Econometrics: Economic Data and Econometric Modeling

Burcu Eke

UC3M

What is econometrics?

- Econometrics $\Rightarrow$ Economic measurements
- It is a discipline based on the development of probabilistic models and statistical inference methods for the study of economic relations, the contrast of economic theories, or the evaluation and implementation of policies.
- Econometrics consists of elements from different disciplines: economics, statistics and mathematics.

What does an econometrician do?

- An econometrician is an economist who uses statistics and mathematics to understand, explain, and predict economic variables such as employment, supply and demand, inflation, company profits, health insurance, and many other important economic topics.

- An econometrician attempts to develop accurate economic forecasting and successful policy planning.

- An econometrician qualitatively and quantitatively analyzes how the factors of interest affect a variable associated with an economic question of interest.

Applications

Most common applications of econometrics are

- prediction of macroeconomic variables such as interest rate, GDP, inflation
- Macroeconomic relationships such as unemployment-inflation and inflation-money
- Microeconomic relationships such as wage-education, production-input
- Finance such as stock volatility
- Forecasting

Observational Data vs Experimental Data

- In an **observational** study, researchers simply observe what is happening or what has happened in the past and try to draw conclusions based on these observations. $\Rightarrow$ Observational data

- In an **experimental** study, researchers impose treatments and controls and then observe characteristic and take measures, in a way, the researchers manipulate the variables and try to determine how the manipulation influences other variables. $\Rightarrow$ Experimental data

Observational Data vs Experimental Data

Experimental Study:

- First, a researcher identifies the data he/she wishes to obtain based on the research question
- Second, the researcher designs an experiment that will be used to obtain the required data. The design process is very crucial. it needs to ensure that the experiments measure what was intended to measure
- Last, the data analysis process takes place

Observational Data vs Experimental Data

Observational Study:

- Studies the variables are observed and recorded.
- Cause and effect are hard (often impossible) to establish. But associations and predictabilities among variables can be investigated.

Observational Data vs Experimental Data, Example

- Case 1:
  20 people went for a flu shot to a public hospital. After a month, an independent researcher checked how many of them got flu. 7 of them got flu, and the others didn't.

- Case 2:
  We randomly select 20 people with similar health condition, and randomly assign them to two groups: A, and B. Then, we give the flu shots to group A, and placebo to group B, and observe how many got flu after a month.

Observational Data vs Experimental Data

- Experimental studies attempt to control all factors that may affect the association under study, observational studies cannot
- Experimental studies randomize assignment of factors of interest to subjects, this is not feasible in observational studies
- Observational studies are performed when experimental studies are infeasible due to cost, or ethical concerns.
- Observational studies require more care in analysis and interpretation
- Economic data is almost always observational

Observational Data vs Experimental Data

Experimental Data

- There are two groups: treatment and control
- Subjects are randomly assigned into these groups by the experimental design
- The difference between these two groups $\Rightarrow$ **treatment**
- Hence, the different results in these two groups can be attributed to the treatment

Observational Data vs Experimental Data

Non-experimental Data

- There can still be two groups: treatment and control
- Generally, subjects self-select themselves to the groups, i.e., people choose to get a flu shot
- The differences between two groups: **treatment** AND **the characteristics of those getting the treatment**
- Effects of the treatment are unclear!

Observational Data vs Experimental Data

Recall the flu shot example. Suppose we want to test if the average number of people getting the flu are different in two groups

- The way to test for both cases is the same, i.e., we use the same test statistic
- Our interpretations need to be different
  - Correlation versus causal relationship

Observational Data vs Experimental Data

- In the experimental study example, the only difference between groups is the treatment $\Rightarrow$ we can conclude on a causal relationship
- In observational data, people **choose** to get the shot, hence, they might be more health conscious than the ones who choose not to.
- For observational data, we can only make inferences about whether the results are correlated!

Empirical Analysis

Order of the empirical analysis

1. Economic model
2. Econometric model
3. Data Analysis

Empirical Analysis: Economic Model

- The economic theory proposes models that explain behavior of one or more variables, say $Y_1, Y_2, ..., Y_m$, as a function of some other variables, say $X_1, X_2, ...X_k$, which are determined outside of the model.
- Formal model: Mathematical equations describing the relationship between the variables.
  - Example: utility function, $U = f(c, l)$ or $U - f(c, l) = 0$, where the utility (U) is a function of consumption (c) and leisure (l).
- Informal model: based on the theory and more intuitive aspects

Empirical Analysis: Variables

- **Exogenous** variable: A factor in a causal model whose value is **independent** from the states of other variables in the model; a factor whose value is determined by factors or variables outside the model under study.
- **Endogenous** variable: A factor in a causal model whose value is determined by the states of other variables in the model; contrasted with an exogenous variable.
- In general, an endogenous variable, say Y may depend on multiple exogenous variables in a model, For example, one may have a model with income as endogenous variable, and education and experience as exogenous variables.

Empirical Analysis: Econometric model

- Econometric models are generally algebraic models that are stochastic in including random variables (as opposed to deterministic models which do not include random variables).
- The random variables that are included, typically as additive stochastic disturbance terms, account in part for the omission of relevant variables, incorrect specification of the model, errors in measuring variables, etc.
- Recall the utility function example, the econometric model would be $U = f(c, l) + \varepsilon$ or $U - f(c, l) = \varepsilon$

Empirical Analysis: Econometric model

- In general, the mathematical equations are written for the whole population, and in econometric analysis, we almost always deal with sample data. in order to account for this, and possible measurement errors, or incorrect specification of the model econometric models include a stochastic component that satisfy the following equation:
  - $E\left[Y - f(x_1, x_2, ...x_k)\right] = E\left[\varepsilon\right] = 0$, where Y is the endogenous variable in the model

Empirical Analysis: Econometric model

- In order to quantify the relationship between economic variables, it is necessary to propose a **functional form** that depends on some variables and unknown parameters.
- The econometric model can be expressed as follows: $Y = f(x_1, x_2, ..., x_k; \boldsymbol{\beta}) + \varepsilon$, where $\boldsymbol{\beta}$ is a vector of unknown parameters and ε is the **error term**
- The nature of the model and the interpretation of the parameters depend on the assumptions on the error term

Empirical Analysis: Example

Consider the following model

- Economic model:
  - Wage is a function of educational attainment, and experience $\Rightarrow$ $W = f(Ed, Ex)$
- Econometric model:
  - $W = \beta_0 + \beta_1 Ed + \beta_2 Ex + \beta_3 Ex^2 + \varepsilon$

Empirical Analysis: Data Analysis

- After determining an economic model, and corresponding econometric model to answer the questions of interest, we analyze the data, i.e., estimate the unknown parameters
- We answer some questions based on the estimated parameters, such as are the estimates β's different from 0?, what sign do the have? and so on.

Types of econometric models

- Single variable versus multiple variables
- Single equation versus simultaneous equations

Types of data

In econometrics there are three main types of data (not necessarily mutually exclusive)

- Cross-sectional data
- Time series data
- Panel (longitudinal) data

All these different data types require specific econometric and statistical techniques for data analysis

Cross-Section

- A type of one-dimensional data set
- Collected by observing many subjects (such as individuals, firms or countries/regions) at the same point of time, or without regarding the differences in time
- In general used to compare the differences among the subjects
- Order does **not** matter
- Examples: Explaining people's wages by reference to their education level

Cross-Section

Example: Data for a sample of individuals in a country

Individual	Income	Marital Status	Educational Attainment
1	1500	single	university degree
2	2500	married	graduate degree
3	2000	separates	university degree
...	...	...	...

Time Series

- A sequence of data points, measured typically at successive times spaced at uniform time intervals i.e., annual, semi-annual, quarterly, monthly, daily and so on.
- Time series models often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values
- Have a natural temporal ordering
- Examples: Annual inflation rates, daily closing value of a certain stock

Time Series

A time series data example

Time	Inflation Rate
1990	1.5%
1991	2%
1992	2.02%
1993	2.4%
...	...

Panel (Longitudinal)

- Date that involve repeated observations of the same items over long periods of time
- Not necessarily cohort study $\Rightarrow$ Different cohorts may have different subjects
- Panel data (Longitudinal) studies track the same subject (people, countries, same set of stocks)
- Measurements are observed or taken on the same subjects repeatedly

Panel (Longitudinal)

Company	Year	Profit
1	2000	1.2 billion
1	2001	1.3 billion
1	2002	2 billion
2	2000	0.2 billion
2	2001	0.3 billion
2	2002	0.1 billion
3	2000	3 billion
3	2001	3.5 billion
3	2002	4 billion
...	...	...

Causality and "ceteris paribus" in Econometrics

- We sometimes study causality between two variables, not just correlation
- Correlation does not guarantee causality
  - This is because our data is economic data, not experimental
  - This prevents us from inferring causality from co-movement, since we have not controlled from other factors that might influence the variables

Causality and "ceteris paribus" in Econometrics

- In the rare case that our data fits exactly the experiment we test for, we can infer causality
- Thus, when studying an empirical case, it may be worth to question "what the right experiment should be" and how our data can address and/or mimic it
- When looking for causality, the "ceteris paribus" concept (all other factors remain constant) plays a crucial role: it allows us to infer the partial effect of one variable on another one
- Econometric techniques allow us to estimate "ceteris paribus effects" and infer causal relations between variables, by simulating a situation close to the experiment we are interested in