

Lecture 7

Administrative Miscellanea

- Grades as of Nov 18 Available on Blackboard
- Pset 3 fully available
- Homework 9 on Blackboard
- Quiz Wednesday

Quasiexperimental Approaches in Econometrics

- When we run a linear regression, we only get a causal estimate when we have exogenous variation - variation that is not jointly correlated with our dependent and independent variable
- Adding controls allow us to correct for most endogeneity that we observe, but if we don't have data on something we can't control for it
- Randomized trials force exogenous variation: because we randomly assigned treatment, it is not related to our existing covariates

Quasiexperimental Approaches in Econometrics

- Without a randomized trial, we can instead use some “natural experiments” - some change that impacted some individuals but not others in a way that is as good as random assignment. Examples include:
 - People were drafted into the Vietnam war based in part on their day of birth, which is unrelated to future employment outcomes (Angrist 1990)

Quasiexperimental Examples

- The tea party organized nationwide protests on April 15th, 2009. In some areas bad weather caused small turnout, leading to worse outcomes for Republican candidates in the 2010 election (Madestam et al 2013)
- In 1980 Fidel Castro allowed anyone to leave Cuba, causing a mass migration of 125,000 cubans to America, most of whom settled in Florida due to proximity

Difference-In-Differences as a Quasiexperiment

- The Mariel Boatlift experiment is an example of a difference-in-differences (DiD) design
- The concept is simple: compare the outcome you're interested in (employment of natives) before and after this event
 - The problem with this is that many other things could affect your outcome (e.g. recessions).
 - We need to know what would have happened without the experiment (the counterfactual)

Difference-In-Differences as a Quasiexperiment

- Take a city that was unaffected by the migration from cuba, but is otherwise similar to the affected city (in this case Miami).
- The difference in employment for your counterfactual city (say, Atlanta) before and after is taken as what would have happened without intervention due to natural trends or other fluctuations

Difference-In-Differences Steps

- Take the difference in employment before and after the boatlift (year 1980) in Miami (difference 1)
- Do the same thing for Atlanta (difference 2)
- Finally, take the difference between these to get your causal impact (difference in differences)

Difference-In-Differences as a Quasiexperiment

- This can be done in a regression as $y = \beta_0 + \beta_1 post + \beta_2 treat + \beta_3 treat * post + \varepsilon$
 - β_3 is our coefficient of interest
- This is also equivalent to $y_{st} = \beta_0 + \beta_1 D_{st} + \lambda_s + \mu_t + \varepsilon_{st}$
 - D is the indicator for the policy being in effect, and λ_s, μ_t are state and time fixed effects.
 - This is a two-way-fixed-effects estimate

Difference-In-Differences as a Quasiexperiment

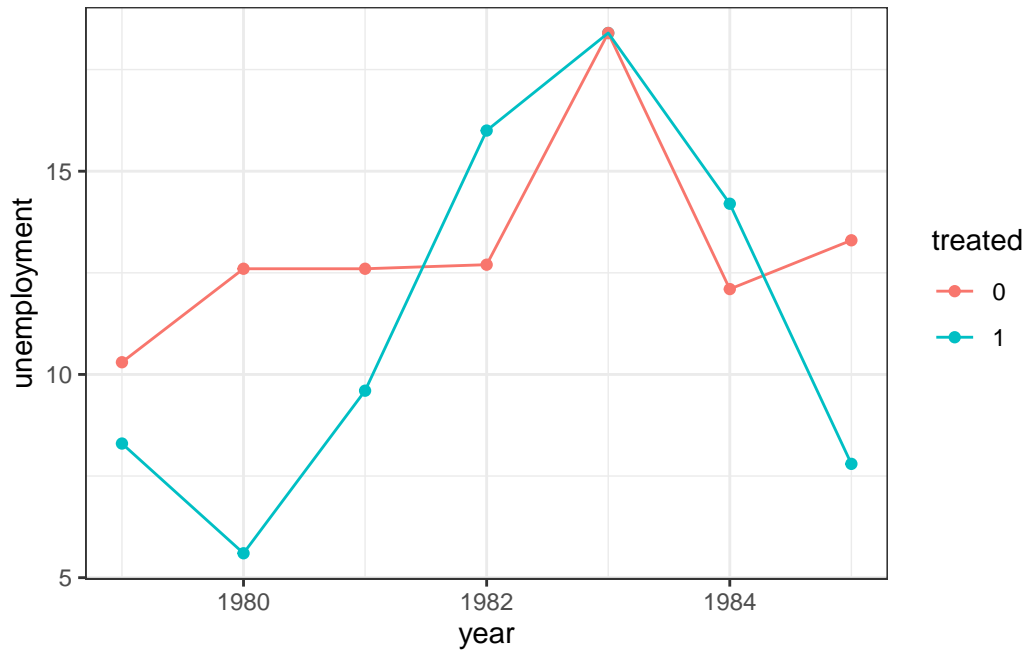
- We only look at variation within a state and within a year.
- The only variation that remains are changes that occur within a state over time that are not attributed to global trends
 - i.e. only the differential trend within a state is picked up.
- We're attributing the entire differential trend between the Miami and Atlanta to be the causal effect
 - This is our identification assumption

Table 4. Unemployment Rates of Individuals Age 16–61 in Miami and Four Comparison Cities, 1979–85.
(Standard Errors in Parentheses)

<i>Group</i>	<i>1979</i>	<i>1980</i>	<i>1981</i>	<i>1982</i>	<i>1983</i>	<i>1984</i>	<i>1985</i>
<i>Miami:</i>							
Whites	5.1 (1.1)	2.5 (0.8)	3.9 (0.9)	5.2 (1.1)	6.7 (1.1)	3.6 (0.9)	4.9 (1.4)
Blacks	8.3 (1.7)	5.6 (1.3)	9.6 (1.8)	16.0 (2.3)	18.4 (2.5)	14.2 (2.3)	7.8 (2.3)
Cubans	5.3 (1.2)	7.2 (1.3)	10.1 (1.5)	10.8 (1.5)	13.1 (1.6)	7.7 (1.4)	5.5 (1.7)
Hispanics	6.5 (2.3)	7.7 (2.2)	11.8 (3.0)	9.1 (2.5)	7.5 (2.1)	12.1 (2.4)	3.7 (1.9)
<i>Comparison Cities:</i>							
Whites	4.4 (0.3)	4.4 (0.3)	4.3 (0.3)	6.8 (0.3)	6.9 (0.3)	5.4 (0.3)	4.9 (0.4)
Blacks	10.3 (0.8)	12.6 (0.9)	12.6 (0.9)	12.7 (0.9)	18.4 (1.1)	12.1 (0.9)	13.3 (1.3)
Hispanics	6.3 (0.6)	8.7 (0.6)	8.3 (0.6)	12.1 (0.7)	11.8 (0.7)	9.8 (0.6)	9.3 (0.8)

Note: Entries represent means of unemployment indicator variable for individuals age 16–61 in Miami and four comparison cities: Atlanta, Houston, Los Angeles, and Tampa–St. Petersburg. Samples are based on individuals in the labor force. See notes to Table 3 for definitions of groups and data sources.

Mariel Boatlift (Card 1990)



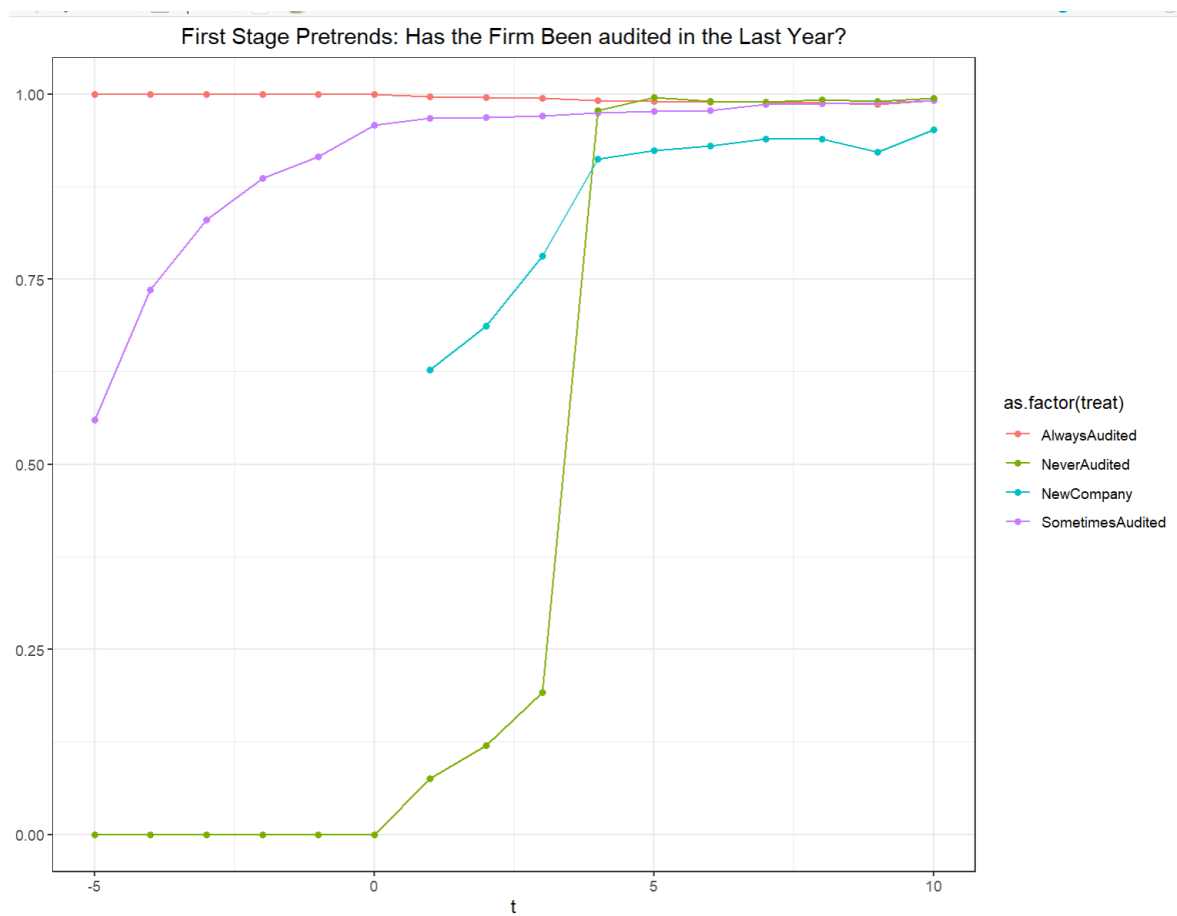
Mariel Boatlift (Card 1990)

- Card concludes that there were no effects on natives employment or earnings. Are there issues with this?

Mariel Boatlift Issues (Card 1990)

- 1) Parallel pretrends do not seem to hold - the comparison cities might not be representative of Miami
- 2) Selection into treatment: natives can move, attenuating the effect. This is a specific case of a “spillover effect”.
 - Another example would be if migrants affected demand for goods produced in Atlanta, changing their employment rate
- 3) General equilibrium response: firms can shift from capital to labor, attenuating the effect
- Takeaway: it can be hard to argue DiD approaches are truly capturing a causal effect

Pretrends

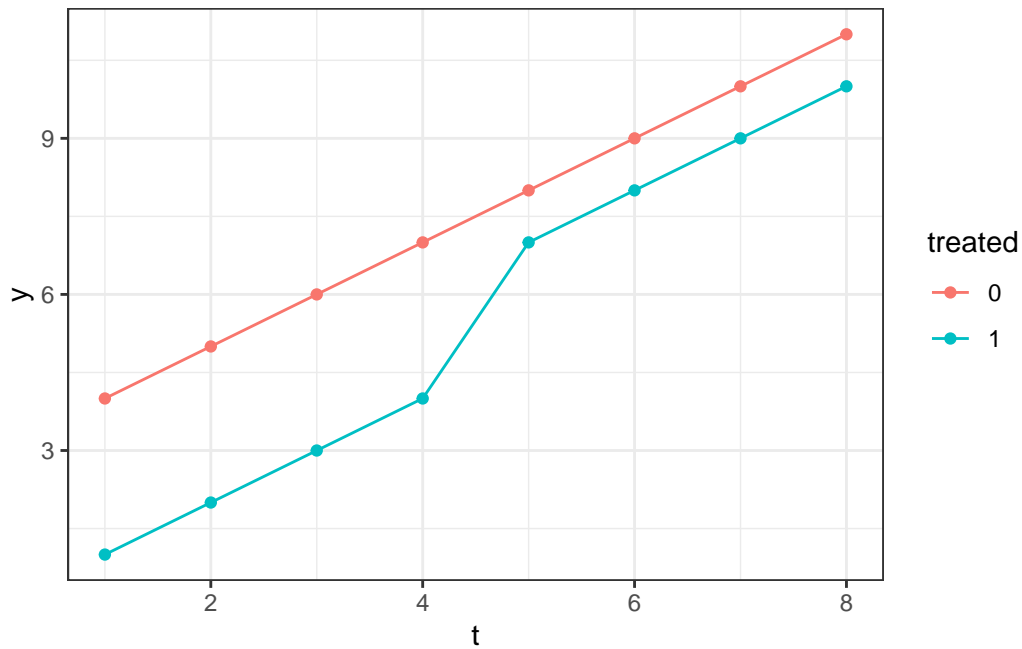


Pretrends

- If treatment and control were trending differentially before the policy was implemented it is hard to argue that we have a valid control
- Even if there is no differential trend (“Parallel Pretrends”), there could be another common shock that happened at the same time that ruins our assumption (SOX was enacted in 2002 following the collapse of Enron and the September 11th attacks)

Example

- Treatment occurs at the end of year 4 (first shows up in year 5)



```

treated 1 2 3 4 5 6 7 8
1:      0 4 5 6 7 8 9 10 11
2:      1 1 2 3 4 7 8 9 10

```

Example

- Diff-in-Diff value using 1 year: $(7-4) - (8-7) = 3-1 = 2$
- DiD using all years: $[(7+8+9+10)/4 - (8+9+10+11)/4] - [(1+2+3+4)/4 - (4+5+6+7)/4] = (8.5-9.5) - (-1 - (-3)) = 2$
- OLS Using all years: $9-4=5$
- DiD takes out the common time trend

Example

```

dt[,post:= as.numeric(t>4)]
lm(data=dt, y ~ treated + post + treated:post)

```

Call:

```
lm(formula = y ~ treated + post + treated:post, data = dt)
```

```

Coefficients:
  (Intercept)      treated1      post  treated1:post
          5.5          -3.0          4.0          2.0

```

```
lm(data=dt, y ~ post)
```

```

Call:
lm(formula = y ~ post, data = dt)

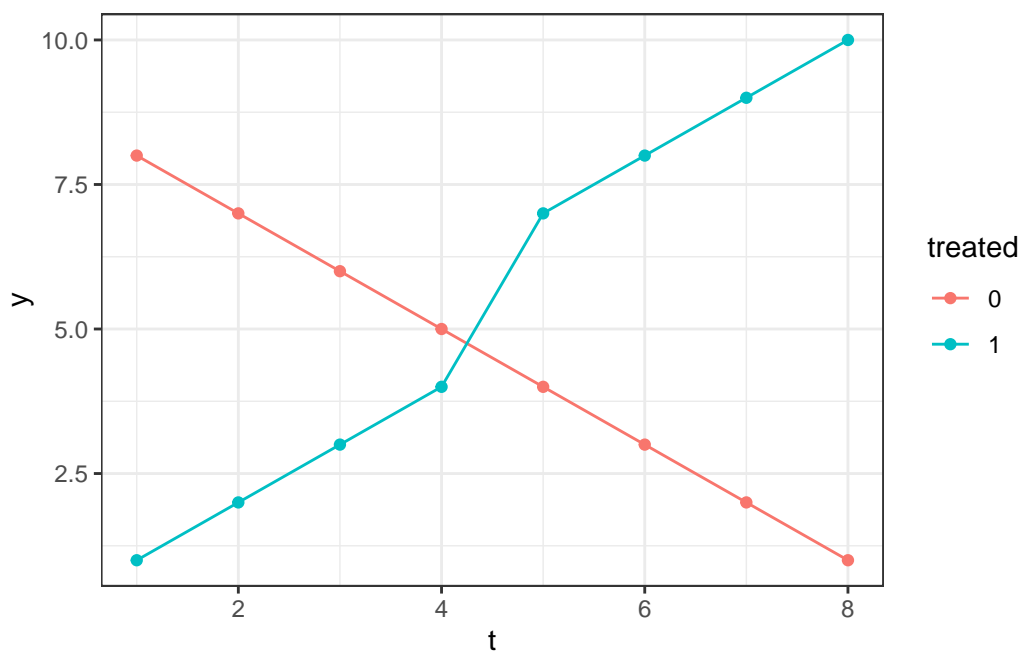
```

```

Coefficients:
  (Intercept)      post
           4          5

```

Example: Not parallel



```

treated 1 2 3 4 5 6 7 8
1:      0 8 7 6 5 4 3 2 1
2:      1 1 2 3 4 7 8 9 10

```


Example: Not parallel

- Diff-in-Diff value using 1 year: $(7-4) - (4-5) = 3 - (-1) = 4$
- Diff-in-Diff value using all years: $(8.5-2.5) - (2.5-6.5) = 6 - (-4) = 10$
- OLS using all years: $11/2 - 9/2 = 1$

Example: Not parallel

```
dt[,post:= as.numeric(t>4)]  
lm(data=dt, y ~ treated + post + treated:post)
```

Call:

```
lm(formula = y ~ treated + post + treated:post, data = dt)
```

Coefficients:

(Intercept)	treated1	post	treated1:post
6.5	-4.0	-4.0	10.0

```
lm(data=dt, y ~ post)
```

Call:

```
lm(formula = y ~ post, data = dt)
```

Coefficients:

(Intercept)	post
4.5	1.0

Dynamic Treatment Effects

- In the prior example the effect was different depending on whether we used 1 year or multiple years. This is a common problem in DiD methods where the treatment effect differs over time
- There is a similar issue where treatment is staggered: e.g. you have a policy that rolls out in multiple different states in different years. When this happens your early-treated states act as a counterfactual for your late-treated states
- These are complex issues that have several solutions, but is not yet settled as to which approach is best

Event Studies: a DiD Variant

- We may have different states adopt the same policy but in different years.
- Recall that one issue with DiD is that it assumes that the only change to occur is the policy, rather than something else that happened at the same time
 - e.g. cannabis was legalized in Illinois in the same year as covid, so it's hard to disambiguate the effect
- Staggered rollouts help mitigate this concern since we now have some variation over time.
- How do we actually implement the model?

Event Studies: a DiD Variant

- Set $t=0$ as the year of implementation of the state, $t=-1$ as the year before, $t=+1$ as the year after, etc.
- This aligns every state on the same axis
- $y_{st} = \beta_0 + \sum \gamma_t T_t + \lambda_t + \mu_s + \varepsilon_{st}$. Where T_t is constructed as above
 - This gives us an effect for each year before and after the policy, which is averaged over every state that implemented
 - What should γ_{-1} be?

Event Studies: a DiD Variant

- $y_{st} = \beta_0 + \sum \gamma_t T_t + \lambda_t + \mu_s + \varepsilon_{st}$
 - What should γ_{-1} be?
 - Every year before the policy should have a coefficient of 0. If they don't this is a violation of parallel pretrends

Staggered Rollout Issues

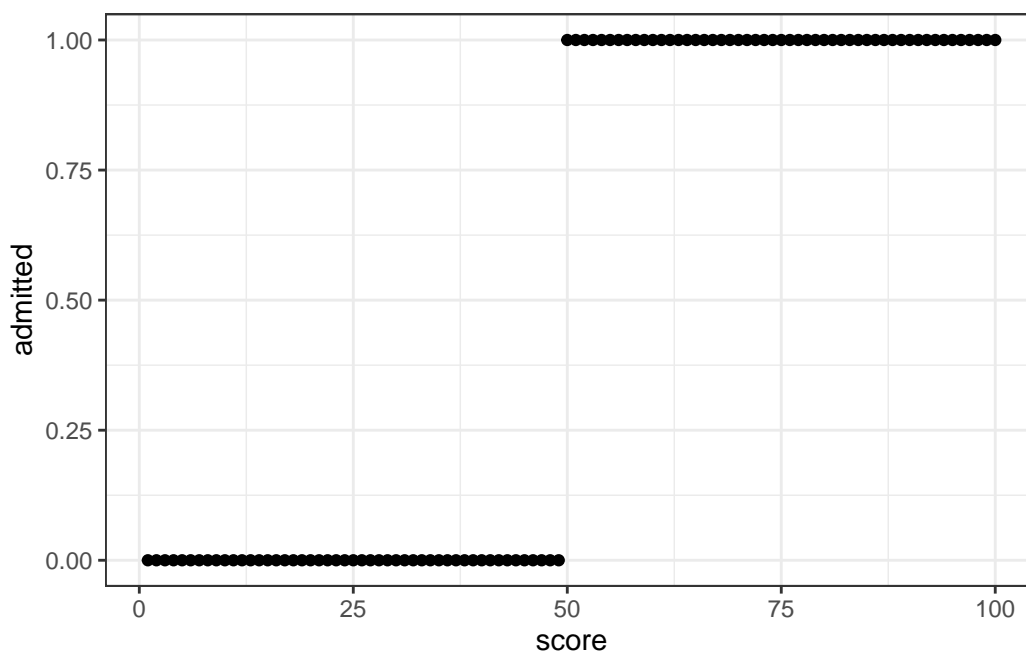
- Note that, e.g. γ_0 for the effect in the first year may be the average of Colorado in 2012, California in 2016, and Illinois in 2020
- γ_4 , on the other hand, will include Colorado and California, but not Illinois since we don't have data yet
 - This can cause wonky effects. See Goodman-Bacon (2021)
- Note also that In 2012 California acts as a counterfactual for Colorado, but in 2020 California does not act as a counterfactual for Illinois
 - Very weird effects if states change treatment status

Regression discontinuity: motivation

- Does attending college increase future earnings, and by how much?
- People who attend college have drastically different characteristics that also impact earnings. We can control for some of these, but many are unobservable
- Consider a low-ranking college (Florida International University) that denies applicants below a certain GPA threshold. Low performing high school students will be denied entry, but high scoring students are accepted. Assume for now that everyone above the cutoff is admitted, and no one below is.

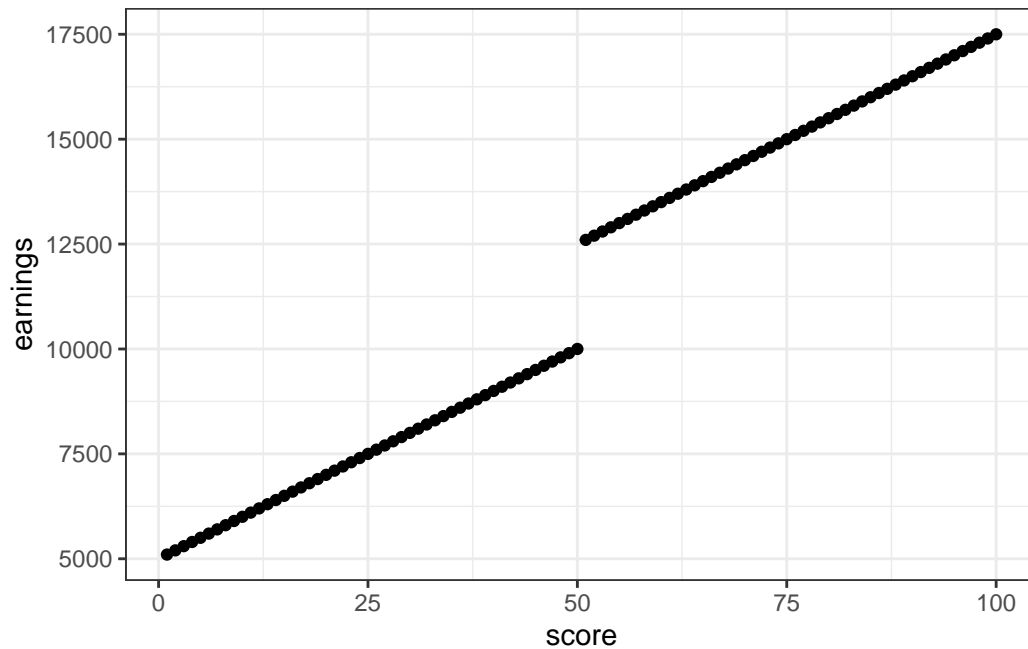
Regression Discontinuity: Motivation

- Our admissions graph will look like this



Regression Discontinuity: Motivation

- Suppose we see that future earnings look like this. What can we infer?



Regression Discontinuity: Motivation

- Score and future income are strongly correlated (perfectly linear), but we see a jump at a score of 50 - exactly when students are accepted to college.
- We can make the inference that this is the causal impact of attending college on future earnings

Regression Discontinuity: Motivation

- We identified the causal effect using the graph before. How do we use a regression equation to solve this?
- Note that there's a linear function of score (We call our independent variable the "running variable" in RD designs) that needs to be modeled, and also a discontinuity, or jump:
- $y_i = \beta_0 + \beta_1(x_i - 50) + \beta_2(x_i > 50)$
- $x_i - 50$ is the distance from the cutoff point of 50
- $x_i > 50$ is a dummy variable equal to 1 if we are above the cutoff
- What is our coefficient of interest?

Regression Discontinuity: Motivation

- β_2 , since this is the jump that we see

Regression discontinuity: Identifying Assumptions

- No fine manipulation of our running variable (e.g. students can't take the ACT until they get a score above the cutoff)
- No other discontinuities at the cutoff that would affect our outcome (e.g. people just above the cutoff are wealthier than those just below)

RD Modifications

- There may be different slopes before and after the cutoff. How do we handle this?
- Just interact the running variable with the cutoff indicator:
 - $y_i = \beta_0 + \beta_1(x_i - c) + \beta_2(x_i \geq c) + \beta_3(x_i \geq c)(x_i - c)$
- Our coefficient of interest is still the jump: β_2
- In practice this is the only modification that is required, but more are available

RD Modifications: Nonlinearities vs Cutoff Window

- Choosing a narrow cutoff window (e.g. GPA from 1.9 to 2.1 instead of 0.0 to 4.0) is much cleaner, but gives less data
- Nonlinear functions can also be chosen (polynomials), but can lead to unstable estimates

How to test that RD is appropriate

- Histogram test: histogram of running variable should be smooth around the cutoff (a jump implies selection into treatment)
- Replace y in our RD regression with some covariate (like family income). These should give $\hat{\beta}_3 \approx 0$
- Permutation test: use a cutoff other than $x = c$. You should obtain a null result

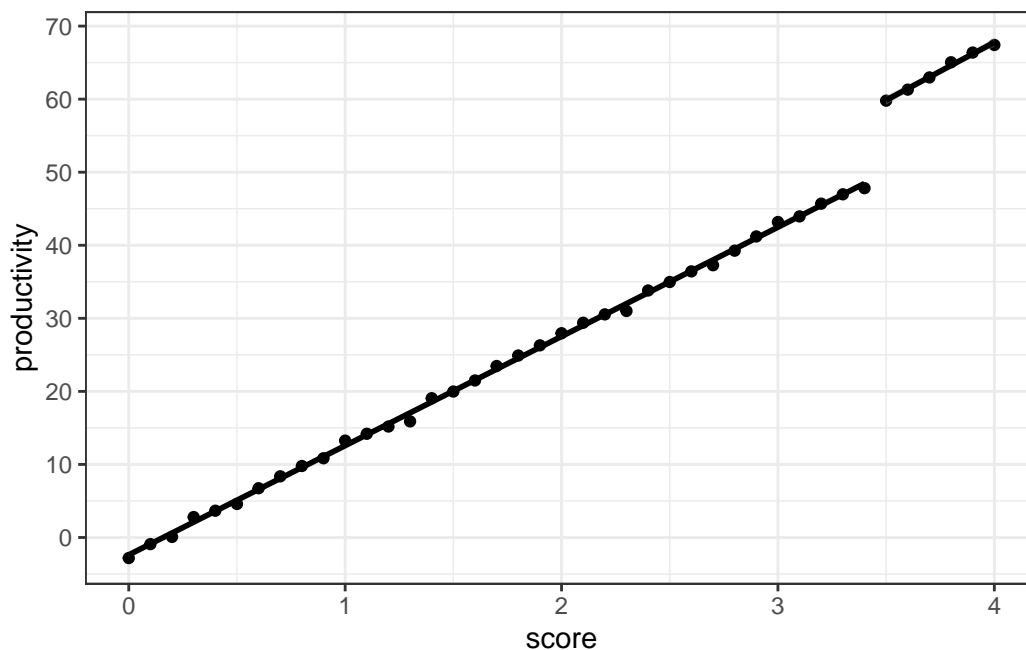
What does RD measure?

- RD measure a causal effect, but it's specifically the causal effect for individuals near the cutoff
- This is called a local average treatment effect (LATE)

Example: Graph

- Employees are rated on a 1-4 scale (Does not Meet Expectations-Exceeds Expectations) which are weighted across categories. Early career actuaries are promoted when they have a score of 3.5 or higher. You wish to study the effect of promotion on productivity and obtain the following graph. What is your estimated effect of promotion on productivity?

Example: Graph



Example: Graph

- Why might this be invalid? How might we check for this?

Graph Answer

- This jumps by 10, so this is our causal impact
 - (we have to extend out our regression line slightly - we jump from 50 to 60).
- There are a lot of reasons this might be invalid

- If scores were subjective and if people were put directly at a score of 3.5 based on non-performance reasons (e.g. they just interviewed at Google).
- You could test if there are jumps in other variables by replacing y, or you could check the density of the score histogram around the 3.5 cutoff.

Example: Table

- Suppose we have a policy kick in at $x=5$. From the table below, what is the causal impact of this policy for people near the cutoff?

	x	y
1:	1	5
2:	2	7
3:	3	9
4:	4	11
5:	5	15
6:	6	21
7:	7	27
8:	8	33

Table Answer

- On the left hand side this grows by 2 each time, yielding a predicted $y=13$ when $x=5$. On the right hand side our estimate grows by 6 each time, and in this case we predict that $y=15$ when $x=5$. Comparing the left hand estimate of 13 with the right hand estimate of 15, we get a causal impact of 2.

Table Answer

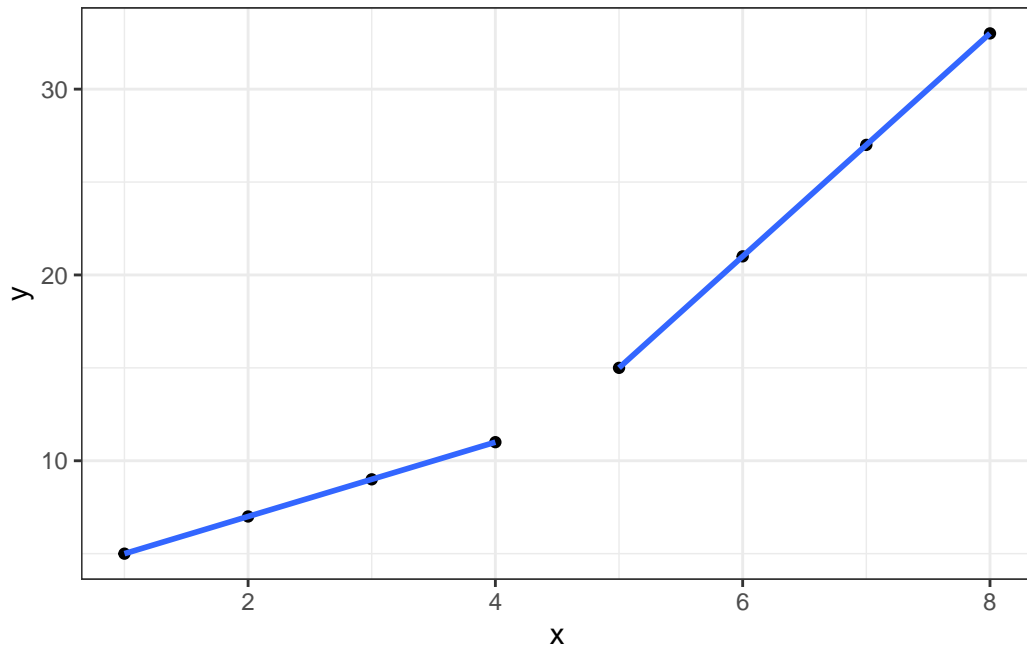
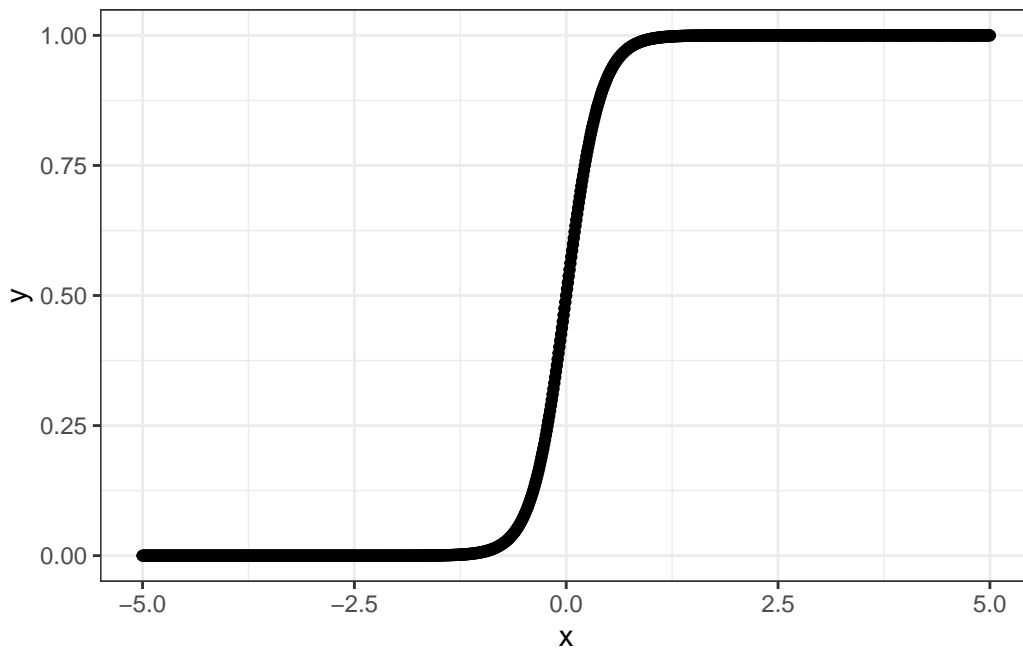


Table Answer

- Note that you could probably reasonably draw a smooth curve that fits these points, in which case RD would not be valid. The sigmoid plot below is a common function shape that becomes very sharp at the inflection point despite being perfectly smooth (it's infinitely differentiable)

Table Answer



Example: Setup

- You have the table from the previous question and wish to find the causal impact using a regression equation. What regression do you run and which coefficient(s) are your causal impact? (the cutoff is at $x \geq 5$)

	x	y
1:	1	5
2:	2	7
3:	3	9
4:	4	11
5:	5	15
6:	6	21
7:	7	27
8:	8	33

Setup Answer

- We need to first standardize the running variable so that the cutoff is at 0 by subtracting 5, then we run $y = \beta_0 + \beta_1 s + \beta_2 (s \geq 0) + \beta_3 s * (s \geq 0)$

Call:

```
lm(formula = y ~ run + ind + run * ind, data = dt)
```

Coefficients:

(Intercept)	run	indTRUE	run:indTRUE
13	2	2	4

Example: Design

- Suppose in 2020 India changed its corporate tax rate, but only for large companies (say, for firms with revenue above 1 billion dollars). You are interested in the effect of this tax change on capital expenditures. You decide to run a difference-in-difference design, but someone points out that you can also run a RD design off of this.
- Your difference in difference is $capex = \beta_0 + \beta_1 large + \beta_2 post + \beta_3 large * post$
- Your RD design is $capex = \beta_0 + \beta_1 * (size - 1000000000) + \beta_2 * large + \beta_3 * (size - 1000000000) * large$
- Since these both arrive at a causal estimate, you think you should arrive at the same answer, but your results differ. Why might this occur?

Example: Design

- A Difference-in-Differences may be biased if large and small companies have differential trends in their capital expenditures
- B Regression Discontinuity may be biased if firms manipulate their revenue to avoid the tax
- C Difference-in-differences captures an average treatment effect, while RD captures an average treatment effect for firms with a billion dollars in revenue
- D Differences-in-Differences may be biased if small and large companies have different levels of capital expenditures
- E Regression discontinuity may be biased if larger firms tend to have higher levels of capital expenditures

Design Answers

- A, B, and C are correct. DiD requires parallel pretrends, RD requires a lack of manipulation of the running variable, and in general RD measures a local average treatment effect of the individual near the cutoff.

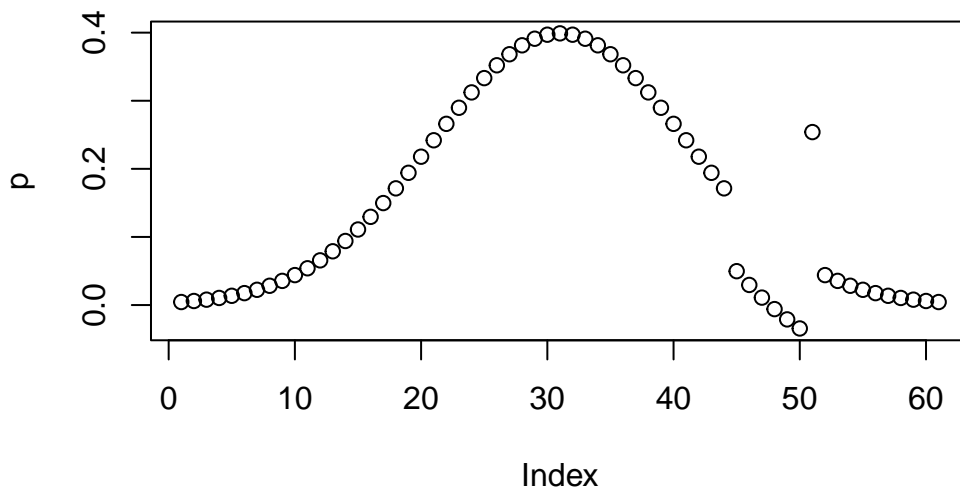
- Level differences do not bias DiD, and omitted variables in RD only matter to the extent that they jump discontinuously

Assumption 1 Violation

Suppose people with a score of 51 or higher get promoted and you obtain the following distribution of scores in a company. Is an RD design valid?

- A Yes, as long as you find a discontinuity in your outcome variable
- B No, because people with high scores tend to have different characteristics on average
- C Yes, since there is a jump in probability of attaining the cutoff score
- D No, the running variable is being manipulated

Assumption 1 Violation



Assumption 1 answer

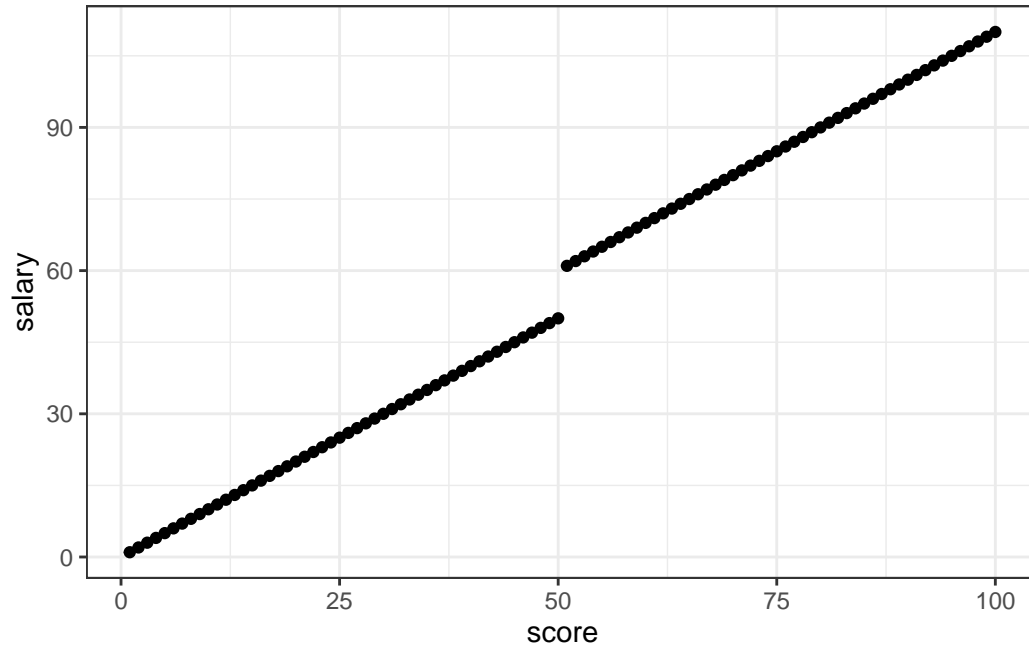
- D - this is a sign that the running variable was manipulation (it's not smooth at the cutoff - histogram test)

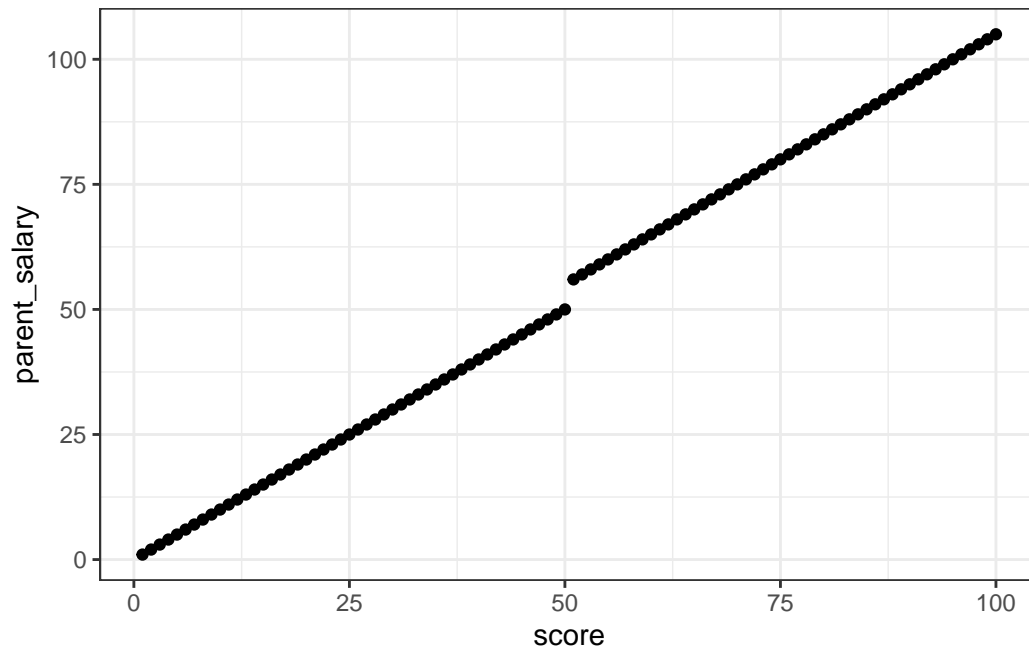
Assumption 2 violation

- You are interested in the effect of promotion on salary, and in this setting people with a score above 50 are promoted. You know that you can study multiple outcomes with the same RD setup, so you decide to study the effect of promotion on future salary as well as parent's salary. You obtain the following results. What does this imply?

- A Promotion causes both own and parents income to rise
- B Promotion causes income to rise, but parents income is probably biased and cannot be inferred to be causal
- C The result on parents income invalidates the entire design
- D Nothing can be said about the results

Assumption 2 violation





Assumption 2 Answer

- C - this is a strong sign of an omitted variable