Use IV in your research

Labor economics

Instructor: Haoran LEI

Hunan University

Class presentation

- The final presentation makes half of your total grades in this course.
- I initially thought it could be in-class presentation if there are about 20 students taking this course. But the wechat group now has 30 students enrolled.

I'll tell the second story of IV.

Keywords are **causality** and **research design**.

- You should have a basic understanding of what *causality* is even though we have not discussed it formally.
- But what is *research design*?

Research design

- Research design: a plan or framework for conducting a study that aims to test a specific hypothesis or answer a question.
- A good research design involves answers to:

1. What is the *causal relationship* of interest?
2. What is the *(quasi-)experiment* that could ideally be used to capture the causal effect of interest?
3. What is your *identification* strategy?
4. What is your mode of *statistical inference*?

Q1: What is the causal relationship of interest?

- A **causal relationship** is useful for making predictions about the consequences of changing circumstances or policies:
  - it tells us what would happen in *alternative* (or *counterfactual*) worlds
- Eg: suppose you are interested in investigating human capital
  - the causal effect of *schooling* on wages: the increment to wages an individual would receive if he or she got more schooling (Card, 1999)
  - the causal effect of *class size* on wages (Angrist and Lavy, 1999)

An ideal experiment

- The most credible and influential research designs use *random assignment*.
- Suppose we are interested in a causal "if-then" question:
    - **Do hospitals make people healthier?**
- One empirical method is to compare the health status of *those who have been to the hospital* to *those who have not*.

Data from National Health Interview Survey (NHIS)

Group,	Sample Size,	Mean health status,	Std. Error
Hospital	7,774	2.79	0.014
No Hospital	90,049	2.07	0.003

Taken at face value, this result suggests that **going to the hospital makes people sicker**.

Data from National Health Interview Survey (NHIS)

Group,	Sample Size,	Mean health status,	Std. Error
Hospital	7,774	2.79	0.014
No Hospital	90,049	2.07	0.003

- However, the two groups (`Hospital`, `No Hospital`) are not comparable:
- Even if after hospitalization **people who have sought medical care** are not as healthy as **those who were never hospitalized**, they may be better off than they otherwise would have been.

Treatments and potential outcomes

- Think about **hospital treatment** as a binary random variable:

$$D_i \in \{0, 1\}.$$

- The outcome is Y_i. For any individual there are two **potential outcome** variables:

$$\text{Potential outcom} = \begin{array}{ll} Y_{1i} & \text{if } D_i = 1; \\ Y_{0i} & \text{if } D_i = 0. \end{array}$$

The observed outcome Y_i is Y_{0i} if $D_i = 1$ or otherwise Y_{1i}.

The **observed difference in average health between two groups**, $E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0]$, can be decomposed into two parts:

- $E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 1]$ (=$E[Y_{1i} - Y_{0i} \mid D_i = 1]$)
  - *Average treatment effect* on the treated
- $E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]$
  - *Selection bias*

The **causal effect** we are interested in:

- $E[Y_{1i} - Y_{0i}]$, the (unconditional) *average treatment effect*

Random Assignment Solves the Selection Problem

- *Random assignment* of D_i solves the selection problem as it makes D_i independent of potential outcomes.

  1. ATE on the treated is the same as the unconditional ATE: $E[Y_{1i} - Y_{0i} \mid D_i = 1] = E[Y_{1i} - Y_{0i}]$.
  2. No selection bias: $E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0] = 0$.

- Notable proponents of this method are Banerjee and Duflo, who won the Nobel prize in 2019 for the usage of Randomized Controlled Trials in policy evaluations.

IV solves the Selection Problem

- If the causal effect of interest forbids experiments, another approach is to use IV.