UGA M1: Econometrics 1
Introduction

Michal W. Urdanivia*

*Université de Grenoble Alpes, Faculté d'Économie, GAEL,
e-mail: michal.wong-urdanivia@univ-grenoble-alpes.fr

September 18, 2017

What is Econometrics?

Econometrics is concerned with the development of quantitative methods for:

- Estimation of economic relationships
- Testing of economic theories
- Forecasting of important economic variables
- Evaluation of government and business policy

Why statistics?

- Economic theory is used to construct models characterizing relationships between variables of interest
- Economic models are only approximations
- A model can take into account a number of important factors but there will be many factors left out that also affect outcomes
- We therefore replace the exact (deterministic) model with a probabilistic model

Example: Becker's (1968) economic model of crime

- Economic Model: we have the relation,

$$Y = f(X_1, X_2, X_3, X_4, ...)$$

  - Y : hours spent in criminal activities (crime)
  - X_1: wage for an hour spent in criminal activities
  - X_2: wage in legal employment (wagem)
  - X_3: income other than from crime or employment (othinc)
  - X_4: probability of getting caught (freqarr)

- An standard econometric specification is:

$$\text{crime} = \beta_0 + \beta_1 \text{wagem} + \beta_2 \text{othinc} + \beta_3 \text{freqarr} + U$$

Where the term U captures unobserved factors such as:

  - the reward for criminal activity,
  - family background,
  - measurement error

Example: Mincer's (1974) wage regression

- Economic Model: wage depends on human capital

$$\text{wage} = f\,(\text{educ}, \text{exper}, \text{training})$$

- Econometric Model:

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 (\text{exper})^2 + \beta_4 \text{training} + u$$

  The term u captures unobserved factors:
  - innate ability,
  - family background,
  - quality of education

- Hypothesis Testing: whether training affects wage $H_0 : \beta_4 = 0$

Data and causal effects

- While we are interested in causal relations, statistics only allows us to establish correlations
- In order to say that one variable has a causal effect on another, other factors affecting the outcome must be held fixed
- In natural sciences can use controlled experiments
- Experiment are often impossible in economics (too costly and/or for ethical reasons)
- Must rely on observational data

Example: effect of health insurance coverage on health

- Question: what is the effect of health insurance on health?
- Ideal experiment: randomly people health insurance or not, compare their health afterward
- Observed survey data:

Group	Sample Size	Mean Health	Std.Dev.
Some insurance	8114	4.01	0.93
No insurance	1281	3.70	1.01

Source: 2009 NHIS data reported by Angrist and Pischke (2014)

- Is 4.01 - 3.70 = 0.31 the causal effect of health insurance on health?
- Not likely, people with and without insurance differ in other ways that might affect health

Group	Mean Education	Mean income
Some insurance	14.31	106,467
No insurance	11.56	45,656

Potential Outcomes

- Potential outcomes (aka the Rubin causal model) are one powerful way of thinking about causality
- Let $D_i = 1$ if person i has health insurance, 0 otherwise
- Let Y_{i0} and Y_{i1} be the **potential outcomes**
  - Y_{i0} = health if person i does not have health insurance
  - Y_{i1} = health if person i has health insurance
- We observe

$$Y_i = \begin{cases} Y_{i0} & \text{if } D_i = 0 \\ Y_{i1} & \text{if } D_i = 1 \end{cases}$$

Hence,

$$\begin{aligned} Y_i &= (1 - D_i) Y_{i0} + D_i Y_{i1} \\ &= Y_{i0} + D_i \times \underbrace{(Y_{i1} - Y_{i0})}_{\text{causal effect of health insurance for person } i} \end{aligned}$$

Potential Outcomes

- Observed average difference in health

$$\mathbb{E}[Y_{i1}|D_i = 1] - \mathbb{E}[Y_{i0}|D_i = 0] = \\ = \underbrace{\mathbb{E}[Y_{i1} - Y_{i0}|D_i = 1]}_{\text{Average treatment effect on the treated}} + \\ + \underbrace{\mathbb{E}[Y_{i0}|D_i = 1] - \mathbb{E}[Y_{i0}|D_i = 0]}_{\text{Selection bias}}$$

- Insurance may help you or hurt you (on average), but wealthier people are more likely to afford insurance and might be healthier for other reasons; conversely people who are sick might be more willing to pay for insurance
- Random assignment of D_i removes selection bias because D_i and Y_{i0} are then independent
- Often cannot randomize, but can
  - Use available data to approximate desired experiment
  - Use economic theory to impose restrictions

Examples

- **Education**

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

U represents other factors, for example, ability.

Since it is very hard to control for ability, one can overestimate the return to education by relying on usual correlations.

- **Minimum wage and employment** (Card and Krueger *AER* 1994)

$$\text{Unemployment} = \beta_0 + \beta_1 \times \text{Minium Wage} + U$$

Reverse causality: High employment may lead to political pressure for higher minimum wage.

Examples

- **Size of the police force and crime**

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

Usually, cities with a lot of criminal activity have a bigger police force. Simple correlations can spuriously indicate that the size of the police force has a positive effect on the crime rates.

References

Angrist, Joshua D and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.