Econometrics 2 (Part 2)

Arieda Muço

Central European University

Spring 2020

Contact

- Sergey Lychagin: LychaginS@ceu.edu
  ~~Office: Budapest campus, Nador 13, 513~~
- Arieda Muço: MucoA@ceu.edu
  ~~Office: Budapest campus, Nador 13, 507~~
- Boldizsár Juhász (TA): juhasz_boldizsar@phd.ceu.edu

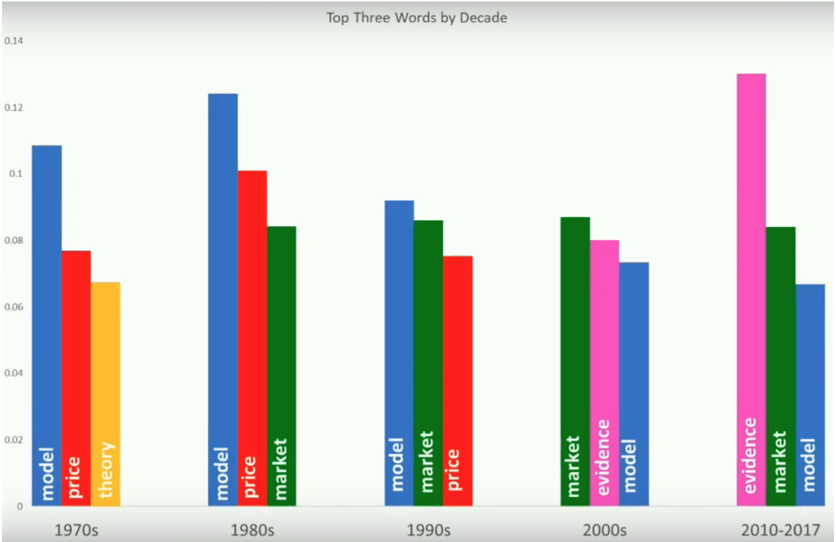
Textbooks

- Introductory Econometrics: A Modern Approach by Wooldridge
- Mostly Harmless Econometrics by Angrist and Pischke
- Mastering Metrics by Angrist and Pischke
- Casual Inference: The Mixtape by Cunningham
- Reading list of applied papers

Outline: Tentative Schedule

1. Economic research questions: causality
2. The experimental ideal
3. Linear regression
4. Instrumental Variables
5. Panel Data, Fixed Effects, Differences-in-Differences
6. Matching
7. Program Evaluation: Nonparametric Methods
8. Regression Discontinuity

Della Vigna and Card


Top Three Words by Decade
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
model
price
theory
model
price
market
model
market
price
market
evidence
model
evidence
market
model
1970s
1980s
1990s
2000s
2010-2017

Applied Microeconometrics Papers

Statistical Models of Shoe and Leather

Freedman, David A. (1991) "Statistical Models and Shoe Leather", Sociological Methodology, 21, 291-313

John Snow studies of the cholera epidemics in Europe in the 19th century and proves that cholera is a waterborne infectious disease

- In the 19^{th} century no microbiology, limited microscopes
- Theory: diseases result from "poison in the air" - miasma
- Cholera Europe in epidemic waves
- Snow studied spatial pattern of epidemics along tracks of human commerce
- Influence of water supply on incidence of Cholera?

Is cholera a waterborne or an airborne disease?

London in the 1800's: different water companies serve different areas

- Some companies take water from the Thames polluted by sewage
- 2 companies
  - Southwark & Vauxhall: downstream from sewage discharges
  - Lambeth: intake point upstream
- Both companies served the same parts of London during the 1853-54 cholera epidemic
- Sometimes houses next to each other in the same street were served by the 2 different companies
  - Each company supplies rich and poor, large and small houses, no difference in condition or occupation
- Idea: compare number of cholera victims

Method of Shoe and Leather

- Snow surveyed houses in large parts of London
- Water company
- Cholera victims
- 300,000 households involved
- Reward: clear result


GREAT MARLBOROUGH STREET
WORK HOUSE
PORTLAND STREET
BREWERY
PUMP
PETER STREET
GOLDEN SQUARE

TABLE 1
Snow's Table IX

	Number of Houses	Deaths from Cholera	Deaths Per 10,000 Houses
Southwark and Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

The Experimental Ideal

Social experiment is the most influential research design. Why?

Solves Selection Problem

Potential Outcome Model

$D_i = \{0, 1\}$ treatment variable (hospital care)
For each population unit i we consider two <u>potential outcomes</u> (health status)

$$\begin{array}{ll} Y_{1i} & \text{outcome with treatment} \\ Y_{0i} & \text{outcome without treatment} \end{array}$$

The gain from treatment or <u>causal effect</u> for unit i is

$$Y_{i1} - Y_{i0}$$

Problem: For each i, only one of Y_{i1} or Y_{i0} is observed.

Observed outcome

We observe

$$Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases} = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

In the population distribution of Y_{1i} and Y_{0i}, we can compare the average health of <u>treated</u> and <u>non-treated</u>

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{\text{observed difference}} =$$

$$\underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{\text{average treatment effect}} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{selection bias}}$$

Observed outcome

Remind that the observed difference is $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$
Which can be rewritten as:

$$E[\underbrace{Y_{0i} + (Y_{1i} - Y_{0i})D_i}_{Y_i}|D_i = 1] - E[\underbrace{Y_{0i} + (Y_{1i} - Y_{0i})D_i}_{Y_i}|D_i = 0]$$

From the properties of the conditional expectation we can rearrange the above equation as:

$$E[Y_{0i}|D_i = 1] + E[(Y_{1i} - Y_{0i})D_i|D_i = 1] - E[Y_{0i}|D_i = 0] - E[(Y_{1i} - Y_{0i})D_i|D_i = 0]$$

Which can be rewritten as:

$$E[Y_{0i}|D_i = 1] + E[(Y_{1i} - Y_{0i})|D_i = 1] - E[Y_{0i}|D_i = 0]$$

And is equivalent to:

$$E[Y_{0i}|D_i = 1] + E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

Rearranging we get:

$$\underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{\text{average treatment effect}} \quad + \quad \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{selection bias}}$$

Random assignment as a solution

Random assignment makes D_i independent of the potential outcome. If D_i is independent of Y_i then $E[Y_i|D_i] = E[Y_i|D_i = 1] = E[Y_i|D_i = 0] = E[Y_i]$

$$\begin{aligned} E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\ &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \\ &= E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}] \end{aligned}$$

The observed difference in mean outcomes equals the average treatment effect. Examples:

- health treatments
- government sponsored training programs
- education production: effect of class size, teacher quality, etc. on student achievement

Regression Analysis of Experiments

Assume $Y_{i1} - Y_{i0} = \rho$ constant treatment effect

$$Y_i = \alpha + \rho D_i + \eta_i$$

$$\begin{aligned}
E[Y_i|D_i = 1] &= \alpha + \rho + E[\eta_i|D_i = 1] \\
E[Y_i|D_i = 0] &= \alpha + E[\eta_i|D_i = 0] \\
E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= \rho + \underbrace{E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0]}_{\text{selection bias}}
\end{aligned}$$

Selection bias amounts to correlation between regression error η_i and D_i.

Regression Analysis of Experiments

We know about the selection bias

$$E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0] = E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

If D_i is randomly assigned, the selection bias is equal to zero. Thus estimating the regression model results in the causal effect ρ.

Regression model with covariates

$$Y_i = \alpha + \rho D_i + \beta X_i + \eta_i$$

If X_i uncorrelated with D_i, including them will not affect estimate of ρ, but increase precision.