Econometrics 1

Lecture 1. Introduction

Alfa Farah

Department of Economics
Diponegoro University

Updated February 15, 2022

Fakultas Ekonomika dan Bisnis
UNIVERSITAS DIPONEGORO

What we are learning

- Assessing Causal Relationship between Two Economic Variables
- What is Econometrics
  - Definition
  - Components of Econometrics
  - Typical goals of econometric analysis
  - Typical Questions
- Steps in Empirical Econometrics Analysis
- The Structure of Econometrics Data
  - Cross-sectional Data
  - Time Series Data
  - Pooled Cross Sections
  - Panel or Longitudinal Data
- Causality and the Notion of Ceteris Paribus in Econometric Analysis

Reading

Gruber, Public Finance and Public Policy, 5ed (Ch. 3)

Wooldridge, Jeffrey M. 2019. *Introductory Econometrics: A Modern Approach, 6ed.* (Ch. 1)

Ch. 1. The Nature of Econometrics and Economic Data

1-1 What is Econometrics?

1-2 Steps in Empirical Econometrics Analysis

1-3 The Structure of Economic Data

1-4 Causality and the Notion of Cateris Paribus in Econometric Analysis

Example 1.1: Economic Model of Crime

Assessing Causal Relationship

- In courses such as macroeconomics and microeconomics, we learn some theoretical models to understand how individuals and firms behave.
- By making relatively straightforward assumptions about how individuals and firms behave, we are able to address complicated questions such as:
  - How changes in prices affect the quantity demanded
  - How BLT program affect labour supply
  - How an increase in cigarette tax lowers smoking among youth
- While theoretical models help us to answered these questions, we have been very imprecise about the potential size of the effect
  - We can say: when the government increases income tax, consumption decreases. However, we cant say: how many percent consumption decreases when the government increases tax by 5%.

Assessing Causal Relationship

- In formulating a policy, it is very crucial to be able to precisely predict the causal effect of the policy
- To measure the potential size of the causal effect, we use data and statistical methods $\Rightarrow$ This means that we conduct an empirical analysis or we employ empirical methods to measure the effect
- The general problem that empirical economists face in trying to use existing data to assess the causal effect of one factor on another is that one might mistake a correlation as a causation.
  - Correlation: Two economic variables are correlated if they move together (example: height and weight across individuals)
  - Causality: Two economic variables are causally related if the movement of one causes movement of the other (example: good nutrition as an infant increases adult height)

Assessing Causal Relationship

- For any correlation between two variables A and B, there are three possible explanations, one or more of which could result in the correlation:
  1. A is causing B
  2. B is causing A
  3. Some third factor is causing both
- The question is: how can we draw causal conclusions about relationships between correlated variables?

Assessing Causal Relationship

- The gold standard of testing for causality: randomized trials.
- Randomized trials involve taking a group of volunteers and randomly assigning them to either a treatment group or a control group.
  - Treatment group: the set of individuals who are subject to an intervention being studied, such government benefit.
  - Control group: the set of individuals who are not given the intervention

Assessing Causal Relationship

Example: Benjamin A. Olken (2007) "Monitoring Corruption: Evidence from a Field Experiment in Indonesia"

- The question: "Does increasing government audit on village projects reduce corruption?"
- The field data was collected between September 2003 and August 2004 from selected 608 villages in Jawa Tengah and Jawa Timur
- At the time the study started, each village in the study was about to start building a village road as part of a nationwide village-level infrastructure project.
- Some of villages were told that their project would be audited by BPKP (this is the treatment group) and some were not (this is the control group)

Assessing Causal Relationship

Example: Benjamin A. Olken (2007) "Monitoring Corruption: Evidence from a Field Experiment in Indonesia"

- Result: increasing the probability of an external government audit in those villages from a baseline of about 4 percent to essentially 100 percent reduced missing expenditures, as measured by discrepancies between official project costs and an independent engineers' estimate of costs

Assessing Causal Relationship

Problems with the randomized trials

1. External validity: the results are only valid for the sample of individuals who volunteer to be either treatments or controls, and this sample may be different from the population at large (e.g., randomized experiment in Sweden or US would not necessarily generate the same results)
2. Attrition: Individuals may leave the experiment before it is complete.
   - Reduction in the size of samples over time, which, if not random, can lead to bias estimates.
3. Costly!

With these problems, is there any other methods available that can allow us to approach the gold standard of randomized trials?

Assessing Causal Relationship

- We can use observational data to asses causal relationship
- Observational data: data generated by individual behaviour observed in the real world, not in the context of deliberately designed experiments (e.g. data collected from BPS such as GDP, population, education attainment)
- With observational data, we can do
  - Quasi-experiment (also called natural experiments): changes in the economic environment that create nearly identical treatment and control groups for studying the effect of that environmental change
  - Regression: statistical tools to asses the relationship between variables $\Rightarrow$ the focus of this course!
- Economists study Econometrics to learn how to use regression analysis to understand economic phenomena

What is Econometrics?

$$\begin{aligned} \text{Econometrics} &= \text{Econo} + \text{Metrics} \\ &= \text{Economics} + \text{Measurements} \end{aligned}$$

"is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy." (Wooldridge, 2013, pg.1)

- Economics answers the questions of how and why whereas Econometrics answers the questions of how much
  - Economics: If the government increases income tax, consumption decreases.
  - Econometrics: If the government increases income tax by 5%, consumption decreases by 1%.

What is Econometrics

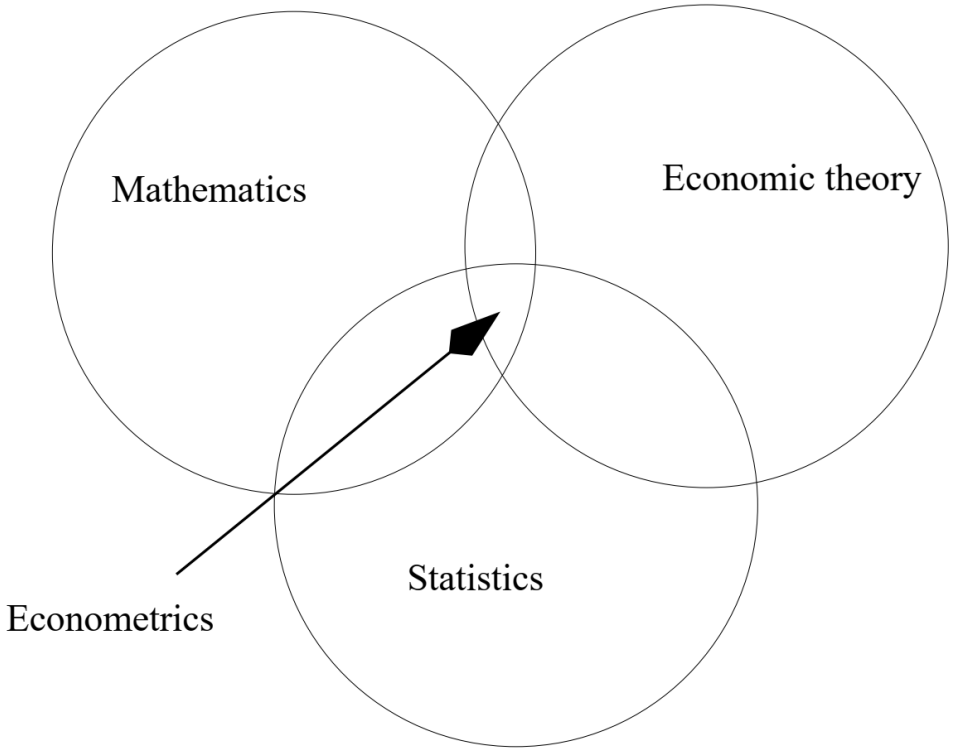
- Since regression is a statistical tool, why do we need econometrics? why is statistics not sufficient?

> "Econometrics by no means the same as economic statistics" (Ragnar Frisch, 1895–1973, one of the founders of the Econometric Society)

- Neither theory nor measurement on its own is sufficient to further our understanding of economic phenomena. We need both theory and measurement. Measurement without theory is unlikely to provide a satisfactory explanation of the way economic forces interact with each other.
- What further distinguishes Econometrics from statistics is that Econometrics focuses to establish causal relationship. So, econometrics emphasizes endogeneity problem and identification problem

What is Econometrics?

Components of Econometrics



- We derive testable hypotheses from economic theories.
- We then collect data and use Econometrics and Statistics to test the hypothesis and draw inference
- If a hypothesis cannot be rejected, the theory can be viewed as not refuted

What is Econometrics?

Typical goals of econometric analysis

Typical goals of econometric analysis

- Stating (causal) relationships between economic variables
- Estimating economic theories and hypotheses
- Forecasting economic variables
- Evaluating and implementing government and business policy

What is Econometrics?

Typical Questions

Descriptive questions

- How much do men and women earn annually on average in Indonesia?
- How does district's primary school enrolment on average vary with the district's income?
- How does a party's vote share vary with campaign spending?

What is Econometrics?

Typical Questions

Forecasting questions

- What will the global temperature be in 2050?
- What will be the Golkar's vote share in the next election?
- What will the stock price of Samsung be on 2nd February?

Applications of econometrics: forecasts of macroeconomic indicators (interest rates, inflation, GDP etc.) $\Rightarrow$ Macroeconometrics

What is Econometrics?

Typical Questions

Causality questions

- What is the effect of political campaign expenditures on voting outcomes?
- If Bank Indonesia lowers interest rates today, what will happen to inflation tomorrow?
- How much more money will you earn as a result of taking this course?

Note: the cause and effect elements in this type of questions. The presence of a causal link is suggested by economic theory.

Application of Econometrics: empirically verify or quantify the causal link $\Rightarrow$ Microeconometrics.

Steps in Empirical Econometrics Analysis

1. Questions of Interest
   - formulating the question(s) of interest
2. Economic Model
   - building a new economic model or finding suitable economic models to analyze the questions
   - An economic model informs formal relationships between economic variables
3. Econometrics Model
   - formulating an econometric models according to the economic model
4. Data
5. Econometric Analyses
   - Using Econometrics method to estimate the Econometrics model
6. Inference
   - using statistical methods to draw inferences

Steps in Empirical Econometrics Analysis

Example 1.1: Economic Model of Crime

Step 1: Formulating the question of interest

Does the **wage** that can be earned in legal employment affect the decision to engage in **criminal activity**? (crime vs. wage)

Steps in Empirical Econometrics Analysis

Step 2: Building or finding an economic model

Using a model developed by Gary Becker (1968) to analyse the relationship between crime and wage:

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7) \tag{1}$$

y hours spent in criminal activity
x_1 criminal "hourly wage"
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

Steps in Empirical Econometrics Analysis

Step 3: Formulating an Econometric Model

3.1 Solve quantification issues

- How can we measure hours spent in criminal activity?
- How do we approximate the probability of being caught with an observable economic variable?

3.2 Specify the functional form of the economic relationships:

$$\begin{aligned} crime =& \beta_0 + \beta_1 wage + \beta_2 othinc + \beta_3 freqarr \\ &+ \beta_3 freqconv + \beta_4 avgsen + \beta_5 age + u \end{aligned} \tag{2}$$

crime: Measure of criminal activity
wage: wage of legal employment
othinc: other income
freqarr: frequency of prior arrest
freqconv: frequency of conviction
avgsen: average sentence length after conviction
age: age
u: unobserved determinant of criminal activity, such moral moral character, family background

Steps in Empirical Econometrics Analysis

Step 4: Collecting Data

Econometric analysis requires Data

We can use

- Experimental data $\Rightarrow$ randomized trials
- Observational data

According to source:

- Primary data: collect data directly from the primary source (doing a fieldwork).
- Secondary data: collect data from secondary source, such as from World Bank, KPU, BPS, etc.

The Structure of Econometrics Data

The structure of econometrics data

1. Cross section
2. Time series
3. Pooled cross section
4. Panel/Longitudinal Data

The Structure of Econometrics Data

Cross-sectional Data

Cross-sectional Data

- Based on individuals, households, firms, cities, states, countries, or other units interest at a given point of time or in a given period
- Cross sectional observations are more or less independent of sample
- For example, pure random sampling from a population
- Sometimes pure random sampling is violated, e.g. units refuse to respond in surveys, or if sampling is characterized by clustering
- Cross sectional data typically encountered in applied microeconomics

The Structure of Econometrics Data

Cross-sectional Data

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

The Structure of Econometrics Data

Cross-sectional Data

TABLE 1.2 A Data Set on Economic Growth Rates and Country Characteristics

obsno	country	gpcrgdp	govcons60	second60
1	Argentina	0.89	9	32
2	Austria	3.32	16	50
3	Belgium	2.56	13	69
4	Bolivia	1.24	18	12
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
61	Zimbabwe	2.30	17	6

The Structure of Econometrics Data

Time Series Data

Time Series Data

- Observations of a variable or several variables over time
- For example stock prices, money supply, consumer price index, gross domestic product, annual homicide rates, automobile sales,
- Time series observations are typically serially correlated
- Ordering of observations conveys important information
- Data frequency: daily, weekly, monthly, quarterly, annually
- Typical features of time series: trends and seasonality
- Typical application: applied macroeconomics and finance

The Structure of Econometrics Data

Time Series Data

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

The Structure of Econometrics Data

Pooled Cross Sections Pooled Cross Sections

- One or more cross sections are combined in one data set
- Cross sections are drawn independently of each other
- Pooled cross sections often used to evaluate policy changes
- Example
  - Evaluate effect of change in property taxes on house prices
  - A random sample of house prices in 1993
  - A new random sample of house prices in 1995
  - Compare before/after (1993: before reform, 1995: after reform)

The Structure of Econometrics Data

Pooled Cross Sections

TABLE 1.4 Pooled Cross Sections: Two Years of Housing Prices

obsno	year	hprice	proptax	sqrft	bdrms	bthrms
1	1993	85,500	42	1600	3	2.0
2	1993	67,300	36	1440	3	2.5
3	1993	134,000	38	2000	4	2.5
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
250	1993	243,600	41	2600	4	3.0
251	1995	65,000	16	1250	2	1.0
252	1995	182,400	20	2200	4	2.0
253	1995	97,500	15	1540	3	2.0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
520	1995	57,200	16	1100	2	1.5

Panel or Longitudinal Data

- The same cross sectional units are followed over time
- Panel data have a cross sectional and a time series dimension
- Panel data can be used to account for time invariant unobservables
- Panel data can be used to model lagged responses
- Example
  - City crime statistics; each city is observed each year
  - Time invariant unobserved city characteristic may be modelled
  - Effect of police on crime rates may exhibit time lag.

The Structure of Econometrics Data

Panel or Longitudinal Data

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

obsno	city	year	murders	population	unem	police
1	1	1986	5	350,000	8.7	440
2	1	1990	8	359,200	7.2	471
3	2	1986	2	64,300	5.4	75
4	2	1990	1	65,100	5.5	75
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
297	149	1986	10	260,700	9.6	286
298	149	1990	6	245,000	9.8	334
299	150	1986	25	543,000	4.3	520
300	150	1990	32	546,200	5.2	493

Causality and the Notion of Ceteris Paribus in Econometric Analysis

- From introductory economics, we learn that most economic questions are ceteris paribus by nature
- The notion of ceteris paribus is also very important in empirical analysis addressing causal relationship between two variables
- In most cases, hypotheses in the social sciences are ceteris paribus in nature: all other relevant factors must be fixed when studying the causal relationship between two variables.
  - The question is: Have enough other factors been held fixed to make a case for causality?
- Because of the non-experimental nature of most data collected in the social sciences, uncovering causal relationships is very challenging.

General Plan for the Semester

1. First Quarter (before mid): cross-section regression analysis
2. Second Quarter (after mid): time-series regression analysis