

# Causal Data Analysis

## Introduction

Álmos Telegdy

Corvinus University of Budapest

# Professors

- Álmos Telegdy: lectures (with lots of interaction - that is my hope)
- László Tőkés: seminar I (applications in R)
- Xiaoru Lin: seminar II (R programming)

## Evaluation

- Short empirical paper (30%)
  - Students work in pairs
  - Outline (question, data, method – 1 Dec.)
  - Short written analysis (app. 5 pages – 15 Dec.)
- Midterm exam (35% – 24 Nov.)
  - Theory and R programming
- Second exam (35% – 15 Dec.)
  - Theory and R programming

To pass the course, you must have at least 50 percent in each assignment

## Prerequisites, textbook

- The course is based on OLS regression analysis
- Textbook: Gábor Békés – Gábor Kézdi (2021). Data Analysis for Business Economics and Policy. US: Cambridge University Press, Part IV

## Aim of the course

- 1 How can we derive causal conjectures from data?
- 2 Econometric methods which let us make causal statements
- 3 More on "everyday" methods (= those which can be used in most circumstances). Advanced methods also mentioned so you can read the literature
- 4 Emphasis on understanding how to do causal analysis and how to interpret the results, less on statistical features

## Topic covered

- A framework for causal analysis (Ch. 19)
- Experiments (Ch. 20)
- Regressions with matching (Ch. 21)
- Difference-in-differences (Ch. 22)
- Panel data regressions (Ch. 23)
- Synthetic controls and event studies (Ch. 24)
- Regression discontinuity
- Instrumental variables

# Causal data analysis

What we do?  $\Rightarrow$  run regressions

$$y_i = \beta_0 + \beta x_i + \epsilon_i \quad (1)$$

What is our goal?

- Under what **conditions** can we say that the relation between  $x$  and  $y$  is causal?
- What is the **meaning** of  $\beta$ ?

# OLS = correlation

- So far: statistical connections
- This course: under what conditions can we declare that one variable **causes** the other variable's variation?
- Causal questions
  - Does the consumption of alcohol cause lung cancer?
  - The the death of the increase the chance of death of the other spouse (The widowhood effect)?
  - Do export subsidies increase productivity improvement?

# Why not obvious?

We study causal data analysis, to avoid the *Post hoc, ergo propter hoc – after it, therefore because of it* type mistakes

A correlated with B

- Maybe both A and B were caused by something else
  - both spouses died in an accident
- Maybe B causes A
  - productive firms start exporting
- Maybe C causes A, but C correlated with B
  - alcohol consumption correlated with smoking which causes lung cancer

The optimization behavior of agents can make all kinds of correlations in the data

## Complete endogeneity: The King's orders

"It is contrary to etiquette to yawn in the presence of a king," the monarch said to him.

"I forbid you to do so."

"I can't help it. I can't stop myself," replied the little prince [...]

"Ah, then," the king said. "I order you to yawn. [...] Come, now! Yawn again! It is an order."

"That frightens me . . . I cannot, any more . . ." murmured the little prince, now completely abashed.

"Hum! Hum!" replied the king. "Then I – I order you sometimes to yawn and sometimes to – " He sputtered a little, and seemed vexed.

For what the king fundamentally insisted upon was that his authority should be respected. He tolerated no disobedience. He was an absolute monarch. But, because he was a very good man, he made his orders reasonable.

*Antoine de Saint-Exupéry: The Little Prince*

# Why are causal relations important?

Policy decisions based on correlations may be very costly

Example: effect of privatization on employment

- Simple correlation: -47%

Problems with the simple correlation

- What is the effect of the business environment?
  - In the 90s most firms decreased their size → downward bias
- Maybe privatized firms are different from those which stayed under state ownership
  - Politicians are willing to privatize the good companies → upward bias

# Why are causal relations important?

Policy decisions based on correlations may be very costly

Example: effect of privatization on employment

- Simple correlation: -47%

Problems with the simple correlation

- What is the effect of the business environment?
  - In the 90s most firms decreased their size → downward bias
- Maybe privatized firms are different from those which stayed under state ownership
  - Politicians are willing to privatize the good companies → upward bias

If we attenuate these biases, the estimated effect becomes -1,7%.

## Correlation not causation



## Central Bank interventions

- Good times coming: decrease the money supply
- Recession expected: buy bonds, increase liquidity

Is countercyclical monetary policy effective? An OLS regression would suggest it is not.  
⇒ Whatever the Central Bank does, the economy behaves the same way.

# Contemplation about reverse causality



# Correlation and Causality

Correlation is not causality, but we can observe only correlation

→ a regression coefficient measures correlation

But we can say more

- If we see evidence of correlation, it's because...
  - ...there is actual causality
  - ...there is no direct causality, something in the background is going on
  - ...it's just randomness, and really, no correlation

Our goal: under what assumptions can we say that a regression coefficient quantifies a causal effect?

We can never say that there is causality. All we can do is to make as simple as possible assumptions under which there is causality.

# What do you need for good analysis?

## 1 Good question

- Sometimes you get it from your boss, sometimes you go for it

## 2 Hypotheses

- What are agent's motivations?
- what is the institutional framework?

## 3 Data

- *"In God we trust. All others must bring data."* (W. Edwards Deming statistician)

## 5 Method

- Under what conditions can we view a correlation to be causal?

## 6 Writing: write down your analysis such that it is understandable and honest

- *Lies, damned lies, and statistics* – true to the extent that with a bad methodology one can derive anything from the data

Admin  
ooooo

Correlation is not causality  
oooo

Causal data analysis  
oooooo●

# Final warning

# Final warning

