Introduction to Econometrics

Carlos Brunet Martins-Filho

University of Colorado at Boulder

Spring 2025

Introduction

- The principal objective of any science is the generation of testable hypotheses and conjectures regarding its object of study
- Reality is complex $\Longrightarrow$ simplification is needed
- Theoretical models and their variables

How are theoretical models evaluated?

- Internal mathematical consistency
- Contrast refutable hypotheses generated with observed reality
- Example from microeconomic theory:
  - Collection of assumptions regarding utility and consumer behavior
  - Obtain $q = f(p, \rho, w)$
  - Is the conjecture that q is a function of p, ρ, and w supported by the reality we observe?

What is Econometrics?

- Econometrics has been defined as the set of concepts, methods and procedures used to summarize and analyze economic data that correspond to the economic variables that appear in economic models.

But, it is more than that:

- Its methods and procedures can be used in other social and natural sciences
- It helps theoretical economists develop better theoretical models
- It can be used for forecasting and policy evaluation

Economic data

- stochastic and mostly observational/non-experimental
- stochastic data that emerge from stochastic phenomena have certain regularities
- stochastic regularity $\Longrightarrow$ we can make probabilistic statements about generating phenomena
- Observational data produced by processes not controlled by researcher

Representing data

If X represents a variable of interest. We represent $n$observations on X as $\{X_k\}_{k=1}^n$:

- When k is an index representing time, we say that $\{x_k\}_{k=1}^n$ is a time series on X.
- When k represents something other than time, e.g., a specific economic agent, a region of the country, etc. we say that $\{X_k\}_{k=1}^n$ is a cross-section on X.
- When $k = (i, t)$ with $i = 1, 2, ...N$ and $t = 1, 2, ...T$ indexing sections and time respectively, we say that $\{X_{it}\}_{i=1,t=1}^{N,T}$ is a panel on X.

Types of data

- X is said to be measured in **ratio scale** if for any two observation x_k and x_s:
  1. $\frac{x_k}{x_s}$ is meaningful
  2. $x_k - x_s$ is meaningful
  3. there exists an order relation between x_k and x_s, i.e., $x_k \geq x_s$ or $x_k \leq x_s$ or both.
- X is said to be measured in **interval scale** if 2 and 3 are valid, but 1 is not. Example: time.
- X is said to be measured in **ordinal scale** if 3 is valid, but not 1 and 2. Examples: income level (upper, middle, low), educational level (High School degree, Bachelor's degree, Graduate degree), etc.
- The variable X is said to be measured in **nominal scale** if neither 1, 2 or 3 are valid. Examples: marital status, gender and employment status.
- Variables measured in ordinal or nominal scales are normally referred to as **categorical** data.

Statistical Models

- A statistical model is a set of general assumptions regarding the stochastic nature of the phenomenon which produced the observed data.
- These assumptions should be general enough to account for the stochastic regularity patterns exhibited by the data.
- A precise description of a statistical model requires the definition of concepts that we will study later, such as **probability**, **stochastic variables**, **probability distributions**, and **statistical independence**.

Example of a statistical model

- Stochastic phenomenon of interest is the toss of a coin.
- The toss produces a result R which takes the value H for heads or T for tails. Results on n consecutive tosses are represented by the time series $\{R_t\}_{t=1}^{n}$.
- A statistical model for R can be described by the following assumptions,
  1. $R_t = H$ with probability p and $R_t = T$ with probability $1 - p$ for all $t = 1, ..., n$. p is an unknown **parameter** of the model.
  2. R_t and R_s are statistically independent for all $t \neq s$.

This very simple statistical model contains the basic elements that are present in many (parametric) statistical models.

1. A parameter (p) that takes value in a pre specified set $[0, 1]$ that defines a class of probabilities (all pairs $(p, 1-p)$ for $p \in [0, 1]$) associated with the occurrence of the variable of interest (R)
2. An assumption about how the data is obtained. In this case the time series is assumed to come from n independent tosses with the same chance of observing T and H in each toss.