

# ECON 340

## Economic Research Methods

Div Bhagia

Lecture 23  
Experiments & Quasi-Experimental Methods

# It's hard to answer causal questions!

- Say, I am interested in whether going to college will increase Sam's income.
- Ideally, what I need:
  - Income of Sam if he goes to college
  - Income of Sam if he does not go to college

But a Sam with and without college doesn't exist!

Hence, this is a situation of a missing *counterfactual*.

# What's the counterfactual?

Suppose Sam does not end up going to college.

Can we answer what would happen if he did?

$$\underbrace{Income_{Sam,college=1} - Income_{Sam,college=0}}_{\text{Missing counterfactual}}$$

What can we do? Is there a good proxy for the counterfactual outcome?

# What's the counterfactual?

Suppose Sam does not end up going to college.

Can we answer what would happen if he did?

$$\underbrace{Income_{Sam,college=1} - Income_{Sam,college=0}}_{\text{Missing counterfactual}}$$

What can we do? Is there a good proxy for the counterfactual outcome?

We want someone just like Sam, who went to college.

# Randomization

- Gold standard: *Randomized controlled trials (RCTs)* or *experiments*
- Common in medicine and sciences, limited in social sciences
- Assign *treatment* randomly, then on average individuals in the *treated* group similar to individuals in the *control* group

# Randomized controlled trials

Define,

$$D = \begin{cases} 1 & \text{if received treatment} \\ 0 & \text{if did not receive treatment} \end{cases}$$

Outcome:

$$Y = \beta_0 + \beta_1 D + u, \quad E(u) = 0$$

If  $D$  is randomized, the exogeneity assumption is satisfied:

$$E(u|D = 1) = E(u|D = 0) = 0$$

# Treatment Effect

$$Y = \beta_0 + \beta_1 D + u$$

Under exogeneity  $E(u|D) = 0$ , the difference in the dependent variable between individuals is due to  $D$ :

$$TE = E(Y|D = 1) - E(Y|D = 0) = \beta_1$$

This effect is called the treatment effect (TE).

# Example

- 300 individuals participated in an RCT to evaluate the effect of caffeine on sleep.

- Treatment:

$$D = \begin{cases} 1 & \text{if given a cup of caffeinated coffee before bed} \\ 0 & \text{if given a cup of decaf coffee before bed} \end{cases}$$

- Flipped a coin to decide which individuals got the real coffee vs decaf (*placebo*)
- Measure the quality of sleep using a score

Aside: Should the control group get decaf or no coffee?

# Example

$$\text{SleepScore} = \beta_0 + \beta_1 D + u, \quad E(u) = 0$$

- We expect different individuals to have different sleep scores due to several factors, which is captured by  $u$
- But since we randomized treatment, these factors are uncorrelated with treatment, i.e.,  $E(u|D) = E(u)$  or equivalently  $\text{Cov}(D, u) = 0$ .
- In other words, some individuals may sleep poorly, but there is an equal chance for these individuals to end up in the treatment or the control group

# Example

$$\text{SleepScore} = \beta_0 + \beta_1 D + u, \quad E(u) = 0$$

- With randomization, it is no longer the case that individuals who had the real coffee might be systematically different from those who didn't
- Hence, the difference in outcomes between the two groups captures the impact of coffee on sleep and not the omitted factors that affect sleep and are correlated with coffee consumption.

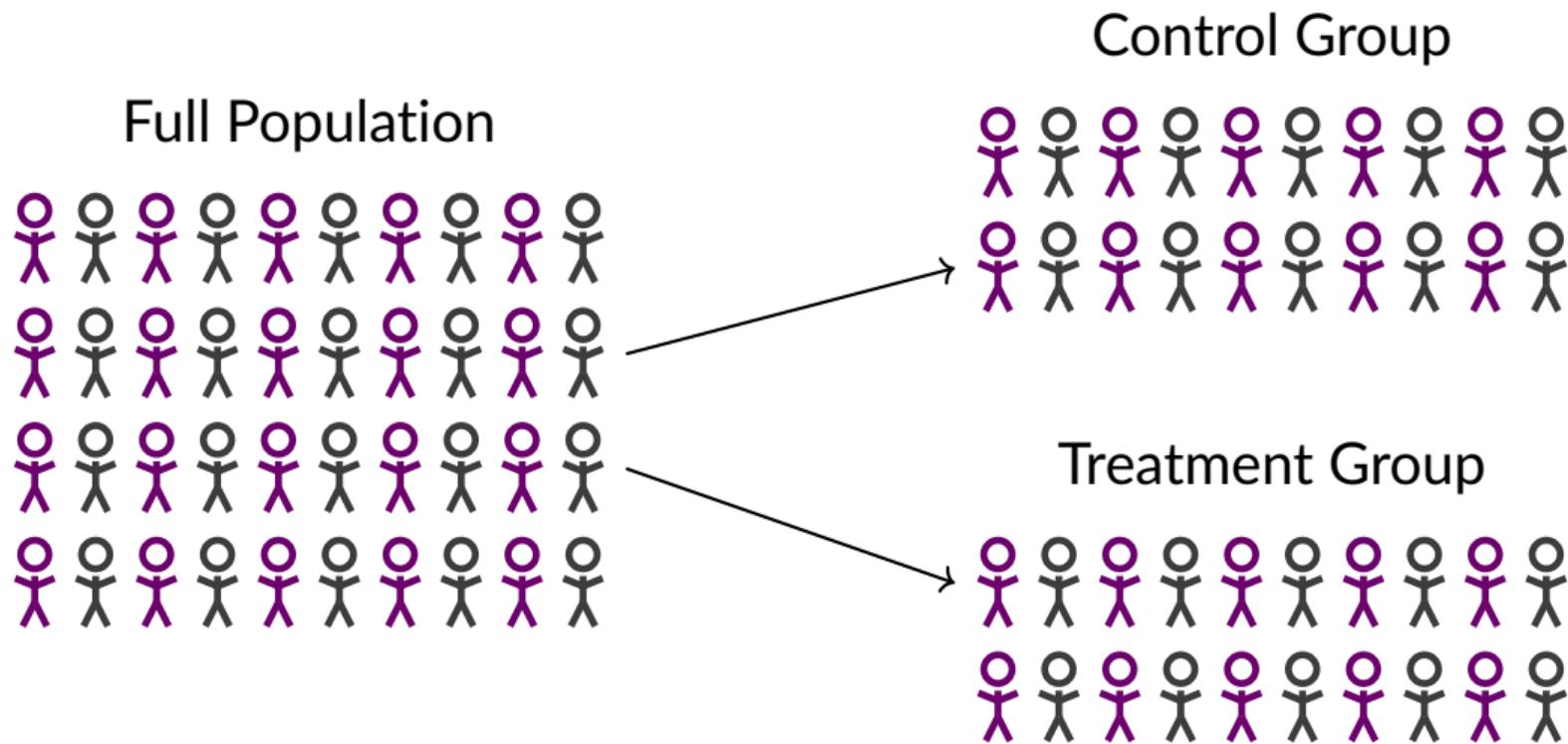
# A Simple Example

- Say half of the people are good sleepers with  $u = 10$ , while the other half are poor sleepers with  $u = -10$
- We randomly put half of the individuals into the control group and half into the treatment group
- Then we should expect a 50-50 distribution of good and poor sleepers in both treatment and control groups

$$E(u|D = 1) = 0.5(10) + 0.5(-10) = 0$$

$$E(u|D = 0) = 0.5(10) + 0.5(-10) = 0$$

# Randomization



# Experiments are not always feasible

- It's often infeasible to run experiments to answer interesting economic questions. We can't just randomize college attendance for a group of individuals!
- Economists have to be clever. Use *quasi*-experimental variation. Need to look for *natural experiments*.
- **2021 Nobel Prize** awarded to David Card, Joshua Angrist and Guido Imbens for their contributions to quasi-experimental methods in economics. (Reading on course website)

# Angrist and Krueger (1991)

- In the US, children typically enter school in the fall of the year they turn six
- Individuals are required by law to stay in school until their sixteenth birthday
- Because all children who are born in a particular calendar year start school on the same date, children who are born early in the year can leave school sooner than children born later in the year.

# Angrist and Krueger (1991)

Example:

- Sam born in Jan 1991, starts school in Aug 1997, can drop in Jan 2007
- Jack born in Dec 1991, starts school in Aug 1997, can drop in Dec 2007

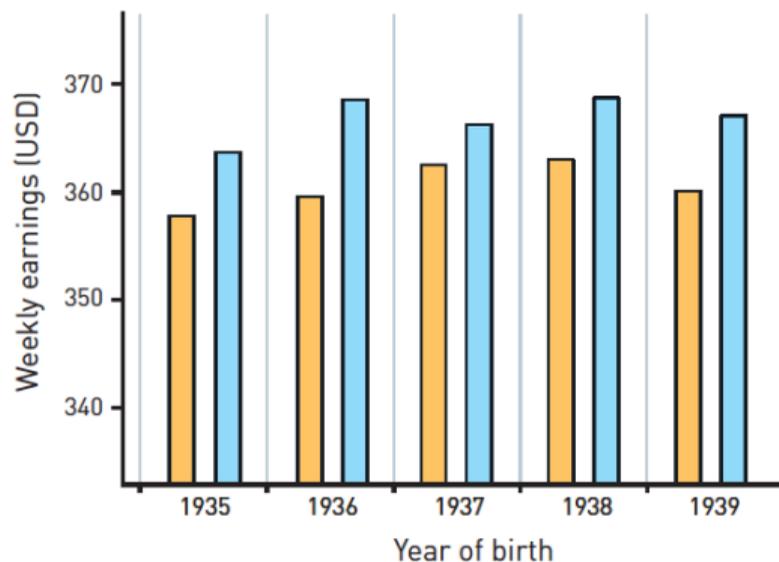
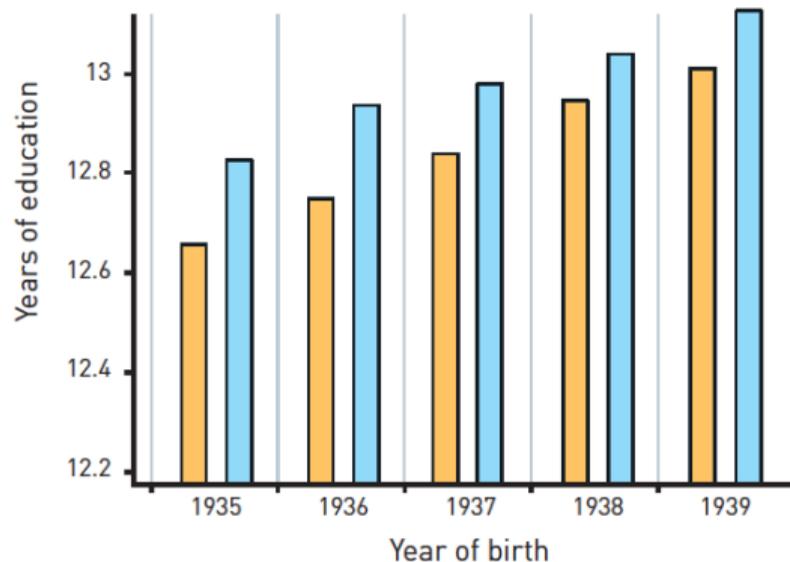
Jack can drop out almost a year earlier out of school than Sam, just because he was born in Dec rather than Jan.

As long as the quarter of birth is random, which is plausible, we could use this variation to estimate the causal impact of education on income.

# Angrist and Krueger (1991)

Additional years of education have a positive effect on income. The figure uses data from Angrist and Krueger (1991).

■ Born in first quarter ■ Born in fourth quarter



# Quasi-Experimental Methods

There is a growing toolkit of quasi-experimental methods. Here are three popular ones:

- (1). *Propensity Score Matching*: Match individuals in the treatment group with individuals in the control group who are similar in all relevant characteristics except for the treatment.

# Quasi-Experimental Methods

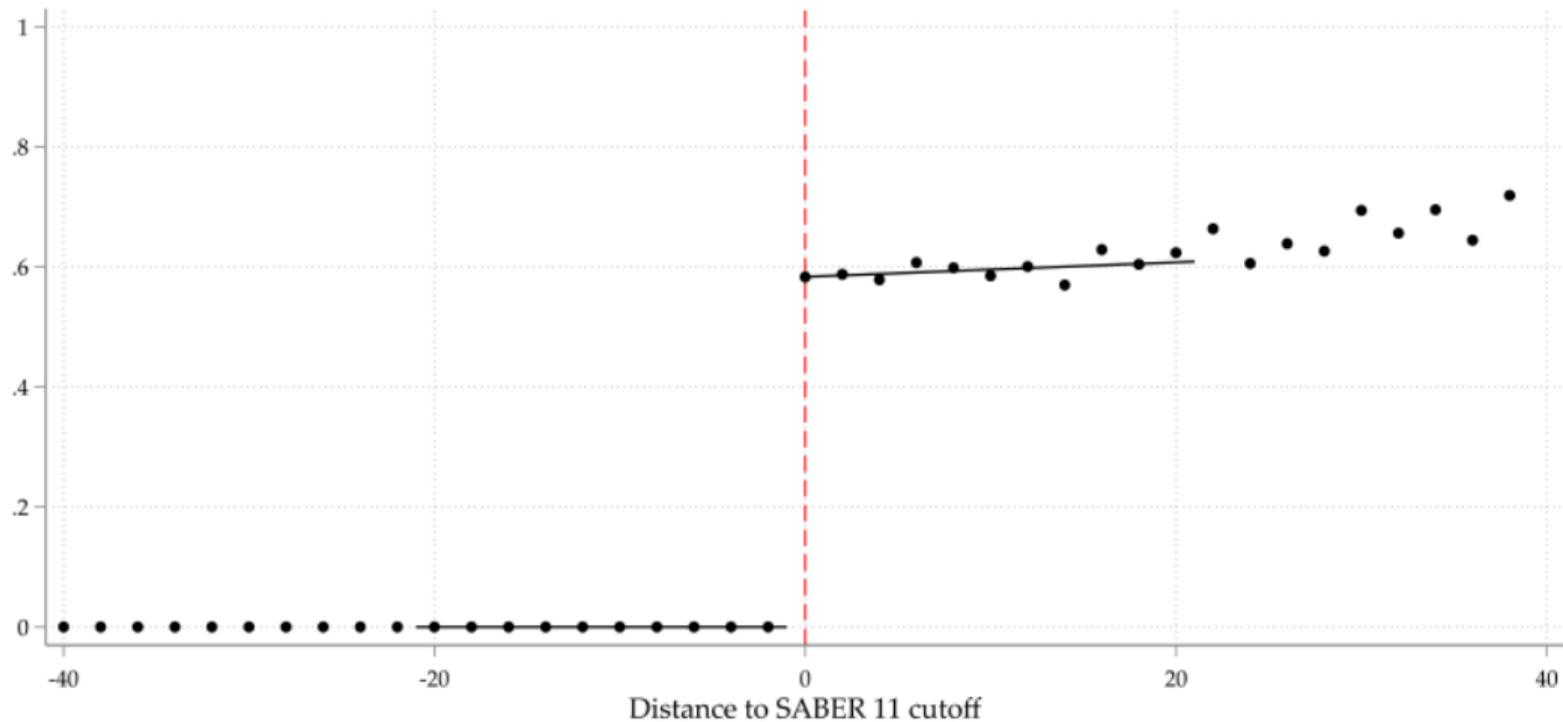
There is a growing toolkit of quasi-experimental methods. Here are three popular ones:

- (2). *Regression Discontinuity*: This design is used when an intervention is assigned based on a continuous score or threshold, and participants on either side of the threshold are assumed to be similar except for the treatment. The difference in outcomes between the two groups is used to estimate the causal effect of the treatment.

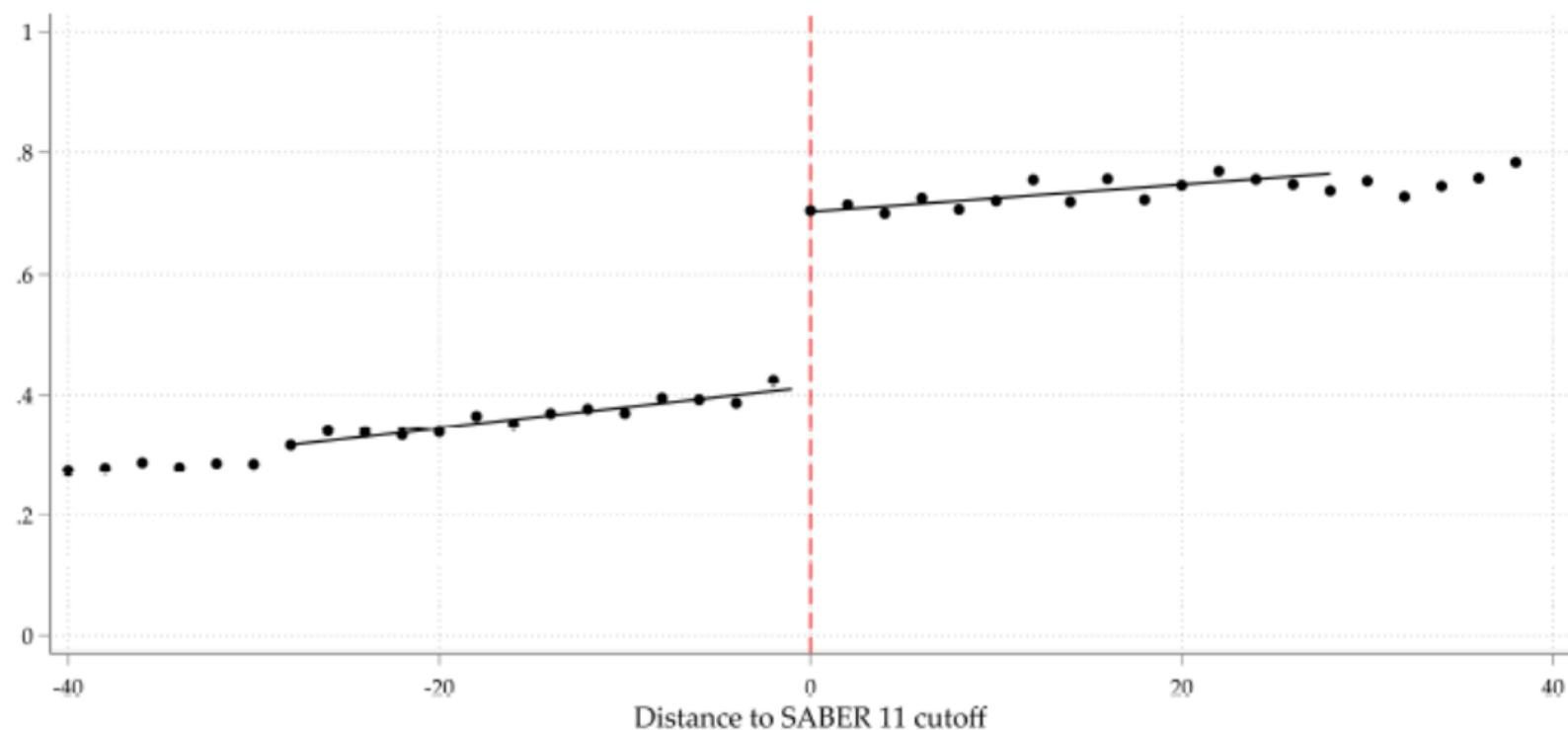
## Londoño-Vélez et. al. (2024)

- In 2014, Colombia implemented a nationwide financial aid program covering the tuition of four-year undergraduate programs at 33 “high-quality” universities
- Beneficiaries required to satisfy two conditions: (1) score in the top decile of Colombia’s national standardized high school exit exam, SABER 11, and (2) come from households in the poorest 52.8%
- Leverage the program’s discontinuous assignment rules based on test scores and household wealth using a regression discontinuity design

# Eligibility



# Share enrolling in college



# Quasi-Experimental Methods

There is a growing toolkit of quasi-experimental methods. Here are three popular ones:

- (3). *Differences-in-Difference*: This design estimates the causal effect of an intervention by comparing the change in outcomes between a treatment group and a control group over time, assuming similar trends in the absence of the intervention.

# Differences-in-Differences

