EDUC 263: Introduction to Econometrics, Lecture 1

Introduction

Ozan Jaquette
ozanj@ucla.edu

University of California, Los Angeles
Higher Education & Organizational Change

What we will do today

1. Introductions
2. Overview of econometrics/causal inference/program evaluation research
3. Syllabus
4. Rubin's causal model: The potential outcomes framework
5. Introduction to Stata

Next class

- Why experiments work
- Core concepts in experimental and non-experimental research design

Introductions

Overview of econometrics/causal inference/program evaluation research

The goal of causal inference research

Descriptive research questions

- Can investigate the magnitude of a problem (univariate):
  - What percentage of high school graduates attend college?
- Investigate correlational relationship between variables:
  - Relationship between buying felt furniture pads and credit score?
  - Relationship between avg. income at a high school and the number of off-campus recruiting visits by universities?

Causal research questions

- Want to know the "causal effect" of independent variable (X) on outcome (Y); If you change value of X, causal effect is the change in Y due to the change in X
- Have the form "what is effect of X on Y?" Examples:
  - What is the effect of class size on math scores?
  - What is effect of grant aid on graduation?

What is the "true" causal effect of an treatment? Actual minus counterfactual

Example: What is effect of participating in Mexican American Studies (MAS) program (X) on high school graduation (Y)

- Actual: Julie participated in MAS (X=1) and graduated from HS (Y=1)
- Counterfactual: for a person that received treatment, counterfactual is what outcome would have been if person had not received treatment
  - If Julie didn't participate in MAS would she graduate?

True causal effect of an intervention:

- Causal effect= (actual outcome for treated) minus (counterfactual outcome for treated)
- Note: causal effect could be different for each person
- What is the problem with this approach to calculating causal effects?

The primary challenge in program evaluation/causal inference research

- Causal effect= (actual outcome for treated) minus (counterfactual outcome for treated)
- Primary challenge in program evaluation methods is finding a substitute for the counterfactual
  - This substitute is called the "comparison group"
- Creating comparison groups for the counterfactual
  1. treated vs. untreated research designs (cross-sectional)
     - use "untreated" groups as "comparison group" for treated
  2. before vs. after research designs (longitudinal)
     - use "outcome before treatment" groups as comparison group for "outcome after treatment"
  3. Some research designs use both cross-sectional and longitudinal variation (e.g., "difference-in-difference")

Treated vs. untreated research designs

- Use "untreated" people as "comparison group" for treated
  - We use non-participants to represent the counterfactual for participants (i.e., what would have happened to participants if they hadn't participated)
- Example:
  - What is effect of participation in MAS on HS graduation?
- Estimate of causal effect based on "cross-sectional" variation (also called "between" variation)
  - Outcome measured at one point in time
  - Uses "between" variation: variation in Y between treated and untreated at time outcome variable is measured
- Sample:
  - People who participate in MAS **and** people who do not participate in MAS
- This course will focus on treated vs. untreated designs

Before vs. after research designs

- Use "outcome before treatment" as comparison group for "outcome after treatment"
  - Assumes outcome prior to treatment is counterfactual for outcome after treatment (i.e., what would have happened to participants if they hadn't participated)
- Example
  - What is the effect of participating in MAS (X) on days absent (Y)?
- Sample:
  - Only students who participate in MAS
- Calculate causal effect from longitudinal ("within")rather than cross-sectional ("between") variation:
  - "within" variation: change over time in Y within each person
  - Must observe outcome before treatment **and** after treatment
- Most of my research uses before vs. after designs because I study change in organizational behavior over time

Types of treated vs. untreated designs

Types of treated vs. untreated designs

1. Random assignment experiment designs
2. Observational (i.e., non-experimental) "selection on observables" designs
   - These designs control for variables that affect outcome and treatment.
   - Multivariate regression
   - Matching estimatros (e.g., propensity score matching)
3. Observational "natural experiments" designs
   - These designs utilize experimental variation in X in real world settings (e.g., access to school determined by lottery)
   - Regression discontinuity
   - Instrumental variables

Random assignment experiments: The "gold standard" treated vs. untreated designs

How experiments work:

- Randomly assign people to values of X (MAS or no MAS)
  - group randomly assigned to "no MAS" serves as counterfactual to group assigned to MAS
- On average, "treated" group is identical to control group on variables (e.g., parental education, math achievement) that affect outcome (e.g., graduation)
- Only difference between "treated" and "control" group is participation in treatment
- Therefore, can say that difference in outcome between treatment and control is due to treatment

Experiments can only identify **average** causal effect

- If we knew the true counterfactual for each person, we could identify causal effect for each person

Random assignment experiment vs. observational designs

Observational (i.e., non-experimental) design

- People not randomly assigned to values of treatment (X)
- Rather, people self-select into treatment, or some other assignment mechanism

Methods for observational data (e.g., matching, regression discontinuity) attempt to recreate experimental conditions

- Important to understand why experiments work so you can assess whether the observational method is recreating experimental conditions

Primary difference between observational methods: "Exogenous" vs. "endogenous" variation

"Exogenous" variation

- Means "determined outside the system"
- In causal inference research: means people in the sample have no influence over the value of X
- In experiment, all variation in X is exogenous

"Endogenous" variation

- Means "determined inside the system"
- In causal inference research: means people in the sample control the value of X (e.g., choose to be in MAS)

X usually contains exogenous and endogenous variation

- "natural experiment" methods isolate exogenous variation
- "selection on observables" methods controls for factors related to X that affect Y

"Selection on observables" methods (e.g., regression, matching)

X of interest contains endogenous variation; control for variables that affect the outcome and related to X

Example: effect of MAS (X1) on graduation (Y)

- Attendance (X2)positively affects graduation (Y)
- People with higher attendance (X2) more likely to choose MAS (X1)
- By including attendance (X2) in regression/matching model, we are "holding attendance constant": compare MAS students to non-MAS students with same attendance

Major assumption (usually false)

- After including control variables, no omitted variables that affect outcome and are related to value of X
- If false, treatment effect contains correlational variation

"Quasi experimental" methods that take advantage of "natural experiment" conditions

X of interest contains endogenous and exogenous variation; isolate exogenous variation, and use that variation to estimate effect of X on Y

Example: effect of serving in Vietnam War (X) on subsequent earnings (Y)

- Some people volunteered to serve (X endogenous)
- Low "draft lottery" number made Vietnam service more likely
- Lottery number (Z) is randomly assigned;
- Isolate variation in Vietnam service (X) due to draft lottery number (Z) to estimate effect of Vietnam service (X) on earning (Y)

Common quasi experimental methods

- Instrumental variables; regression discontinuity; difference-in-difference

Quasi experiments stronger than "selection on observables" methods

- Isolating experimental (exogenous) variation rather than trying to control for all the factors that affect Y and related to X

Why take a class on econometrics?

- This is a different language/culture; helpful to have a guide
- Change policy at the local level
- Fight for issues you care about
- Get a seat at the table at state/national level
- Learn language of power; then critique language of power
- On causal inference, economists are often right
- Learn how to "muddle through"; helpful for getting published

Syllabus

Rubin's causal model: The potential outcomes framework

Overview of potential outcomes framework (Rubin's causal model)

In this overview, I present slides from the presentation "Causal Inference, Potential Outcomes, and RCTs" by Vivian C. Wong

- Dr. Wong is an assistant professor of at University of Virginia
- The presentation is from a 2017 Institute for Education Sciences workshop on quasi-experimental designs

Notation for the potential outcomes framework

Write this down on separate sheet of paper

- let $i = 1...N$ be units (people) in sample
- D_i indicates receipt of treatment (e.g., Head Start)
  - $D_i = 1$ for treated units; $D_i = 0$ for untreated units
- "Potential outcomes"
  - $Y_i(1)$: outcome for person i if i receives treatment $D_i = 1$
  - $Y_i(0)$: outcome for i if i doesn't receives treatment $D_i = 0$
  - In real world, we only observe one outcome; missing outcome is the counterfactual
- "Observed outcome"
  - $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$
  - if $D_i = 1$ (treated):
    - $Y_i = Y_i(1) * 1 + Y_i(0)(1 - 1) = Y_i(1)$
  - if $D_i = 0$ (untreated):
    - $Y_i = Y_i(1) * 0 + Y_i(0)(1 - 0) = Y_i(0)$

Introduction to Stata

Introduction to Stata

- I'm assuming that all of you have experience with some statistical software package (e.g., SAS, SPSS, R)
- Some of you have experience with Stata; others not
- This is not a class about statistical programming
- I will try to give you as much relevant code as I can
- People with Stata experience: please help students without Stata experience until they feel comfortable

We will now work with some lecture material from my *Data Management Using Stata* course