

INTRODUCTION

Hüseyin Taştan¹

¹Yıldız Technical University
Department of Economics

These presentation notes are based on
Introductory Econometrics: A Modern Approach (2nd ed.)
by J. Wooldridge.

14 Ekim 2012

What is econometrics?

- ▶ Literal meaning: **economic measurement**: “econo-metrics”.
- ▶ But the scope of econometrics is much wider.
- ▶ Two popular definitions of econometrics:
- ▶ “*Econometrics may be defined as the social science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena*”. (A.S. Goldberger, 1964).
- ▶ “*...econometrics may be defined as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of science*”. (P. Samuelson)

Scope of Econometrics

- ▶ Developing statistical methods for the estimation of economic relationships,
- ▶ Testing economic theories and hypothesis,
- ▶ Evaluating and applying economic policies,
- ▶ Forecasting,
- ▶ Collecting and analyzing nonexperimental or observational data.

Components of Econometrics

- ▶ Econometric inputs:
 - ▶ Economic Theory
 - ▶ Mathematics
 - ▶ Statistical Theory
 - ▶ Data
 - ▶ Computers (CPU power)
 - ▶ Interpretation
- ▶ Econometric outputs:
 - ▶ Estimation - Measurement
 - ▶ Inference - Hypothesis testing
 - ▶ Forecasting - Prediction
 - ▶ Evaluation

Why Do We Need Econometrics?

- ▶ We learned statistical methods so why do we need a separate discipline?
- ▶ The reason is as follows: econometrics focuses on the analysis of nonexperimental economic data.
- ▶ **Nonexperimental data** (or **observational data**) are not obtained through controlled experiments on economic agents (consumers, firms, households, sectors, countries, etc.)
- ▶ **Experimental data** are collected in laboratory environments in natural sciences.
- ▶ Although some social experiments can be devised it is usually impossible to conduct economic experiments.
- ▶ Unlike statistical methods employed in natural sciences, econometrics develops special methods to handle nonexperimental data.

Classical Methodology in Econometrics

- ▶ Formulation of theory or hypothesis,
- ▶ Specification of economic (mathematical) model,
- ▶ Specification of econometric model,
- ▶ Collecting data,
- ▶ Estimation of parameters,
- ▶ Hypothesis tests,
- ▶ Forecasting/Prediction),
- ▶ Evaluation of results for policy analysis or decision making. (Gujarati, p.3)

ECONOMIC MODEL

Example 1 - Economic Model of Crime

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

f functional form (not yet specified)

Description of variables

- ▶ y : hours spent in criminal activities,
- ▶ x_1 : "earnings" for an hour spent in criminal activity,
- ▶ x_2 : hourly wage in legal employment,
- ▶ x_3 : other income,
- ▶ x_4 : probability of getting caught,
- ▶ x_5 : probability of being convicted if caught,
- ▶ x_6 : expected sentence if convicted,
- ▶ x_7 : age

ECONOMIC MODEL vs. ECONOMETRIC MODEL

Economic Model Example 2 - Job Training and Worker Productivity

$$wage = f(educ, exper, training),$$

$wage$: hourly wage (in dollars)

$educ$: level of education (in years)

$exper$: level of workforce experience (in years)

$training$: weeks spent in job training.

Econometric Model: f Linear specification

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 training + u$$

ECONOMETRIC MODEL: Linear Specification

Econometric Model Example 2 - Job Training and Worker Productivity

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 training + u$$

Components of econometric model:

- ▶ **u : random error term or disturbance term**
- ▶ Random error term u contains influence of factors that are not included in the model. It also contains unobserved factors such as innate ability or family background.
- ▶ No matter how comprehensive the specified model there will always factors that cannot be included in the econometric model. We can never eliminate u entirely.

ECONOMETRIC MODEL: Linear Specification

Econometric Model Example 2 - Job Training and Worker Productivity

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 training + u$$

Components of econometric model:

- ▶ $\beta_0, \beta_1, \beta_2, \beta_3$: parameters of the econometric model
- ▶ These are unknown constants.
- ▶ They describe the directions and strengths of the relationship between wage and factors affecting wage included in the model.
- ▶ For example, we may be interested in testing $H_0 : \beta_3 = 0$ which says that job training has no effect on wage.

ECONOMETRIC DATA TYPES

- ▶ **Cross-sectional data**
- ▶ **Time series data**
- ▶ **Pooled cross-section**
- ▶ **Panel data (longitudinal data)**

ECONOMETRIC DATA TYPES

Cross-sectional data:

- ▶ consists of a sample of individuals, households, firms, cities, states, countries, or a variety of other units, taken at a given point in time
- ▶ Significant feature: random sampling from a target population
- ▶ Generally obtained through official records of individual units, surveys, questionnaires (data collection instrument that contains a series of questions designed for a specific purpose)
- ▶ For example, household income, consumption and employment surveys conducted by the Turkish Statistical Institute (TUIK/TURKSTAT)

Cross-sectional data example: Wage Data (GRETl data set: wage1.gdt)

A Cross-sectional data set on wages and individual characteristics

Obs. No	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
6	8.75	16	9	0	1
.
.
.
524	4.67	15	13	0	1
525	11.56	16	5	0	1
526	3.50	14	5	1	0

ECONOMETRIC DATA TYPES

Time series data:

- ▶ consists of observations on a variable or several variables over time.
- ▶ Chronological ordering
- ▶ Frequency of time series data: hour, day, week, month, year
- ▶ Time length between observations is generally equal
- ▶ Examples of time series data include stock prices, money supply, consumer price index, gross domestic product, annual homicide rates, and automobile sales figures.

A Time Series Data Example: GRETl: prminwage.gdt

Minimum Wage, Unemployment, and Related Data for Puerto Rico

<i>obsno</i>	<i>year</i>	<i>avgmin</i>	<i>avgcov</i>	<i>unemp</i>	<i>gnp</i>
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

ECONOMETRIC DATA TYPES

Pooled cross-section:

- ▶ consists of cross-sectional data sets that are observed in different time periods and combined together
- ▶ At each time period (e.g., year) a different random sample is chosen from population
- ▶ Individual units are not the same
- ▶ For example if we choose a random sample 400 firms in 2002 and choose another sample in 2010 and combine these cross-sectional data sets we obtain a pooled cross-section data set.
- ▶ Cross-sectional observations are pooled together over time.

A Pooled Cross-sectional Data Example

Pooled Cross Sections: Two Years of Housing Prices

<i>obsno</i>	<i>year</i>	<i>hprice</i>	<i>proptax</i>	<i>sqft</i>	<i>bdrms</i>	<i>bthrms</i>
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

18

ECONOMETRIC DATA TYPES

Panel Data (longitudinal data):

- ▶ consists of a time series for each cross-sectional member in the data set.
- ▶ The same cross-sectional units (firms, households, etc.) are followed over time.
- ▶ For example: wage, education, and employment history for a set of individuals followed over a ten-year period.
- ▶ Another example: cross-country data set for a 20 year period containing life expectancy, income inequality, real GDP per capita and other country characteristics.

A Panel Data Example

A Two-Year Panel Data Set on City Crime Statistics

<i>obsno</i>	<i>city</i>	<i>year</i>	<i>murders</i>	<i>population</i>	<i>unem</i>	<i>police</i>
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

20

Causality and the Notion of Ceteris Paribus

- ▶ In testing economic theory usually our goal is to infer that one variable has a **causal effect** on another variable.
- ▶ Correlation may be suggestive but cannot be used to infer causality.
- ▶ Fundamental notion: **Ceteris paribus**: “other relevant factors being equal”
- ▶ Or “holding all other factors fixed”
- ▶ Most economic questions are ceteris paribus by nature.
- ▶ For example, in analyzing consumer demand, we are interested in knowing the effect of changing the price of a good on its quantity demanded, while holding all other factors (such as income, prices of other goods, and individual tastes) fixed.
- ▶ If other factors are not held fixed, then we cannot know the causal effect of a price change on quantity demanded.

Causality and the Notion of Ceteris Paribus

- ▶ Therefore, the relevant question in econometric analysis is “do we control sufficient number of factors?”
- ▶ Are there other factors that are not included in the model?
- ▶ Can we say that other components are held fixed?
- ▶ In most serious applications the number of factors is immense so the isolation of the effect of any particular variable may seem hopeless. But, if properly used, econometric methods can help us determine ceteris paribus effects.

Ceteris Paribus Example: Effects of Fertilizer on Crop Yield

- ▶ Suppose the crop is wheat. We are interested in measuring the impact of fertilizer on wheat yield (production).
- ▶ Obviously there are several factors that affect the production of wheat such as rainfall, quality of soil and presence of parasites.
- ▶ We need to control these factors in order to determine the ceteris paribus impact of fertilizers.
- ▶ To do this we can devise the following experiment: divide the land into equal pieces (such as one acre) and apply different amounts of fertilizer to each land plot and then measure the wheat yield.
- ▶ This gives us a cross-sectional data set where observation unit is land plot.
- ▶ We can apply statistical methods to this data set to measure the impact of fertilizers on crop yield.

Ceteris Paribus Example: Effects of Fertilizer on Wheat Yield

- ▶ How do we know the results of this experiment can be used to measure the ceteris paribus effect of fertilizer?
- ▶ Can we be sure that all other factors (quality of land plots for example) are held fixed?
- ▶ It is generally very difficult to observe the quality of soil.
- ▶ But we can still use ceteris paribus notion
- ▶ Amounts of fertilizers should be assigned to land plots independently of other plot features such as quality of plots
- ▶ In other words, other characteristics of plots should be ignored when deciding on fertilizer amounts.

Ceteris Paribus Example: Measuring the Return to Education

- ▶ Question: How can we measure the return to education?
- ▶ If a person is chosen from the population and given another year of education, by how much will his or her wage increase?
- ▶ This is also a ceteris paribus question: all other factors are held fixed while another year of education is given to the person.
- ▶ There are several factors other than education that affect wages: experience, tenure, innate ability, gender, age, region, marital status, etc.

Example: Measuring the Return to Education

- ▶ Similar to fertilizer example we can design the following hypothetical experiment:
- ▶ Social planner has the ability to assign any level of education to any person.
- ▶ The planner chooses a group of individuals from population and randomly assign each person an amount of education: some are given high school education, some are given 4-year college education, etc.
- ▶ Subsequently the planner measures wages for each individual.
- ▶ If levels of education are assigned independently of other characteristics that affect productivity (such as innate ability or experience) then we can measure the impact of education on wages correctly.
- ▶ Of course such an experiment is impossible to conduct.
- ▶ Even though we cannot obtain an experimental data, we can obtain observational data set that contains information on wages, education, experience and other personal characteristics (e.g. from TUIK household employment surveys)

Ceteris Paribus Example: The Effect of Law Enforcement on City Crime Levels

- ▶ Does the presence of more police officers on the street deter crime?
- ▶ Ceteris paribus question: If a city is randomly chosen and given, say, ten additional police officers, by how much would its crime rates fall?
- ▶ Or: If two cities are the same in all respects, except that city A has ten more police officers than city B, by how much would the two cities' crime rates differ?

Ceteris Paribus Example: Measuring the Return to Education

- ▶ People choose their education levels. Thus, individual characteristics will be correlated with the level of education.
- ▶ For example, people with more innate ability tend to have higher levels of education.
- ▶ Workers with higher levels of education tend to have higher wages.
- ▶ It becomes difficult to isolate the impact of education from the impact of innate ability on wages.
- ▶ How much of this effect comes from education? How much from innate ability?

Ceteris Paribus Example: The Effect of Law Enforcement on City Crime Levels

- ▶ It almost impossible to find two cities identical in all respects.
- ▶ But this is not necessary in econometric analysis.
- ▶ We just need to know if the data on crime rates and number of police officers can be viewed as experimental.
- ▶ In most cases this is not the case, data is observational.
- ▶ The size of police force is determined by city authorities who probably take into account several other city characteristics.
- ▶ The problem is a little bit more complex: Does the size of police force affect the amount of crime or vice versa?
- ▶ The amount of crime and police force are simultaneously determined.