Lecture 1:
Welcome to Econometrics II

August 30, 2023

Welcome

Goal of this course is twofold

Help you read research

- Are these the right data?
- Is this the right method?
- Can the author causally identify what he asserts?

Welcome

Goal of this course is twofold

Help you read research

- Are these the right data?
- Is this the right method?
- Can the author causally identify what he asserts?

Help you create research

- Ask a causal question
- Develop a technique to approximate a causal analysis
- Create tests for the validity of your work

Course Administration

1. Syllabus
2. In-class discussion
3. TA Genevieve
4. TA sessions on days indicated
5. Handouts
   - summary instructions
   - replication proposal handout
   - problem set 1
6. Random assignment to presentation dates
   - write an alias if you don't want your name public
   - trade with others as you prefer
   - and let me know

Expectations

- PhD level class
- responsible for a fair amount of learning on your own
  - choosing relevant paper
  - learning coding as you need: we do not teach Stata
  - intellectual creativity
- you must do the reading – class does not work if you do not
- use the resources I've linked online to learn some Stata

Introductions

Tell us

- name
- school
- degree
- work, if any
- why this class
- what you want to do when you're done

Lecture 1 Plan

1. The causal problem
2. Randomization
3. Some key terms
4. Experimental vs observational evidence
5. Regression as a conditional expectation function
6. Omitted variable bias in terms of regression coefficients

Defining the Problem

The Whole Semester

We talk about the **causal impact** of a **treatment** on an **outcome**.

The Whole Semester

We talk about the **causal impact** of a **treatment** on an **outcome**.

Causal impact: an impact on the outcome **because** of the treatment.

The Reason it is Hard to Credibly Identify a Causal Impact

- Define Mr. i's outcome in the world where he gets treatment as Y_{1i}
- Define Mr. i's outcome in the world where he does not get treatment as Y_{0i}
  - we call these "potential outcomes"

The Reason it is Hard to Credibly Identify a Causal Impact

- Define Mr. i's outcome in the world where he gets treatment as Y_{1i}
- Define Mr. i's outcome in the world where he does not get treatment as Y_{0i}
  - we call these "potential outcomes"
- We are interested in $Y_{1i} - Y_{0i}$

The Reason it is Hard to Credibly Identify a Causal Impact

- Define Mr. i's outcome in the world where he gets treatment as Y_{1i}
- Define Mr. i's outcome in the world where he does not get treatment as Y_{0i}
    - we call these "potential outcomes"
- We are interested in $Y_{1i} - Y_{0i}$
- Why is this tough?

The Reason it is Hard to Credibly Identify a Causal Impact

- Define Mr. i's outcome in the world where he gets treatment as Y_{1i}
- Define Mr. i's outcome in the world where he does not get treatment as Y_{0i}
  - we call these "potential outcomes"
- We are interested in $Y_{1i} - Y_{0i}$
- Why is this tough?
- We observe Mr. i in only one world

The Reason it is Hard to Credibly Identify a Causal Impact

- Define Mr. i's outcome in the world where he gets treatment as Y_{1i}
- Define Mr. i's outcome in the world where he does not get treatment as Y_{0i}
  - we call these "potential outcomes"
- We are interested in $Y_{1i} - Y_{0i}$
- Why is this tough?
- We observe Mr. i in only one world

Bottom line: $Y_{1i} - Y_{0i}$ is fundamentally unobservable.

Do Lots of People Solve the Problem?

Maybe we care more generally about the average difference:

$$\text{Avg}_n[Y_{1i} - Y_{0i}]$$

Do Lots of People Solve the Problem?

Maybe we care more generally about the average difference:

$$\text{Avg}_n[Y_{1i} - Y_{0i}] \;=\; \frac{1}{n}\sum_{i=1}^{n}[Y_{1i} - Y_{0i}]$$

Do Lots of People Solve the Problem?

Maybe we care more generally about the average difference:

$$\begin{aligned} \text{Avg}_n[Y_{1i} - Y_{0i}] &= \frac{1}{n}\sum_{i=1}^{n}[Y_{1i} - Y_{0i}] \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_{1i} - \frac{1}{n}\sum_{i=1}^{n} Y_{0i} \end{aligned}$$

Do Lots of People Solve the Problem?

Maybe we care more generally about the average difference:

$$\begin{aligned}\text{Avg}_n[Y_{1i} - Y_{0i}] &= \frac{1}{n}\sum_{i=1}^{n}[Y_{1i} - Y_{0i}] \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_{1i} - \frac{1}{n}\sum_{i=1}^{n} Y_{0i}\end{aligned}$$

No, lots of people don't solve the problem.

Defining Treatment

- Let D_i be the treatment
- Let $D_i \in \{0, 1\}$
- We usually assume a binary treatment for ease of analysis

$$D_i = \begin{cases} 1 & \text{if treated} \\ 0 & \text{if not treated} \end{cases}$$

Defining the Conditional Expectation

The expectation function:

$$E[Y_i]$$

- population average of outcomes for all units i
- We say "the expectation of Y_i"

Defining the Conditional Expectation

The expectation function:

$$E[Y_i]$$

- population average of outcomes for all units i
- We say "the expectation of Y_i"

A conditional expectation is the population average for the subset that follows the condition.

- $E[Y_i|D_i = 1]$
  - the average outcome for units for which $D_i = 1$

Defining the Conditional Expectation

The expectation function:

$$E[Y_i]$$

- population average of outcomes for all units i
- We say "the expectation of Y_i"

A conditional expectation is the population average for the subset that follows the condition.

- $E[Y_i|D_i = 1]$
  - the average outcome for units for which $D_i = 1$
- $E[Y_i|X = x]$
  - the population average outcome for units where variable X is equal to x
  - for example, average height (Y_i) for people whose hair is curly ($X = x$)

Re-writing the Problem of Interest

$$\text{Avg}_n[Y_i|D_i = 1] - \text{Avg}_n[Y_i|D_i = 0]$$

Re-writing the Problem of Interest

$$\text{Avg}_n[Y_i|D_i = 1] - \text{Avg}_n[Y_i|D_i = 0]$$

But note!

$\text{Avg}_n[Y_i|D_i = 1] = \text{Avg}_n[Y_{1i}|D_i = 1]$ and $\text{Avg}_n[Y_i|D_i = 0] = \text{Avg}_n[Y_{0i}|D_i = 0]$

Re-writing the Problem of Interest

$$\text{Avg}_n[Y_i|D_i = 1] - \text{Avg}_n[Y_i|D_i = 0]$$

But note!
$\text{Avg}_n[Y_i|D_i = 1] = \text{Avg}_n[Y_{1i}|D_i = 1]$ and $\text{Avg}_n[Y_i|D_i = 0] = \text{Avg}_n[Y_{0i}|D_i = 0]$

Therefore

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

Re-writing the Problem of Interest

$$\text{Avg}_n[Y_i|D_i = 1] - \text{Avg}_n[Y_i|D_i = 0]$$

But note!
$\text{Avg}_n[Y_i|D_i = 1] = \text{Avg}_n[Y_{1i}|D_i = 1]$ and $\text{Avg}_n[Y_i|D_i = 0] = \text{Avg}_n[Y_{0i}|D_i = 0]$

Therefore

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

That is, we only see treated outcomes for treated people and untreated outcomes for untreated people.

Now Assume a Constant Treatment Effect

Assume that the effect of treatment D is the same for every person, treated or not. When might this be the case?

$$Y_{1i} - Y_{0i} = \kappa$$

Now Assume a Constant Treatment Effect

Assume that the effect of treatment D is the same for every person, treated or not. When might this be the case?

$$Y_{1i} - Y_{0i} = \kappa$$

To be clear, this implies

$$Y_{1i} = \kappa + Y_{0i}$$

Assuming a Constant Treatment Effect Makes Selection Bias Easy to See

$$\text{Avg}_n[Y_{1i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0] \quad =$$

Assuming a Constant Treatment Effect Makes Selection Bias Easy to See

$$
\begin{aligned}
\text{Avg}_n[Y_{1i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0] &= \\
\text{Avg}_n[\kappa + Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0] &=
\end{aligned}
$$

Assuming a Constant Treatment Effect Makes Selection Bias Easy to See

$$
\begin{aligned}
\text{Avg}_n[Y_{1i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0] &= \\
\text{Avg}_n[\kappa + Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0] &= \\
\kappa + \text{Avg}_n[Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0] &
\end{aligned}
$$

Assuming a Constant Treatment Effect Makes Selection Bias Easy to See

$$
\begin{aligned}
\text{Avg}_n[Y_{1i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0] &= \\
\text{Avg}_n[\kappa + Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0] &= \\
\kappa + \text{Avg}_n[Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0]
\end{aligned}
$$

Red term is difference in outcome Y for treated relative to untreated in the absence of treatment: **selection bias**.

How Randomization Solves the Problem

Remember the Problematic Term

$$\text{Avg}_n[Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0]$$

Remember the Problematic Term

$$\text{Avg}_n[Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0]$$

How can we assure ourselves that there are no systematic differences between treated and untreated people?

Remember the Problematic Term

$$\text{Avg}_n[Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0]$$

How can we assure ourselves that there are no systematic differences between treated and untreated people? Randomize.

Remember the Problematic Term

$$\text{Avg}_n[Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0]$$

How can we assure ourselves that there are no systematic differences between treated and untreated people? Randomize. Enough people.

Remember the Problematic Term

$$\text{Avg}_n[Y_{0i}|D_i = 1] - \text{Avg}_n[Y_{0i}|D_i = 0]$$

How can we assure ourselves that there are no systematic differences between treated and untreated people? Randomize. Enough people.

For a specific definition of “enough” you need a power analysis – take a different class.

But How Can You Be Sure Your Randomization Worked?

We think of *i*s (people, neighborhoods, etc) as having two qualities

1. those observed by researchers
2. those unobserved by researchers

But How Can You Be Sure Your Randomization Worked?

We think of *i*s (people, neighborhoods, etc) as having two qualities

1. those observed by researchers
2. those unobserved by researchers

Randomization works when these characteristics do not vary across the treated and untreated groups. This is also known as when the sample is "balanced."

But How Can You Be Sure Your Randomization Worked?

We think of *i*s (people, neighborhoods, etc) as having two qualities

1. those observed by researchers
2. those unobserved by researchers

Randomization works when these characteristics do not vary across the treated and untreated groups. This is also known as when the sample is "balanced."

But we only see the observed characteristics.

But How Can You Be Sure Your Randomization Worked?

We think of *i*s (people, neighborhoods, etc) as having two qualities

1. those observed by researchers
2. those unobserved by researchers

Randomization works when these characteristics do not vary across the treated and untreated groups. This is also known as when the sample is "balanced."

But we only see the observed characteristics.

So we compare observed characteristics across the two groups and **assume** that if those do not differ, the unobserved characteristics do not differ.

But How Can You Be Sure Your Randomization Worked?

We think of *i*s (people, neighborhoods, etc) as having two qualities

1. those observed by researchers
2. those unobserved by researchers

Randomization works when these characteristics do not vary across the treated and untreated groups. This is also known as when the sample is "balanced."

But we only see the observed characteristics.

So we compare observed characteristics across the two groups and **assume** that if those do not differ, the unobserved characteristics do not differ.
Put differently,

$$\text{Avg}_n[Y_{0i}|D_i = 1] = \text{Avg}_n[Y_{0i}|D_i = 0]$$

Likelihood of treatment is independent of outcome in the untreated state.

Some Key Terms

Key Terms and Concepts

1. unit of observation
2. identification
3. matrix algebra notation

The Unit of Observation

The unit of observation is

- the level of the individual unit that you analyze
- can be person, country, neighborhood, person-year, neighborhood-hour
- time series data always have a time component

The Unit of Observation

The unit of observation is

- the level of the individual unit that you analyze
- can be person, country, neighborhood, person-year, neighborhood-hour
- time series data always have a time component

The unit of observation is a **big deal**

- results differ by unit of observation
- flawed understanding of unit of observation can ruin analysis
- Social Security: you can estimate between 13 and 22 percent of people over 55 get their income solely from social security, depending upon the unit of observation
- sidebar: which gives the lower number, persons or families? See nice handout here.

Identification

You will hear

- What is the identification strategy?
- What identifies this result?
- Where is the identification coming from?
- Is this result identified?

Identification

You will hear

- What is the identification strategy?
- What identifies this result?
- Where is the identification coming from?
- Is this result identified?

"Identification" means the random variation that yields a causal result.

Identification

You will hear

- What is the identification strategy?
- What identifies this result?
- Where is the identification coming from?
- Is this result identified?

"Identification" means the random variation that yields a causal result.

"Identification strategy" is a strategy for finding a causal effect with a non-randomized treatment

Matrix Algebra Notation

- You do not need to know matrix algebra for this class
  - though if you want to be more serious, it's not a bad idea
- But some papers will use this notation
- I'd like you to be slightly familiar with it

A Vector of Outcomes

- In your previous class, you probably talked a lot about y_i, or an outcome for Mr. i.
- You may have written $\sum_{i=1}^{N} y_i$, where N is the number of people in the population
- in this class, you will frequently see just Y, which is

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}$$

A Vector of Covariates

You may remember that Mr. i has a vector of covariates associated with him: $(x_{1i}, x_{2i}, \ldots, x_{Ki})$, where K is the number of covariates

$$X = \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{K1} \\ x_{12} & x_{22} & \ldots & x_{K2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1N} & x_{2N} & \ldots & x_{KN} \end{bmatrix}$$

Putting These Together

We can therefore also rewrite the regression equation you know as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki} + \epsilon_i$$

as

$$Y = X\beta + e$$

where

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

Next Lecture

- Start looking for an article to replicate – be aware that this can take some time
- Read *MM* Chapter 2
- Read Black article on webpage, pages as noted