

# Lecture 1

## Introduction

Nurgul Tilenbaeva

American University - Central Asia

12.09.2022

# Contents

- 1 What is Econometrics?
- 2 Structure of Economic Data
- 3 Causality and the Notion of Ceteris Paribus

# What is Econometrics?

## Econometrics

the application of statistical methods to the study of economic data and problems (Merriam-Webster)

the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference (Samuelson, Koopman and Stone (1954))

Focus on **nonexperimental** / **observational** / **retrospective** data

# What is Econometrics?

Examples of problems that can be studied:

- 1 Evaluate the effectiveness of "Bezopasnyi gorod" program  
What, if any, effect the installation of cameras has on the number of car accidents in Bishkek?
- 2 Forecast the inflation rate for the next quarter in Kyrgyzstan
- 3 Test if the recipients of "Gold certificate" in the Republic Testing have higher earnings than their counterparts

# What is Econometrics?

## Empirical analysis

uses **data** to test a theory or to estimate a relationship

## Economic model

a simplified representation of economic processes through mathematical equations

## Econometric model

a statistical model that helps testing an economic model

# What is Econometrics?

## Economic model of crime (Gary Becker)

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

where

$y$ =hours spent in criminal activities

$x_1$ ="wage" for an hour spent in criminal activity

$x_2$ =hourly wage in legal employment

$x_3$ =income other than from crime or employment

$x_4$ =probability of getting caught

$x_5$ =probability of being convicted if caught

$x_6$ =expected sentence if convicted

$x_7$ =age

# What is Econometrics?

## Econometric model of crime

*crime* =

$$\beta_0 + \beta_1 \text{wage}_m + \beta_2 \text{othinc} + \beta_3 \text{freqarr} + \beta_4 \text{freqconv} + \beta_5 \text{avgsen} + \beta_6 \text{age} + u$$

where

*crime* = some measure of the frequency of criminal activity

*wage<sub>m</sub>* = the wage that can be earned in legal employment

*othinc* = the income from other sources (assets, inheritance, and so on)

*freqarr* = the frequency of arrests for prior infractions (to approximate the probability of arrest)

*freqconv* = the frequency of conviction, and

*avgsen* = the average sentence length after conviction

# What is Econometrics?

$u$  - **error term** or **disturbance term**

- unobserved factors
- errors in measurement

$\beta_0, \beta_1, \dots, \beta_6$  - **parameters**

- describe the directions and strengths of the relationship between crime and the factors used to describe crime in the model



# What is Econometrics?

## Hypothesis:

$wage_m$  has no effect on criminal behavior:  $\beta_1 = 0$

# Structure of Economic Data

- 1 Cross-sectional data
- 2 Time series data
- 3 Pooled cross sections
- 4 Panel or longitudinal data

# Structure of Economic Data

## Cross-sectional data set

consists of a sample of individuals, households, firms, cities, states, countries, or a variety of other units, taken at a given point in time

**TABLE 1.1 A Cross-Sectional Data Set on Wages and Other Individual Characteristics**

obsno	wage	educ	exper	female	married
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
525	11.56	16	5	0	1
526	3.50	14	5	1	0

# Structure of Economic Data

## Cross-sectional data

- we can often assume **random sampling** from the underlying population, which simplifies the analysis of cross-sectional data.
- ordering of the data does not matter for econometric analysis.

# Structure of Economic Data

## Time series data set

consists of observations on a variable or several variables over time

**TABLE 1.3** Minimum Wage, Unemployment, and Related Data for Puerto Rico

obsno	year	avgmin	avgcov	prunemp	prgnp
1	1950	0.20	20.1	15.4	878.7
2	1951	0.21	20.7	16.0	925.0
3	1952	0.23	22.6	14.8	1015.9
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
37	1986	3.35	58.1	18.9	4281.6
38	1987	3.35	58.2	16.8	4496.7

# Structure of Economic Data

## Time series data

- economic observations are **not** independent across time.
- there may be **seasonality** in the data.
- the chronological ordering of observations matters.

# Structure of Economic Data

## Pooled cross sections

- have both cross-sectional and time series features.
- use **different** samples at different periods of time

TABLE 1.4 Pooled Cross Sections: Two Years of Housing Prices

obsno	year	hprice	proptax	sqrft	bdrms	bthrms
1	1993	85500	42	1600	3	2.0
2	1993	67300	36	1440	3	2.5
3	1993	134000	38	2000	4	2.5
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
250	1993	243600	41	2600	4	3.0
251	1995	65000	16	1250	2	1.0
252	1995	182400	20	2200	4	2.0
253	1995	97500	15	1540	3	2.0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
520	1995	57200	16	1100	2	1.5

# Structure of Economic Data

## Pooled cross sections

- often used to increase sample size.
- usually analyzed like a standard cross section.
- useful way to analyze the effects of a new government policy by comparing outcomes before and after the change.



# Structure of Economic Data

## Panel or longitudinal data set

- consists of a time series for **each** cross-sectional member in the data set
- the **same** cross-sectional units (individuals, firms, or countries) are followed over a given time period

TABLE 1.5 A Two-Year Panel Data Set on City Crime Statistics

obsno	city	year	murders	population	unem	police
1	1	1986	5	350000	8.7	440
2	1	1990	8	359200	7.2	471
3	2	1986	2	64300	5.4	75
4	2	1990	1	65100	5.5	75
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
297	149	1986	10	260700	9.6	286
298	149	1990	6	245000	9.8	334
299	150	1986	25	543000	4.3	520
300	150	1990	32	546200	5.2	493

# Structure of Economic Data

## Panel or longitudinal data

- the ordering in the cross section of a panel data set does not matter.

### **BUT!**

- the ordering in the time series of a panel data set matters.
- allows controlling for certain unobserved characteristics of individuals, firms, and so on.
- possibility for **causal** inference.
- allows studying the importance of lags in behavior.
- difficult to get such data.

# Causality and the Notion of Ceteris Paribus

We are often interested in the **causal effect** of one variable (such as education) on another variable (earnings).

- No problem in experiments.
- But experiments are hard to perform in economics.
- Econometric methods can simulate a ceteris paribus experiment.

## Ceteris paribus

"other (relevant) factors being equal", or "holding other things constant"