummy. The cumulative drop in such coefficients, first from including the initial density and concentric population controls and then from including the supplemental historical controls, suggests that the 1920–1960 increase in Great Lakes and ocean coast population density does partly reflect delayed adjustment and steady-state path dependence. But even after exhaustively controlling for initial conditions, the ocean and Great Lakes coast coefficients continue to statistically differ from zero at the 0.05 level. So the increase also partly reflects an increase in coastal proximity's contribution to productivity and quality of life.

Quantitatively, counties located within 80 km of an ocean or Great Lakes coast have respective expected 1920–1960 annual growth rates at least 0.35 and 0.54 percentage points higher than those of more inland counties. Thse differences are relatively large compared to the mean annual county population growth rate of 0.40 percent over the period 1920–1960 but relatively small compared to the associated 1.25 percent standard deviation of such county growth rates. Reflecting that coastal proximity is only a small source of the wide variation of realized 1920–1960 county population density growth rates, the marginal contribution from the coastal dummies towards accounting for such variation after including the various sets of controls ranges from 1 to 4 percent; on their own, the coastal dummies account for 14 percent.

Table 4 (columns (5)–(8)) report comparable results for regressions using the change in population density from 1960 to 2000 as the dependent variable. For this latter period, including the weather and topography controls cuts the magnitude of the positive coefficient on the ocean coast dummy by nearly three quarters (Table 4, column (6) versus (5)); however, even after controlling for initial population density and initial concentric total population, the positive ocean coast coefficient remains statistically significant at the 0.05 level (Table 4, column (7)). For the positive coefficient on the Great Lakes coast dummy, controlling for the weather and topography causes a substantial increase in magnitude. So for the period 1960–2000, changing tastes and technology with respect to the weather made location at the Great Lakes coast relatively less attractive. Additionally controlling for initial population density and concentric total population, the Great Lakes coast dummy increases further.

Again, it is possible that we are still not sufficiently controlling for initial conditions. Table 4 (column (8)) reports results from the 1960–2000 change regression controlling for a supplemental set of 29 variables measuring initial urbanization, age, education, industry mix, industrial density and bank deposit density (enumerated in the right panel of Table 5). These additional historical controls cause a further decrease in the magnitude of the coefficient on the ocean coast dummy, which no longer statistically differs from zero. So exhaustively controlling for initial conditions, we cannot reject that there was no increase circa 1960 in the contribution to productivity and quality of life from location near an ocean coast. But as emphasized in the econometric specification section above, this inability to reject in no way implies any decrease in the ocean coast contribution to productivity and quality of life. Rather, the alternative hypothesis is simply that such contribution may have remained constant since approximately 1960.[13]

On the other hand, even after exhaustively controlling for initial conditions, the positive coefficient on the Great Lakes dummy remains statistically significant. Counties with centers within 80 km of a Great Lakes coast have expected 1960–2000 growth rates at least 0.37 percentage points higher than those of more inland counties. This magnitude is relatively moderate compared to a mean annual county population growth rate of 0.72 over this period; it is relatively small compared to the 1.14 percent standard deviation of county growth rates.

Regardless of specification, the 1960–2000 change regressions admit a negative coefficient on the navigable river dummy, though in no case does this statistically differ from zero. More strongly, decade-by-decade change regressions controlling for initial population density and concentric total population admit a negative, statistically significant coefficient on the navigable river dummy for six of the twelve decades—the four consecutive decades from 1880 to 1920 along with the 1980s and 1990s (Rappaport and Sachs, 2002). For four of these six decades, such a result is despite the positive bias from navigability being determined by actual usage at a later date. So there indeed exists a reasonable amount of evidence that the navigable river contribution to productivity and quality of life has decreased since 1880 and perhaps also since 1960. But at least for the latter period, the statistical result is fragile.[14]

The statistically significant, positive partial correlations of the 1920–1960 change in population density with ocean and Great Lakes coast proximity are extremely robust. They hold regardless of the distance demarcating the ocean and Great Lakes coast dummies (Rappaport and Sachs, 2002) as well as for an alternate set of weather controls based on Mendelsohn et al. (1994). The statistically significant, positive partial correlation of the 1960–2000 change in population density with Great Lakes coast proximity is moderately robust. It also holds regardless of the distance demarcating the Great Lakes coast dummy but no longer statistically differs from zero when controlling for the Mendelsohn et al. set of weather variables.[15]

The results from the previous subsection along with those immediately above together strongly argue for a large continuing contribution from coastal proximity to present-day productivity and quality of life. Present-day population and employment density are highly positively correlated with coastal proximity. This correlation has been increasing since 1920, even after extensively controlling for historical forces. And finally, there is little evidence suggesting a decrease in the partial correlation between population density and coastal proximity.[16] Hence we unequivocally reject the hypothesis that the high coastal concentration of population and employment just reflects historical forces.

On the other hand, the small positive coefficient on the navigable river dummy in the level regressions along with the corresponding negative, though not statistically significant, coefficient in the change regressions suggest that there may no longer be any contribution to productivity and quality of life from location near a navigable river.

It remains for us to try to distinguish how much of the coastal population concentration derives from a contribution to productivity (historical and present-day) and how much derives from a contribution to quality of life (historical and present-day). Before doing so, however, we turn to a brief analysis of the the partial correlations of the level and change in population density with dummies for each of the North Atlantic, South Atlantic, Gulf of Mexico, and Pacific coasts.

5.3. Population Density Level and Change by Separate Ocean Coast

Table 6 reports results from regressions analogous to those in Tables 3 and 4 based on the specific coast to which a county is closest: the North Atlantic, the South Atlantic, the Gulf of Mexico, the Pacific, or the Great Lakes. The resulting five-way partition of the continental United States is shown in Map 3. All regressions include separate intercepts for the five regions along with the weather, topography, and navigable river proximity controls used in Tables 3 and 4. The coefficients reported in Table 6 are therefore determined solely by the counties in the respective shaded region.

We focus first on the level regressions (Table 6, columns (1) and (2)). Not controlling for the weather and topography only the positive coefficient on the South Atlantic coast dummy does not statistically differ from zero; but with such controls, only the North Atlantic and Great Lakes coast coefficients remain statistically significant. In other words, weather and topography largely account for high observed population density at the Pacific and Gulf of Mexico coasts. At the Pacific coast in particular, the weather's contribution to

Table 6. Coastal proximity by closest coast.

Dependent Variable →	2000 Pop Density		ΔPop Density (1920–1960)				ΔPop Density (1960–2000)			
RHS Variables ↓	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Closest Coast fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather/topography controls	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Initial density/ concentric pop.	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Supplemental historical controls	No	No	No	No	No	Yes	No	No	No	Yes
North Atlantic coast dummy	***1.827***	***1.174***	***1.064***	***0.665***	*0.427*	***0.477***	***0.518***	0.236	*0.378*	0.131
	(0.338)	***(0.313)***	***(0.171)***	***(0.170)***	*(0.227)*	***(0.197)***	***(0.161)***	(0.191)	*(0.199)*	(0.169)
South Atlantic coast dummy	0.298	0.025	***0.720***	0.364	***0.599***	0.191	*0.566*	0.309	0.413	0.258
	(0.193)	(0.211)	***(0.285)***	(0.231)	***(0.228)***	(0.191)	*(0.291)*	(0.262)	(0.251)	(0.187)
Gulf of Mexico coast dummy	***1.017***	0.154	***1.563***	***0.725***	***0.785***	***0.508***	***1.187***	0.181	0.023	0.114
	(0.188)	(0.190)	***(0.217)***	***(0.240)***	***(0.231)***	***(0.193)***	***(0.240)***	(0.203)	(0.213)	(0.170)
Pacific coast dummy	***2.259***	0.336	***1.752***	***1.070***	***1.019***	0.359	***0.739***	*0.525*	0.218	−0.071
	(0.355)	(0.293)	***(0.300)***	***(0.276)***	***(0.280)***	(0.241)	***(0.182)***	*(0.309)*	(0.318)	(0.284)
Great Lakes coast dummy	***1.207***	***0.814***	***0.956***	***0.755***	***0.632***	***0.507***	***0.597***	***0.662***	***0.612***	***0.453***
	(0.217)	***(0.179)***	***(0.156)***	***(0.156)***	***(0.123)***	***(0.119)***	***(0.124)***	***(0.135)***	***(0.143)***	***(0.104)***
Observations	3,069	3,069	3,014	3,014	3,014	3,013	3,063	3,063	3,063	3,063
Number of indep. variables	10	24	10	24	38	60	10	24	38	67
Sum of squared residuals	4,233.2	2,815.8	3,876.7	3,240.0	2,825.5	2,319.2	3,296.0	2,883.2	2,448.8	1,767.5
R^2	0.330	0.554	0.181	0.315	0.403	0.510	0.174	0.277	0.386	0.557
Control variables R^2	0.173	0.531	0.058	0.287	0.381	0.500	0.110	0.261	0.370	0.550

Notes: Ocean coast and Great Lakes coast dummy variables are one if county center is within 80 km of the respective coast, zero otherwise. All regressions additionally include a navigable river dummy and closest-coast-specific intercepts. Standard errors in parenthesis are robust to spatial correlation using the Conley spatial estimator discussed in the text with a weighting that declines quadratically to zero for counties with centers 200 km apart. Bold type signifies coefficients statistically different from zero at the 0.05 level; italic type signifies coefficients statistically different from zero at the 0.10 level.

productivity and quality of life is especially dramatic. Not controlling for weather and topography, the 55 counties with centers within 80 km of the Pacific coast have expected 2000 population density 9.6 times that of the remaining 290 counties for which the Pacific is the closest coast. But controlling for weather and topography, the Pacific coastal counties' expected 2000 population density is just 1.4 times that of the remaining 290 counties, nor can we statistically reject that all counties for which the Pacific is the closest coast have identical expected 2000 population density.[17]

In the 1920–1960 change regressions (Table 6, columns (3)–(6)), the North Atlantic, Gulf of Mexico, and Great Lakes coast dummies admit positive, statistically significant coefficients across all four specifications. So for these three coasts, we can definitively reject that there was no increase in the contribution to productivity and quality of life from coastal proximity during the early-to-mid twentieth century.

The 1920–1960 regressions also admit positive coefficients on the Pacific and South Atlantic coast dummies, but statistically these are more fragile. The Pacific coast coefficient statistically differs from zero with no additional controls, with the weather and topography controls, and additionally with the initial density and concentric population controls. So there is very strong evidence that the increase in Pacific coast population density derives from an increasing contribution to productivity and quality of life. So long as Pacific coast proximity did not initially detract from productivity and quality of life, such evidence implies a postive present-day contribution. But as the Pacific coast coefficient no longer statistically differs from zero after controlling for the 22 supplemental historical controls, we cannot reject that there was no such increase in productivity and quality of life. For the South Atlantic coast, a positive coefficient on the dummy statistically differs from zero with no additional controls and when controlling for the weather, initial density, and concentric population but not in the two other specifications. Along with the lack of a positive, statistically significant coefficient on the South Atlantic coast dummy in either of the level regressions, there is little evidence supporting a positive contribution to productivity and quality of life from South Atlantic coast proximity.

In the 1960–2000 change regressions (Table 6, columns (7)–(10)), only the Great Lakes coast dummy admits a positive, statistically significant coefficient across all four specifications. As in the previous subsection, we reject that there was no increase over the period 1960 to 2000 in the contribution to productivity and quality of life from Great Lakes coast proximity.[18] The four ocean coast dummies admit positive, statistically significant coefficients with no additional controls but not otherwise. So only weak evidence supports the increase subsequent to 1960 in the contribution to productivity and quality of life from proximity to each of the United States ocean coasts. But as with the unified ocean coast change regressions discussed in the previous section, there is virtually no evidence suggesting a decrease in ocean coast proximity's contribution to productivity and quality of life.

Overall, the separate-ocean-coast regressions reaffirm the unified-ocean-coast ones. Ocean and Great Lakes coast proximity bestows a large productivity and quality-of-life advantage that increased subsequent to 1920 and remains positive today. Only at the South Atlantic coast is this conclusion suspect.

We next try to disentangle coastal proximity's contribution to productivity versus its contribution to quality of life.

5.4. Productivity versus Quality of Life

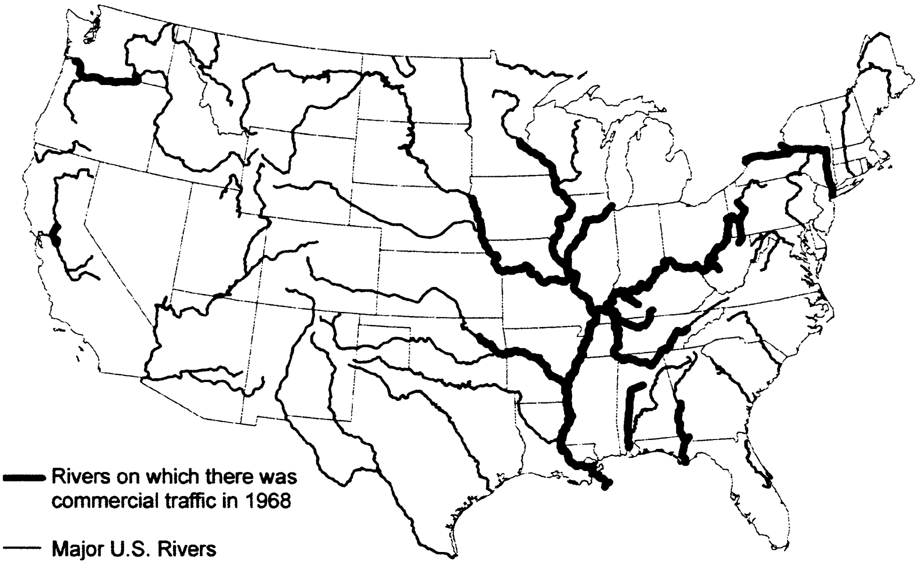
The results above establish a causal relationship from coastal location to high population density that continues well into the twentieth century. But so far we have not been able to distinguish whether the causal mechanism works via a contribution to productivity or via a contribution to quality of life. In this section we attempt to do so by using coastal proximity measures that we believe a priori influence productivity or quality of life but not both.

For oceans and Great Lakes, we augment our base specification to include an analogous harbor proximity measure as well as the ratio of a county's shoreline to its total area. We argue that harbors measure access to low-cost bulk transportation and so primarily raise productivity but not quality of life. On the other hand, shoreline measures access to recreational and scenic amenities and so primarily increases quality of life but not productivity. Controlling for harbor proximity and shoreline, any remaining partial correlation of population density with coastal proximity should continue to reflect a combination of productivity and quality of life as above.

An ideal harbor measure would be all coastal geological formations affording shelter for seagoing vessels above a certain size threshold. In practice, we identify harbors as a subset of actual seaports included in the *World Port Index* (US Naval Oceanographic Office, 1971). This classifies seaports by four size categories—very small, small, medium, and large—based on several applicable factors including area, facilities, and wharf space. We define ''harbors'' as medium or large seaports. To minimize the possibility of reverse causality, we further exclude from our harbor measure any seaports that rely on constructed breakwaters or tide gates rather than natural barriers for shelter. Map 3 shows the resulting ''natural harbors'' as well as the excluded medium and large seaports relying on constructed shelter. A selection bias remains in that geological formations affording the necessary shelter but that did not actually develop into seaports will be excluded. Hence we explore the sensitivity of results to using alternative harbor measures that are likely to encompass such excluded geological shelters.

For navigable rivers, we augment our base specification to include distance to the nearest ''major'' river. Major rivers are defined as a superset of navigable rivers to include the longest North American rivers as well as shorter rivers that connect lakes to the ocean (see data appendix). Map 4 illustrates. Our prior is that controlling for proximity to major rivers, any residual correlation of economic activity with proximity to navigable rivers is likely to be picking up a productivity effect. On the other hand, to the extent that population density is correlated with the presence of major rather than navigable rivers, the underlying mechanism may be either productivity (e.g. drinking water, hydroelectric power) or quality of life (e.g. fishing, canoeing).

We focus first on the level regressions (Table 7, columns (1) and (2)). Not controlling for the weather and topography, the ocean coast dummy admits a positive, statistically significant coefficient slightly smaller than that on the ocean natural harbor dummy; and the ocean shoreline measure admits a positive, statistically significant coefficient as well. But controlling for weather and topography, the ocean coast coefficient falls dramatically, the ocean natural harbor dummy falls slightly, and the shoreline coefficient no longer statistically differs from zero. Comparing the magnitude of the ocean coast and natural



Map 4. Navigable and major US rivers.

harbor coefficients in this latter regression suggests that more than two thirds of the present-day ocean coast concentration of US population derives from a positive contribution, historical or continuing, to productivity from location near natural harbors. The Great Lakes coefficients similarly suggest that the larger part of present-day Great Lakes population concentration derives from a productivity rather than a quality-of-life effect.

Comparing the navigable river versus major river coefficients, a large positive, statistically significant coefficient on the former when not controlling for weather and topography no longer statistically differs from zero after adding such controls. Controlling for the weather and topography, the river coefficients are equal in magnitude; so only about half the higher expected population density near navigable rivers derives from such rivers' historical navigability *per se*.

Focusing next on the 1920–1960 change regressions (Table 7, columns (3)–(6)), both the ocean coast and Great Lakes coast dummies admit a positive, statistically significant coefficient across all four specifications. The ocean natural harbor dummy admits a positive, statistically significant coefficient so long as the supplemental historical controls are not included; the Great Lakes natural harbor dummy admits a positive, statistically significant coefficient so long as no initial condition controls are included. And both the ocean and Great Lakes shoreline measures admit a negative, statistically significant coefficient in the regression that controls only for weather and topography (column 4); but neither holds up to controlling for initial conditions. Similarly, a positive coefficient on the major river dummy statistically differs from zero only when not controlling for initial conditions.

Table 7. Coast versus harbor proximity.

Dependent Variable →	2000 Pop Density		ΔPop Density (1920–1960)				ΔPop Density (1960–2000)			
RHS Variables ↓	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Weather/topography controls	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Initial density/ concentric pop.	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Supplemental historical controls	No	No	No	No	No	Yes	No	No	No	Yes
Oceans										
Coast dummy	***0.935***	***0.412***	***1.198***	***0.539***	***0.528***	*0.231*	***1.118***	*0.314*	0.264	*0.211*
	(0.151)	***(0.149)***	***(0.184)***	***(0.132)***	***(0.127)***	*(0.137)*	***(0.176)***	*(0.167)*	(0.173)	*(0.127)*
Natural harbor dummy	***1.032***	***0.872***	***0.392***	***0.558***	*0.227*	0.219	− 0.104	0.057	− 0.064	− 0.108
	(0.235)	***(0.197)***	***(0.185)***	***(0.130)***	*(0.137)*	(0.133)	(0.184)	(0.151)	(0.151)	(0.125)
Shorclinc/km^2	***4.032***	− 0.136	0.148	− 2.182	0.570	0.562	− 0.237	− 1.692	***2.474***	0.614
	(1.727)	(1.727)	(1.281)	(1.152)	(1.082)	(0.975)	(1.107)	(1.238)	***(1.123)***	(1.158)
Great Lakes										
Coast dummy	***0.998***	***0.594***	***0.809***	***0.707***	***0.663***	***0.467***	***0.387***	***0.587***	***0.527***	***0.361***
	(0.200)	***(0.160)***	***(0.144)***	***(0.143)***	***(0.123)***	***(0.122)***	***(0.148)***	***(0.153)***	***(0.162)***	***(0.109)***
Natural harbor dummy	***1.196***	***0.891***	***0.833***	***0.735***	0.405	0.282	0.124	0.033	− 0.149	0.019
	(0.336)	***(0.282)***	***(0.343)***	***(0.315)***	(0.281)	(0.284)	(0.195)	(0.199)	(0.190)	(0.190)
Shoreline/km^2	− 2.720	− 2.646	*− 6.033*	***− 6.222***	− 3.939	1.978	***− 5.714***	***− 5.391***	1.176	0.917
	(3.894)	(2.823)	*(3.294)*	***(2.722)***	(2.452)	(1.929)	***(2.538)***	***(2.517)***	(2.403)	(2.655)
Rivers										
Navigable river dummy	***0.848***	0.177	− 0.084	0.139	0.025	0.026	*− 0.213*	*− 0.195*	***− 0.217***	*− 0.139*
	(0.143)	(0.114)	(0.099)	(0.094)	(0.091)	(0.080)	*(0.115)*	*(0.114)*	***(0.107)***	*(0.080)*
"Major" river dummy	− 0.018	***0.176***	***0.214***	*0.118*	0.085	0.039	***0.191***	***0.162***	*0.123*	0.086
	(0.110)	***(0.066)***	***(0.077)***	*(0.066)*	(0.064)	(0.057)	***(0.089)***	***(0.078)***	*(0.070)*	(0.054)
Observations	3,069	3,069	3,014	3,014	3,014	3,013	3,063	3,063	3,063	3,063
Number of indep. variables	8	22	8	22	36	58	8	22	36	65
Sum of squared residuals	4,869.6	2,958.9	4,003.2	3,222.6	2,829.3	2,329.6	3,597.7	2,941.8	2,479.1	1,778.3
R^2	0.229	0.531	0.154	0.319	0.402	0.508	0.098	0.263	0.379	0.554
Control variables R^2	—	0.481	—	0.271	0.376	0.497	—	0.247	0.362	0.548

Notes: Ocean and Great Lakes dummy variables are one if county center is within 80 km of the respective coast or natural harbor, zero otherwise. River dummy variables are one if county center is within 40 km of a navigable or major river as defined in the text, zero otherwise. Standard errors in parenthesis are robust to spatial correlation using the Conley spatial estimator discussed in the text with a weighting that declines quadratically to zero for counties with centers 200 km apart. Bold type signifies coefficients statistically different from zero at the 0.05 level; italic type signifies coefficients statistically different from zero at the 0.10 level.

For both the ocean and Great Lakes coasts, we cannot reject that the entire increase in coastal proximity's contribution to productivity and quality of life was exclusively an increase in its contribution to quality of life. But the 1920–1960 results are also consistent with the entire increase deriving from an increasing contribution to productivity (e.g., coastal proximity is an obvious prerequisite to the location of a harbor relying on constructed shelter). Casting doubt on an exclusively quality of life interpretation are the negative coefficients on the shoreline measures.

The 1960–2000 change regressions (Table 7, columns (7)–(10)) suggest that the late twentieth century continuing increase in ocean coast concentration may derive more from an increasing ocean coast contribution to quality of life than to productivity. Three of the four specifications result in a small negative coefficient on the natural harbor dummy, though in no case does this statistically differ from zero. In contrast, three of the four specifications result in a positive, statistically significant coefficient on the ocean coast dummy. For the one specification that does not (Table 7, column (9)), a positive coefficient on ocean shoreline statistically differs from zero at the 0.05 level (Table 7, column (9)). Quantitatively, the combined magnitudes of the ocean coast dummy and ocean shoreline coefficients imply only a moderate increase in expected 1960–2000 growth. Not holding constant the supplemental historical controls, counties with centers within 80 km of an ocean coast and that have the average ratio of shoreline to area among counties that border an ocean have expected 1960–2000 annual growth 0.44 percentage points higher than that of remaining inland counties. Holding constant the supplemental historical controls, expected higher annual growth falls to 0.25 percentage points. For comparison, average (across all continental counties) expected 1960–2000 annual growth is 0.72 percentage points with corresponding standard deviation 1.14 percentage points.

The 1960–2000 change regressions similarly suggest that the late twentieth-century remaining high concentration of US population near the Great Lakes coasts may increasingly derive from a quality-of-life effect. The coefficient on the Great Lakes natural harbor dummy never statistically differs from zero; but all four specifications admit a moderate positive, statistically significant coefficient on the Great Lakes coast dummy. A negative, statistically significant coefficient on Great Lakes shoreline becomes positive (though not statistically significant) after controlling for initial conditions.

Finally, the 1960–2000 change regressions clearly show a decreasing contribution from navigable river proximity to productivity. A negative coefficient on the navigable river dummy statistically differs from zero across all four specifications. The regressions also show an increasing contribution from major river proximity to productivity and quality of life. A positive coefficient on the major river dummy statistically differs from zero in three of the four specifications; but as the coefficient no longer statistically differs from zero after controlling for the 29 supplemental 1960 historical controls, we cannot reject no such increase. Quantitatively, the negative coefficient on the navigable river dummy slightly exceeds (in absolute value) the positive coefficient on the major river dummy. But both coefficients are small. For counties near navigable rivers, the two effects approximately cancel each other out.

The signs and statistical significance of coefficients on the coast and harbor dummies are fairly robust to alternative natural harbor proxies and weather controls. The regressions in Table 7 classify 199 counties as having centers within 80 km of a medium or large naturally sheltered ocean seaport and 30 counties as having centers within 80 km of a medium or large naturally sheltered Great Lakes seaport (versus 389 counties with centers within 80 km of an ocean coast and 170 with centers within 80 km of a Great Lakes coast). Expanding the harbor proxy to include "small" natural seaports (resulting in 405 ocean natural harbor and 124 Great Lakes natural harbor counties) or "very small" and "small" natural seaports (514 ocean natural harbor and 147 Great Lakes natural harbor counties) leads to a negative, statistically significant coefficient on the Great Lakes harbor dummy in several of the 1920–1960 and 1960–2000 change regressions; but results otherwise are essentially unchanged (Rappaport and Sachs, 2002).[19] Using the Mendelsohn et al. (1994) weather controls, the main change in results is that the positive coefficient on the Great

Lakes coast dummy in the 1960–2000 change regression controlling for initial conditions (Table 7, columns (9) and (10)) no longer statistically differs from zero.

Broken out by separate ocean coast, Table 8 shows that the positive partial correlation between the 1920–1960 change in population density and ocean natural harbor proximity in Table 7 appears to derive largely from an increase in the contribution to productivity from proximity to Gulf of Mexico natural harbors (Table 8, columns (3)–(6)). Counties

Table 8. Coast versus harbor proximity by closest coast.

Dependent Variable→	2000 Population Density			ΔPop Density (1920–1960)				ΔPop Density (1960–2000)		
RHS Variables↓	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Closest coast fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather/topography controls	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Initial density/ concentric pop.	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Supplemental historical controls	No	No	No	No	No	Yes	No	No	No	Yes
North Atlantic Counties										
Coast dummy	0.579	0.256	**0.806**	**0.528**	0.323	0.299	**0.797**	**0.591**	**0.422**	0.295
	(0.412)	(0.395)	**(0.255)**	**(0.242)**	(0.239)	(0.256)	**(0.231)**	**(0.243)**	**(0.252)**	(0.222)
Natural harbor dummy	**1.415**	**1.244**	**0.495**	*0.404*	0.080	0.127	−0.357	*−0.441*	−0.342	**−0.381**
	(0.405)	**(0.409)**	**(0.246)**	*(0.228)*	(0.207)	(0.225)	(0.239)	*(0.235)*	(0.212)	**(0.183)**
Shoreline/km^2	2.890	0.388	−1.209	**−2.787**	1.601	**2.074**	0.001	0.035	**4.932**	*2.606*
	(2.857)	(2.295)	(1.075)	**(1.194)**	(1.037)	**(1.022)**	(1.470)	(1.621)	**(1.040)**	*(1.372)*
South Atlantic Counties										
Coast dummy	0.222	0.153	**0.920**	**0.574**	**0.782**	*0.395*	**0.622**	0.320	0.322	0.274
	(0.258)	(0.251)	**(0.396)**	**(0.273)**	**(0.265)**	*(0.231)*	**(0.371)**	(0.313)	(0.317)	(0.224)
Natural harbor dummy	−0.186	−0.149	−0.389	−0.136	−0.216	−0.159	−0.149	0.029	−0.086	−0.175
	(0.248)	(0.230)	(0.370)	(0.249)	(0.244)	(0.227)	(0.341)	(0.264)	(0.284)	(0.194)
Shoreline/km^2	3.322	−1.548	−0.731	*−3.413*	−2.055	**−2.941**	0.123	−0.513	3.586	1.400
	(2.584)	(2.540)	(2.387)	*(2.009)*	(1.838)	**(1.460)**	(2.296)	(2.138)	(2.363)	(2.020)
Gulf of Mexico Counties										
Coast dummy	**0.563**	−0.108	**0.918**	0.253	0.332	0.176	**1.113**	0.141	0.016	0.103
	(0.177)	(0.189)	**(0.191)**	(0.228)	(0.209)	(0.188)	**(0.275)**	(0.202)	(0.217)	(0.160)
Natural harbor dummy	**0.776**	**0.752**	**0.917**	**1.031**	**0.929**	**0.747**	0.318	0.468	0.248	0.283
	(0.249)	**(0.212)**	**(0.264)**	**(0.266)**	**(0.284)**	**(0.281)**	(0.389)	(0.309)	(0.318)	(0.278)
Shoreline/km^2	*4.537*	0.309	**8.889**	3.036	3.744	2.085	−1.049	−3.876	−2.308	−3.046
	(2.591)	(2.556)	**(3.022)**	(2.758)	(2.722)	(2.559)	(2.630)	(2.711)	(2.417)	(2.131)
Pacific Counties										
Coast dummy	**1.631**	0.250	**1.699**	**1.030**	**1.104**	0.394	**0.731**	*0.468*	0.283	0.057
	(0.355)	(0.291)	**(0.364)**	**(0.281)**	**(0.293)**	(0.244)	**(0.215)**	*(0.280)*	(0.268)	(0.238)
Natural harbor dummy	**0.929**	*0.603*	0.409	**0.618**	0.177	0.191	0.098	0.282	−0.135	−0.166
	(0.370)	*(0.313)*	(0.321)	**(0.230)**	(0.265)	(0.255)	(0.242)	(0.308)	(0.350)	(0.363)
Shoreline/km^2	6.003	−2.026	−2.137	*−5.292*	−2.336	−1.038	−0.672	−0.482	3.817	−0.826
	(4.350)	(4.144)	(2.247)	*(2.798)*	(2.721)	(2.593)	(3.905)	(4.530)	(3.041)	(3.261)
Observations	3,069	3,069	3,014	3,014	3,014	3,013	3,063	3,063	3,063	3,063
Number of indep. variables	21	35	21	35	49	71	21	35	49	78
Sum of squared residuals	4,095.1	2,733.5	3,776.1	3,159.6	2,780.6	2,291.8	3,285.2	2,858.8	2,416.3	1,751.6
R^2	0.351	0.567	0.202	0.332	0.412	0.516	0.177	0.284	0.394	0.561
Control variables R^2	0.173	0.531	0.058	0.287	0.381	0.500	0.110	0.261	0.370	0.550

Notes: Dummy variables are one if county center is within 80 km of the respective coast or natural harbor, zero otherwise. All regressions control for Great Lakes, navigable river, and major river proximity as in Table 7. Standard errors in parenthesis are robust to spatial correlation using the Conley spatial estimator discussed in the text with a weighting that declines quadratically to zero for counties with centers 200 km apart. Bold type signifies coefficients statistically different from zero at the 0.05 level; italic type signifies coefficients statistically different from zero at the 0.10 level.

with centers within 80 km of a Gulf of Mexico natural harbor had expected 1920–1960 annual population growth at least 0.75 percentage points faster than remaining counties for which the Gulf of Mexico was the closest coast.

For the 1960–2000 change regressions, the coefficients on the North Atlantic proximity measures suggest a decreasing contribution to productivity but an increasing contribution to quality of life (Table 8, columns (7)–(10)). All four specifications admit a negative coefficient on the North Atlantic natural harbor dummy; but only two of these statistically differ from zero at the 0.10 level or below. In the regression controlling for everything but the supplemental historical controls (column (9)), the p-value for rejecting there was no change in the North Atlantic natural harbor contribution to producivity is 0.11. Using the Mendelsohn et al. (1994) weather controls (and continuing to exclude the supplemental historical controls), the p-value rises to 0.21. Both regressions that control for initial conditions also admit a positive, statistically significant coefficient on North Atlantic shoreline. But again there is some statistical fragility. With the combination of Mendelsohn et al. (1994) weather controls and the supplemental historical variables, the p-value for rejecting no increase in the North Atlantic shoreline's contribution to quality of life rises to 0.17.

Overall, the regressions discussed in this section establish that at least historically, the larger contribution from ocean and Great Lakes coastal proximity is to productivity rather than quality of life. But the 1920–2000 migration of US population to ocean and Great Lakes coasts is consistent with an increasing contribution to either productivity or quality of life. And especially for the latter part of the twentieth century, the results suggest that coastal proximity's contribution to US productivity may be starting to wane while its contribution to US quality of life may be rising.

6. Conclusions

Economic density in the United States is overwhelmingly concentrated at its ocean and Great Lakes coasts. This concentration has been increasing throughout much of the twentieth century. Extensively controlling for historical conditions suggests that the coastal concentration captures a present-day contribution to productivity and quality of life. Stronger partial correlations of the level of economic density with proximity to harbors rather than with proximity to the coast *per se* suggest that the larger coastal contribution is to productivity. Stronger partial correlations of the change in economic density with proximity to the coast per se rather than with proximity to harbors suggest an increasing coastal contribution to quality of life.

We leave it for future research to better understand how coastal proximity increases productivity and quality of life. With respect to productivity, proximity to harbors obviously lowers transportation costs for many tradable goods. To the extent that international trade accounts for a rising proportion of US consumption and output, it should not be surprising that easy access to sea-borne transport continues to be an important component of local productivity. Indeed, the ocean leg of international shipping has consistently accounted for only one third of door-to-door shipping costs since at least 1950 (Hummels, 1999); hence the large benefit to minimizing the remaining port and overland costs. Reinforcing this productivity advantage, many coastal cities' airports have emerged as the primary US gateways for international air travel. Certainly, high initial populations contributed to such emergence; but so too did the coastal geographic advantage of shorter travel times.

With respect to quality of life, coastal proximity offers several obvious advantages including recreation and scenic beauty. Moreover, as underscored by the regression analysis, coastal proximity is highly correlated with favorable weather. An increasing relative demand for quality of life arising from long-term rising incomes serves as a force that should continue to concentrate US population at its coasts.

For developing nations, our results reinforce the present consensus on the importance of openness to trade in promoting economic growth. Countries blessed with large ocean ports stand to benefit from increasing world trade. For other countries, "getting a port" may not be a policy option. Instead, development policy in these countries needs to take account of what is likely to be a substantial productivity disadvantage. Of course, doing so requires a better understanding of how coastal proximity affects productivity. So again, more research is needed.

Appendix

The Compensating Differential Framework

Assume a large number of localities across which there is high labor and capital mobility. In a long run spatial steady state, no individual should be able to increase their utility by moving to a different locality; nor should any firm be able to increase their profitability by doing so. Any variations in exogenous local attributes which affect utility and profits must be offset by compensating wage and nontradable price differentials.

The equating of utility levels across localities is captured by

$$\begin{aligned} V(p, w; \text{quality of life}) = &\left\{ \max_{c,n} u(c, n; \text{quality of life}) \text{ s.t. } c + pn \leq w \right\} = \overline{V}, \\ &u_c(\cdot) > 0; u_{cc}(\cdot) < 0, \\ &u_n(\cdot) > 0; u_{nn}(\cdot) < 0, \\ &u_{\text{quality}}(\cdot) > 0; u_{c,\text{quality}}(\cdot) = u_{n,\text{quality}}(\cdot). \end{aligned} \tag{A.1}$$

Here, $V(\cdot)$ represents an indirect utility function with the price of land services, p, and the wage level, w, as its arguments and quality of life as a shift parameter. The underlying (direct) utility function, $u(\cdot)$, is increasing in consumption of a tradable good, c, and nontradable land services, n. With the tradable good as numeraire and with the per capita quantity of inelastically supplied labor normalized to one, individuals face the budget constraint that their tradable consumption plus their expenditure on land services cannot exceed the wage rate. The first two sets of derivative restrictions just establish that utility is strictly increasing and concave with respect to each of the tradable and nontradable goods. The third set of derivative restrictions establishes that a higher quality of life indeed raises individual utility but that it does not alter the relative utility tradeoff between the tradable and nontradable goods.

The equal profit condition is captured by

$$\Pi(w, \bar{r}; \text{productivity}) = \left\{ \max_{K,L} F(K, L; \text{productivity}) - wL - \bar{r}K \right\} = \overline{\Pi}, \qquad \text{(A.2)}$$
$$F_K(\cdot) > 0; F_L(\cdot) > 0; F_{\text{productivity}}(\cdot) > 0.$$

$\Pi(\cdot)$ represents a firm profit function which, given local wages and an exogenous interest rate, is the maximized value of firm production less its wage and interest bill. The derivative assumptions establish that the marginal products of capital and labor always remain positive and that higher productivity indeed raises output.

Normalizing the quantity of land to one, and assuming a unit flow of land services from each unit of land, a representative locality's resource constraint gives

$$nL = 1. \qquad \text{(A.3)}$$

Note that for the representative locality, L measures both population and population density. Generalizing to localities with different (fixed) quantities of land, L should be interpreted only as population density. For the analysis which follows, the key theoretical results are that

$$\frac{dw}{d\ \text{productivity}} > 0; \frac{dp}{d\ \text{productivity}} > 0; \frac{dL}{d\ \text{productivity}} > 0, \qquad \text{(A.4)}$$

$$\frac{dw}{d\ \text{quality of life}} = 0; \frac{dp}{d\ \text{quality of life}} > 0; \frac{dL}{d\ \text{quality of life}} > 0. \qquad \text{(A.5)}$$

To establish that dw/d productivity > 0, recognize that $\Pi(\cdot)$ is a profit function. Hence its derivatives with respect to input prices are negative: $\Pi_{w'}(\cdot) < 0$ and $\Pi_{\bar{r}}(\cdot) < 0$. Using the envelope theorem, we know that

$$\frac{d\Pi(\cdot)}{d\ \text{productivity}} = \frac{\partial\Pi(\cdot)}{\partial \text{productivity}} + \Pi_{w'}(\cdot)\frac{dw}{d\ \text{productivity}} = 0,$$
$$= F_{\text{productivity}}(\cdot) + \Pi_{w'}(\cdot)\frac{dw}{d\ \text{productivity}} = 0. \qquad \text{(A.6)}$$

By assumption $F_{\text{productivity}}(\cdot) > 0$. Hence d$w$/d productivity > 0.

To establish that

$$\frac{dp}{d\ \text{productivity}} > 0$$

recognize that $V(\cdot)$ is an indirect utility function. Hence its derivative with respect to its resource constraint will be positive, $V_{w'}(\cdot) > 0$, and its derivative with respect to the prices of utility arguments will be negative, $V_p(\cdot) < 0$. Taking the total derivative of $V(\cdot)$, setting this equal to zero, and rearranging gives

$$\frac{dp}{d\ \text{productivity}} = -\frac{V_{w'}(\cdot)}{V_p(\cdot)}\frac{dw}{d\ \text{productivity}} > 0. \qquad \text{(A.7)}$$

To establish that dw/d quality of life $= 0$, totally differentiate $\Pi(\cdot)$ and rearrange.

To establish dp/d quality of life > 0, the envelope theorem gives

$$\frac{dV(\cdot)}{d\text{ quality}} = \frac{\partial V(\cdot)}{\partial\text{ quality}} + V_p(\cdot)\frac{dp}{d\text{ quality}} = 0,$$
$$= u_{\text{quality}}(\cdot) + V_p(\cdot)\frac{dp}{d\text{ quality}} = 0. \tag{A.8}$$

By assumption $u_{\text{quality}}(\cdot) > 0$. Hence dp/d quality of life > 0.

Finally, to show that population density rises with increases in productivity and quality of life,

$$\frac{dL}{d\text{ productivity}} > 0 \quad \text{and} \quad \frac{dL}{d\text{ quality of life}} > 0.$$

By the economy resource constraint, (A.3), this is equivalent to showing that per capita land consumption drops with such changes,

$$\frac{dn}{d\text{ productivity}} < 0 \quad \text{and} \quad \frac{dn}{d\text{ quality of life}} < 0,$$

$$\frac{dV(\cdot)}{d\text{ productivity}} = u_c(\cdot)\frac{dc}{d\text{ productivity}} + u_n(\cdot)\frac{dn}{d\text{ productivity}} = 0. \tag{A.9}$$

$$\frac{dV(\cdot)}{d\text{ quality}} = u_{\text{quality}}(\cdot) + u_c(\cdot)\frac{dc}{d\text{ quality}} + u_n(\cdot)\frac{dn}{d\text{ quality}} = 0. \tag{A.10}$$

Individual utility maximization gives,

$$p = \frac{u_n(\cdot)}{u_c(\cdot)}. \tag{A.11}$$

Suppose that dn/d productivity > 0. By (A.9) it follows that dc/d productivity < 0. $u(\cdot)$ is such that $u_n(\cdot)/u_c(\cdot) = p$ must fall. But this violates that dp/d productivity > 0. Hence dn/d productivity < 0. The same argument using (A.10) establishes that dn/d quality of life < 0.

A caveat to the partial derivatives in (A.4) and (A.5) is that several rely on the exclusion of land from the production function, $F(\cdot)$. When land is included in the production function as in Roback (1982) and Gyourko and Tracy (1989, 1991), the derivative of the output-denominated wage with respect to quality of life, dw/d quality of life, is negative: in order to attain their reservation level of profits, firms pay a lower output-denominated wage as compensation for the higher output-denominated land price. With land excluded from the production function, the derivative with respect to quality of life of the output-denominated wage is zero but the derivative with respect to quality of life of the Hicksian, consumption-denominated real wage is negative.

More importantly, the positive derivative of population density with respect to productivity may not follow. Higher productivity causes an outward shift in both firms' and individuals' demand for land services (due, respectively, to an increase in the marginal product of land and the income effect of higher output-denominated wages). Together with the resulting increase in the price of land services, higher productivity may cause the actual aggregate quantity of land services purchased by firms and the per capita quantity of land services purchased by individuals to either increase or decrease. When land is absent from the production function, the price effect dominates the income effect implying that per capita land service consumption drops and hence population must increase. But if firms increase their aggregate use of land, then even a decrease in per capita land service consumption may not be sufficient to prevent a decrease in population. Hence the aggregate framework used herein may not be appropriate for examining the contribution from attributes that primarily increase the productivity of land-intensive industries. Numerical solutions, however, suggest that a positive derivative of population density with respect to productivity continues to hold even when production is quite land intensive (Rappaport, 2002a).

Data

Our choice of counties as the unit of observation is motivated by the near constancy of their borders across time. Constant borders allow for intertemporal comparisons between geographically fixed areas. Municipal and metropolitan area borders, in contrast, show considerable variation across time. A second benefit of constant borders is that they can be considered historically determined and therefore exogenous relative to most data generating processes.

To be sure, occasional changes in county borders do occur. Most frequently such changes take the form of the splitting of a county into two or more counties. Wherever possible, we have recombined such ''split'' counties to allow for intertemporal comparisons (based primarily on Horan and Hargis, 1995, and Thorndale and Dollarhide, 1987). The need for combining counties applies especially within US territories that had not yet been admitted to the union. North Dakota, South Dakota, Montana, and Washington were admitted as US states subsequent to the 1880 census; Idaho, Wyoming, and Utah were admitted subsequent to the 1890 census; Oklahoma, subsequent to the 1900 census; and New Mexico and Arizona subsequent to the 1910 census. The 1920 census is the first to include all 48 continental US states. (We include Washington DC as a county equivalent but exclude counties within the states of Alaska and Hawaii.) The combining of counties is limited to only those regressions for which it is needed. So for a county that subdivided in 1925, we combine the constituent parts for the 1920–1960 regressions but not for the 1960–2000 regressions.

A second type of adjustment we have made is the combining of counties to achieve geographic contiguity. Particularly in Virginia, there exist a number of ''independent cities'' completely surrounded by counties from which they are formally separate. We have merged these back into their surrounding counties.

Combined county weather, topography, and centroid values are calculated as a land weighted average of present-day constituent values.

Ocean and Great Lakes coasts and county boundaries are based on the 1 : 1.25 million ArcUSA Map constructed and distributed by ESRI Corporation (www.esri.com). For each county, the ESRI software package ArcView was used to calculate the distance to the nearest shoreline from the county's centroid (a mathematical approximation of "the center" of an irregular polygon). Note, therefore, that even counties with long coastal borders will generally have a strictly positive distance to the coast. Two Oregon counties, Douglas and Lane, actually border the ocean but have respective centroids 95 and 96 km inland and so are not classified as coastal.

Population and land area data are derived from various years of the US Department of Commerce's decennial census. These are disseminated in electronic form from several different sources listed in the bibliography. Employment and income data listed in Table 1 is from the US Department of Commerce's Bureau for Economic Analysis *Regional Economic Information System*, Tables CA-05 and CA-25. Capital income is the sum of dividends, interest, and rent received by individuals based on where they live. Age, education, and immigrant status data listed in Table 1 is from the 2000 Decennial Census, Summary File 3.

"Navigability" of rivers is based on a 1968 academic study of inland waterway commercial traffic and requires a minimum channel depth of 9 feet (Southern Illinois University at Carbondale, 1968). The inclusion of man-made canals within the navigable river category highlights one potential source of endogeneity. More generally, maintaining a river's navigability is a challenge requiring the continual attention of the US Army Corps of Engineers. To the extent that the funding for the maintenance of navigability may be correlated with population density, a reverse causal link will exist from population density to navigable river proximity. A second concern is selection bias: any navigable or potentially navigable river on which there was no commercial traffic in 1968 would be excluded from our navigable classification. Both of these concerns bias upward the coefficient on the navigable river dummy; hence the negative coefficient on this variable in many of the change regressions is likely to understate the decrease in the contribution to productivity and quality of life from navigable river proximity.

Major rivers (regardless of navigability) are made up of all rivers in the 1 : 25 million North America map from ESRI Corporation combined with a few navigable rivers that were not included. The ESRI map is constructed to include the longest rivers as well as shorter rivers which connect lakes to the ocean. The ESRI map additionally seeks "that the visual density of the rivers reflect, to a degree, the amount of flowing water present in a region." (ESRI, email correspondence with author, January 1, 1998.)

Historically navigable rivers shown in Rappaport and Sachs (2002) are based on the map of commercially navigated waterways in 1890 included in Fogel (1964). This map was then used to edit the "Major Water" shape file (distributed by ESRI, produced by Geographic Data Technology Incorporated) to remove river portions that are not part of the Fogel set. We further removed a handful of rivers included by Fogel that were only locally navigable (i.e., that did not afford the ability to navigate continuously to an ocean or Great Lakes coast).

Natural ports represent a subset of the seaports included in the World Port Index (US Naval Oceanographic Office, 1971). This catalogs all US Great Lakes and ocean seaports as well as some ports on navigable rivers. As the physical prerequisites for establishing a port on a navigable river are minimal, we exclude ports located more than 100 km from an ocean or Great Lakes coast. Proximity to the more inland ports is captured instead by the navigable river measure. The 100-km boundary allows cities that are usually considered to be seaports to be classified as such (e.g., Houston, 20 km inland; Philadelphia, 43 km inland; Portland Oregon, 85 km inland) while excluding cities more commonly considered to be river ports (e.g., Albany, 172 km inland; Memphis, 521 km inland).

The World Port Index classifies seaports by four size categories—very small, small, medium, and large—based on several applicable factors including area, facilities, and wharf space. We define "harbors" as medium or large seaports. To minimize the possibility of reverse causality we further exclude from our harbor measure any seaports that rely on constructed breakwaters or tide gates for shelter. The resulting "natural harbors" instead are distinguished by being sheltered from the wind and sea by virtue of a location within a natural coastal indentation or in the protective lee of an island, cape, or other natural barrier or by being located on a river adjoining the ocean. For the robustness check in Table 10, we continue to exclude seaports more than 100 km inland or that rely on constructed shelter.

Our weather variables are derived from data we have purchased from www.climatesource.com. The Climate Source data, in turn, is based on detailed weather observations over the period 1961 to 1990 from the more than 5,000 meteorological stations managed by the US National Oceanographic and Atmospheric Administration. A peer-reviewed hybrid statistical-geographical methodology developed by researchers at the Spatial Climate Analysis Service at Oregon State University is applied to such data to fit surfaces over a 1.25 arc minute (approximately 2 km) grid of the continental United States. The methodology includes considerable attention to accurately measuring highly-varying weather near coasts and mountains. County weather values are then constructed as the mean over all grid cells that lie within a county.

The specific six weather controls, each entered linearly and quadratically, were chosen from a much larger group of potential weather variables that *ex ante* seemed likely to affect productivity and quality of life (Rappaport, 2002b). January minimum temperature is the mean of January daily minimum temperature. July maximum heat index is a discomfort index combining the mean of July daily maximum temperature with the mean of July daily humidity. The remaining weather control variables—mean annual precipitation, mean annual days with precipitation of at least 0.1 inch, mean annual days temperature falls below 32 °F, and mean annual days temperature rises above 90 °F—are relatively self-explanatory.

Alternative weather variables discussed with respect to robustness are borrowed from Mendelsohn, Nordhaus, and Shaw (1994) who derive these based on meteorological station observations over the period 1951 through 1980. The temperature variables represent the average over these 30 years of mean daily temperature ((minimum temperature + maximum temperature)/2) in the months of January, April, July, and October. The precipitation variables represent average monthly precipitation in these same months. The actual county observations are fitted values for county geographic centers based on data from surrounding weather stations.

Compared to the 16 Mendelsohn et al. (1994) weather controls, the Climate Source weather controls used herein have slightly higher explanatory power for the population and employment density level regressions (Table 3) and the 1920–1960 change regressions (Table 4, columns (1)–(4)) but slightly lower explanatory power for the 1960–2000 change regressions (Table 4, columns (5)–(8)).

Our topography variable is constructed based on a 1.25 arc minute grid of United States altitude. The standard deviation of altitude across the grid cells within a county is divided by total county land area. The result, shown in Rappaport and Sachs (2002), nicely picks up mountainous topography, which is likely to serve as a hindrance to dense development.

Notes

1. Supporting the more general proposition that geography is an important determinant of economic activity, Davis and Weinstein (2001) argue that Japanese settlement patterns have remained relatively constant over millennia. Bloom and Sachs (1998) document the substantial drag on central African development posed by weather highly conducive to parasitic disease transmission and location at a latitude with low photosynthetic potential and hence low agricultural productivity.
2. Implicitly underpinning the two derivatives and the associated intuition is the empirical observation that net migration swamps natality and mortality as the primary source of US county population change. For the four decades 1950–1990, the pairwise correlation between county net migration and county population growth ranges from 0.948 to 0.959.
3. Color versions of maps are available for download from www.kc.frb.org/Econres/staff/jmr.htm.
4. So counties with centers within both 80 km of an ocean coast and 40 km of a navigable river are included in the ocean coast category; those with centers within both 80 km of a Great Lakes coast and 40 km of navigable river are included in the Great Lakes coast category.
5. In contrast to the classification underlying Figure 2 (panel (a)), such categorical dummies are not mutually exclusive. 389 counties have centers within 80 km of an ocean coast, 170 have centers within 80 km of a Great Lakes coast, and 508 have centers within 40 km of a river on which there was navigation in 1968 (Table 2). Of the latter navigable river counties, the centers of 49 also lie within 80 km of an ocean coast; the centers of 16 also lie within 80 km of a Great Lakes coast.
6. Note that designating rivers as navigable based on actual commercial usage in 1968 (rather than intrinsic capability of being navigated by commercial vessels) imparts an upward selection bias to the partial correlation between population density and such proximity, especially for the population density circa 1970. However, designating navigability based on Fogel's (1964) enumeration of rivers on which there was commercial navigation in 1890 results in a nearly identical time-series pattern of partial correlation coefficients, albeit one that is shifted down from that shown in Figure 2 (panel (b) (Rappaport and Sachs, 2002).
7. The change regression may even estimate $\widehat{d\beta_t^k}$ to be the opposite sign of $d\beta_t^k$. Consider the decreasing relative population density but increasing correlation coefficients associated with the navigable river counties in Figure 2. One reconciliation of the disparate trends goes as follows: Navigable river counties previously enjoyed a locational productivity advantage that no longer exists. In addition, such navigable river counties on average may have weather conditions that changing tastes and technology have made increasingly disadvantageous. Finally, the delayed population density adjustment downward by the navigable river counties may be proceeding slower than the delayed population density adjustment upward for counties with newly advantageous weather conditions. (Glaeser and Gyourko, 2001, suggest the durability of housing as the underlying mechanism for such an asymmetric response.) Lower fitted population density near navigable rivers due to changing coefficients on the weather variables would be offset by a rising coefficient on the navigable river's dummy (and hence $\widehat{d\beta_t^k} > 0$). Controlling for initial conditions as in (3b) helps to eliminate such a sign reversal (Table 4 below, columns (3) and (4) versus column (1)).
8. Implied expected population density magnitudes apply multiplicatively for counties proximate to both an ocean coast and a navigable river (49 such counties) or to both a Great Lakes coast and navigable river (16 such counties).
9. An advantage of using the change in log(1 + population density) rather than the change in log(population density) is that the former implicitly underweights counties with low population densities. We believe this is desirable given that idiosyncratic shocks (e.g. the migration choices of only a few individuals) could otherwise disproportionately affect results.

10. We choose to focus on the period 1920–2000, dividing this into two equal subperiods of forty years each. Figure 2 (panel (b)) along with decade-by-decade change regressions reported in Rappaport and Sachs (2002) show only a weak partial correlation between the change in population density and ocean and Great Lakes coastal proximity for the period 1880 and 1920. Dramatically fewer observations along with lower data quality prevent extending our analysis prior to 1880. To extend their analysis back further, Beeson et al. (2001) limit themselves to examining counties in the 27 states that constitute the US in 1840.
11. Coefficients on the weather and topography variables are shown in Rappaport and Sachs (2002).
12. Coefficients on the initial population densities spline and concentric total population variables are shown in Rappaport and Sachs (2002).
13. Supplementing the Table 4 (column (7)) regression with just the four 1960 education variables, the ocean coast coefficient falls only to 0.269 and statistically differs from zero at the 0.10 level; the Great Lakes coast coefficient rises to 0.515 and remains significant at the 0.05 level. Analogously, supplementing the Table 4 (column (3)) regression with just the six 1920 education variables, the ocean coast coefficient falls only to 0.586 and the Great Lakes coast coefficient rises to 0.651; both remain significant at the 0.05 level.
14. For four of these six decades, the statistical significance of a negative coefficient depends on the inclusion of the initial population density and concentric total population variables. The ability of the change regressions with these additional controls to pick up an intuitively appealing decreasing contribution to productivity and quality of life from navigable river proximity suggests that they should also be able to do so if such contributions from ocean and Great Lakes proximity were decreasing.
15. The Mendelsohn et al. (1994) variables—linear and quadratic temperature and precipitation in each of January, April, July, and October—were used to control for weather in an earlier version of the present paper (Rappaport and Sachs, 2001). We believe the weather variables in the current version better correspond to characteristics that may affect productivity and quality of life and are therefore less likely to absorb variation that should properly be attributed to coastal location (see data appendix).
16. The 1890–1900 change regression controlling for initial population density and concentric total population admits a quantitatively large, statistically significant negative coefficient on the ocean coast dummy (Rappaport and Sachs, 2002). Several other of the decade-by-decade change regressions admit quantitatively small negative coefficients on the ocean and Great Lakes dummies. For the most part such coefficients are swamped by their standard errors. But in a few cases, p-values for rejecting that coefficients equal zero fall below 0.15 (the 1890–1900 and 1930–1940 regressions controlling only for weather and topography and the 1880–1890 regression additionally controlling for initial density and concentric population.
17. Hence we infer that counties near the Pacific Coast enjoy extremely favorable weather and topography.
18. Compared to the regressions reported in Table 4 (columns (5)–(8)), the present coefficients on the Great Lakes coast dummy are somewhat larger in magnitude; the difference stems largely from the inclusion of the separate intercepts for each of the five coastal regions shown in Map 3. In other words, the positive coefficients in Table 4 show the additional expected annual growth rate from Great Lakes proximity compared to all continental US counties; the positive coefficients in Table 6 show the additional expected annual growth rate from Great Lakes proximity compared to remaining counties for which the Great Lakes is the closest coast.
19. As discussed in the data appendix, ocean natural harbor proxies can be located slightly inland from the piecewise linear boundary we are using to define ocean coasts. As a result, the number of counties with centers within 80 km of such proxies may exceed the number of counties with centers within 80 km of an ocean coast.

References

Beeson, P. E., D. N. DeJong, and W. Troesken. (2001). "Population Growth in U.S. Counties, 1840–1990," *Regional Science and Urban Economics* 31(6), 669–699.

Black, D., and V. Henderson. (2001). "Urban Evolution in the USA." Mimeo, University of California Irvine (July).

Blomquist, G. C., M. C. Berger, and J. P. Hoehn. (1988). "New Estimates of Quality of Life in Urban Areas," *American Economic Review* 78(1), 89–107.

Bloom, D., and J. Sachs. (1998). "Geography, Demography, and Economic Growth in Africa," *Brookings Papers on Economic Activity* 2.

Caselli, F., G. Esquivel, and F. Lefort. (1996). "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics," *Journal of Economic Growth* 1(3), 363–389.

Ciccone, A., and R. E. Hall. (1996). "Productivity and the Density of Economic Activity," *American Economic Review* 86(1), 54–70.

Conley, T. G. (1999). "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics* 92(1), 1–45.

Davis, D. R., and D. E. Weinstein. (2001). "Bones, Bonds, and Break Points: The Geography of Economic Activity." Mimeo, Columbia University.

Dixit, A. (1973). "Models of Dual Economies." In J. A. Mirrlees and N. H. Stern (eds.), *Models of Economic Growth*. New York: Wiley & Sons.

Drazen, A., and Z. Eckstein. (1988). "On the Organization of Rural Markets and the Process of Economic Development," *American Economic Review* 78(3), 431–443.

Fogel, R. W. (1964). *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins Press.

Fujita, M., P. Krugman, and A. J. Venables. (1999). *The Spatial Economy*. Cambridge, MA: The MIT Press.

Fujita, M., and T. Mori (1996). "The Role of Ports in the Making of Major Cities: Self-Agglomeration and Hub-Effect," *Journal of Development Economics* 49, 93–120.

Gallup, J. L., and J. D. Sachs. (1998). "Geography and Economic Development," Consulting Assistance on Economic Reform II Discussion Paper No. 39 (March). Available for download from http://www2.cid.harvard.edu/cidpapers/caer/paper39.pdf.

Glaeser, E. L., and J. Gyourko. (2001). "Urban Decline and Durable Housing" National Bureau for Economic Research Working Paper 8598, November.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer. (1992). "Growth in Cities," *Journal of Political Economy* 100(6), 1126–1152.

Glaeser, E. L., J. A. Scheinkman, and A. Shleifer. (1995). "Economic Growth in a Cross-Section of Cities," *Journal of Monetary Economics* 36, 117–143.

Gyourko, J., and J. Tracy. (1989). "The Importance of Local Fiscal Conditions in Analyzing Local Labor Markets," *Journal of Political Economy* 97(5), 1208–1231.

Gyourko, J., and J. Tracy. (1991). "The Structure of Local Public Finance and the Quality of Life," *Journal of Political Economy* 99(4), 774–806.

Hanson, G. H. (2001). "Market Potential, Increasing Returns, and Geographic Concentration." Mimeo, University of California San Diego (December).

Harris, J. R., and M. P. Todaro. (1970). "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review* 60(1), 126–142.

Haurin, D. R. (1980). "The Regional Distribution of Population, Migration, and Climate," *Quarterly Journal of Economics* 95(2), 293–308.

Henderson, V. (1974). "The Sizes and Types of Cities," *American Economic Review* 64(4), 640–656.

Henderson, V. (1988). *Urban Development: Theory, Fact, and Illusion*. New York: Oxford University Press.

Henderson, V., A. K. Ari, and M. Turner. (1995). "Industrial Development in Cities," *Journal of Political Economy* 103(5), 1067–1085.

Horan, P. M., and P. G. Hargis. (1995). "County Longitudinal Template, 1840–1990." ICPSR Study No. 6576 [computer file]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. Corrected and amended by Patricia E. Beeson and David N. DeJong, Department of Economics, University of Pittsburgh, 2001.

Hummels, D. (1999). "Have International Transportation Costs Declined?" Mimeo, University of Chicago Graduate School of Business (November).

Islam, N. (1995). "Growth Empirics: a Panel Data Approach," *Quarterly Journal of Economics* 110(4), 1127–1170.

Jorgenson, D. W. (1961). "The Development of a Dual Economy," *Economic Journal* 71(282), 309–334.

Krugman, P. (1991). "Increasing Returns and Economic Geography," *Journal of Political Economy* 99, 483–399.

Lewis, A. W. (1954). "Economic Development with Unlimited Supplies of Labour," *The Manchester School of Economics and Social Studies* 22, 139–191.

Mendelsohn, R., W. D. Nordhaus, and D. Shaw. (1994). "The Impact of Global Warning on Agriculture: a Ricardian Analysis," *American Economic Review* 84(4), 753–771.
Poor, H. V. (1860). *History of the Railroads and Canals of the United States of America*. Reprint, New York: Augustus M. Kelley: 1970.
Ranis, G., and J. C. H. Fei. (1961). "A Theory of Economic Development," *American Economic Review* 51(4), 533–565.
Rappaport, J. (2000a). "Why Are Population Flows So Persistent?" Federal Reserve Bank of Kansas City Working Paper 99-13 (August).
Rappaport, J. (2000b). "Is the Speed of Convergence Constant?" Federal Reserve Bank of Kansas City, Working Paper 00-10 (December).
Rappaport, J. (2002a). "Why Are Some Cities so Crowded?" Mimeo, Federal Reserve Bank of Kansas City.
Rappaport, J. (2002b). "Moving to Nice Weather." Mimeo, Federal Reserve Bank of Kansas City.
Rappaport, J., and J. D. Sachs. (2001). "The U.S. As a Coastal Nation," Federal Reserve Bank of Kansas City Working Paper 01-11 (November), first working version. Available by request from jordan.m.rappaport@kc.frb.org
Rappaport, J., and J. D. Sachs. (2002). "The U.S. As a Coastal Nation: Supplemental Tables, Figures, and Maps," Available for download from www.kc.frb.org/Econres/staff/jmr.htm
Roback, J. (1982). "Wages, Rents, and the Quality of Life," *Journal of Political Economy* 90(6), 1257–1278.
Rosen, S. (1979). "Wage-Based Indexes of Urban Quality of Life." In Miezkowski and Straszheim (eds), *Current Issues in Urban Economics*. Baltimore: Johns Hopkins University Press.
Sokoloff, K. L. (1988). "Inventive Activity in Early Industrial America: Evidence from Patent Records, 1790–1846," *Journal of Economic History* 48(4), 813–850.
Southern Illinois University at Carbondale, Transportation Institute. (1968). "A Study of River Ports and Terminals," Washington D.C.: U.S. Department of Commerce, National Bureau of Standards.
Stover, M. E., and C. L. Leven. (1992). "Methodological Issues in the Determination of the Quality of Life in Urban Areas," *Urban Studies* 29(5), 737–754.
Tanner, H. S. (1840). *A Description of the Canals and Railroads of the United States*. Reprint, New York: Augustus M. Kelley: 1970.
Thorndale, W., and W. Dollarhide. (1987). *Map Guide to the Federal Censuses, 1790–1920*. Baltimore: Genealogical Publishing Company.
Tiebout, C. M. (1956). "A Pure Theory of Local Expenditures," *Journal of Political Economy* 64(5), 416–424.
US Department of Commerce, Bureau of the Census. Decennial Census reports, 1890–1960 [investigator]. *Historical, Demographic, Economic, and Social Data: The United States, 1790–1970*, ICPSR Study No. 3 (computer file). Ann Arbor, MI: Inter-university Consortium for Political and Social Research (producer and distributor), 1992. Corrected and amended by Michael R. Haines, Department of Economics, Colgate University, 2001.
US Department of Commerce, Bureau of the Census. *Census of Population and Housing, 1970: Fourth Count Population Summary Tape, File C* (computer file). Washington, DC: US Department of Commerce, Bureau of the Census (producer), 1982. Cambridge, MA: National Bureau for Economic Research (distributor). Also available as ICPSR Study No. 8107.
US Department of Commerce, Bureau of the Census. *Census of Population and Housing, 1970: Fourth Count Housing Summary Tape, File C* (computer file). Washington, DC: US Department of Commerce, Bureau of the Census (producer), 1982. Cambridge, MA: National Bureau for Economic Research (distributor). Also available as ICPSR Study No. 8129.
US Department of Commerce, Bureau of the Census. *Census of Population and Housing, 1980: Summary Tape File 3C*, ICPSR Study No. 8038 (computer file). Washington, DC: US Department of Commerce, Bureau of the Census (producer), 1982. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor), 1992.
US Department of Commerce, Bureau of the Census. *Census of Population and Housing, 1990: Summary Tape File 3C*, ICPSR Study No. 6054 (computer file). Washington, DC: US Department of Commerce, Bureau of the Census (producer), 1992. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor), 1993.
US Department of Commerce, Bureau of the Census. "Census 2000 PHC-T-4. Ranking Tables for Counties: 1990

and 2000'' [computer file]. Internet Release date: April 2, 2001. Available for download from http://blue.census.gov/population/cen2000/phc-t4/tab01.xls

US Department of Commerce, Bureau of the Census. ''2000 Census of Population and Housing'' [national computer file]. Internet Release date: September 25, 2002. Available for download from http://www2.census.gov/census_2000/datasets/Summary_File_3/0_National

US Department of Commerce, Bureau of Economic Analysis. ''CA05: Personal Income by Major Source and Earnings by Industry,'' (computer file). *REIS Regional Economic Information System, 1969-2000* CD-Rom (RCN-0295).

US Department of Commerce, Bureau of Economic Analysis. ''CA25: Full-time and Part-time Employment by Industry,'' (computer file). *REIS Regional Economic Information System, 1969–2000* CD-Rom (RCN-0295).

US Naval Oceanographic Office. (1971). *World Port Index*.Publication 150, 4th edn. Washington: US Government Printing Office.

Wright, G. (1990). ''The Origins of American Industrial Success, 1879–1940,'' *American Economic Review* 80(4), 651–668.