

# Topics in Applied Econometrics

Challenges to identify causal effects

Attila Gyetvai

Duke Economics  
Summer 2020

# How to tackle an empirical project

- 1 What causal effects are we interested in?
- 2 What ideal experiment would capture this effect?
- 3 What is our identification strategy?
- 4 What is our mode of statistical inference?

# Establishing some language

- “Treatments”
  - ▶ Terminology and some tools come from the medical literature
  - ▶ Any causal effect can be thought of as the effect of some “treatment”
- We are interested in the outcome  $Y$
- Consider the effect of the treatment on a potential participant  $i$
- Two states of the world:
  - ▶  $Y_{1i}$  if  $i$  is treated
  - ▶  $Y_{0i}$  if  $i$  is not treated
- $Y_{1i} - Y_{0i}$  is the effect of the treatment on  $i$

# Counterfactual approach

- We can be in only one state of the world
- $i$  is either treated or not
  - ▶ If  $i$  is treated, we can observe  $Y_{1i}$  but not  $Y_{0i}$
  - ▶ If  $i$  is not treated, we can observe  $Y_{0i}$  but not  $Y_{1i}$
- The unobserved outcome is the *counterfactual* outcome

## Fundamental problem of causal inference

$Y_{1i} - Y_{0i}$  cannot be observed

- We can only see the actual but not the counterfactual outcome
- Therefore the effect of the treatment is fundamentally unobservable
- However, not all hope is lost!

# Measuring the counterfactual

- Observation vs. measurement
  - ▶ We cannot observe the counterfactual outcome
  - ▶ ...but we can measure it under certain assumptions
- *Example:*  $Y_{1i} = Y_j$  when  $j$  is treated and  $i$  is not
  - ▶ This assumption is simple but very restrictive
- *Question:* What is the least restrictive set of assumptions under which we can identify the effect of the treatment?
- During this course, we will cover various sets of such assumptions

# The potential outcomes framework

- Potential outcome: what the counterfactual outcome could be
- Treatment variable:  $D_i = 1$  if  $i$  is treated / 0 if  $i$  is not treated
- Observed outcome:

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

# Starting to think about treatment effects

- The effect of the treatment is  $Y_{1i} - Y_{0i}$
- How can we say something meaningful about this effect?
- We need to aggregate outcomes within groups, i.e., take expectations
- *Example:* the average treatment effect (ATE)

$$ATE = \mathbb{E}(Y_{1i} - Y_{0i}) = \mathbb{E}(Y_{1i}) - \mathbb{E}(Y_{0i})$$



# Average treatment effect

$$ATE = \mathbb{E}(Y_{1i} - Y_{0i}) = \mathbb{E}(Y_{1i}) - \mathbb{E}(Y_{0i})$$

- Recall the fundamental problem: we can only observe the actual outcomes
- How can we measure the ATE from these observations?
- We need to make assumptions about the counterfactual outcomes  $\mathbb{E}(Y_{1i} | D_i = 0)$  and  $\mathbb{E}(Y_{0i} | D_i = 1)$
- E.g., assume that the potential outcomes in the treated and untreated groups are the same
- When is this assumption reasonable?

# Challenges arise

- Potential outcomes are not the same
  - ▶ Units “select into treatment” (choose treatment status)
  - ▶ Selection on observable vs. unobservable characteristics
- Treated and untreated groups are not well-defined
  - ▶ Some units opt out of the treatment
  - ▶ Some units fight their way into treatment, even though they are not supposed to
- The treatment effect is different for various subgroups
  - ▶ The effect is larger for units with certain characteristics
  - ▶ The effect is larger for those who are “just treated”
- And so on...

# Evaluating the treatment effect

- Suppose we successfully measure the treatment effect. What does it tell us?
- In social sciences, we are mostly interested in generalizing the findings from a certain program to other settings
- Internal validity: Does the analysis show the effect of the particular program?
- External validity: Can the effects be generalized to other programs?
- Internal vs. external validity:
  - ▶ One can argue that only internally valid results should be generalized
  - ▶ ...but the more tailored the analysis to a certain program, the less it is generalizable

# Looking forward

- **We** will cover the cornerstones of a successful empirical project
  - ① What causal effects are we interested in?
  - ② What ideal experiment would capture this effect?
  - ③ What is our identification strategy?
  - ④ What is our mode of statistical inference?
- **You** will tackle an empirical project along these principles
  - ▶ Pose an empirical question
  - ▶ Design a project that seeks to answer the question
  - ▶ Describe the ideal experiment
  - ▶ Simulate data from a non-ideal experiment
  - ▶ Evaluate the treatment effect

# Empirical project

- You will work in groups
- Groups will be formed after the add/drop date
- Groups will regularly update me on their progress in Zoom meetings
- At the end of the course, each group will present their project