Lecture 1: Introduction

Economics 326 — Methods of Empirical Research in Economics

Vadim Marmer, UBC

What is econometrics?

Econometrics develops **statistical methods** for:

- **Estimating** economic relationships
- **Testing** economic theories
- **Forecasting** important economic variables
- **Evaluating** government and business policy

Why statistics?

- Economic theory motivates models of relationships between variables of interest.
- Economic **models are approximations**, not exact descriptions of reality.
- Even good models omit important factors that affect outcomes.
- We replace a **deterministic** model with a **probabilistic** model.

Examples

- Estimation of demand and supply functions
  Elasticities help evaluate the effects of taxation.

- Mincer (1974), *Schooling, Experience, and Earnings*
  Uses individual data to estimate returns to schooling and experience.

  - Determine an "optimal" amount of schooling
  - Study education in developing countries
  - Study gender and race discrimination
  - Study the impact of immigration on labour markets

- Paarsch (1997), *Journal of Econometrics*
  Estimates optimal reserve prices for BC timber auctions.

- Chandra et al. (2008), *Pediatrics*
  Studies how exposure to sexual content on TV relates to teen pregnancy.

Types of data: cross-section

> **i** Definition
>
> A **cross-sectional** dataset contains observations on individuals (e.g., workers or firms) collected in a single time period.

Example (wages and individual characteristics):

obs	wage	education	experience	female	married
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
...	...	...	...	...	...

- The order of observations is not important.
- It is often reasonable to assume observations are **statistically independent**.

Types of data: time series

> **i** Definition
>
> A **time series** dataset contains observations on one or more variables over time.

Example (Puerto Rico minimum wage, unemployment, and GNP):

obs	year	minimum wage	unemployment	gnp
1	1950	0.20	15.4	878.7
2	1951	0.21	16.0	925.0
3	1952	0.23	14.8	1015.9
...	...	...	...	...

- Data frequency can be daily/weekly/monthly/quarterly/annual; in finance, trade data can be very high frequency.
- The order of observations is important.
- Observations are often correlated (e.g., trends).

Types of data: panel

> **i** Definition
>
> A **panel** dataset combines cross-section and time series: a time series for each cross-sectional unit.

Example (two-year panel on city crime):

obs	city	year	murders	population	unempl	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
...	...	...	...	...	...	...

Causality

- We care about **causal** relationships, but data often only reveal **correlations** (associations).
- To claim a causal effect, **other factors** affecting the outcome must be held fixed (controlled for).
- Controlled experiments help with causality in the natural sciences.
- Experiments are often impossible in economics (cost and ethics).
- We typically work with **observational data**.

Examples (causality)

Education

$$\log(\text{Wage}) = \alpha + \beta \times \text{Years of Schooling} + U$$

U includes other factors (e.g., **ability**). If ability is hard to control for, simple correlations can **overestimate** returns to education.

Police and crime

$$\text{Number of Crimes} = \alpha + \beta \times \text{Size of the Police Force} + U$$

Cities with more crime often hire more police, so simple correlations can **spuriously** suggest police increase crime.