Labor Economics

Intro to Empirics: Causality

Motaz Al-Chanati

Summer 2019

Columbia University

Outline

1. Causality

2. Identification Strategies

- Randomized Control Trials (RCT)
- Regression Discontinuity (RD)
- Instrumental Variable (IV)
- Difference-in-Difference (DID)

Causality

Regressions

- So far: use regressions for *predictions*

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- If X increases by 1 unit, we predict that Y will increase by β_1
  - Alternatively: "1 unit change in X is *associated* with a β_1 change in Y"
  - Does **not** mean that 1 unit change in Xwill result in Y changing by β_1
- Common phrase: **correlation does not imply causation**
  - Regressions just fit data. Can't see the truth in the world

Why Causality?

- Often a key question: "what effect does X have on Y?"
  - $X =$class size, $Y =$ test scores
  - $X =$attending college, $Y =$ earnings
  - $X =$minimum wage, $Y =$ unemployment
- Why should we care about causality?
  - Intellectual curiosity: how does the world work?
  - Policy: how can you fix something without understanding how it works?

Example

- Suppose we have a dataset recording each student's test score (Y) and the number of books they have at home (X)
- We run a regression of Y on X and find a positive coefficient $\beta_1 > 0$
- More books at home associated with higher test scores
  - If you treat this as causal, what is the policy implication?
  - Give everyone lots of books $\rightarrow$ watch test scores rise!
  - But if it is *not* causal, then just implemented a bad policy

Framework

- Suppose that we think D affects Y (call D the **treatment** variable)

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$

  - In the case before, let D = number of books at home, Y = test scores

- If $\beta_1 \neq 0$, then regression tells us that there is an association between D and Y. But it can't distinguish between two possibilities:
  1. D affects $Y \implies$ correlation between D and Y
  2. D correlated with X, and X affects $Y \implies$ correlation between D and Y

- Where X is some other variable. In this case, X could be family income

Framework

- So the *true* equation could be:

$$Y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 X_i + \varepsilon_i$$

- We would have *misspecified* the model because we said it was:

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$

- This results in the wrong coefficient estimate ($\alpha_1 \neq \beta_1$)
  - This creates a **biased estimate** of the coefficients (i.e. it might too high or too low)
  - Intuitively: the regression says that D causes more change in Y than it *actually* does

Example

- Continuing the books example: coefficient on number of books is likely biased *upwards*
  - It is *positively* correlated with family income
  - Higher family income results in higher test scores
- The coefficient attributes *too much* of the changes in test scores to changes in books.
  - In reality, books were changing because income was changing (and that was the real reason scores were changing)

Omitted Variable Bias

- Most common reason why a regression does not give a causal relationship is **omitted variable bias (OVB)**
- Imagine the true dataset recorded every possible piece of information about the observation (i.e. infinite columns)
  - Note: not more observations (i.e. more rows), but more info about each row
  - e.g. amount in savings, blood type, measures of personality etc
- Our dataset is only a subset of the true one - we only see some of the columns
  - Every column we do not have goes into the unobserved error term ε

Omitted Variable Bias

- OVB occurs when the treatment variable D is correlated with the error term ε
- Solution? We can **control** for X by including it in the model (i.e. remove X from ε so that we remove the correlation between D and ε)
  - But that may not work...
  - May not have the data on X
  - How do we know that there isn't some other variable Z in ε that will cause the same problem? (e.g. mother's education)
  - Controlling for more is generally good, but it cannot solve the causality problem (sometimes makes things worse)

Selection

- Even if you collected lots of data and controlled for every observable characteristic, we still worry about **unobservable** characteristics
  - e.g. people *choose* the number of books in their home
  - A *choice* variable is **endogenous** (depends on something else, e.g. preferences)
  - People with more books could value education more, pass on these preferences to their children, leading to higher test scores
  - This causes a **selection bias** issue (an example of OVB)
- Basically: we will never have all the right columns in our dataset

Enter the Economists

- At this point, sounds hopeless
  - For most fields: “correlation is not causation” is simply a warning
  - For economists: it is a challenge
- Economists are obsessed with studying causality, especially in recent years
  - Our advantage over other social sciences
  - Labor economists played a major role in this **credibility revolution**
- Causality is important, but does not mean correlations are useless
  - Identifies possible relationships to explore
  - Useful for predictions (e.g. identifying at-risk population)

Design-Based Studies

- To sum up so far:
  - Problem: **causal inference** (how to draw conclusions about causal effects?)
  - Not the solution: controlling for more
- Solution: **design-based studies**
  - Think about experiments - putting the *science* in social science
  - It's not about more data, it's about good *research design*
  - We will now see the **identification strategies** to isolate the causal effects

Identification Strategies

Terminology

- Our goal is to see how D (the treatment variable) affects Y (the outcome variable)
  - Think of the treatment D as being given some magical pill
- There are two groups of people in the world:
  - The **treated/treatment group**: the people who receive the treatment D
  - The **control group**: the people who do not receive the treatment D
- Variation Y can also be explained by variation in some other variables X that we will call **controls**
  - X and D are both explanatory variables, but D is the only one that we care about finding the causal effect of
  - X can (and usually does) contain many variables

Fundamental Problem

- To see how the treatment affects the outcome, we need to observe two things:
  1. What happens to you if you take the treatment?
  2. What happens to you if you do not take the treatment?

- The only way to do this?
  - Time travel

- Since that is not possible (yet!), we have the **fundamental problem of causal inference**
  - People either are in the treated or control group.
  - We cannot see the **counterfactual**: what would have happened if they were in the other group?

Example

- Let's work with an example throughout this section
- Suppose we wanted to learn the returns to education
  - What is the causal effect of education on earnings?
- Estimate the following equation for person i:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \varepsilon_i$$

  where Y_i is log income, D_i is education, X_i are controls (race, gender, age etc)
  - Note: use log income because income has very skewed distribution
  - Coefficient interpretation: 1 unit increase in D_i causes a $100 \times \beta_1\%$ increase in income

Example: OVB

- Very important question: does more school lead to higher income?
  - Typically, D_i is years of schooling (quantity)
  - Can also think about D_i as quality of education
- As it stands, cannot interpret it casually
  - People who have more education (treated) are very different to people who have less education (control)
  - Different both in observables *and* unobservables

Identification Strategies

Randomized Control Trials (RCT)

Gold Standard

- If a scientist wants to examine the effect of a new medicine, they set up an **experiment** or **randomized control trial (RCT)**
  - Take a group of people, and randomly assign them to the treatment and control groups
  - Give the treatment group the treatment, and the control group nothing
  - Look at how their outcomes differ, on average. This difference is because of the treatment
- This is probably familiar to us, e.g. high school science, media
  - This is called “The Gold Standard”. The best we can do, given the fundamental problem

RCTs and Causality

- Why do RCTs show causality?
  - Treatment assigned **exogenously** (it was not *chosen*)
  - We do not have to worry about selection bias (no selection)
  - We do not have to worry about any OVB (if assigned randomly, should be uncorrelated with anything else)
- What about the fundamental problem (FP)? This is not time travel!
  - FP says we can't know how the treatment affects a particular individual
  - RCTs tell us the **average treatment effect** (what is effect of the treatment *on average*?)
  - Control group (as a whole) provides a counterfactual for the treatment group (as a whole)

Regression for RCT

- Regressions for RCTs are simple. Just run it as written:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \varepsilon_i$$

  - If treatment D_i was randomly assigned, can interpret β_1 causally
  - Including controls X isn't necessary, but helps with precision of estimates (lower SEs)

Issues with RCTs

1. Was treatment truly randomly assigned?
2. Did people comply with their assignment?
    - Did people in the control group seek the treatment in some other way?
    - Did people in the treatment group actually take the treatment?
3. Was there any *unintended* treatment affecting the results?
    - In medicine: treated gets pill, control group gets placebo (sugar pill). Why?
4. How well can we generalize the results?
    - Well-designed RCTs have good **internal validity** (correctly estimate the causal effect for the people in the experiment)
    - Sometimes have questionable **external validity** (will we find the same effects in other settings?)

How Does Your Kindergarten Classroom Affect Your Earnings? *Evidence from Project STAR*

Raj **Chetty**

John **Friedman**

Nathaniel **Hilger**

Emmanuel **Saez**

Diane Whitemore **Schanzenbach**

Danny **Yagan**

***QJE*, 2011**

Paper Overview

- **Question:** Does school quality in early childhood affect future earnings?
  - Measures of school quality: class size, teacher experience
- **Endogeneity Problem:** Cannot just compare people who had small classes to people who had large classes
  - Schools that have smaller class sizes also have more funding/better resources and are in more affluent areas.
- **Ideal Experiment:** Randomize class size and observe what happens to earnings
  - This is exactly what happened in Tennessee!

Project STAR

- Project STAR: 11,571 students in TN and teachers were randomly assigned within their schools from kindergarten to third grade
  - Some students assigned to small classes (~15) and some to large (~22)
  - Experiment was done in 1985-1989. Can observe these people now as adults
- Past studies: class size, teacher quality had impacts on test scores
  - Strong correlation between test scores and future earnings

    Graph
  - Possible that school quality affects earnings

Class Size

- Since class size was randomly assigned, can effectively run this regression

$$Y_i = \beta_0 + \beta_1 Small_i + \varepsilon_i$$

where Y_i is person i's outcome and $Small_i$ is a dummy for being in a small class (treatment)

Table 1: Project STAR: Class Size Effects

Dependent variable	(1) Test score (%)	(3) College by age 27 (%)	(5) Wage earnings ($)
Small class (no controls)	4.81	1.91	4.09
	(1.05)	(1.19)	(327)
Observations	9,939	10,992	10,992
Mean of dep. var.	48.67	45.50	15,912

Source: Chetty et al. (2011), Table 5

Teacher Effects

- Since teachers were randomly assigned, can effectively run this regression

$$Y_i = \beta_0 + \beta_1 Exp_i + \varepsilon_i$$

where Y_i is person i's outcome and Exp_i is a dummy for having an experienced teacher (treatment)

Table 2: Project STAR: Teacher Effects

Dependent variable	(1) Test score (%)	(2) Wage earnings ($)	(3) Test score (%)	(4) Wage earnings ($)
Teacher with >10 years of experience	3.18	1093	1.61	−536.1
	(1.26)	(545.5)	(1.21)	(619.3)
Entry grade	KG	KG	Grade ≥1	Grade ≥1
Observations	5,601	6,005	4,270	4,909

Source: Chetty et al. (2011), Table 6

Results

- Smaller class size leads to higher test scores and more likely to be enrolled in college
  - Increase in earnings at age 27 are not significant (estimates are very *noisy*)
- Students assigned to an experienced KG teacher have higher income
  - Cannot causally say this due to the teacher's experience
  - Teacher experience is correlated with other factors, so cannot identify mechanism
  - But, it does say that teachers in early childhood have an impact on later life

Alternatives to RCTs

- RCTs are great, but not perfect
- We can't use RCTs to answer every problem
  - Some things cannot be randomized for practical/ethical reasons (e.g. randomize people going to prison)
- In real life, there are sometimes randomizations that give us **natural experiments**
  - These are settings that look close to an RCT (e.g. judge assigned to case is random)
- But not everything is assigned randomly, so we sometimes look for **quasi-experiments**
  - No randomization, but follow a similar framework (hence, quasi)

Identification Strategies

Regression Discontinuity (RD)

RD Setup

- RD looks for a treatment (D) that is assigned based on some cutoff
  - Need a **running variable** (X), some continuous characteristic (e.g. a test score, birthday)
  - Need a **cutoff** on the running variable (X^*) that determines treatment (e.g. get scholarship if you score above 80%)
- Treatment ($D = 1$) is assigned as follows:

$$D = \begin{cases} 1 & \text{if } X \geq X^* \\ 0 & \text{if } X < X^* \end{cases}$$

RD Intuition

- Cutoffs are a pretty arbitrary way of making decisions
  - Your friend gets a 91% in a class and gets an $A-$. You get a 89% and get a $B+$
  - Are you and your friend really that different?
- Real life has a lot of randomness
  - Take a group of people. Suppose each one has a "true" test score of X^*.
  - On test day, some people get a little lucky and score above their true value: $X > X^*$. Some people get a little unlucky and score below their true value: $X < X^*$
  - Which side of the cutoff you end up on is random (i.e. it's *almost* like the world runs an RCT for us, so this is a **quasi-experiment**)

RD Requirements

- How to get a group of people with the same "true" score of X^*? We never observe this
  - Someone who scores 99% is very different to someone who scores 20%. Can't attribute all of that to just luck.

- Key idea: if you look at ***just before*** and ***just after*** the cutoff, the treatment is close to random
  - Within a *small* window, we assume that all variation is due nature

- How to choose the window? Pretty arbitrary. There are trade-offs
  - Narrow enough to make sure the people are similar on either side of cutoff

  - Wide enough to make sure we have a big enough sample

RD Requirements

- Big requirement to RD is that you can't "manipulate the running variable"
- If the cutoff for a test is 80%, people can't be able to exactly achieve 80%
- Assignment of treatment is no longer random. Now it is a deliberate choice, which reflects a selection problem
- If you could choose your test score exactly, do you think the people who get 80% are similar to people who got 79%?

RD Regression

- For outcome Y, running variable X, and cut-off X^*, the regression is:

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 (X \times D) + \varepsilon$$

  where $D = \mathbb{1}\{X \geq X^*\}$ (indicator variable)

- If below the cutoff: $D = 0$

$$Y = \beta_0 + \beta_2 X$$

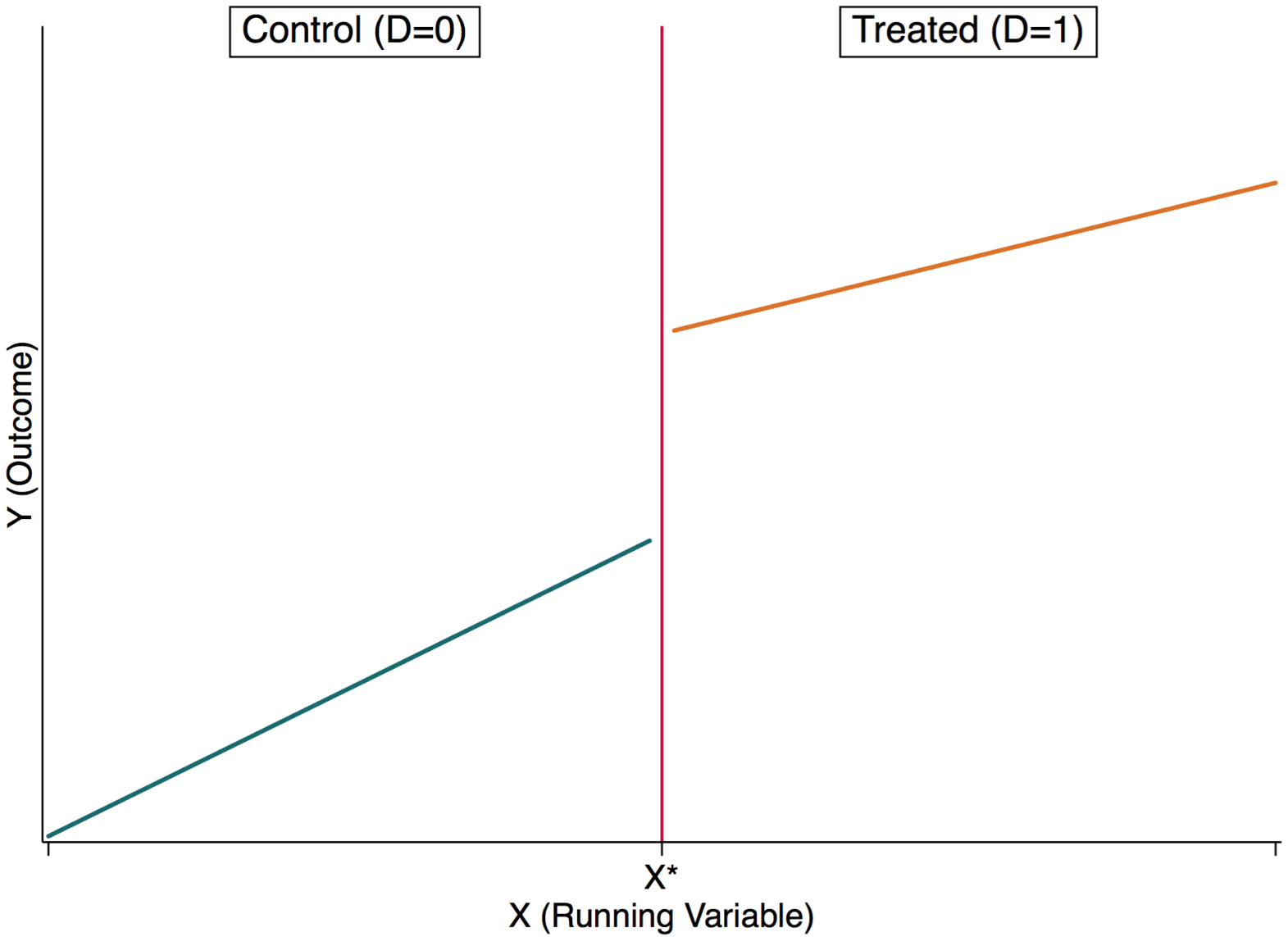
- If above the cutoff: $D = 1$

$$Y = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) X$$

- We are often interested in β_1: the jump that occurs at the cutoff X^*

RD Graph

Figure 1: Regression Discontinuity



Are Some Degrees Worth More than Others?

Evidence from College Admission Cutoffs in Chile

Justine
Hastings

Christopher
Neilson

Seth
Zimmerman

NBER WP, 2014

Paper Overview

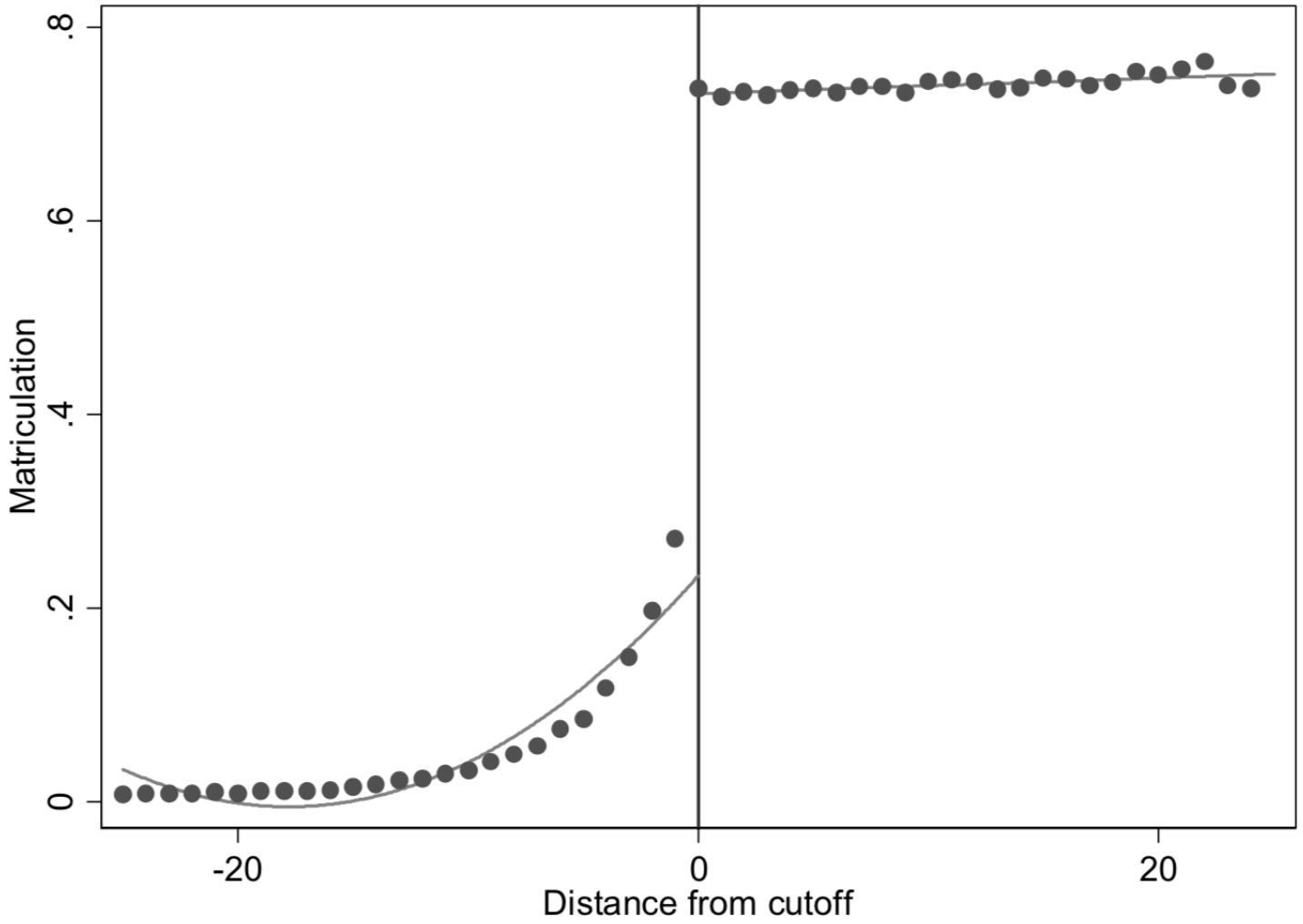
- **Question:** Do different degrees have different labor market returns?
  - Different degrees by selectivity, for example
- **Endogeneity Problem:** Cannot just compare people selective programs to those in non-selective programs
  - They were selected for a reason!
- **Ideal Experiment:** Randomize whether someone gets into a selective program or not
  - Admissions cut-offs give us something close to this

Setting

- In Chile, university admissions determined entirely by grades
- Identification strategy: look at people just below/above the cutoff
- Look at their earnings after college to see the effect of getting into the program
- Look at the RD estimates for high- versus low-selectivity degrees
  - Unlike US, students apply to a major at a university
  - Every degree program gives a cut-off to study (1,103 cut-offs!)

First-Stage Graph

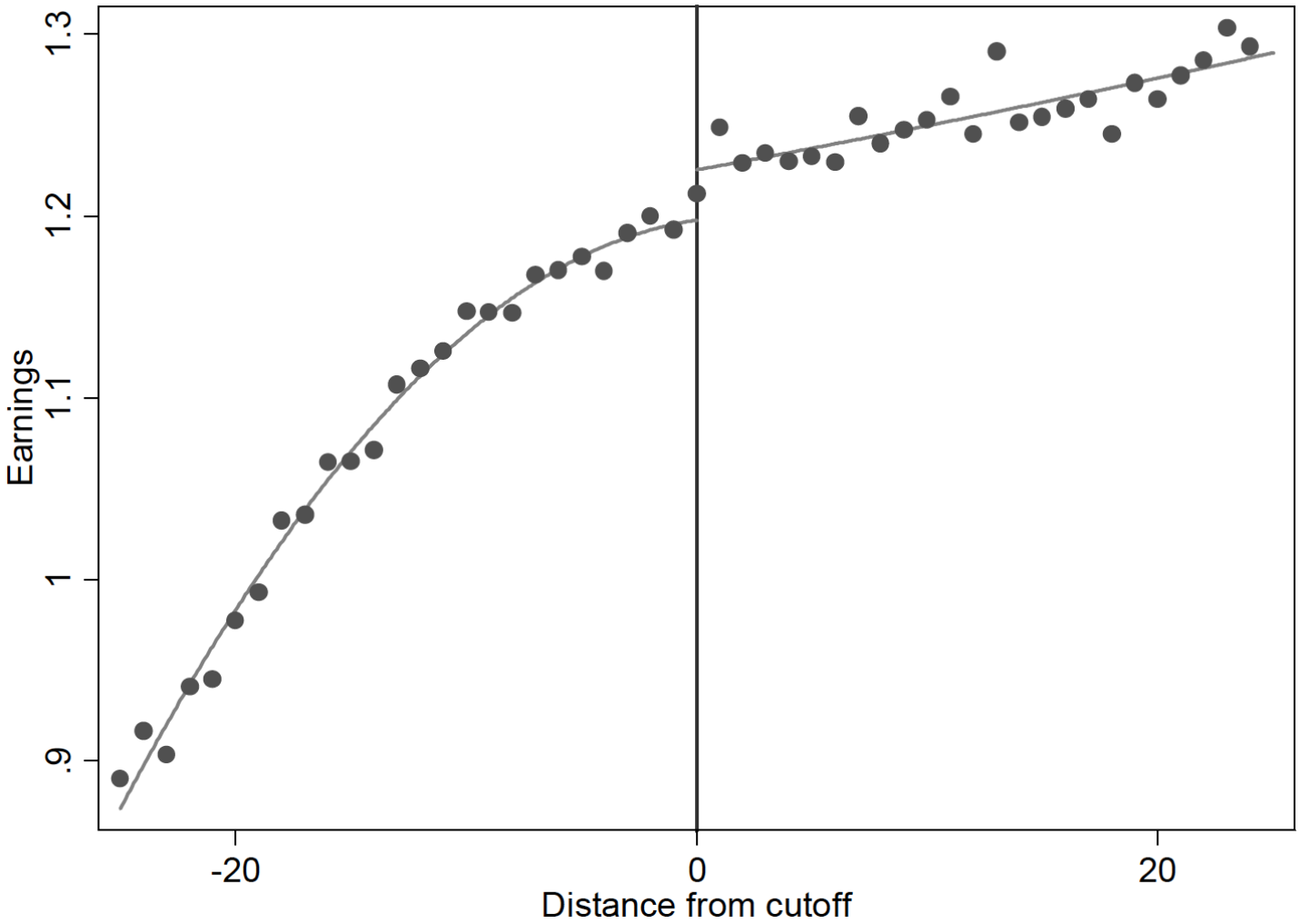
Figure 2: Program Matriculation by Admissions Score



Source: Hastings, Neilson, and Zimmerman (2013), Figure 3

RD Graph

Figure 3: Earnings by Admissions Score



Source: Hastings, Neilson, and Zimmerman (2013), Figure 4

Regression

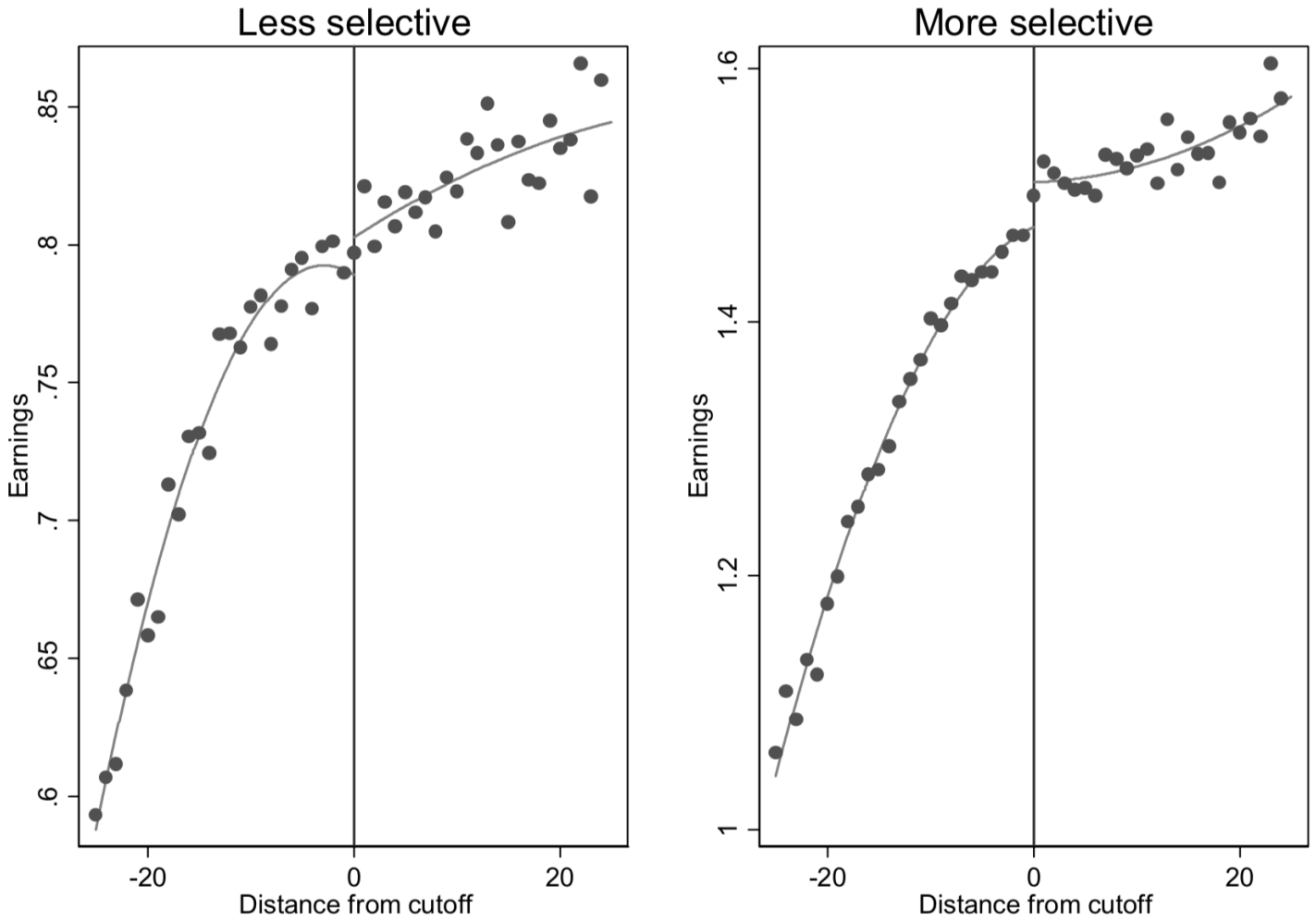
- The running variable is the admission score (X) with a cutoff of 0 (normalized). Let $D = \mathbb{1}\{X \geq 0\}$
- The regression for a person i who applies to program p:

$$Y_{ip} = \beta_0 + \beta_1 D_{ip} + \beta_2 X_i + \beta_3 (D_i \times X) + \varepsilon_{ip}$$

  where Y is log income
- Run this regression for every program, and look at people within 25 points of cutoff
- Define selectivity based on test scores of admitted students

RD Graph

Figure 4: Earnings by Admissions Score and Selectivity



Source: Hastings, Neilson, and Zimmerman (2013), Figure 5

RD Table

Table 3: RD Estimate of Threshold Crossing

	Threshold-crossing
	All years
Pooled	0.045***
	(0.008)
Bottom Quartile	0.020*
	(0.010)
2nd Quartile	0.037***
	(0.013)
3rd Quartile	0.034**
	(0.016)
Top Quartile	0.091***
	(0.021)
N	796,724

Source: Hastings, Neilson, and Zimmerman (2013), Table 4

Identification Strategies

Instrumental Variable (IV)

IV Intuition

- There are many reasons why somebody gets the treatment:
  1. They actively chose it/correlated with other factors (endogeneity issue)
  2. There is random luck in some other factor that leads to the treatment
- We often look for randomization in the treatment variable (D) itself
  - (2) tells us to look for *another* variable (Z) that is close to randomly assigned and determines D (at least partially)
- Variable Z is called an **instrumental variable** (IV)

IV Setup

- Three variables: outcome (Y), treatment (D), instrument (Z)
- First-stage regression:

$$D = \alpha_0 + \alpha_1 Z + e$$

  - When D is a dummy variable: α_1 as the increase in probability of getting the treatment from an additional unit of Z
  - Purpose of first-stage: does the instrument actually affect the treatment?

IV Setup

- Second-stage regression:

$$Y = \beta_0 + \beta_1 \hat{D} + \varepsilon$$

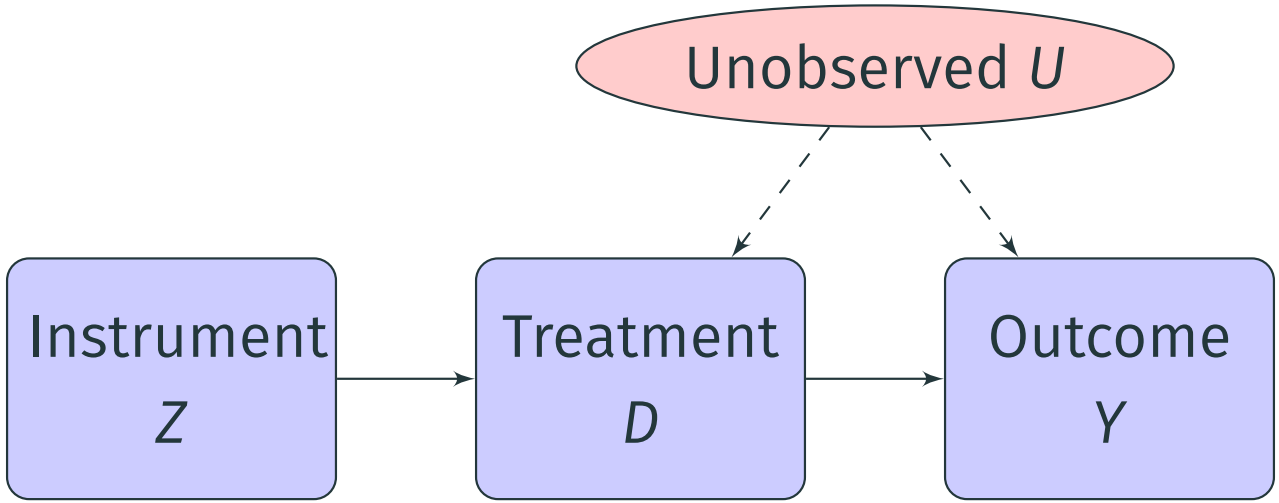
  where $\hat{D}$ is the **predicted treatment** (the variation in treatment that is explained by Z)
    - Purpose of second-stage: does the random variation in D affect the outcome Y?

- The coefficient of interest is β_1
    - Normally, variation in D is not random
    - Since $\hat{D}$ is changing only because of Z, the variation we focus on is random

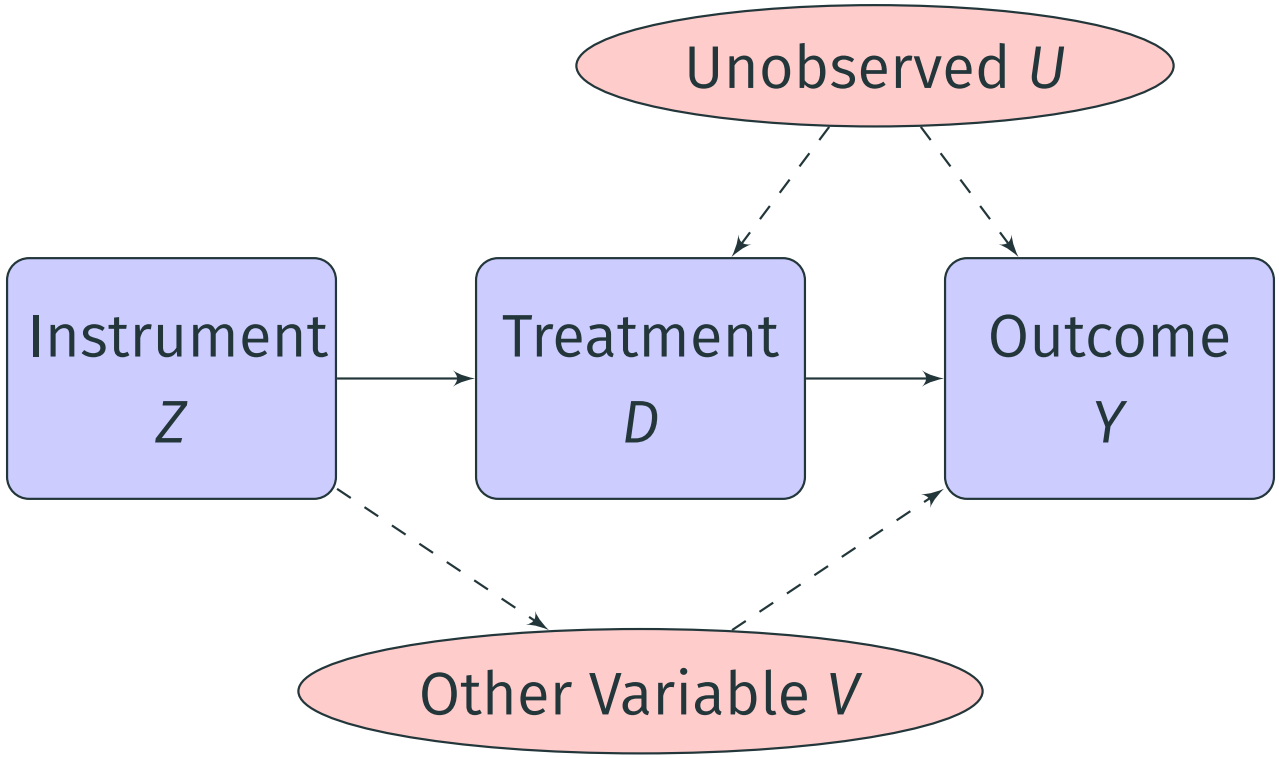
IV Requirements

- There are two key requirements to a valid instrument:
  1. **Relevance** – Z has to affect D substantially
     - This is what the **first-stage** regression tells us
  2. **Exogeneity** – Z has to affect Y only through its impact on D
     - We cannot prove this. Authors have to argue why this is true (**exclusion restriction**)

IV Diagram



IV Diagram


Unobserved U
Instrument
Z
Treatment
D
Outcome
Y
Other Variable V

IV Examples

- RD can be thought in an IV framework
  - Relevance: running variable cut-off partially determines treatment status
  - Exogeneity: which side of the cut-off you land on is somewhat random and probably doesn't determine anything else
- Angrist (1990): Vietnam War draft
  - Relevance: draft lottery number partially determined veteran status
  - Exogeneity: draft number determined by birthday and chosen randomly

Does Compulsory School Attendance Affect Schooling and Earnings?

Joshua
Angrist

Alan
Krueger

QJE, 1991

Paper Overview

- **Question:** Does more time in school lead to higher earnings?
  - In particular, consider people who drop-out of school
- **Endogeneity Problem:** People who drop-out of school are very different to those who continue
- **Ideal Experiment:** Randomize how long people stay in school
  - School attendance laws provide random variation

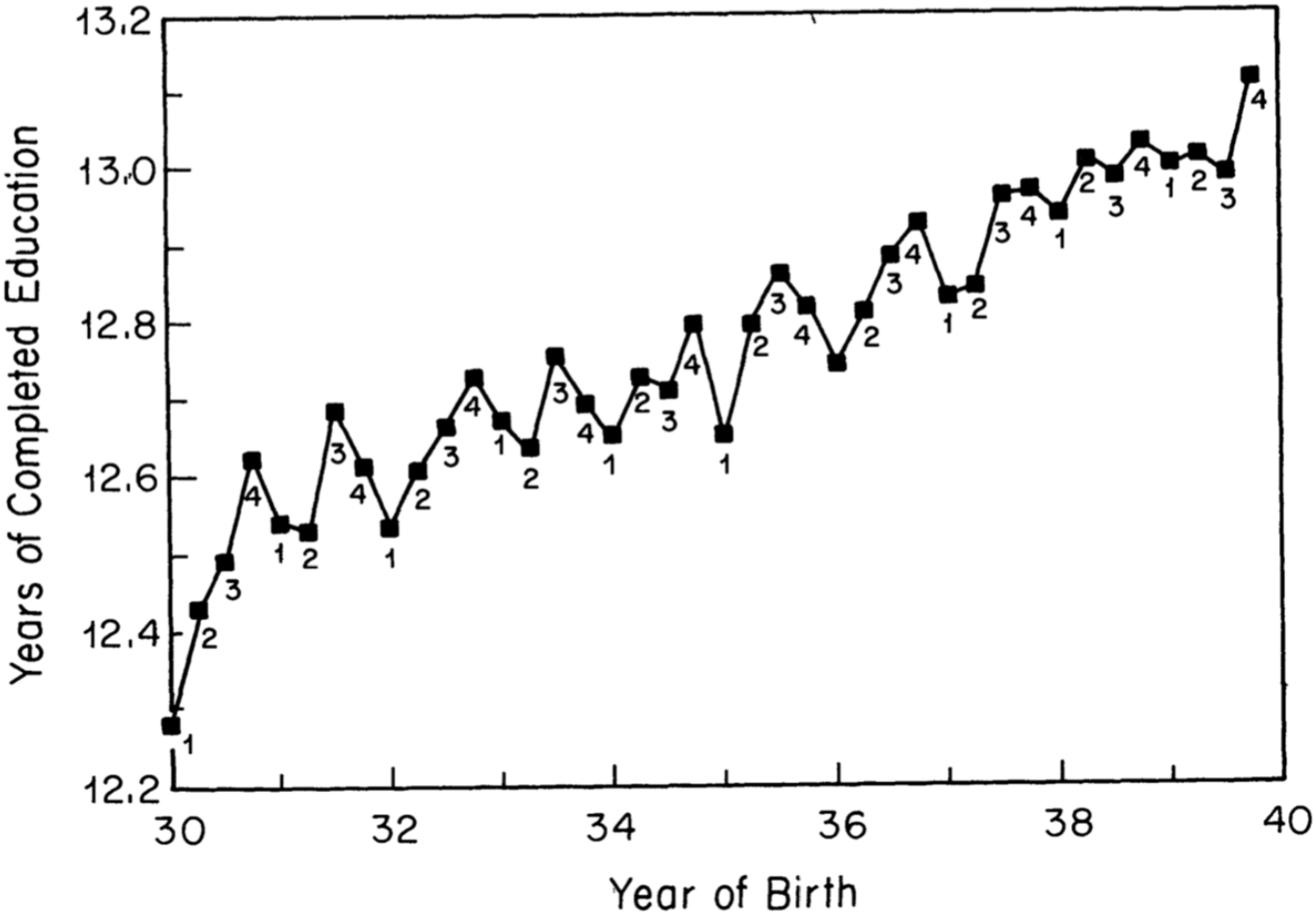
School Attendance Laws

- In the US (and many other countries), there is a compulsory school requirement
  - Stay in school until 16th or 17th birthday
- Additionally, there is also a school entry requirement
  - Must be 6 years old by Jan 1 to enter in that academic year
- For example, consider these two people:
  - Person 1: born on Dec 31, 1950 $\rightarrow$ starts school in Sept 1955 (age $5\frac{3}{4}$)
  - Person 1: born on Jan 2, 1950 $\rightarrow$ starts school in Sept 1956 (age $6\frac{3}{4}$)
- If they both drop out at 16 (in 1966), person 1 gets 10.25 years of schooling, while person 2 only gets 9.25

IV Strategy

- Part of the reason that people have different years of schooling is due to their birthday
  - School starting age varies, but school leaving is fixed. The interaction between the two generates variation in schooling years
- **Instrument:** quarter-of-birth
- **Relevance:** time of birth affects length schooling (we can test this)
- **Exogeneity:** time of birth seems plausibly independent from other individual characteristics

Table 4: Years of Education and Quarter of Birth



Source: Angrist and Krueger (1991), Figure 1

First-Stage

- Can run a first-stage regression of the form:

$$Y_i = \beta_0 + \beta_1 Q_1 + \beta_2 Q_2 + \beta_3 Q_3 + \varepsilon_i$$

  where Q_t is a dummy for being born in quarter t

- The coefficients are relative to quarter 4 (the omitted category)

Table 5: First-Stage Regression

	Quarter of birth effect		
	1	2	3
Total years of education	−.124 (.017)	−.086 (.017)	−.015 (.016)
Years of education for high school graduates	−.004 (.014)	.051 (.014)	.012 (.014)

Source: Angrist and Krueger (1991), Table 1 (CCZ Table 4.2)

Second-Stage

- Recall, we didn't want to regress earnings on education because the estimates would be biased
  - People who stay in school for longer likely have unobserved characteristics that make them have higher earnings
- Authors then run the second-stage regression (2SLS). Interesting to compare this to simple, but probably biased, regression (OLS)
  - Regress log earnings on years of education
  - Coefficient interpretation: an extra year of school increases earnings by $\beta\%$
- OLS coefficient estimate: 0.0632
- IV coefficient estimate: 0.0600

IV Generalizability

- Key question: who does the IV apply to?
- This IV looks at variation in years of schooling caused by variation in birthday
- Therefore, IV estimate only tells us the returns to education for the people whose years of schooling were actually affected by their birthday
- Doesn't tell us anything about people who stay in school no matter their quarter of birth (maybe the returns for them are higher)
- IVs can be hard to generalize (**local average treatment effect**)

IV Validity

- Is the exclusion restriction satisfied?
- Paper will try to convince you, but you need to be very skeptical here
- This is a very famous paper but the instrument is likely not exogenous
- Quarter of birth associated with health outcomes (this could explain lower earnings)
- Women giving birth in winter months tend be younger, less educated, less likely to be married (Buckles and Hungerman, 2013)
- Finding a good instrument is hard!

Identification Strategies

Difference-in-Difference (DID)

Diff-in-Diff Setup

- There are two groups of people: treatment and control
  - e.g. State *T* and State *C*
- We can track these groups over time (**panel data**)
- There is a policy change that only affects the treatment group
  - e.g. State *T* changes its law
- We can look at groups before the change (**pre-period**) and after the change (**post-period**)

Diff-in-Diff Steps

- Let's denote the outcome of interest in group i at time t as Y_{it}
  - $i = T$ or C for treatment or control
  - $t = 0$ or 1 for pre- or post-period
- The DID works as follows:
  1. Take the difference between T and C in the pre-period: $D_0 = Y_{T0} - Y_{C0}$
  2. Take the difference between T and C in the post-period: $D_1 = Y_{T1} - Y_{C1}$
  3. Take the difference of the differences: $DID = D_1 - D_0$
- This last step, under some assumptions, gives us the causal effect of the policy change

Diff-in-Diff Intuition

- Why so many differences? The DID is actually very intuitive
- Suppose state *T* changes its minimum wage policy in 2015. We want to see how this affects employment
  - Should we just compare state *T*'s employment in 2016 to employment in 2014? Can we attribute all this to the policy change?
  - No! For example, the local economy is also changing over time
  - Need to separate out changes due to the policy and changes due to trends over time

Diff-in-Diff Intuition

- The control group provides a group that is *only* affected by the time trends but not the policy change
- Any change to the control group between pre- and post must be some overall time trend (that also affects the treatment group)
- Subtracting the change in the control group leaves only the change due to the policy

$$DID = \underbrace{(Y_{T1} - Y_{C1})}_{\substack{\text{Post-Period Diff} \\ (D_1)}} - \underbrace{(Y_{T0} - Y_{C0})}_{\substack{\text{Pre-Period Diff} \\ (D_0)}} = \underbrace{(Y_{T1} - Y_{T0})}_{\substack{\text{Change in } T \\ \text{(Policy + Trend)}}} - \underbrace{(Y_{C1} - Y_{C0})}_{\substack{\text{Change in } C \\ \text{(Trend)}}}$$

Diff-in-Diff Requirements

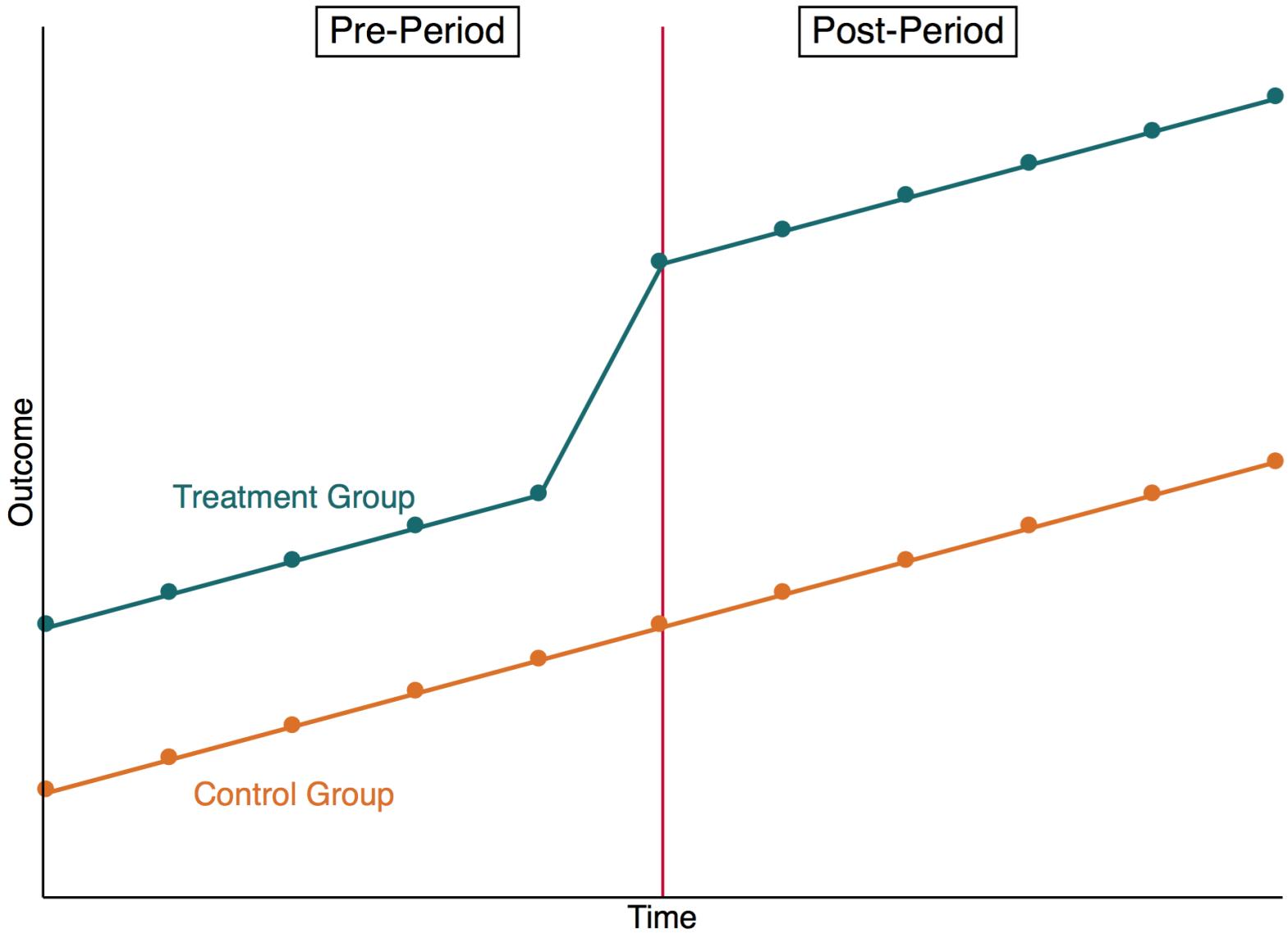
- The beauty of a DID is that the treated and control groups do not have to be the same *before* the policy change
- The two big requirements are:
  1. The policy change does not affect the control group (**no spillovers**)
  2. The treatment and control groups had similar trends before the change (**parallel trends**)
- You want the control group to be similar enough to the satisfy parallel trends, but not so similar that it is actually affected by the treatment

Parallel Trends

- The key requirement to focus on is parallel trends
- The difference between *T* and *C* should be constant over time in the pre-period
- Allows us to extrapolate into the post-period and generate a counterfactual
- Logic: if the difference was constant in the pre-period, then it would have been constant in the post-period too, if the treatment hadn't happened

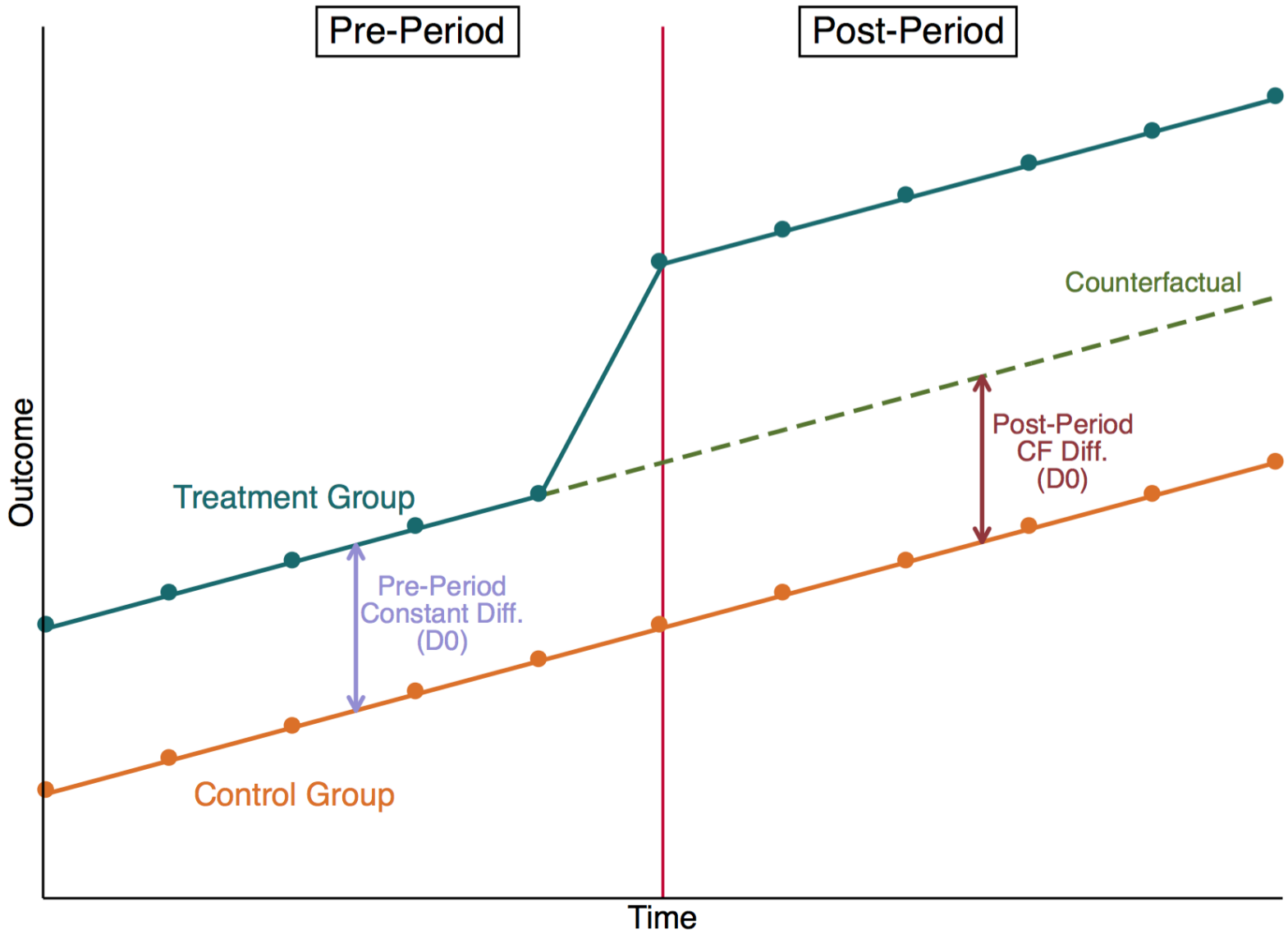
DID Graph

Figure 5: DID Setup



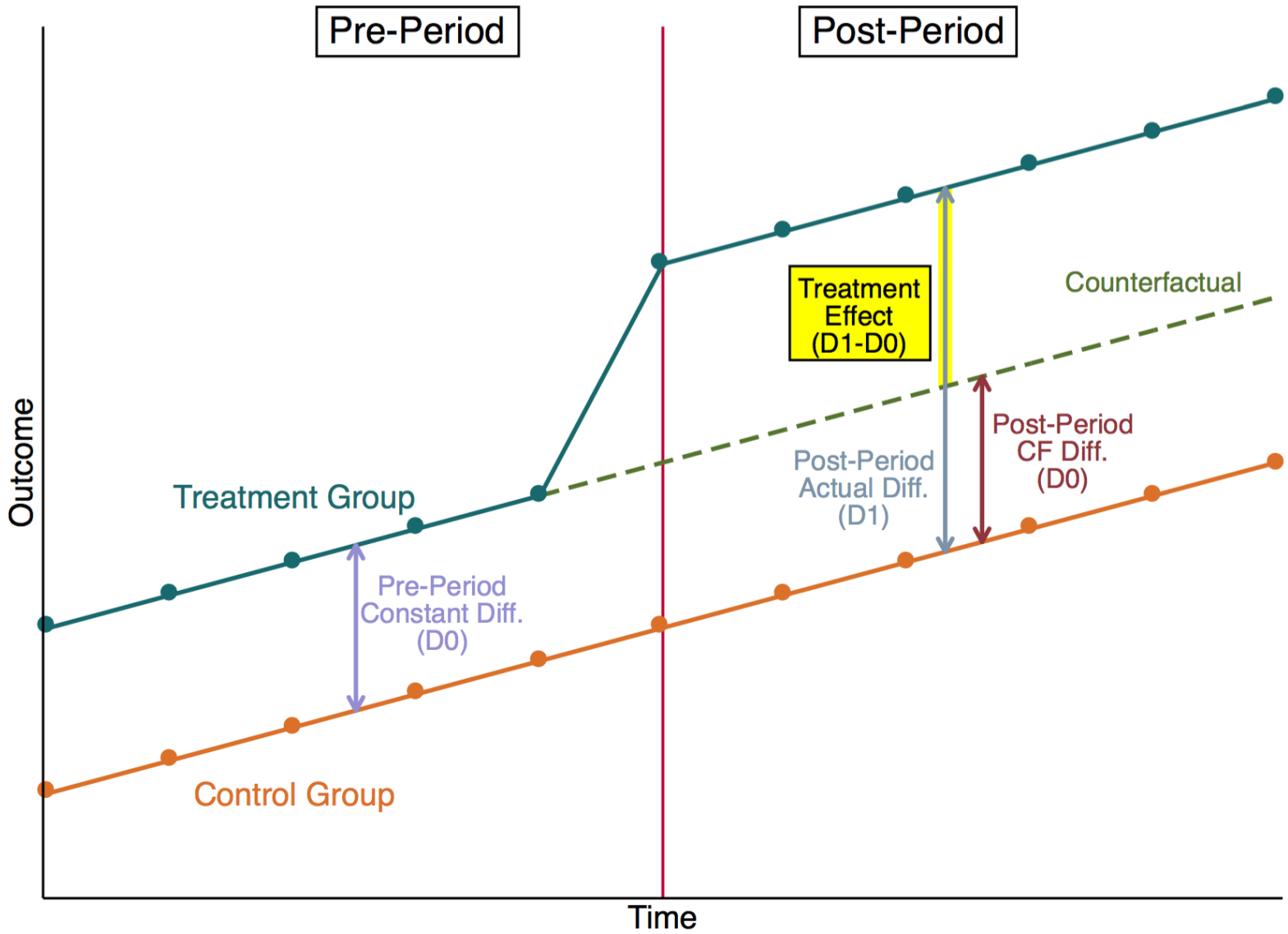
DID Graph

Figure 6: DID Counterfactual



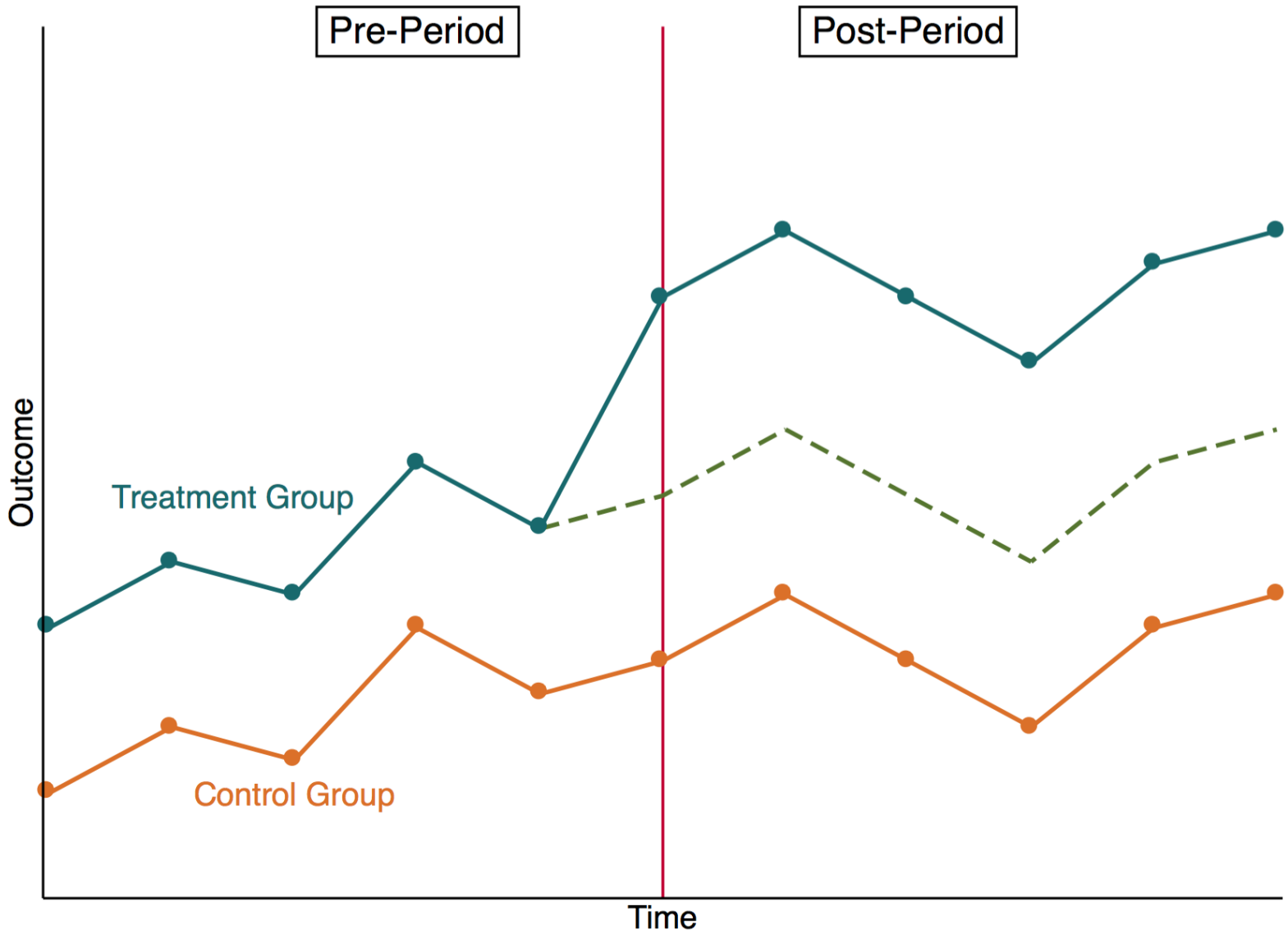
DID Graph

Figure 7: DID Explanation



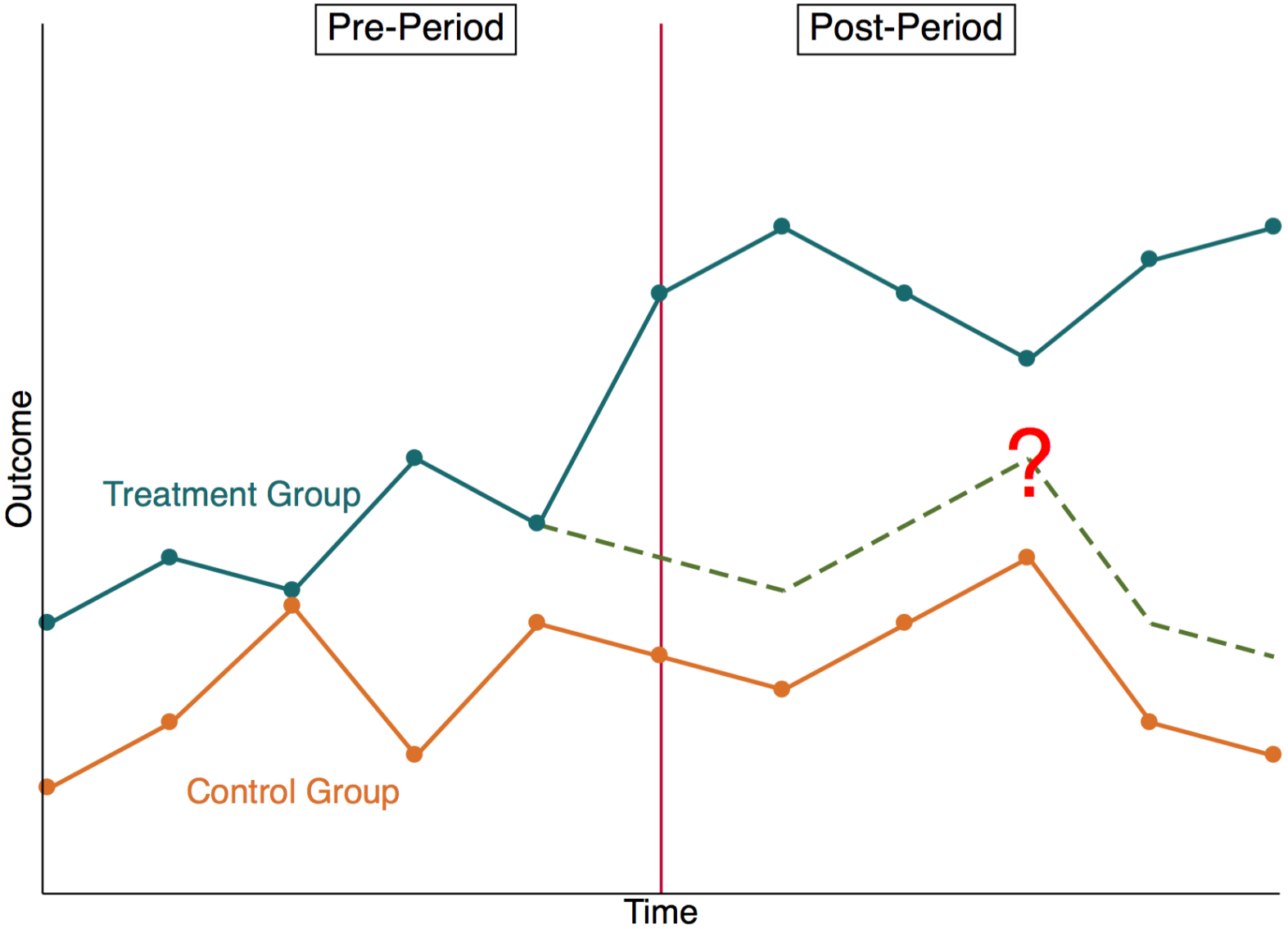
DID Graph

Figure 8: DID Non-Constant Slopes



DID Graph

Figure 9: Violating Parallel Trends



DID Regression

- To estimate a DID, use the following regression for group i in time t:

$$Y_{it} = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Post}_t + \beta_3 \text{Treat}_i \times \text{Post}_t + \beta_4 X_{it} + \varepsilon_{it}$$

  - where Treat_i is a dummy for the treatment group
  - and Post_t is a dummy for being in the post period

DID Regression

$$Y_{it} = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Post}_t + \beta_3 \text{Treat}_i \times \text{Post}_t + \beta_4 X_{it} + \varepsilon_{it}$$

- **Pre-Period Diff:** $D_0 = (\beta_0 + \beta_1) - (\beta_0) = \beta_1$
  - Control, Pre-Period: β_0
  - Treatment, Pre-Period: $\beta_0 + \beta_1$
- **Post-Period Diff:**
  $D_1 = (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_2) = \beta_1 + \beta_3$
  - Control, Post-Period: $\beta_0 + \beta_2$
  - Treatment, Post-Period: $\beta_0 + \beta_1 + \beta_2 + \beta_3$
- **Diff-in-Diff:** $DID = D_1 - D_0 = \beta_3$
  - The coefficient on $\text{Treat}_i \times \text{Post}_t$ gives us the causal effect of the policy

Schooling and Labor Market Consequences of School Construction in Indonesia

Evidence from an Unusual Policy Experiment

Esther

Duflo

AER, 2001

Paper Overview

- **Question:** Does opening more schools lead to higher educational attainment and higher earnings?
  - Key question in developing countries
- **Endogeneity Problem:** Locations of schools are not random
- **Ideal Experiment:** Randomize whether places get a school or not
  - Don’t have this, but a program in Indonesia can be used to setup a DID

INPRES Program

- From 1973-1979, Indonesia built over 61,000 primary schools (average of 1 school per 500 students)
- Program targeted areas that had low school enrollment
  - Children in high enrollment areas should not be affected by the program
  - Children in low enrollment areas should be affected by the program
- Children attend primary school from age 7 to 12
  - Children over age 12 in 1974 should not be affected by the program
  - Children between ages 2-6 in 1974 should be most affected by the program

Regression

- Regression we run is:

$$Y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Post}_i + \beta_3 \text{Treat}_i \times \text{Post}_i + \varepsilon_i$$

- Where:
  - Treat_i = 1 if person *i* born in a high program (low-enrollment) region or not
  - Post_i = 1 if person *i* aged between 2-6 in 1974, 0 if 12-17 in 1974

DID Table

Table 6: DID School Construction

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)

Source: Duflo (2001), Table 3

	$Treat_i = 1$	$Treat_i = 0$	Diff.
$Post_i = 1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_2$	$\beta_1 + \beta_3$
$Post_i = 0$	$\beta_0 + \beta_1$	β_0	β_1
Diff.	$\beta_2 + \beta_3$	β_2	β_3

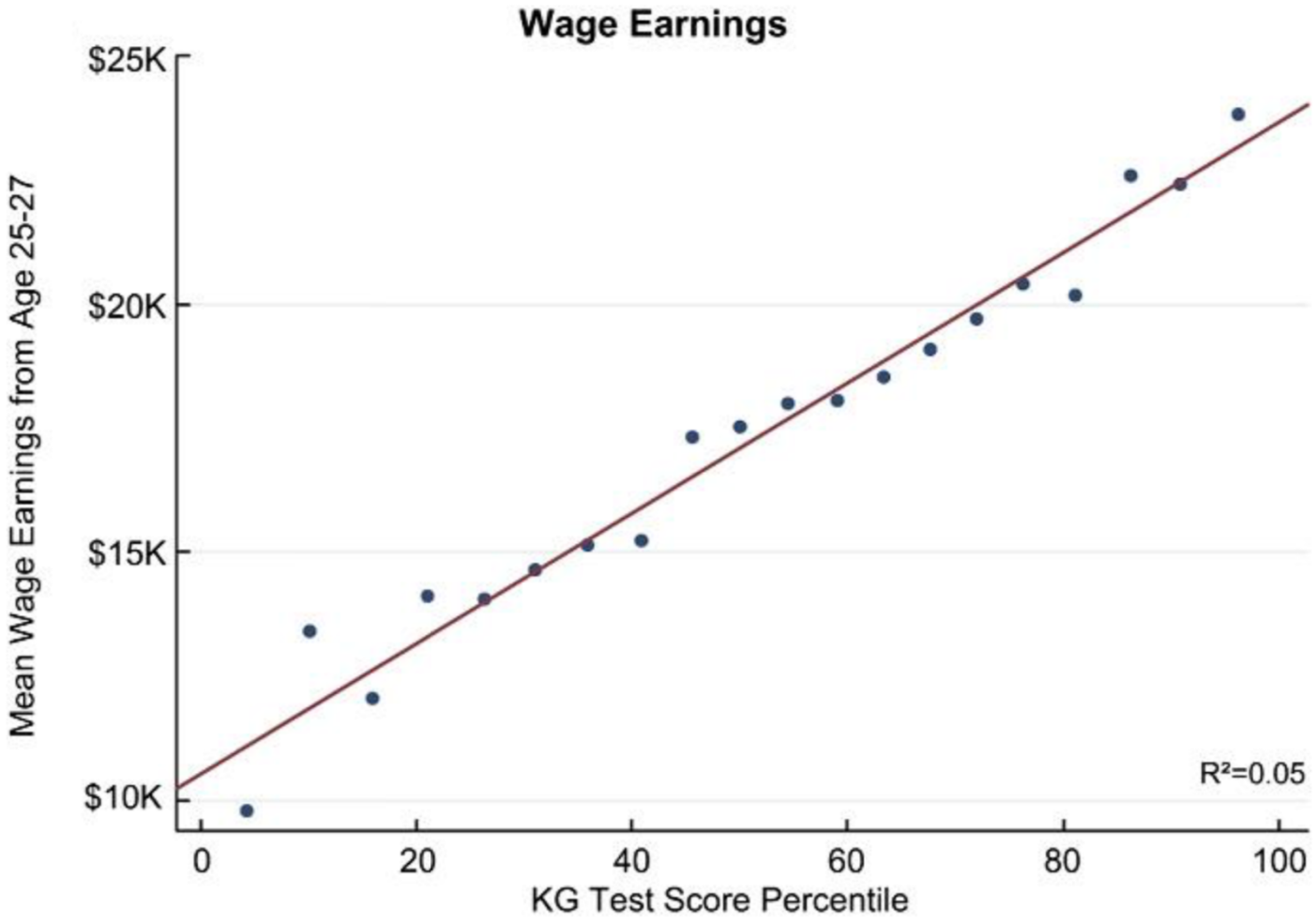
Next Steps

- Now we know the techniques for causal inference
- Papers may not fit exactly into these frameworks
- Key ideas to keep in mind:
  - What is the treatment and control group?
  - Does the experiment allow us to see a reasonable counterfactual?
  - Compare to the ideal RCT you would want to run. How close do we get?

Appendix

KG Test Scores and Wages

Figure A1: Correlation between Kindergarten Test Scores and Wages



Source: Chetty et al. (2011), Figure 1(a)

Back

References

References i

Angrist, J. D. (1990). “Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records”. In: *American Economic Review* 80.3, pp. 313–336.

Angrist, J. D. and A. B. Krueger (1991). “Does compulsory school attendance affect schooling and earnings?”. In: *The Quarterly Journal of Economics* 106.4, pp. 979–1014.

Buckles, K. S. and D. M. Hungerman (2013). “Season of birth and later outcomes: Old questions, new answers”. In: *Review of Economics and Statistics* 95.3, pp. 711–724.

Chetty, R. et al. (2011). “How does your kindergarten classroom affect your earnings? Evidence from Project STAR”. In: *The Quarterly Journal of Economics* 126.4, pp. 1593–1660.

References ii

Duflo, E. (2001). “Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment”. In: *American Economic Review* 91.4, pp. 795–813.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2013). “Are some degrees worth more than others? Evidence from college admission cutoffs in Chile”. In: *NBER Working Paper*.