w241: Experiments and Causality

Causal Inference from Observational Data

David Reiley, David Broockman, D. Alex Hughes
UC Berkeley, School of Information
Updated: 2021-07-02

Introduction to Causal Inference from Observational Data

Natural Experiments

Instrumental Variables

A Few Clarifications

- What does "reduced form" mean?
- We previously talked about regressing treatment X on instrument Z, computing the fitted values of X, then regressing outcome Y on these fitted values $\hat{X}$
  - The reduced-form equation **"reduces" these two equations to a single equation**, the regression of Y on Z
- The initials in the example's names correspond to:

Alvaro <--> **A**lways-takers

Camila <--> **C**ompliers

Normando <--> **N**ever-takers

- The author's Passover analogy is not entirely clear but may be as follows for four types of children:

Wise <--> Compliers

Wicked <--> Defiers

Simple <--> Always-takers

Doesn't know how to ask <--> Never-takers

- "External validity" is a common term in academic social science, but I prefer the term "generalizability"

Example: Vietnam Draft Lottery

and Returns to Schooling

- IV can be used for natural experiments where the treatment is a continuous variable rather than binary
- Angrist and Krueger (1992) used the Vietnam draft lottery as a natural experiment to assess returns to schooling on earnings
    - During the Vietnam War, a fraction of young men from every birth year were prioritized for draft via a lottery system based on birth date
    - Men with lower draft numbers became more likely to enroll in college because an education deferment would allow them to avoid service
    - Thus, the lottery created a random variation in the number of years of schooling, which created an opportunity to estimate returns to schooling

Example: Vietnam Draft Lottery (cont'd)

- We start the IV regression with a person's years of schooling X regressed on instrument Z, the order his birth date was pulled in the lottery
    - Then we regress earnings Y on the predicted values of years of schooling X
    - This tells us how the "clean" variation in number of years of schooling affects earnings
- This IV regression is much better than simply regressing wages on years of schooling, which is fraught with OVB problems

Potential Problem:

Violation of the Exclusion Restriction

- The required assumption is that the random process cannot affect the outcome in any way other than the treatment being examined
  - This assumption is no problem in many examples such as the KIPP lottery
- However, serving in Vietnam may affect earnings in ways unrelated to schooling
  - Eg. PTSD may reduce earnings, which would underestimate returns to schooling
- In the next example, David Broockman will describe a natural experiment where the exclusion restriction might be violated

Two-State Least Squares

Optional Example:

Does Economic Stagnation Cause War?

Does Economic Stagnation Cause War?

- Can foreign aid prevent civil war?
- Need randomly assigned economic growth
  - Rain tends to be associated with economic growth in agricultural areas

Rain -> Economic Growth -> Less Civil War?

- Compare countries with high rainfall to low
- "If there's a correlation between rain and civil war, it could only be caused by lower economic growth"
- Potential policy recommendation: If war looks likely, send money
- But is economic growth the only way rain could affect civil war?
  - "No one likes to fight a war when it's muddy"

Rain -> ~~Economic Growth~~ mud -> Less Civil War

Advantages and Disadvantages of Natural Experiments

Advantages of Natural Experiments

- If you are lucky enough to find one, you save the time of running the experiment
- Estimation is straightforward
  - IV/2SLS is very similar to OLS
  - Estimation packages like "ivreg" in R give correct standard errors

Disadvantages of Natural Experiments

- Administration of randomizing can be difficult
  - People select, randomization fails, etc.
  - In an observational natural experiment, you have no control over the design, data collection, execution, block, or attrition
- You might not find a natural experiment on the topic you care about
  - Good natural experiments can be as rare as unicorns
  - We often have to run the experiments ourselves

Reading Assignment

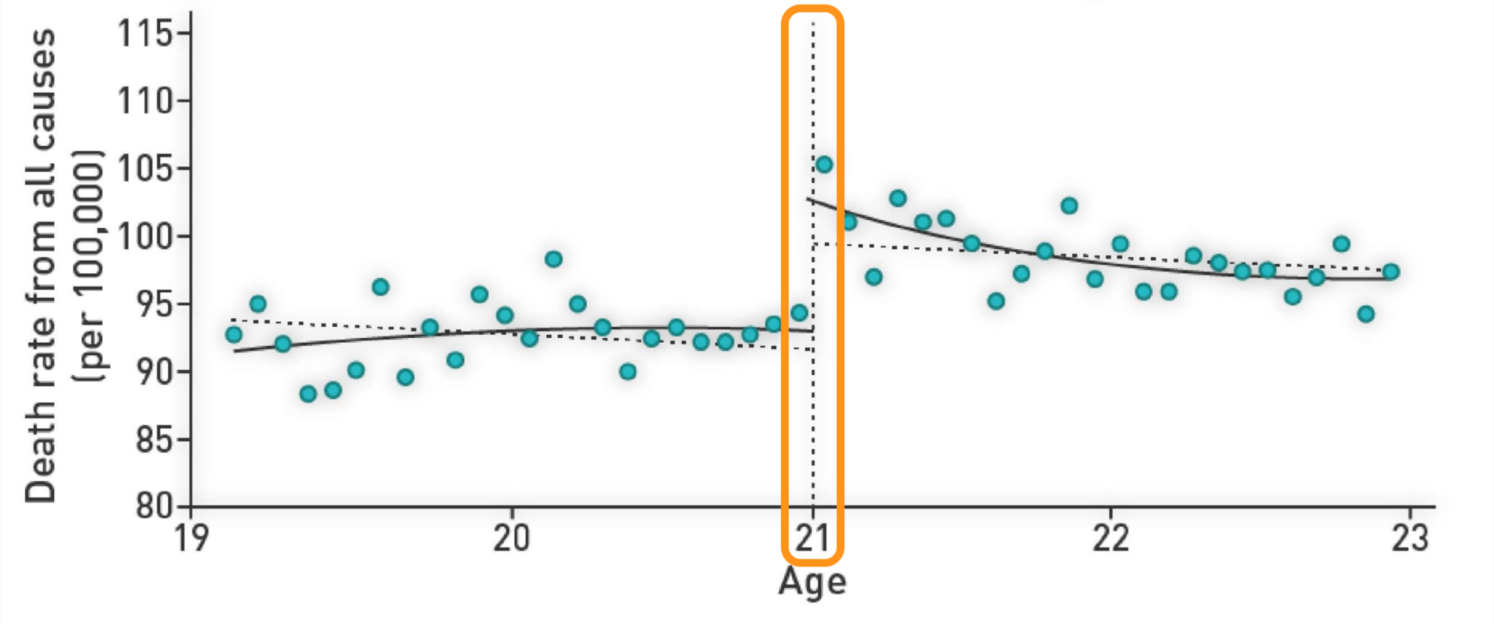
Read *Mastering Metrics*, Section 4.1, pages 147-153

- Measuring the mortality effects of drinking alcohol
- Pay particular attention to figure 4.2: How convincing is this as a demonstration of causal effect?

Introduction to Regression Discontinuity

Regression Discontinuity

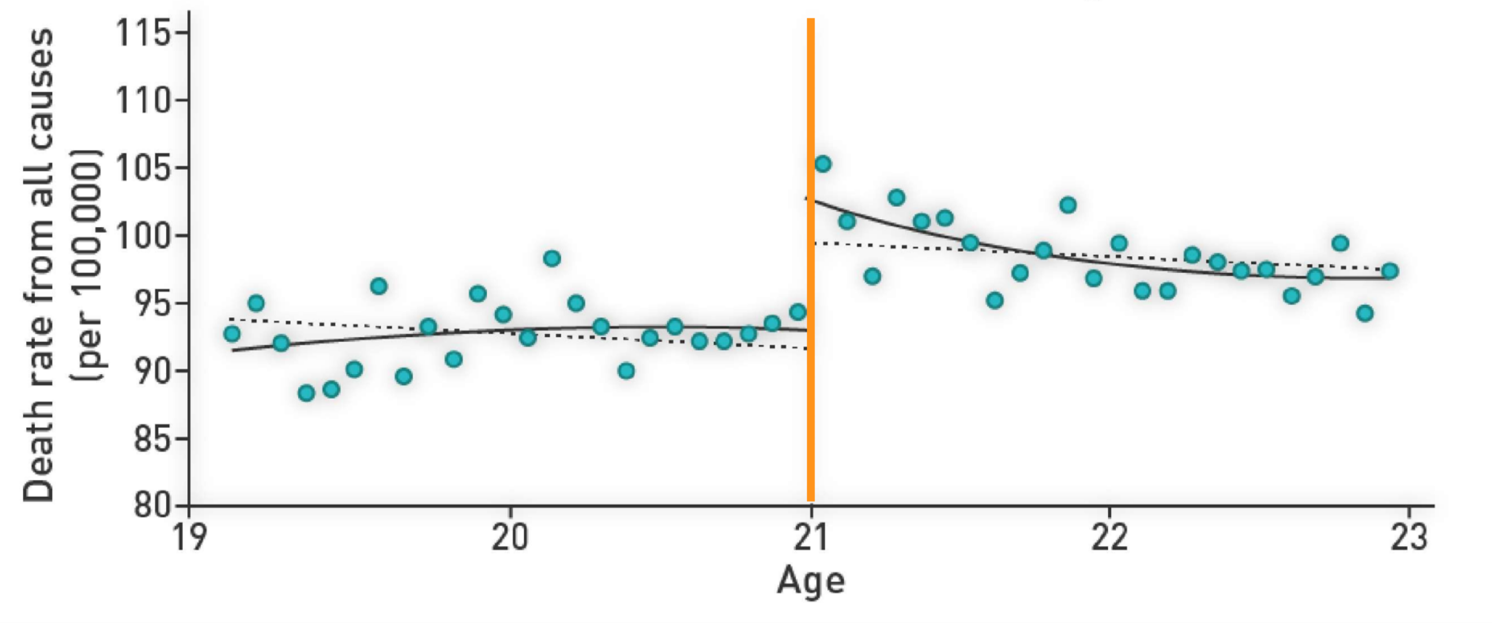
Figure 4.2



- There is a non-random "running variable": age
- There is a "threshold": the minimum legal drinking age of 21

Regression Discontinuity

Figure 4.2 (cont'd)



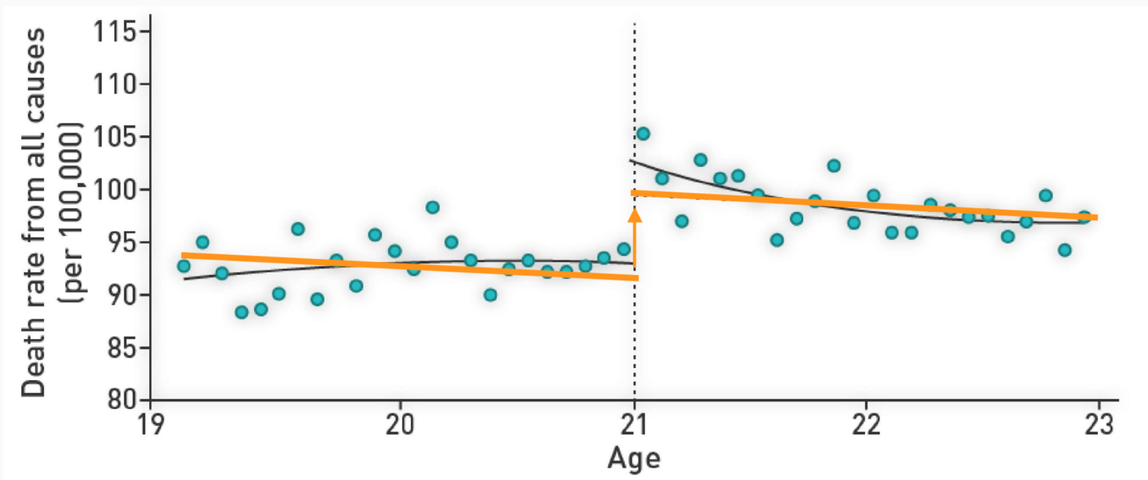
- It seems reasonable to believe that other reasons for mortality are exactly the same for someone who is 20 years, 364 days old, and someone who is exactly 21 years old
  - If we compare people on either side of the discontinuity, the difference should be due only to the access of alcohol

Regression Discontinuity

Equation 4.2

$$ \bar{M}_{a} = \alpha + \rho D_{a} + \gamma a + e_{a} $$

- We regress the outcome (mortality rate) on"
    - A treatment dummy variable (are you over 21?)
    - A linear term for age, in years. This allows for a time trend in mortality



- Figure 4.2 shows the resulting regression line: downward-sloping, except for a discrete jump on the 21st birthday

Regression Discontinuity

Equation 4.2 (cont'd)

- For a "fancy" analysis, we can:
    - Use a quadratic regression (rather than linear) **OR**
    - Allow the regression to change slope right at the discontinuity
- But these are niceties
- We care most about the discrete jump in Y at the age discontinuity

Reading Assignment

Read *Mastering Metrics* page 153 to end of Section 4.1 on page 164

- See if you can recognize the placebo tests

Example:

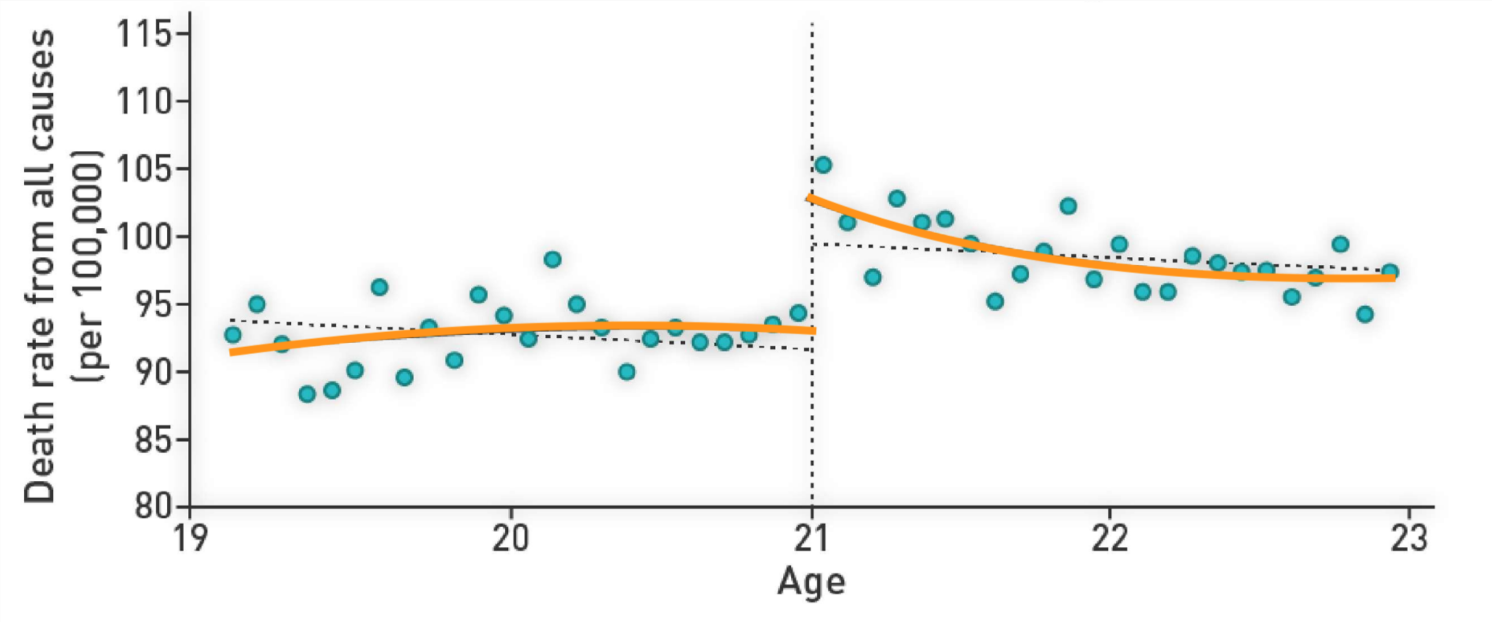
Detailed Analysis of the Effects of the Drinking Age

Equation 4.4

$$\bar{M}_a = \alpha + \rho D_a + \gamma_1 (a - a_0)^2 + \delta_1 [(a - a_0) D_a] + \delta_2 [(a - a_0)^2] D_a + e_a$$

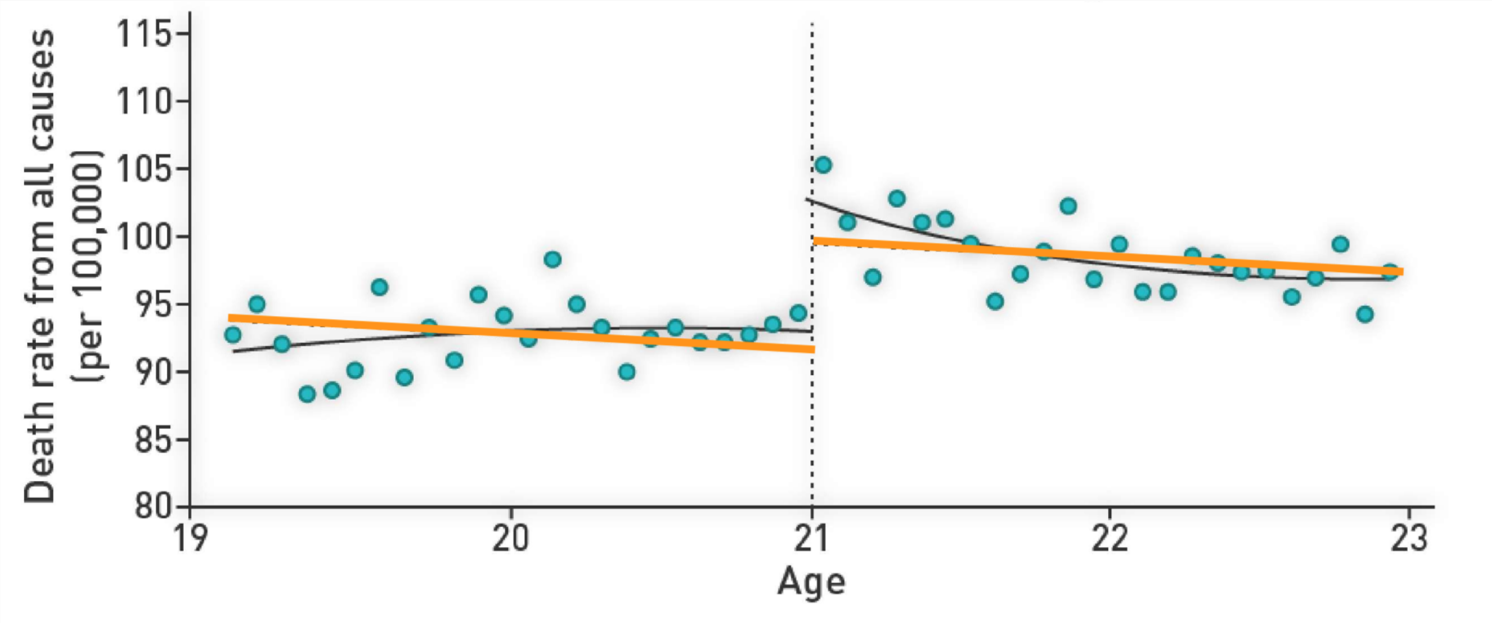
- This "fancy" regression allows the relationship between mortality and age to have:
  - A quadratic shape to the left of the 21st birthday
  - A different quadratic shape to the right

Equation 4.4 Results: Figure 4.4



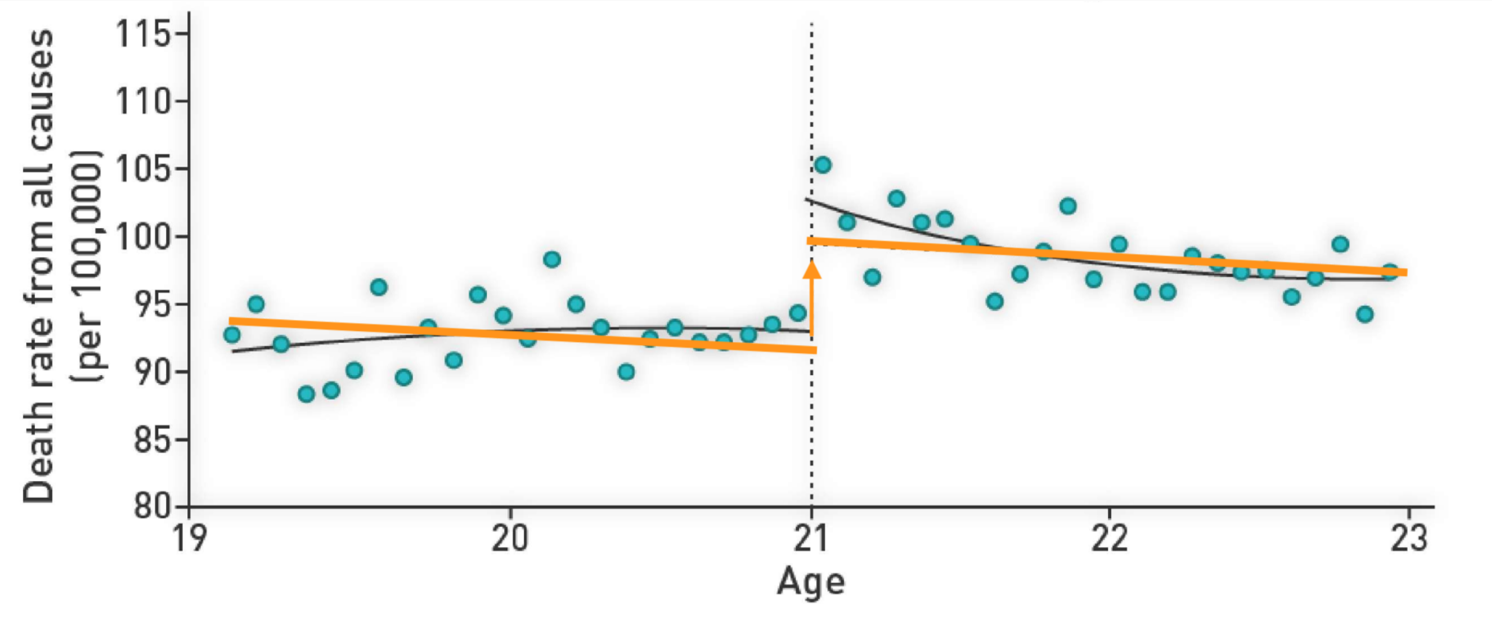
- The solid curved line shows the fancy (quadratic) regression curve

Equation 4.4 Results: Figure 4.4



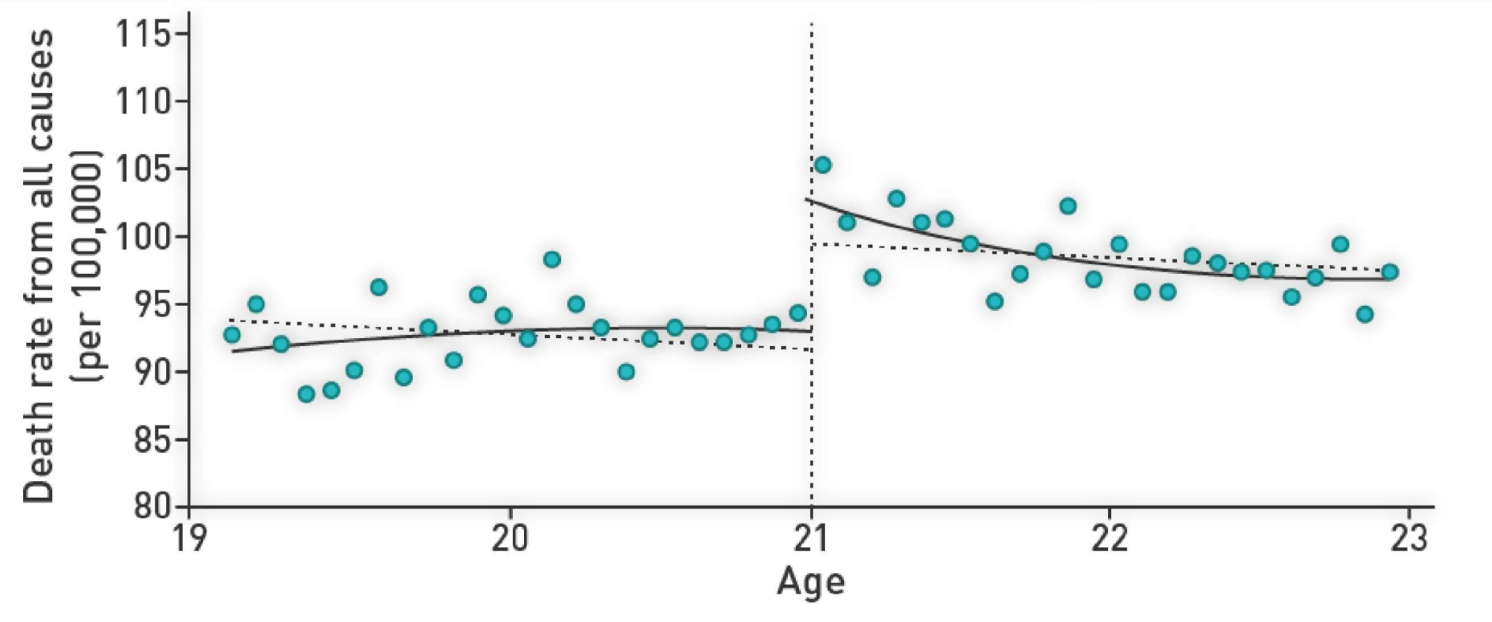
- The solid curved line shows the fancy (quadratic) regression curve
- The dashed straight line shows the simple regression curve
    - The slope on the simple curve must be the same on either side of the discontinuity
    - The fancy curve has a different slope and curvature on each side

Equation 4.4 Results: Figure 4.4



- The solid curved line shows the fancy (quadratic) regression curve
- The dashed straight line shows the simple regression curve
    - The slope on the simple curve must be the same on either side of the discontinuity
    - The fancy curve has a different slope and curvature on each side
    - This particularly captures the jump in mortality rates right after the 21st birthday
    - The curve declines over time, but the mortality rate at age 23 is still a little higher than at 20

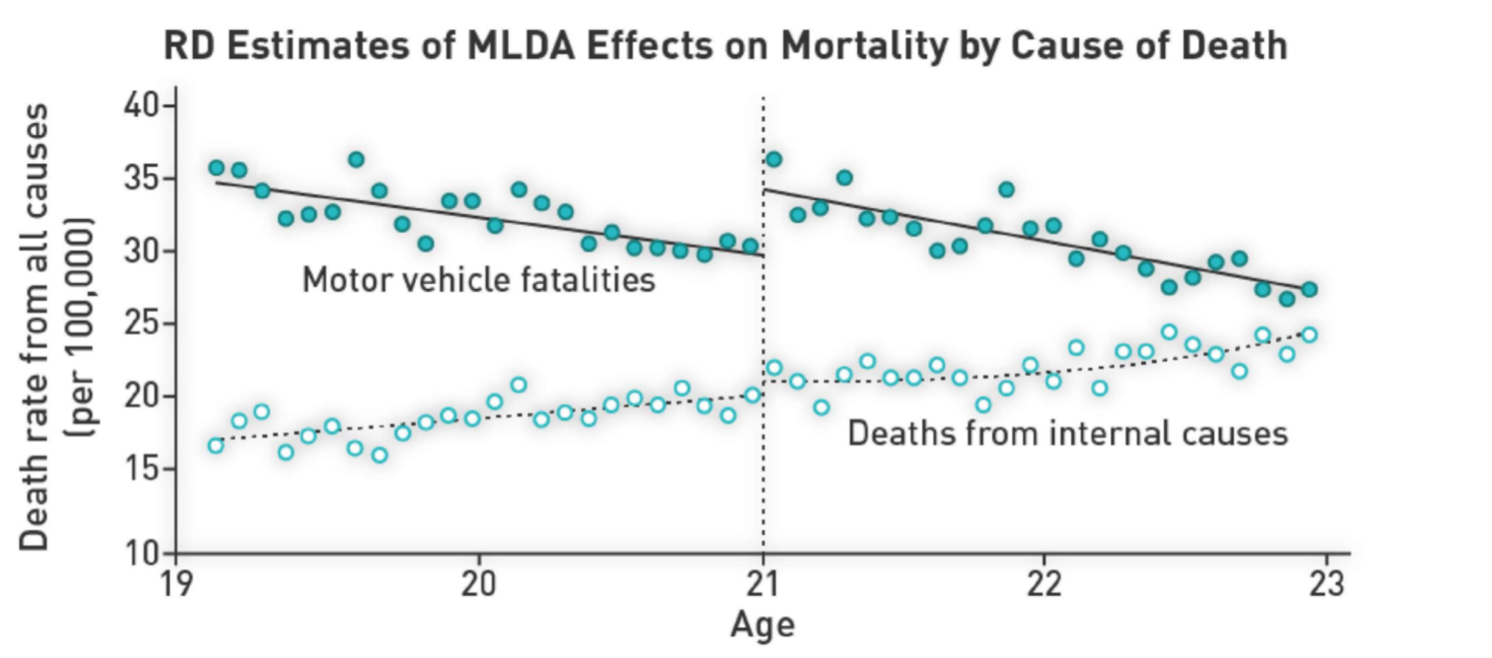
Equation 4.4 Results: Figure 4.4 (cont'd)



- We don't want to extrapolate too far--the power of an RD is right at the break point
- Since the trends are gradual, some extrapolation may be okay
  - Legal access to alcohol appears to produce higher mortality rates, especially at first
  - Continuing for a couple of years, mortality rates return to baseline

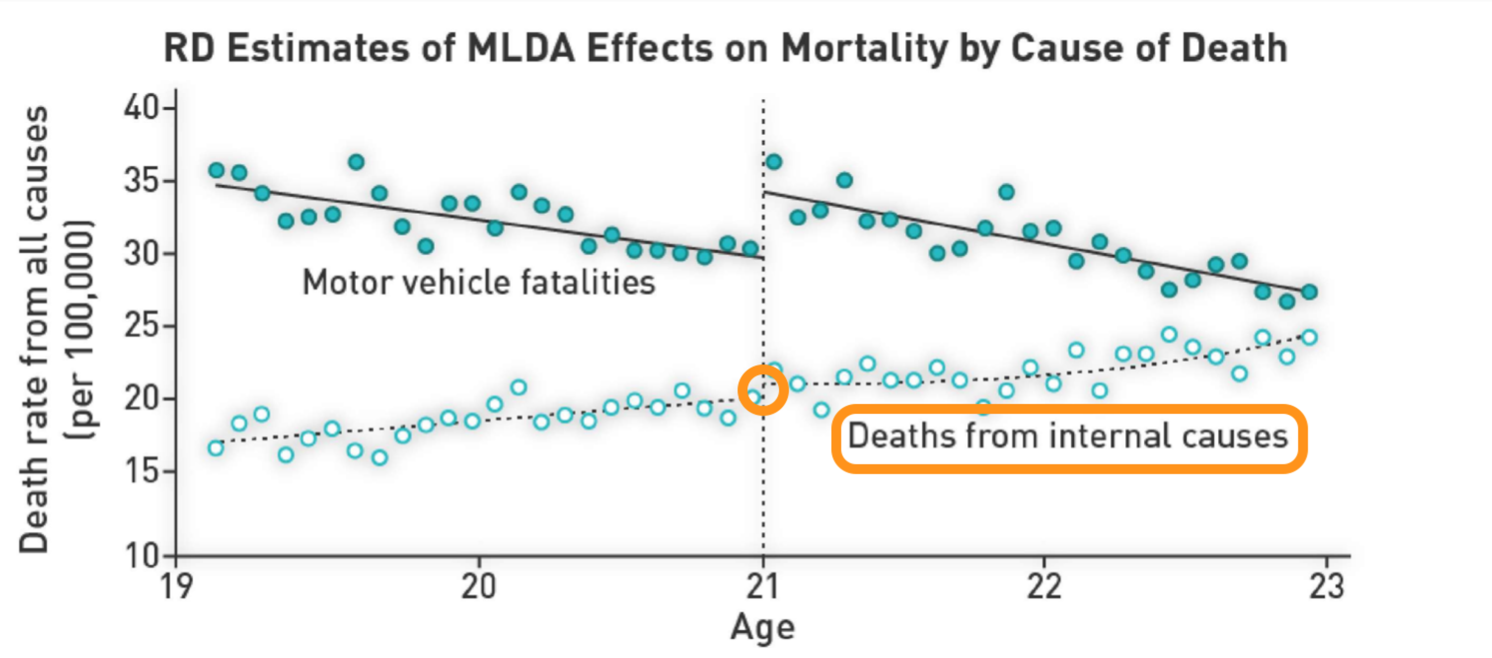
Placebo Tests: Figure 4.5

Table 4.1 and Figure 4.5 show placebo tests



Placebo Tests: Figure 4.5

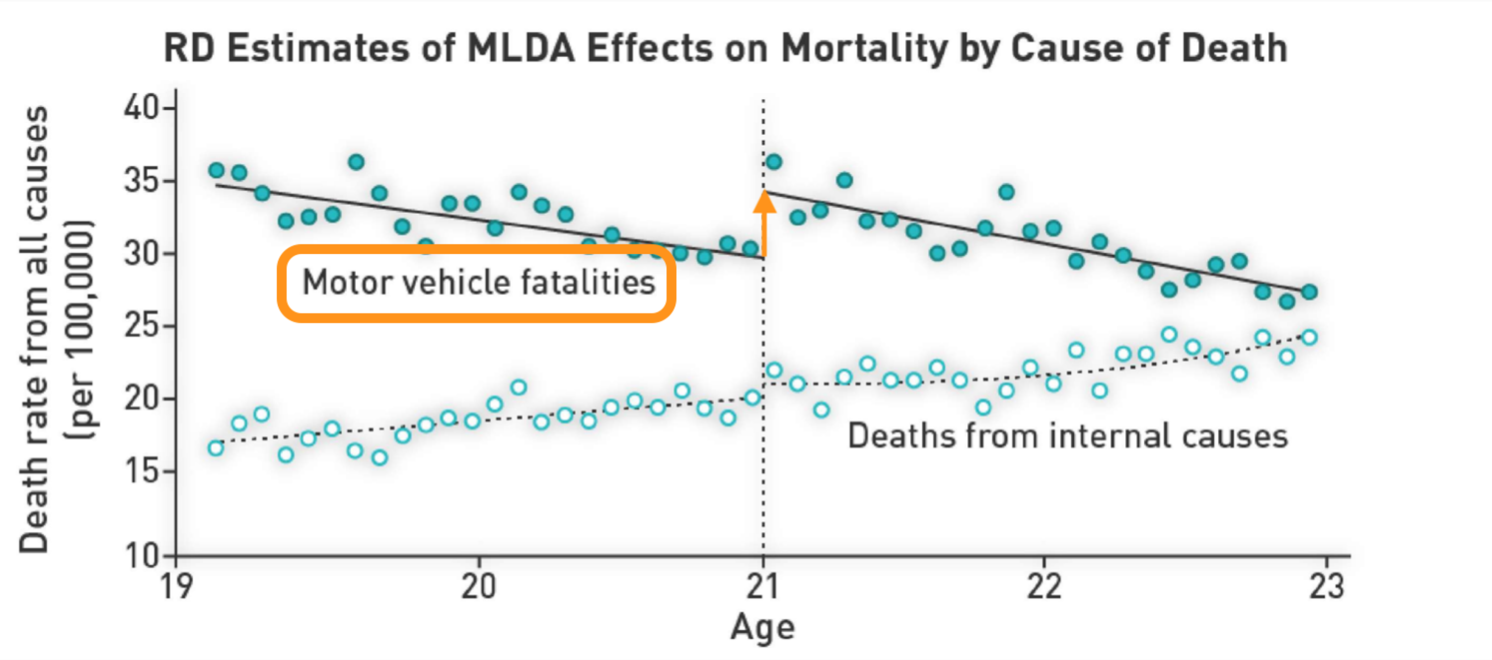
Table 4.1 and Figure 4.5 show placebo tests



- Deaths from disease ("internal causes") show no serious jumps at the discontinuity

Placebo Tests: Figure 4.5

Table 4.1 and Figure 4.5 show placebo tests



- Deaths from disease ("internal causes") show no serious jumps at the discontinuity
- Motor vehicle fatalities show a substantial jump
- This makes us even more confident that the jump we saw in the initial RD is really a causal effect of legal access to alcohol
    - Deaths associated with alcohol show a big jump

Parametric vs. Nonparametric RD

- **Parametric RD:** We estimate regression lines or curves and have a shift due to the treatment dummy variable
- **Nonparametric RD:** we avoid assuming some parametric form, in case that might introduce specification error
  - Instead, we look at the simple change in Y in a narrow window around the threshold
- **Bandwidth** is the width of this window

Parametric vs. Nonparametric RD

Table 4.1

Sharp RD Estimates of MLDA Effects on Mortality

Dependent variable	Ages 19–22		Ages 20–21	
	(1)	(2)	(3)	(4)
All deaths	7.66	9.55	9.75	9.61
	(1.51)	(1.83)	(2.06)	(2.29)
Motor vehicle accidents	4.53	4.66	4.76	5.89
	(.72)	(1.09)	(1.08)	(1.33)
Suicide	1.79	1.81	1.72	1.30
	(.50)	(.78)	(.73)	(1.14)
Homicide	.10	.20	.16	-.45
	(.45)	(.50)	(.59)	(.93)
Other external causes	.84	1.80	1.41	1.63
	(.42)	(.56)	(.59)	(.75)
All internal causes	.39	1.07	1.69	1.25
	(.54)	(.80)	(.74)	(1.01)
Alcohol-related causes	.44	.80	.74	1.03
	(.21)	(.32)	(.33)	(.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

- In Table 4.1, the results are not very sensitive to the choice of window
- We see similar estimates when we compare Column 1 (window ages 19-22) to Column 3 (window ages 20-21)

Parametric vs. Nonparametric RD

Table 4.1 (cont'd)

Sharp RD Estimates of MLDA Effects on Mortality

Dependent variable	Ages 19–22		Ages 20–21	
	(1)	(2)	(3)	(4)
All deaths	7.66 (1.51)	9.55 (1.83)	9.75 (2.06)	9.61 (2.29)
Motor vehicle accidents	4.53 (.72)	4.66 (1.09)	4.76 (1.08)	5.89 (1.33)
Suicide	1.79 (.50)	1.81 (.78)	1.72 (.73)	1.30 (1.14)
Homicide	.10 (.45)	.20 (.50)	.16 (.59)	-.45 (.93)
Other external causes	.84 (.42)	1.80 (.56)	1.41 (.59)	1.63 (.75)
All internal causes	.39 (.54)	1.07 (.80)	1.69 (.74)	1.25 (1.01)
Alcohol-related causes	.44 (.21)	.80 (.32)	.74 (.33)	1.03 (.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

- Trade-off of narrow bandwidth:
  - Gives a much cleaner RD. We think all other things are equal on both sides of the threshold **BUT**
  - Cuts down on amount of data, which decreases precision. Notice how Column 3 has larger standard errors than Column 1

Policy Question

- Chapter 4 opened with the policy question **What should the drinking age be?**
- With RD, we saw that legal access to alcohol has a big causal effect on mortality for 21-year-olds
- Does that tell us what would happen if we reduced the drinking age to 18?
  - Maybe, but we have to think about generalizability
- Earlier we discussed "local average treatment effects". With RD we get a very local ATE
  - We have learned the effects of alcohol access on mortality for 21-year-olds
  - We cannot be certain these results would generalize to 18-year-olds
- Next, David Broockman will discuss a few more examples of RD

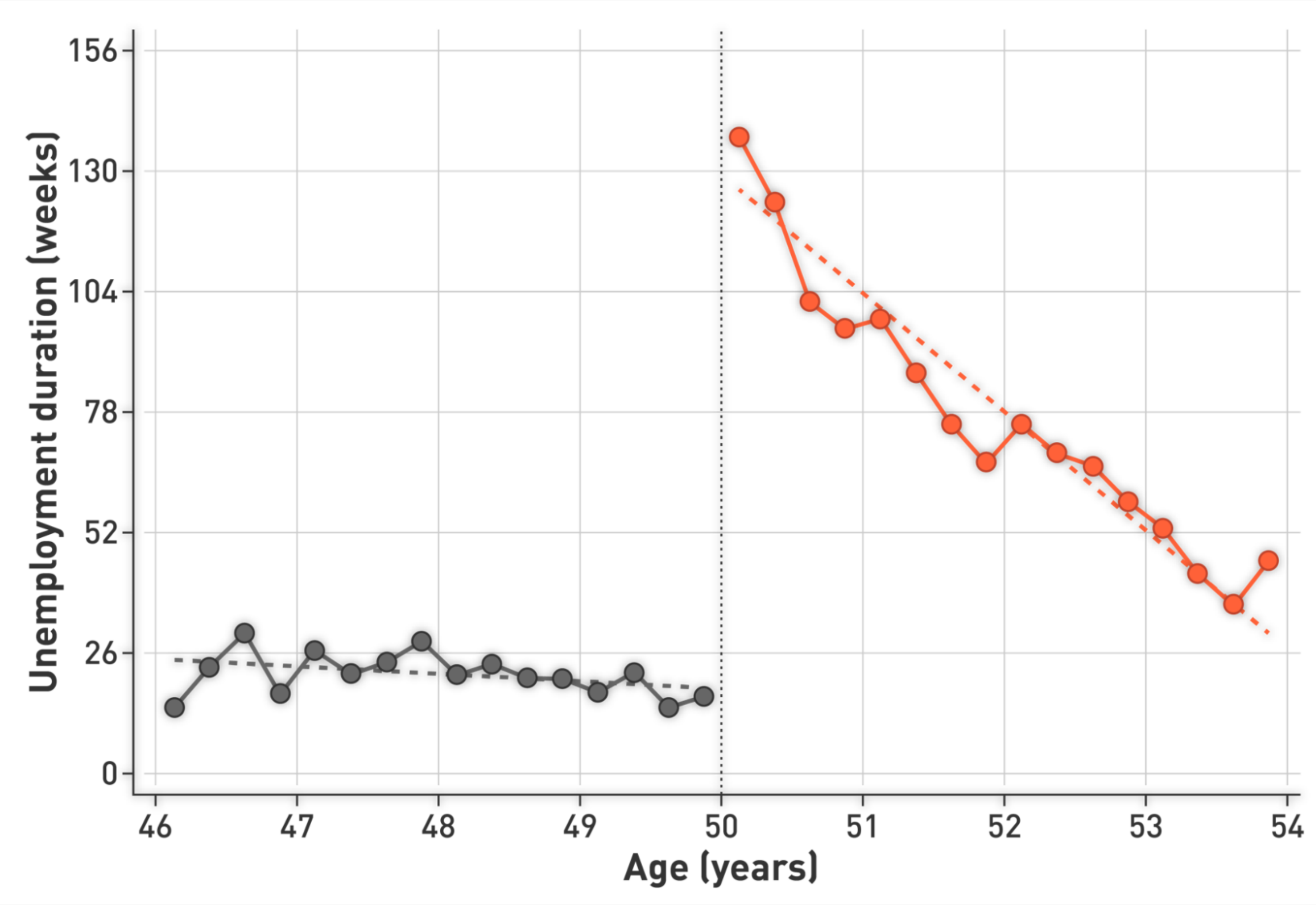
Example:

Effects of Unemployment Insurance

Example of Regression Discontinuity

- Do unemployment benefits lead people to slack off?
- Ideal experiment: randomly assign some countries or persons to strict unemployment regimes
- Find "naturally occurring" coin flip situations
- "Discontinuity" in Austria
  - People over 50 eligible for unemployment benefits
  - People under 50 not eligible
  - Does the age threshold affect worker productivity?

Example of Regression Discontinuity



- Simply comparing young vs. old could lead to false conclusion that being old causes differences in unemployment duration

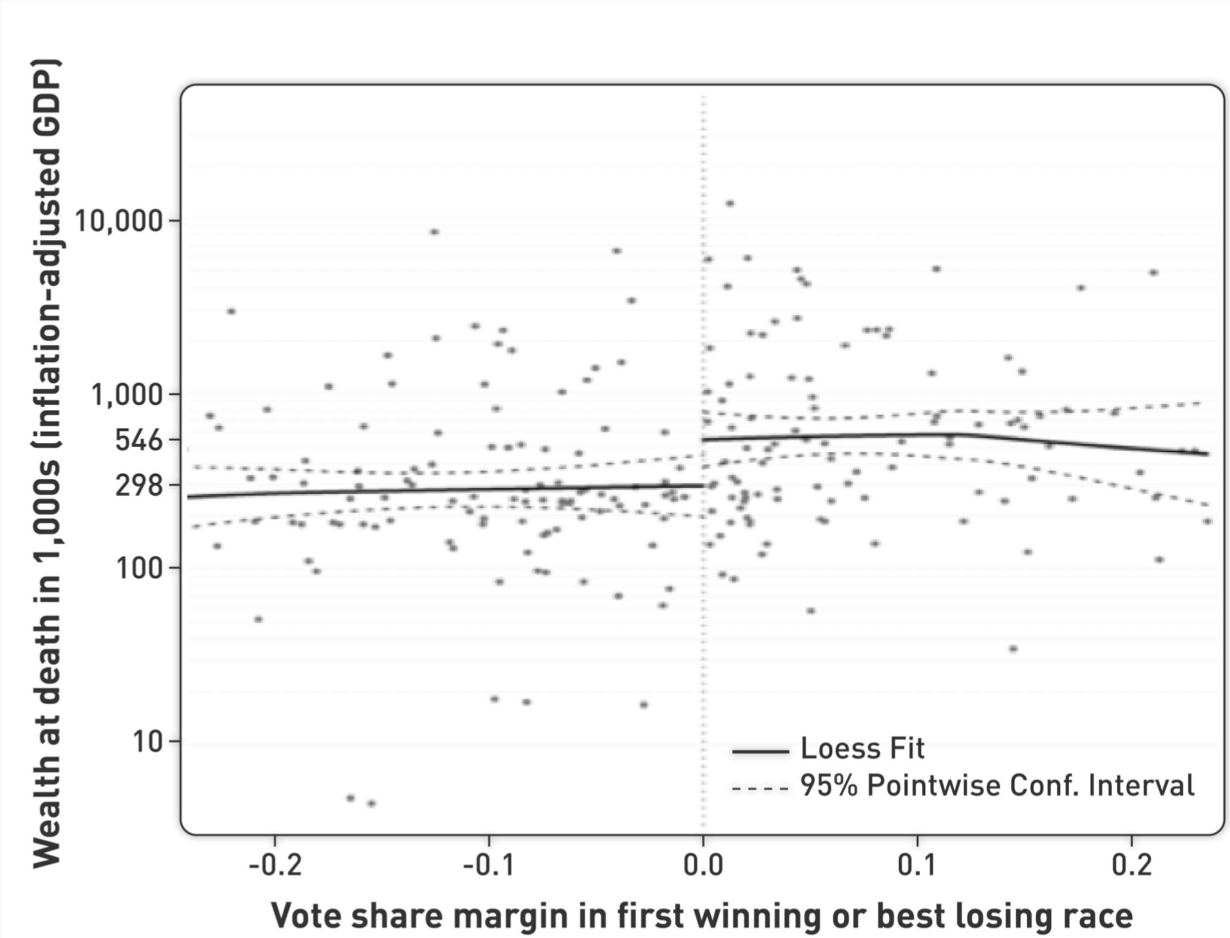
Example:

Does Getting Elected Make You Rich?

Does Getting Elected Make You Rich?

- Ideal experiment: randomly assign person to be elected and measure their wealth when they die
- Eggers and Hainmueller
  - Examined wealth at death of British MPs who won or lost by narrow margins

Does Getting Elected Make You Rich?



Note: the y-axis on this graph is a logarithmic scale, so MPs who won essentially have double the wealth

Does Getting Elected Make You Rich?

- Ideal experiment: randomly assign person to be elected and measure their wealth when they die
- Eggers and Hainmueller
  - Examined wealth at death of British MPs who won or lost by narrow margins
- Interpretation issue: Are results simply due to MPs moving to London?

Strengths and Weaknesses of Regression Discontinuity Designs

Regression Discontinuity Designs

"Sorting" at cut point

Ex ante: before the event

- People are aware of the threshold and can try to get above it
- Eg. Latin honors
  - Effect of honors on success in labor market
  - Awareness of threshold results in better students just above cutoff

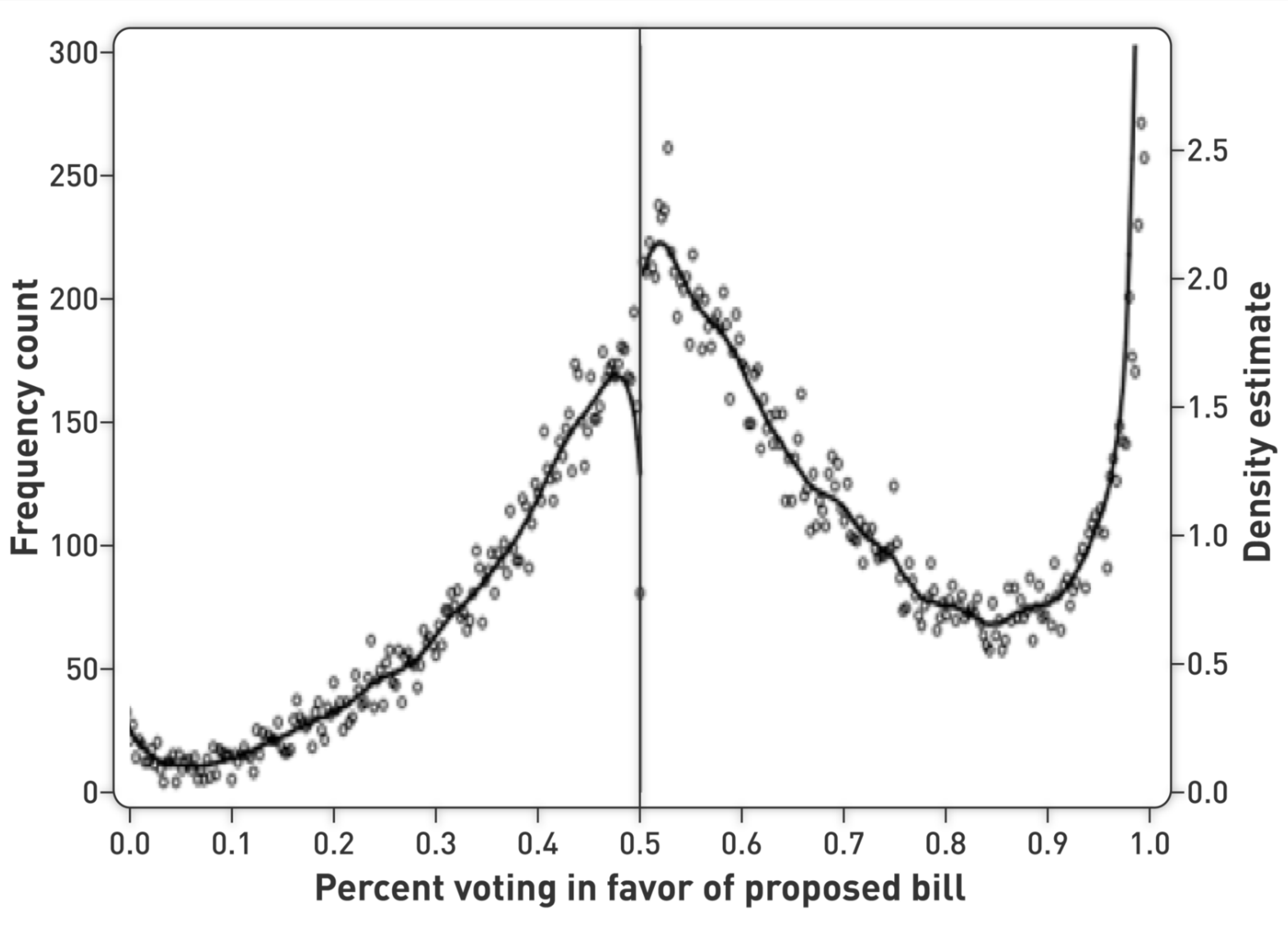
Ex post: after the event

- Eg. Elections of Black mayors in the 1970s
- Black candidates disproportionately lost by slim margins
- Manipulations make analysis difficult

Tests for Sorting

- **Covariate balance** on either side of cut point
  - Does covariate balance hold up based on early measurements?
- **Smoothness (McCrary test)**
  - Eg. effects of minimum-wage laws on unemployment
  - Comparison between states based on margin of vote passing

Tests for Sorting



Tests for Sorting

- **Covariate balance** on either side of cut point
  - Does covariate balance hold up based on early measurements?
- **Smoothness (McCrary test)**
  - Eg. effects of minimum-wage laws on unemployment
  - Comparison between states based on margin of vote passing
- Can't directly compare bills due to differences in kinds of bills that just pass and just fail
- Check for proportion of observations in running variable
  - Eg. fraction of yes votes vs treatment variable
- Passage of regression discontinuity check doesn't guarantee assumptions hold
- Check for lack of covariate balance, lack of smoothness

Yoga:

Conducting Regression Discontinuity Analysis

Difference in Differences

Reading Assignment

- The third observational technique we will consider is called **difference in differences (DID)**
- This technique is used when a natural experiment takes place over time, in a before-after setting, and when we don't have a randomized control group
- Instead, we have a group that is similar to the treated group but not guaranteed to be identical

Read *Mastering Metrics*, Section 5.1, page 178 through the top of page 187

Note: Lecturer meant to say *"...section 5.1"*

Example:

Difference in Differences

Mississippi Credit Bank

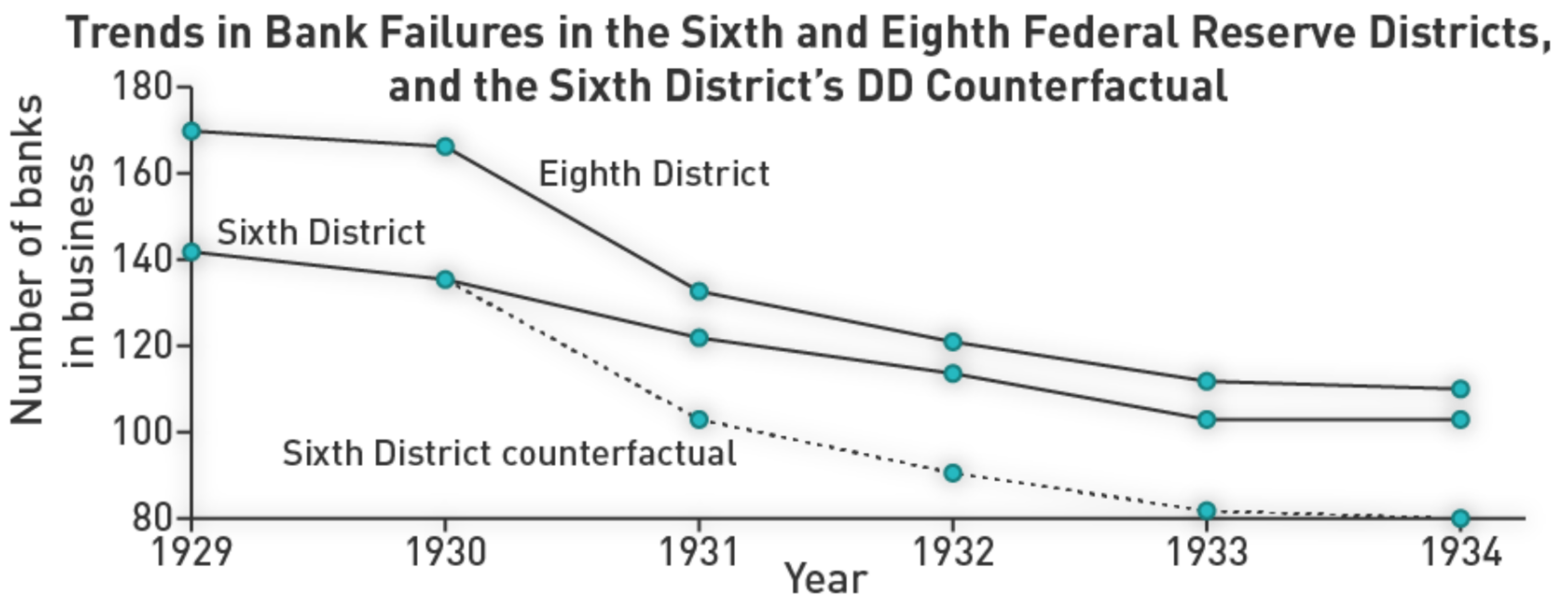
Effects of Federal Reserve Lending Policy to Mississippi Banks

- It is difficult to do a field experiment on monitary policy
- Instead we look for a natural experiment
  - In 1930, the Federal Reserve Bank of Atlanta (6th District) chose to provide much more credit to troubled banks than did the St. Lous Fed (8th District)
  - The state of Mississippi was divided almost equally between the two districts

Effects of Federal Reserve Lending Policy to Mississippi Banks (cont'd)

- The banks are not randomized, so a simple difference can be misleading
  - In 1931, 11 fewer banks were open in the 6th District than the 8th
  - We might be tempted to conclude that easy lending caused more banks to close
  - BUT before the crisis, 30 fewer banks were open in the 6th District
- We must account for this baseline difference
- This is what DID does by subtracting the before-after difference in the treatment group from the before-after difference in the psuedo-control group

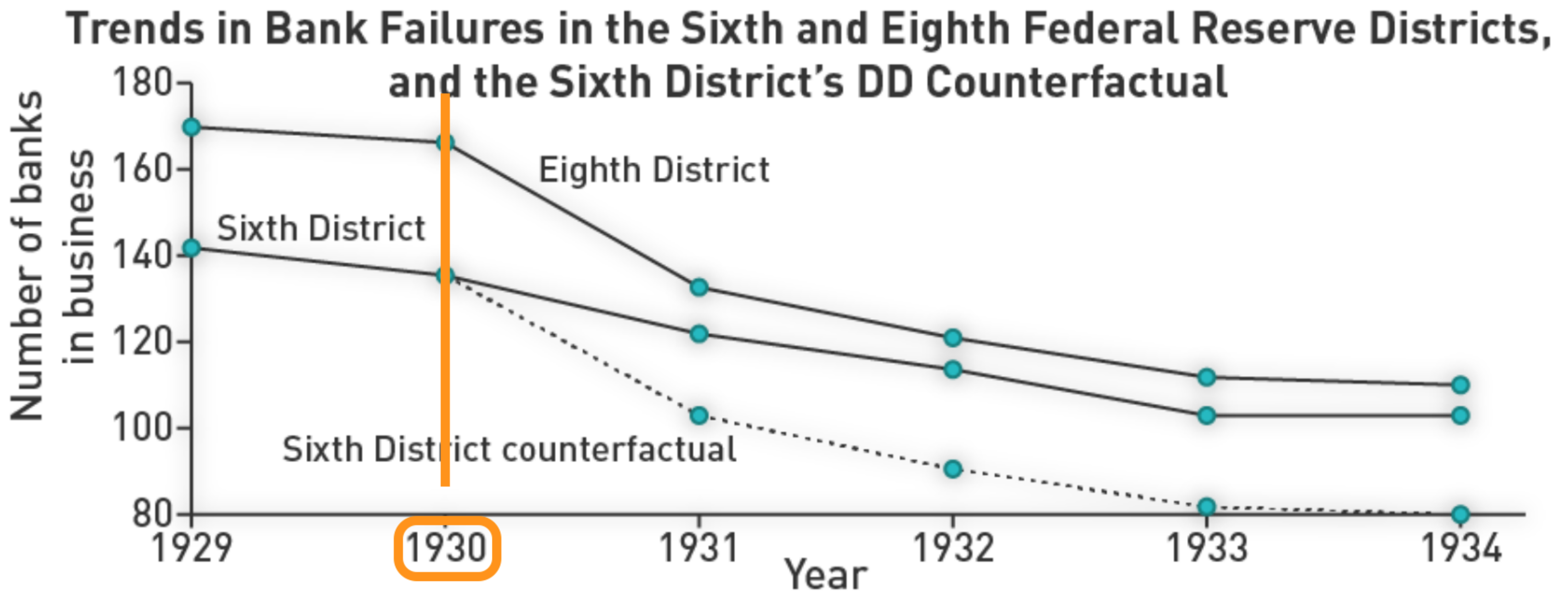
Figure 5.3: DID Strategy Illustration



Notes: This figure adds DD counterfactual outcomes to the banking data plotted in Figure 5.2. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district after 1930 as did in the Eighth.

- To get a valid causal difference with DID, we need to satisfy the **"common trends"** (or "parallel trends" assumption

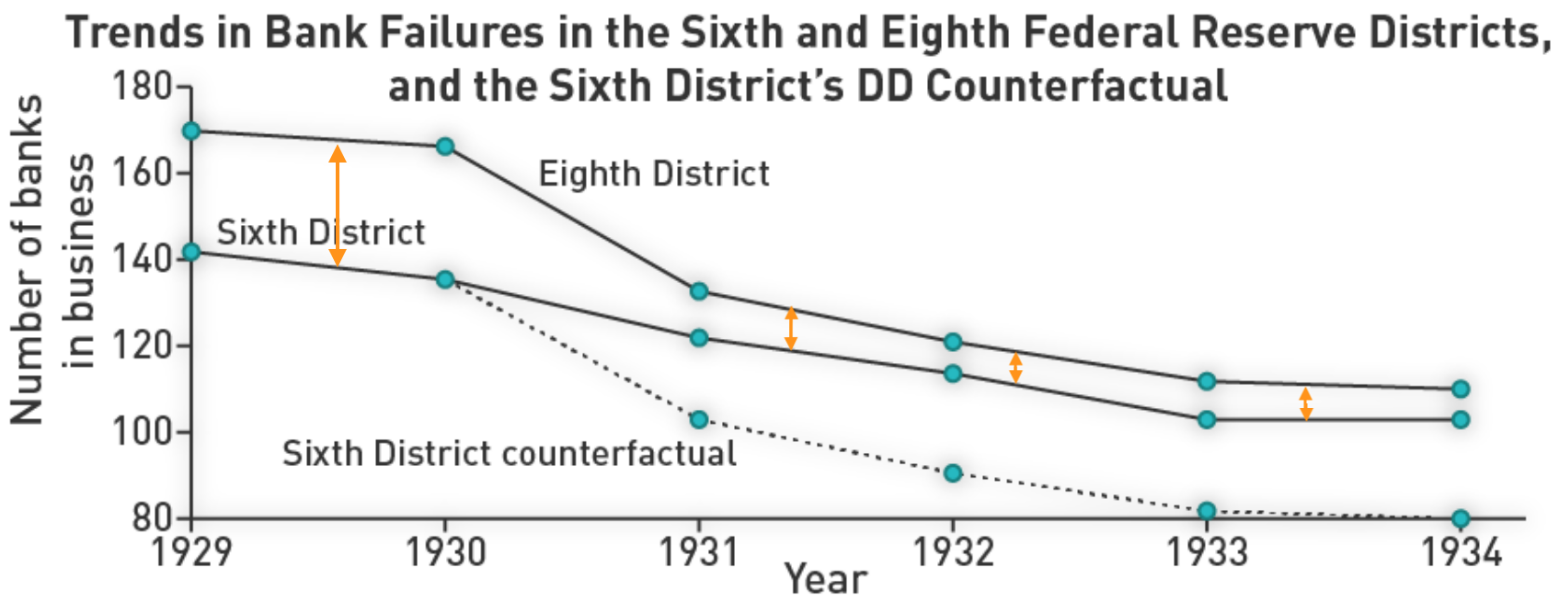
Figure 5.3: DID Strategy Illustration



Notes: This figure adds DD counterfactual outcomes to the banking data plotted in Figure 5.2. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district after 1930 as did in the Eighth.

- Figure 5.3 shows that this assumption is reasonable
- The hypothesized policy difference occurs in 1930

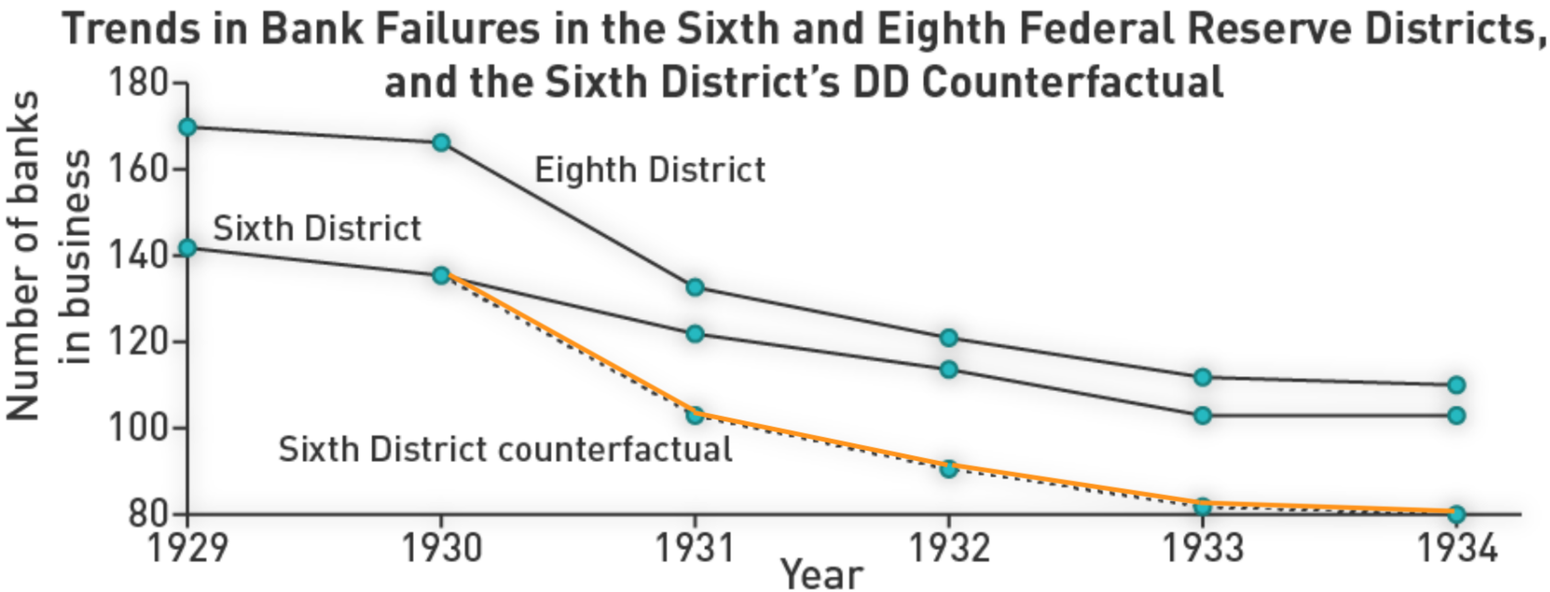
Figure 5.3: DID Strategy Illustration



Notes: This figure adds DD counterfactual outcomes to the banking data plotted in Figure 5.2. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district after 1930 as did in the Eighth.

- In other years the two lines (number of banks open) move in parallel

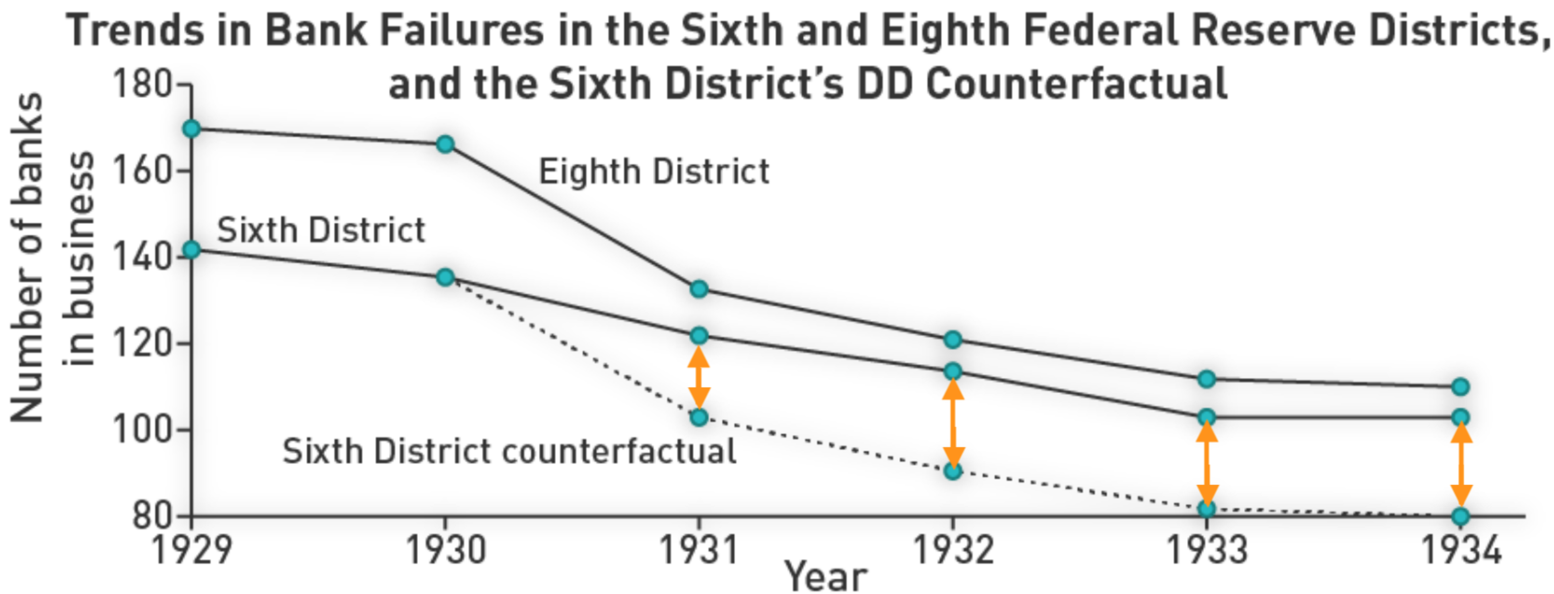
Figure 5.3: DID Strategy Illustration



Notes: This figure adds DD counterfactual outcomes to the banking data plotted in Figure 5.2. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district after 1930 as did in the Eighth.

- The dashed line shows the counterfactual number of banks that would have been open in the 6th District if the Atlanta Fed had not pursued an easy lending policy in 1930

Figure 5.3: DID Strategy Illustration



Notes: This figure adds DD counterfactual outcomes to the banking data plotted in Figure 5.2. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district after 1930 as did in the Eighth.

- The difference between the 6th District actual and hypothetical curves shows the estimated causal effect of the 6th District's lending policy

Reading Assignment

Read the rest of *Mastering Metrics* Section 5.1, pages 187-191

- This section discusses how to implement a DID strategy using regression
- The technique will look familiar if you remember the material on using regression to analyze heterogeneous treatment effects

Difference in Difference Regressions

DID Regressions

$$Y_{dt} = 167 - 29TREAT_d - 49POST_t + 20.5(TREAT_d * POST_t) + e_{dt}$$

- If we have more than two time periods or more than two comparison groups available, we can implement DID via a regression
- **TREAT** is a dummy variable equal to one in the treatment group (the 6th district in this example
    - The coefficient of -29 means that in the pretreatment baseline period there were 29 fewer banks open in the 6th District than in the 8th

DID Regressions (cont'd)

$$Y_{dt} = 167 - 29TREAT_d - 49POST_t + 20.5(TREAT_d * POST_t) + e_{dt}$$

- **POST** is a dummy variable
  - POST = 0 prior to treatment
  - POST = 1 during the treatment time period
  - The coefficient of -49 means there was a downward trend in banks open before and after 1930 (49 fewer banks on average during the post-period than in the pre-period in the 8th District)

DID Regressions (cont'd)

$$Y_{dt} = 167 - 29TREAT_d - 49POST_t + 20.5(TREAT_d * POST_t) + e_{dt}$$

- The coefficient of 20.5 on the interaction term is the estimated causal effect of easy lending policy to troubled banks
    - Eg. the difference in the post-period vs the pre-period in the treatment group, relative to the difference in the comparison group
- Next, David Broockman will share some additional examples of DID

Yoga:

Difference-in-Difference Regressions

Yoga:

Coding the DID for Mississippi Banks

Yoga:

Difference-in-Difference Regressions

Difference in Differences:

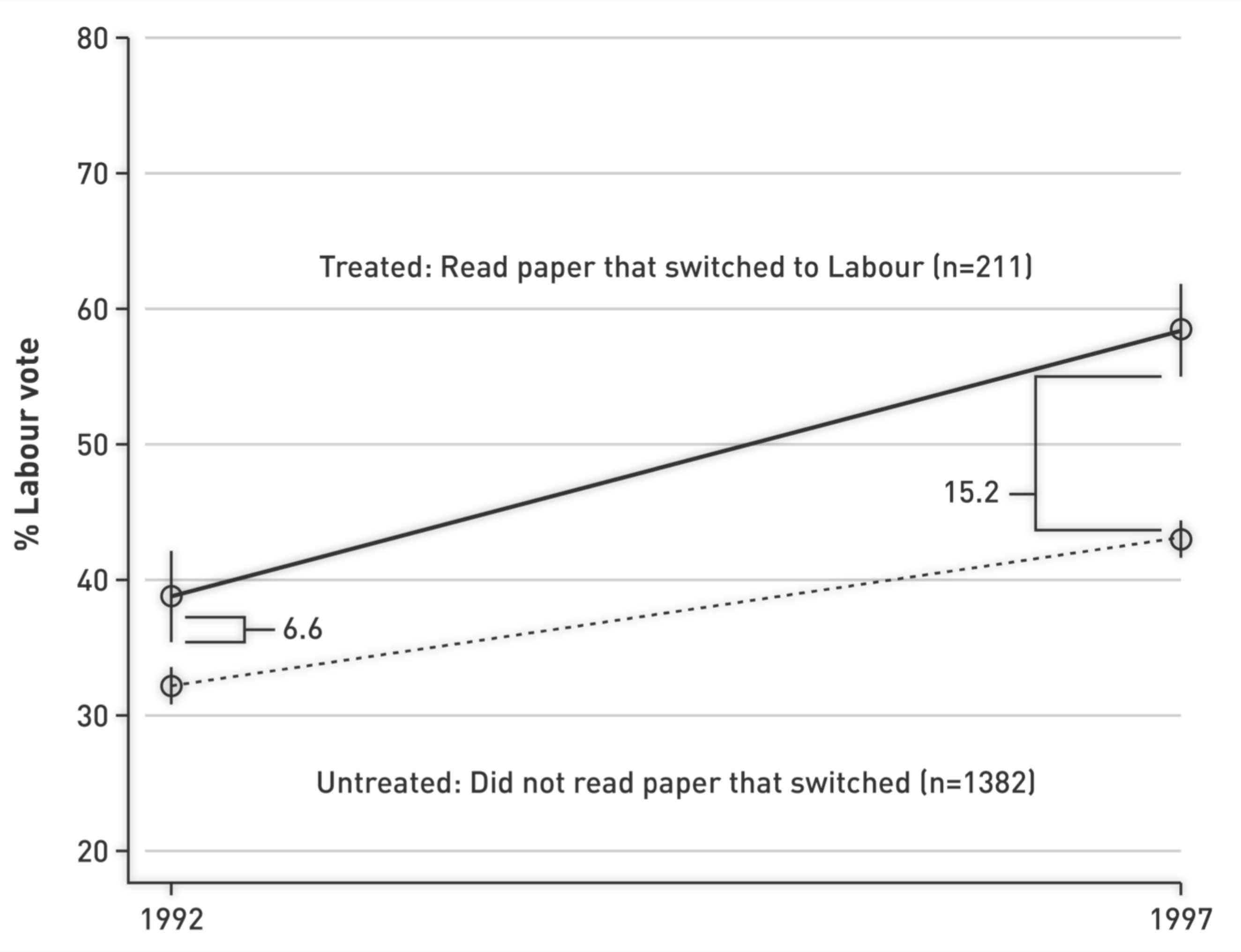
Does Media Affect Politics?

DID Example

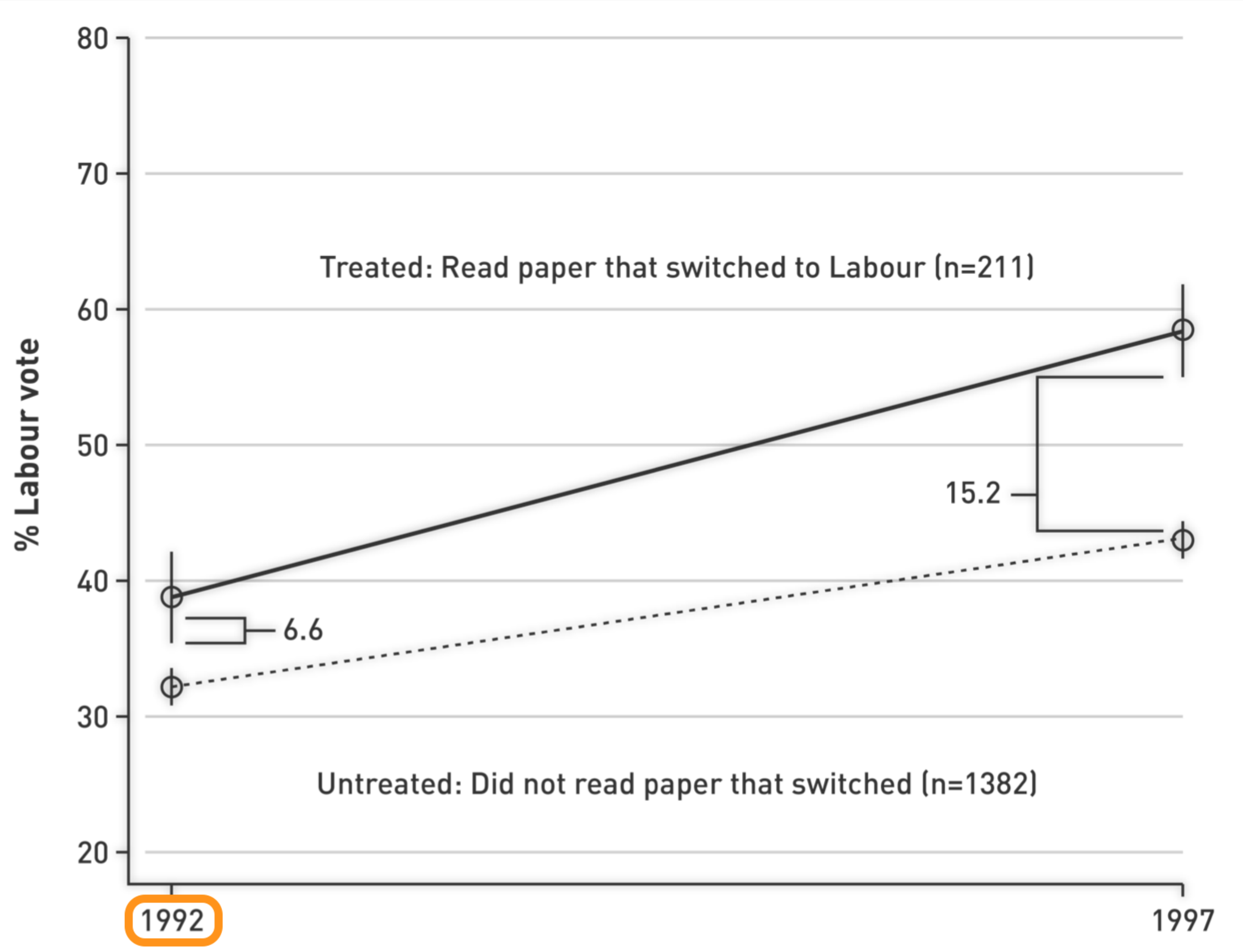
Ladd and Lenz (2009)

- Does the media affect politial outcomes, specifically newspaper endorsements?
- Ideal experiment: Randomly assign newspapers to endorse a candidate
- People often read newspapers that reinforce their views
  - Do readers' views change if newspaper switches endorsement?
- Data on voters' decision to vote for Labour Party in 1992 and 1997
  - Which newspapers voters read
  - Some newspapers switched endorsement to Labour Party

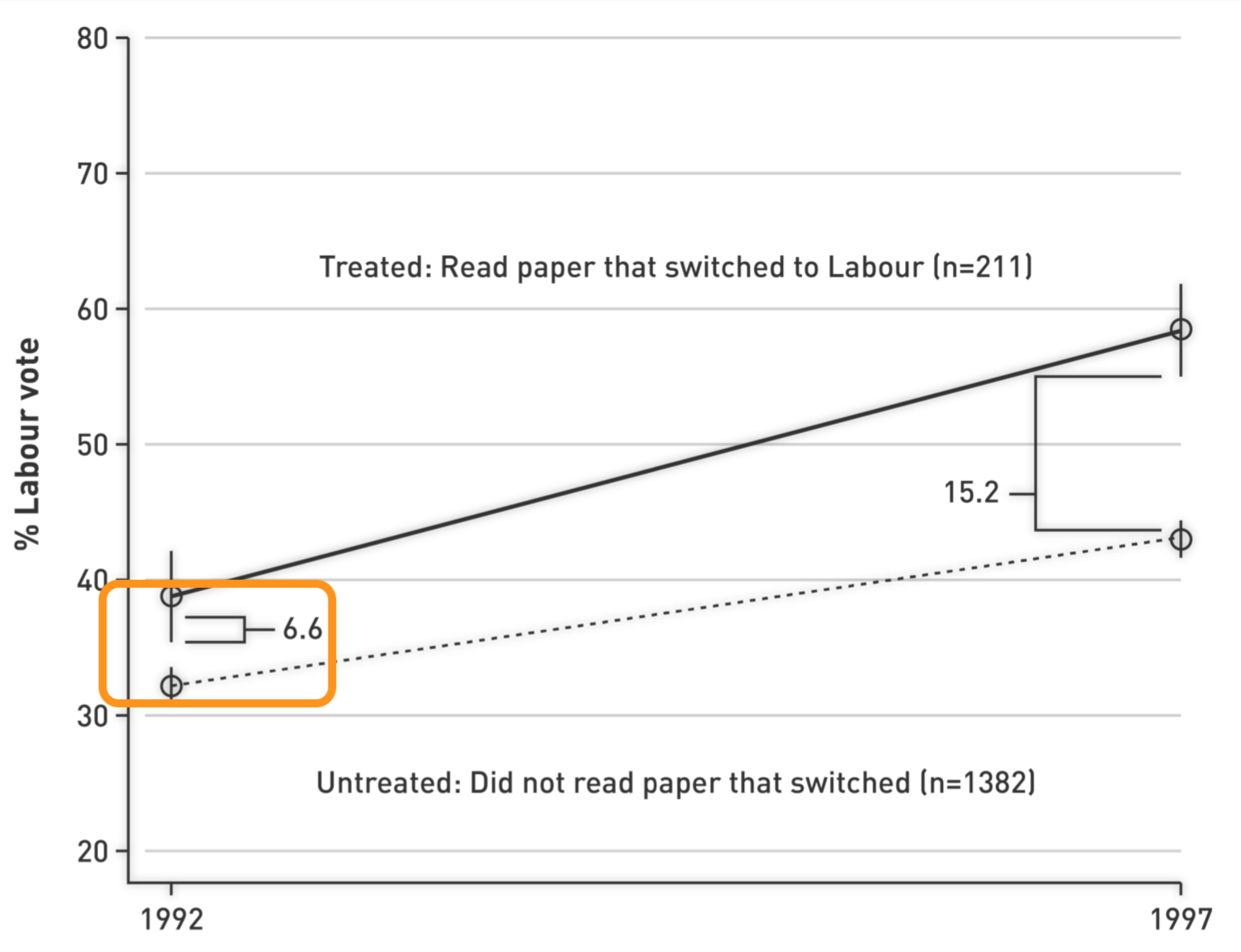
DID Example: Ladd and Lenz (2009)



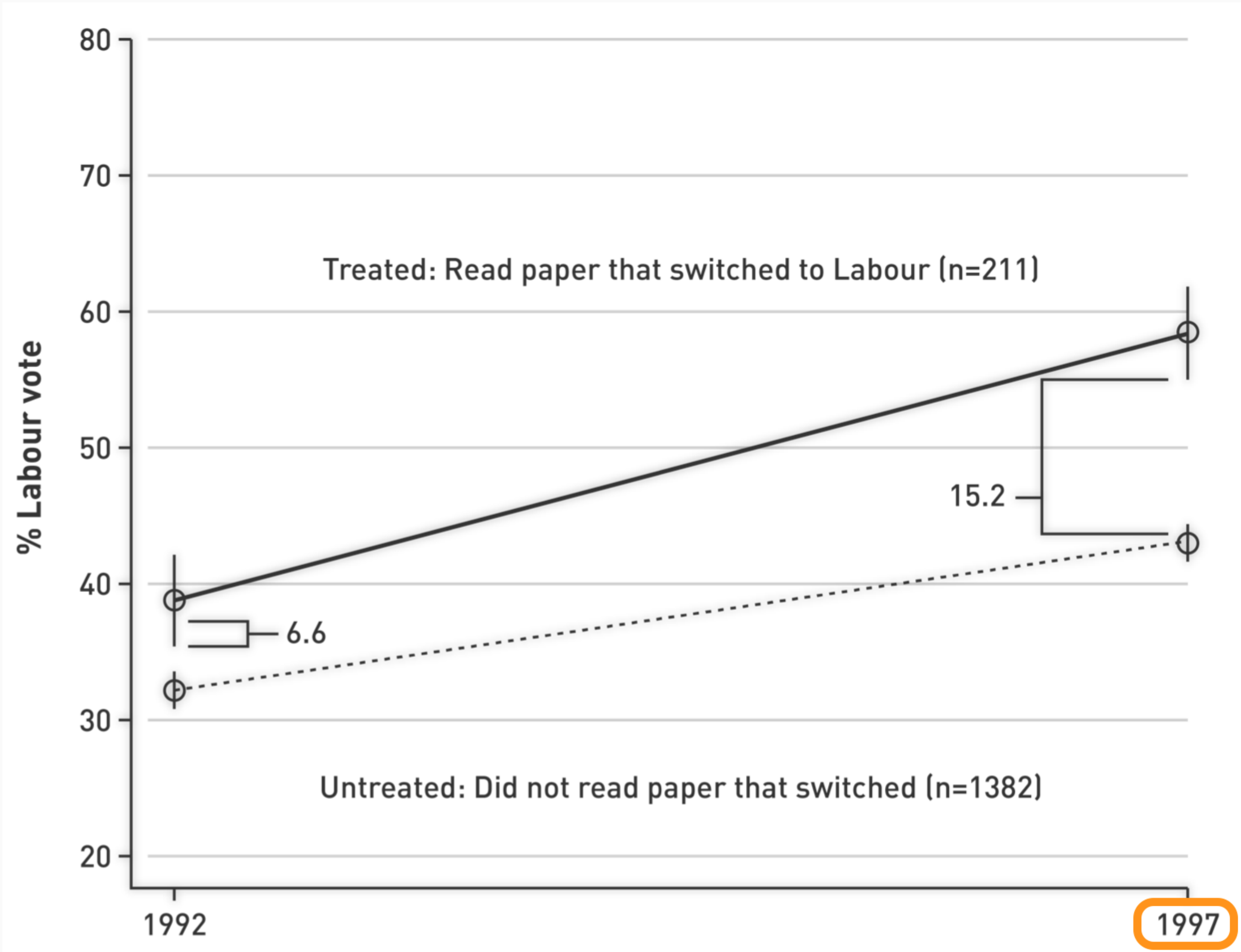
DID Example: Ladd and Lenz (2009)



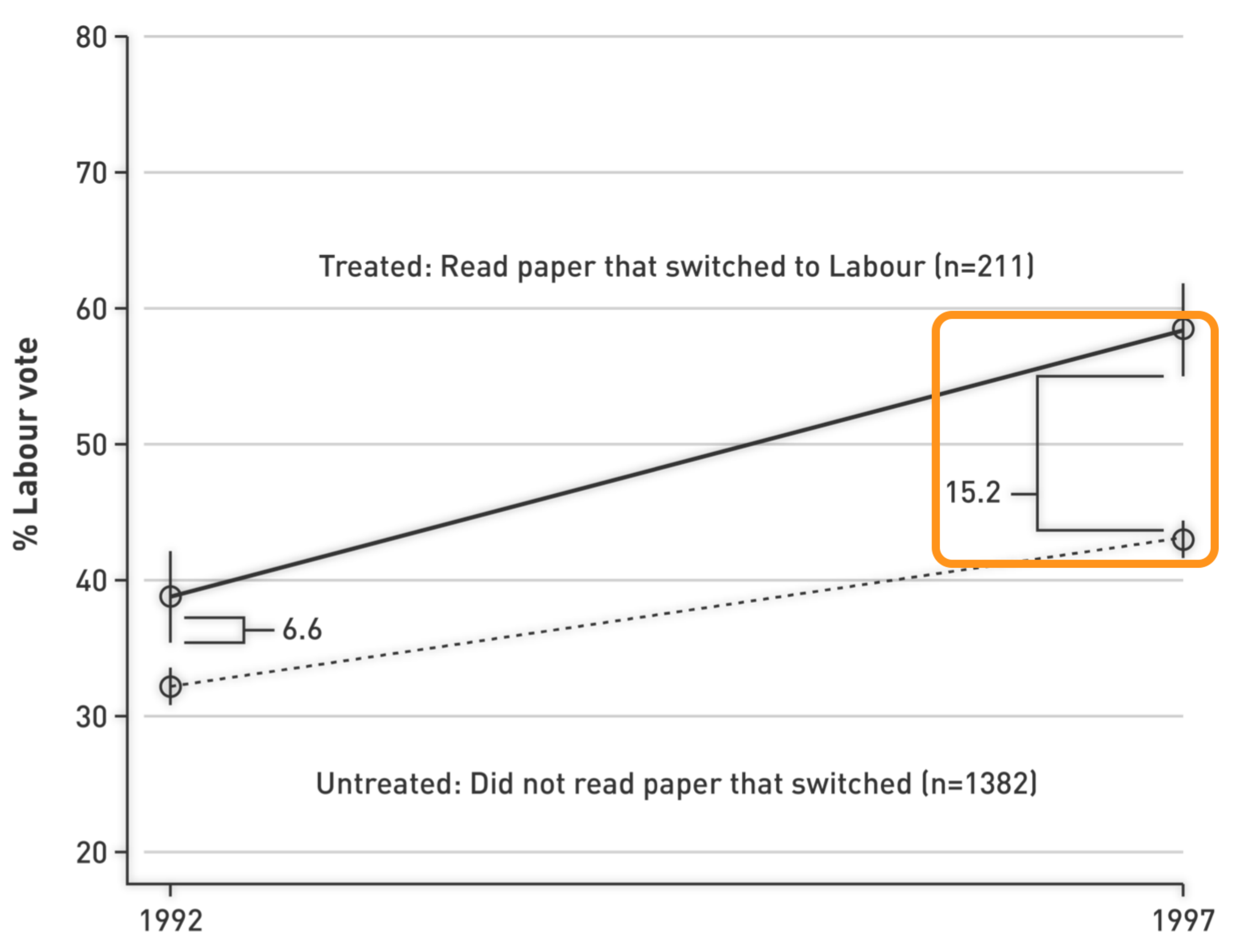
DID Example: Ladd and Lenz (2009)



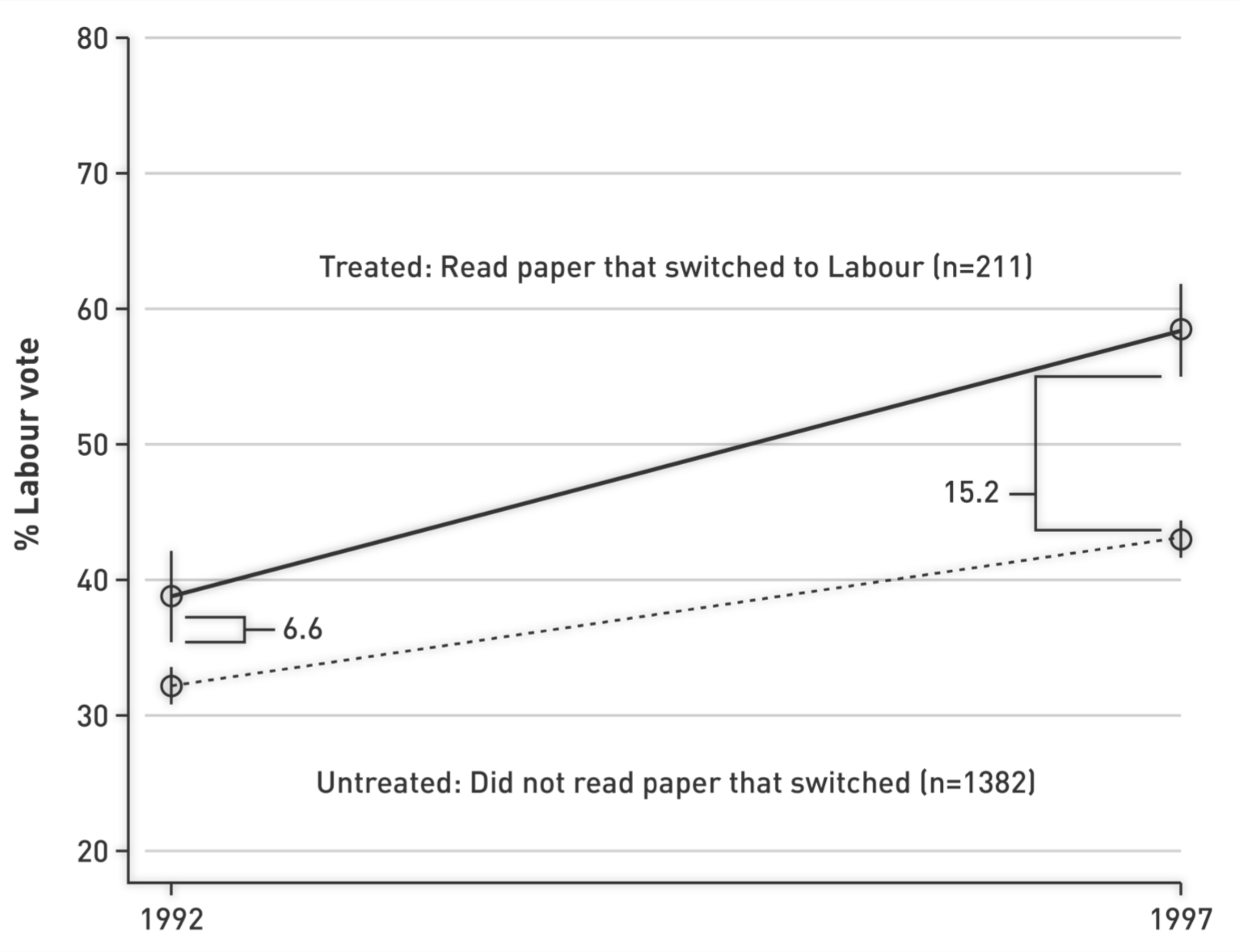
DID Example: Ladd and Lenz (2009)



DID Example: Ladd and Lenz (2009)



DID Example: Ladd and Lenz (2009)



DID Example: Ladd and Lenz (2009)

- Does the media affect politial outcomes, specifically newspaper endorsements?
- Ideal experiment: Randomly assign newspapers to endorse a candidate
- People often read newspapers that reinforce their views
  - Do readers' views change if newspaper switches endorsement?
- Data on voters' decision to vote for Labour Party in 1992 and 1997
  - Which newspapers voters read
  - Some newspapers switched endorsement to Labour Party
- Key assumption: nothing changing over time that leads to treatment and outcome
  - Consider other reasons for results

Potential Problems with Ladd and Lenz

- Big housing crisis in 1997
- People voted for Labour due to frustrations with incumbents
- People with homes were:
  - More upset over housing crisis
  - More likely to get newspapers
  - More likely to receive treatment and vote based on housing crisis
- Ladd and Lenz checked on this using a placebo test
  - DID when treatment didn't change, readers of other papers
  - No effect seen among readers of other papers, only papers which switched their endorsement

Difference in Differences:

Does Price Advertising Change Prices?

DID Example: Advertising and Prices

- Rhode Island (RI) banned advertising prices of alcohol
    - Will firms offer less competetive pricing as a result?
- Could compare RI alcohol prices to those in Connecticut and Massachusetts (MA)
    - Other conditions between states might differ
- RI ended advertising ban
    - Can't simply do a before/after analysis
    - Price change could reflect other factors
- Compare changes in RI over time to changes in MA over time
    - Prices increased 2% in Rhode Island
    - Likely due to inflation
- Milyo and Waldfogel (1999) collected data on liquor prices in RI and MA before and after ban
    - Had foresight to collect data when Supreme Court agreed to hear case on advertising ban
    - Found prices of newly advertised products in RI **decreased** relative to same goods in MA
    - Prices didn't drop for product that were never advertised
    - Prices dropped only for newly advertised products

Review of Difference-in-Differences

- Did heps counteract selection efforts
- Can't always directly compare two entities
- Before/after effects can be misleading

Difference in Differences:

Home Prices and Incinerator Location

DID Example: Home Prices

- How much do property values drop as a result of having an incinerator nearby?
- Can't simply look at home prices in relation to distance from incinerator
  - Why some places have incinerators while others don't?
  - Ideal experiment: randomly assign location of incinerators
  - Incinerators tend to be built in bad neighborhoods where property values are lower to begin with
  - Unobserved heterogeneity (Eg. bad construction or recently built development)

Kiel and McClain (1995)

- DID approach
- Collected data in two time periods:
  - Before incinerator announced
  - After incinerator built
- Compared data in North Andover to other locations
- Found home prices near incinerator increase about 5% less but not significantly significant
  - Other characteristics matter more
  - Extra bathroom: +13% +/- 5%
  - Average prices over time (before/after): +14% +/- 6%

DID Example: Home Prices

Importance of the DID Approach

- Home prices near incinerator could have had lower baseline resulting from pre-existing differences
- DID protects from false conclusions drawn from simple before/after approach
- Vulnerabilities still exist
    - Eg. demographic changes could lead to incinertor being built in the first place

Strengths and Weaknesses of DID

Strengths of DID

- DID is a simple tool: we computer differences directly or do OLS regression with clustered standard errors
- If we have covariates that might predict differences in Y, we can include them to try to improve precision
- Causal effects are much stronger and more believable than with simple before/after differences or simple differences across non-randomized comparison groups

Weaknesses of DID

- Causal inference hinges on the assignment mechanism
- The common trends (parallel trends) assumption is a strong assumption.
- An example where it fails: observational online-advertising
  - People may be gone on vacation
  - People may see simultaneous newspaper advertising campaigns

Synthetic Controls as Generalized Difference in Differences

Yoga:

Coding California Smoking with Synthetic Controls #1

Yoga:

Coding California Smoking with Synthetic Controls #2

Concluding Thoughts