

# Lecture 1: Introduction

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

April 5, 2026

## What is econometrics?

Econometrics develops **statistical methods** for:

- **Estimating** economic relationships
- **Causal** or **counterfactual** analysis: what happens to an outcome when we force a change in a factor.
- **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

## 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.

## Forecasting vs estimation

### Forecasting

- **Forecasting:** predicting **out-of-sample** values of a variable: E.g.,
  - Predict next quarter's GDP
  - Assess the probability of default on a loan given borrower characteristics (income, education, past history of defaults, etc.)
  - Predictive text: given previous words, predict the next word.
  - ChatGPT: given a prompt, generate the desired response.
  - Predict the number of crimes in a city next year given current police force size.
  - Forecasting equation:

$$\widehat{\text{Number of Crimes}} = f(\text{Size of the Police Force})$$

where  $\widehat{\text{Number of Crimes}}$  is the forecasted number of crimes in a city with Size of the Police Force.

- Use data to estimate or “learn” the best forecasting function  $f(\cdot)$ .
- Focus on **prediction accuracy**: minimizing some measure of forecast errors.

### Estimation

- **Estimation:** estimating **structural parameters** of an economic model: E.g.,
  - Effect of education on wages
  - Effect of police force size on crime
  - Structural equation:

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

where  $\beta$  is the structural parameter measuring the effect of police force size on crime, and  $U$  includes other factors affecting crime.

- Focus on **causal or counterfactual interpretation**: what happens to the outcome when we change a variable.
- Best forecasting model may produce misleading counterfactual effects.
- Good structural model may have poor forecasting performance.
- Choose the right tool for the job!
- Causal/counterfactual analysis always requires some **assumptions** about the data generating process.