

Advanced topics & overview

Francisco Villamil

Research Design for Social Sciences
MA Computational Social Science, UC3M
Fall 2023

Roadmap

Generalizing results

Robustness and inference tests

Mechanism test and additional implications

Causal methods in detail

Controlling and matching

Fixed effects

Difference-in-differences

Regression discontinuity

Instrumental variables

Generalizing results

- The Credibility Revolution in social sciences
- What are we really learning?
- References:
 - Egami and Hartman (2023) Elements of External Validity: Framework, Design, and Analysis. *APSR* 117(3): 1070–1088.
 - Munger (2023) Temporal validity as meta-science. *Res&Pol* 10(3).
 - See also this post: Generalizing Knowledge of Twitter to “X”.
 - Esterling, Brady, & Schwitzgebel (2023) The Necessity of Construct and External Validity for Generalized Causal Claims: A Critical Review of the Literature on Quantitative Causal Inference. Preprint.

Generalizing results

- **Construct validity** : are we really measuring what we intent to?
 - is the treatment doing what we theoretically expect it to? and are we measuring the outcome correctly?
- **External validity** : would we get the same results if we replicate this in another context?
 - especially: temporal and spatial validity

Roadmap

Generalizing results

Robustness and inference tests

Mechanism test and additional implications

Causal methods in detail

Controlling and matching

Fixed effects

Difference-in-differences

Regression discontinuity

Instrumental variables

Basic guidelines on robustness tests

Inference tests: placebo tests

- Reference:
 - AC Eggers, G Tuñón, A Dafoe (2023) **Placebo Tests for Causal Inference**. *American Journal of Political Science*, published online.

Roadmap

Generalizing results

Robustness and inference tests

Mechanism test and additional implications

Causal methods in detail

Controlling and matching

Fixed effects

Difference-in-differences

Regression discontinuity

Instrumental variables

Testing the mechanism

Additional implications of the theory

Roadmap

Generalizing results

Robustness and inference tests

Mechanism test and additional implications

Causal methods in detail

Controlling and matching

Fixed effects

Difference-in-differences

Regression discontinuity

Instrumental variables

Controlling in regression

- We probably already know this
- We've seen what it means controlling: just adjusting for the variation already predicted by the other variables
- One limitation though: we need to *observe* those variables

Controlling

- We are going to see two further methods of controlling widely used:
 - Matching
 - Fixed effects

Matching

- Adding control variables is not the only way to control / close back doors
- Imagine we have the following model:
 - $Z \rightarrow X \rightarrow Y \leftarrow Z$
 - Where Z is whether someone is retired or not
- If we select a sample of *only* retired people, we are closing that back door
 - $(X \leftarrow Z \rightarrow Y)$
- Matching is something like this, it's basically about creating groups of comparison where Z (which can be several variables) does not vary

How matching works

- We have a treatment group and a control group (so: *binary* treatment variables)
- The main idea: give different *weights* to treated and control observations, so we eliminate
- We get these weights by using one or more *matching variables* (i.e. confounding variables)

How matching works

- Imagine we have:
 - Treatment: get some specific skill training
 - Outcome: get a job afterwards
 - Confounding variable: gender
- Control group: 80 men and 20 women
 - 75% of men get a job, 60% of women do
- Treatment group: 500 men and 500 women
 - 70% of men get a job, 55% of women do
- Comparing within each group, we know the treatment effect is a 5% increase in the odds of getting a job ($70 \rightarrow 75$, $55 \rightarrow 60$)
- But if we do the global comparison, it's almost 10%
 - $60 \text{ men} + 12 \text{ women out of } 100 = 72/100 = 72\%$, vs $(350 + 275)/1000 = 62.5\%$, so a difference of 9.5 points

How matching works

- The problem is we have 4 times more men in the treated group than women, whereas we have equal proportion on the control group
- So we'll weight the control observations by gender, giving *more* weight to the men observations, so it looks more similar to the treatment group
- $(4 * 350 + 1 * 275) / (4 * 500 + 1 * 500) = 67\%$
- Now, the unweighted difference in the treatment group was 72%, and now the weighted difference in the control group is 67%
- The difference is 5 points, the same as the within-group calculation we did before

Two approaches to matching

1. Distance matching

- We want to create a dataset where treatment and control observations have similar values (distance) in the confounding variables
- If, say, our confounding variable is income, we'll pick control observations that have a similar value on income to each of the treatment observations
- <https://nickchk.com/causalgraphs.html>

Two approaches to matching

1. Distance matching

- We want to create a dataset where treatment and control observations have similar values (distance) in the confounding variables
- If, say, our confounding variable is income, we'll pick control observations that have a similar value on income to each of the treatment observations
- <https://nickchk.com/causalgraphs.html>

Two approaches to matching

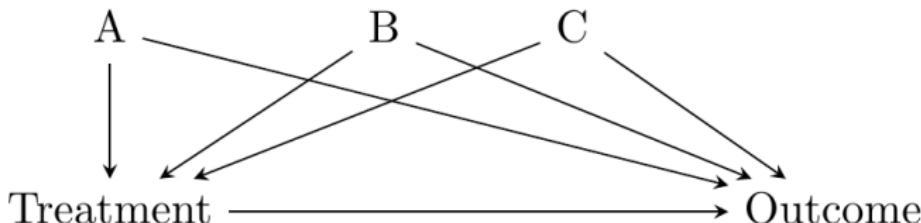
1. Distance matching

- We want to create a dataset where treatment and control observations have similar values (distance) in the confounding variables
- If, say, our confounding variable is income, we'll pick control observations that have a similar value on income to each of the treatment observations
- <https://nickchk.com/causalgraphs.html>

Two approaches to matching

1. Distance matching

- We want to create a dataset where treatment and control observations have similar values (distance) in the confounding variables
- If, say, our confounding variable is income, we'll pick control observations that have a similar value on income to each of the treatment observations
- <https://nickchk.com/causalgraphs.html>



Two approaches to matching

2. Propensity score matching

- We want to account for the differential likelihood in getting into treatment depending on the value of the confounding variables
- We estimate the probability of getting into treatment, usually by doing a regression where the outcome is the treatment and the right-hand variables are the confounders
- We control for the propensity score matching, or select based on it (or both)

Two approaches to matching

2. Propensity score matching

- We want to account for the differential likelihood in getting into treatment depending on the value of the confounding variables
- We estimate the probability of getting into treatment, usually by doing a regression where the outcome is the treatment and the right-hand variables are the confounders
- We control for the propensity score matching, or select based on it (or both)

Two approaches to matching

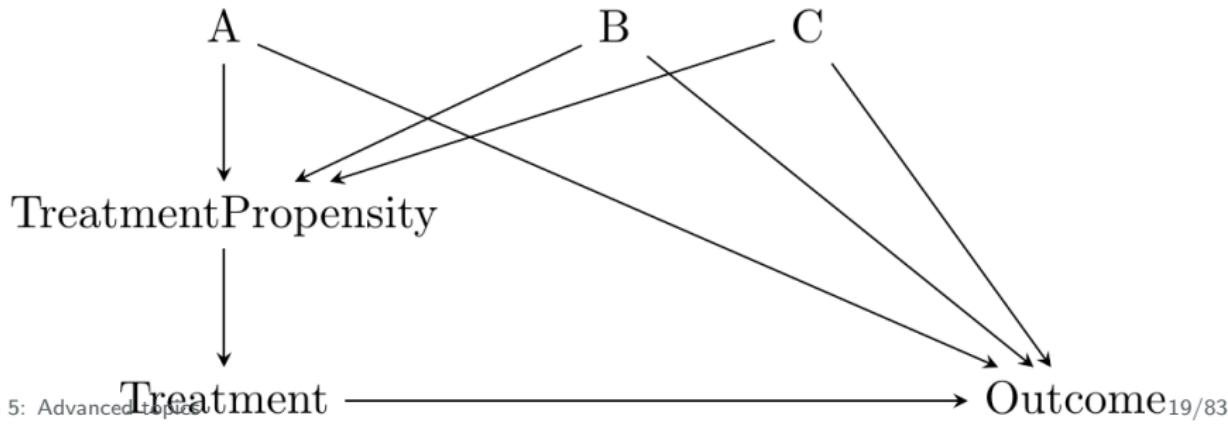
2. Propensity score matching

- We want to account for the differential likelihood in getting into treatment depending on the value of the confounding variables
- We estimate the probability of getting into treatment, usually by doing a regression where the outcome is the treatment and the right-hand variables are the confounders
- We control for the propensity score matching, or select based on it (or both)

Two approaches to matching

2. Propensity score matching

- We want to account for the differential likelihood in getting into treatment depending on the value of the confounding variables
- We estimate the probability of getting into treatment, usually by doing a regression where the outcome is the treatment and the right-hand variables are the confounders
- We control for the propensity score matching, or select based on it (or both)



Two approaches to matching: propensity score

```
df = data.frame(  
    treat = c(rep(1, 100), rep(0, 1000)),  
    gender = c(rep("M", 80), rep("F", 20), rep(c("M","F"), each = 500)),  
    y = NA  
)  
  
df$y[df$treat == 1 & df$gender == "M"] = rbinom(80, 1, 0.75)  
df$y[df$treat == 1 & df$gender == "F"] = rbinom(20, 1, 0.6)  
df$y[df$treat == 0 & df$gender == "M"] = rbinom(500, 1, 0.7)  
df$y[df$treat == 0 & df$gender == "F"] = rbinom(500, 1, 0.55)  
  
m1 = glm(y ~ treat, data = df)  
m2 = glm(y ~ treat + gender, data = df)  
modelsummary(list(m1, m2))  
  
ps = glm(treat ~ gender, data = df)  
df$propensity_score = predict(ps, newdata = df)  
m3 = glm(y ~ treat + propensity_score, data = df)  
modelsummary(list(m1, m2, m3))
```

Two approaches to matching: propensity score

	Model 1	Model 2	Model 3
(Intercept)	0.623	0.550	0.493
	(0.015)	(0.021)	(0.030)
treat	0.007	-0.037	-0.037
	(0.051)	(0.051)	(0.051)
genderM		0.147	
		(0.029)	
propensity_score			1.478
			(0.296)
Num.Obs.	1100	1100	1100

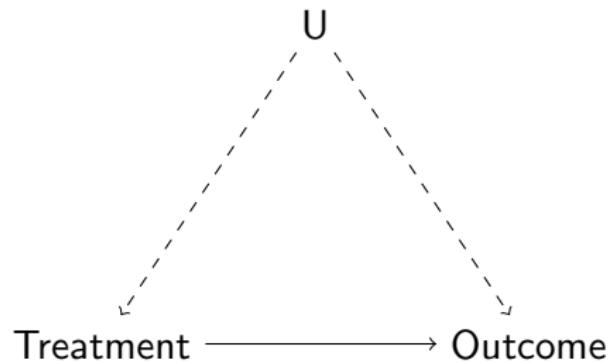
Matching vs regression

- Matching and regression are complementary approaches
- e.g. regression doesn't waste any information, but has a linearity assumption
- It's usual to use both at the same time

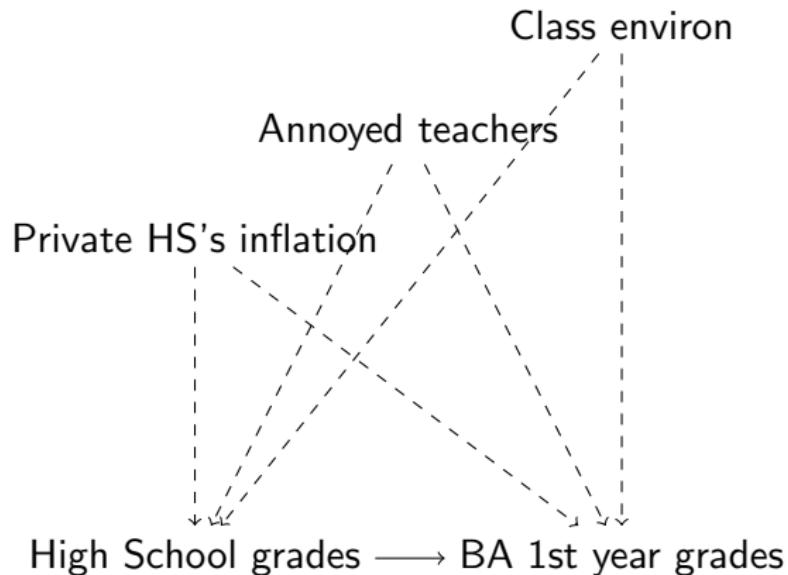
Fixed effects

- The problem with covariate adjustment, regardless of whether we use regression or matching, is that we need to *observe* those variables
- But another strategy when we have unobserved confounders is to try within-group comparisons, which will work when the unobserved variance is contact within some group
- For example, imagine cases when our U variable is:
 - *Country history*, in a cross-national analysis
 - *City of origin*, in an individual-level analysis
 - *Individual background*, in a panel survey analysis
 - *Company effects*, if we look at the effects of English courses on internal promotion using individual data from many different companies
 - etc

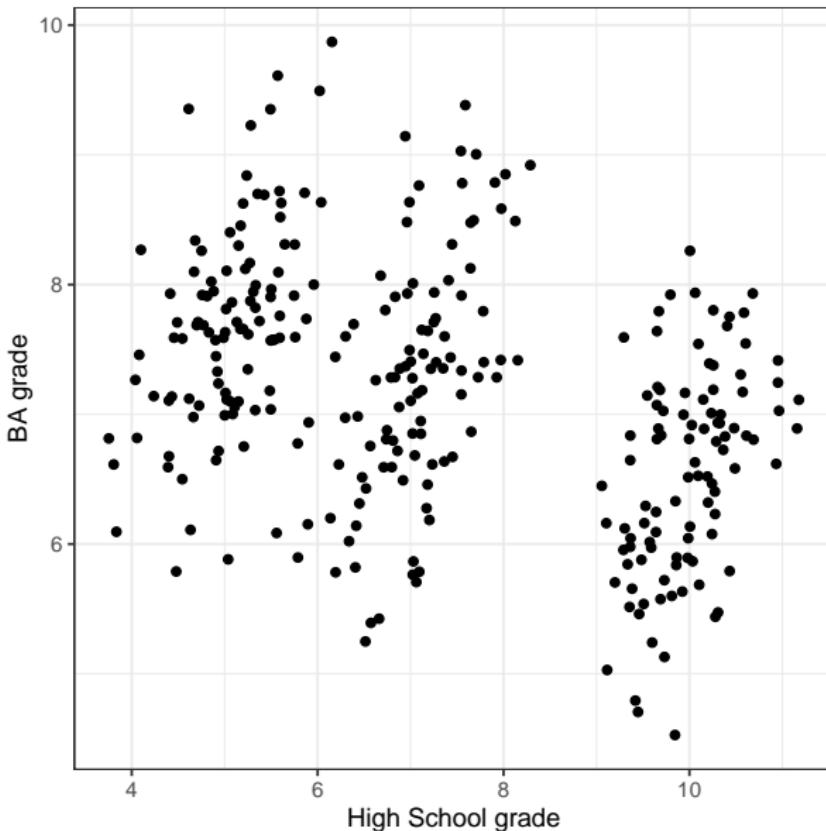
Fixed effects - when do we use them?



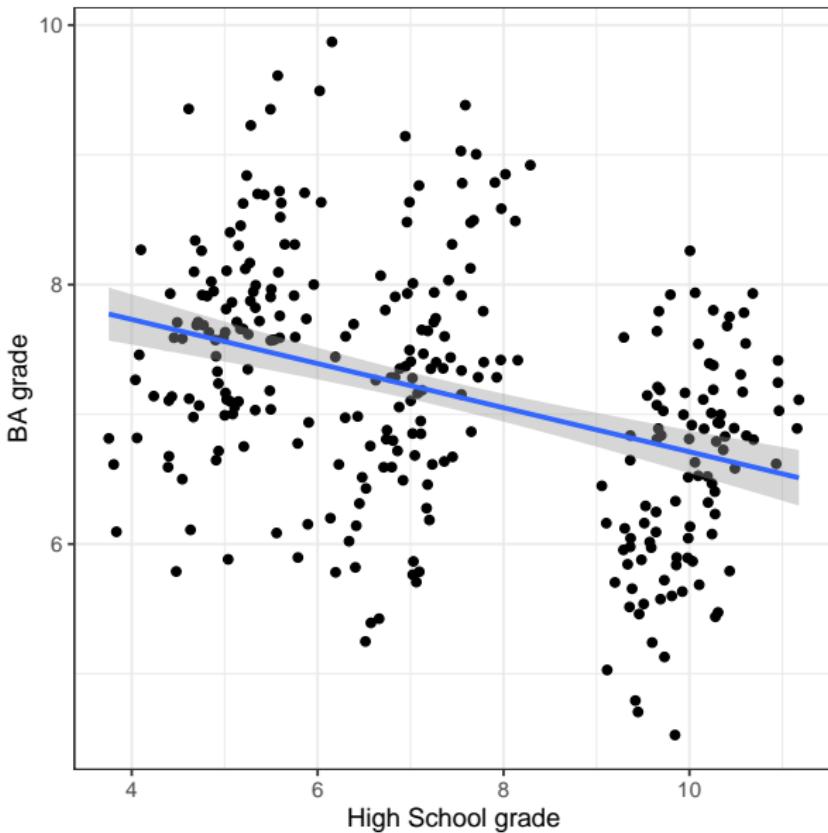
Fixed effects - when do we use them?



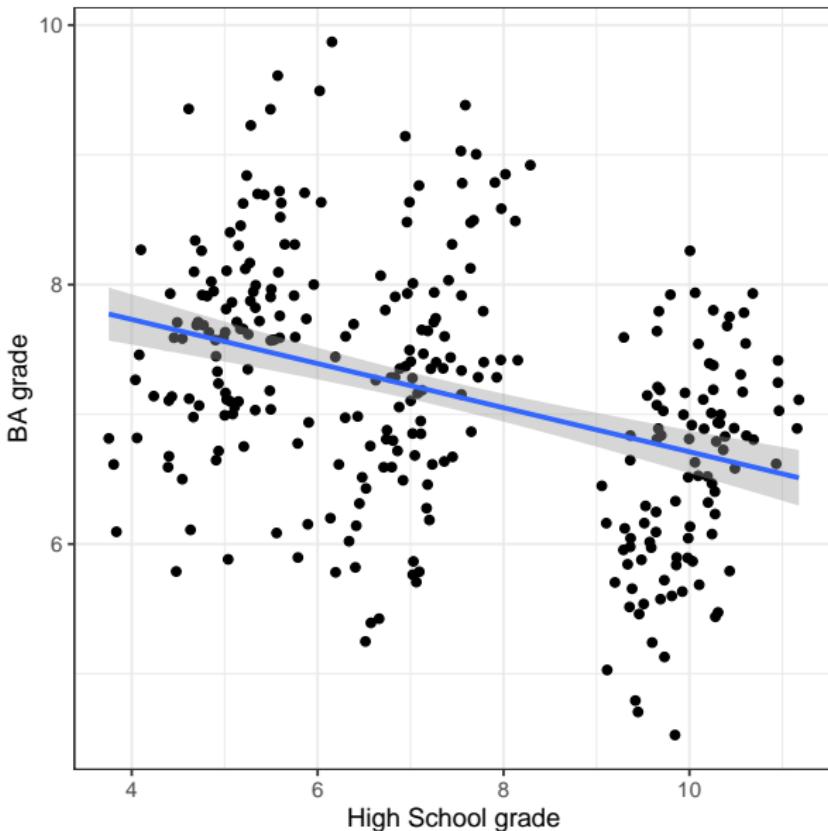
Fixed effects



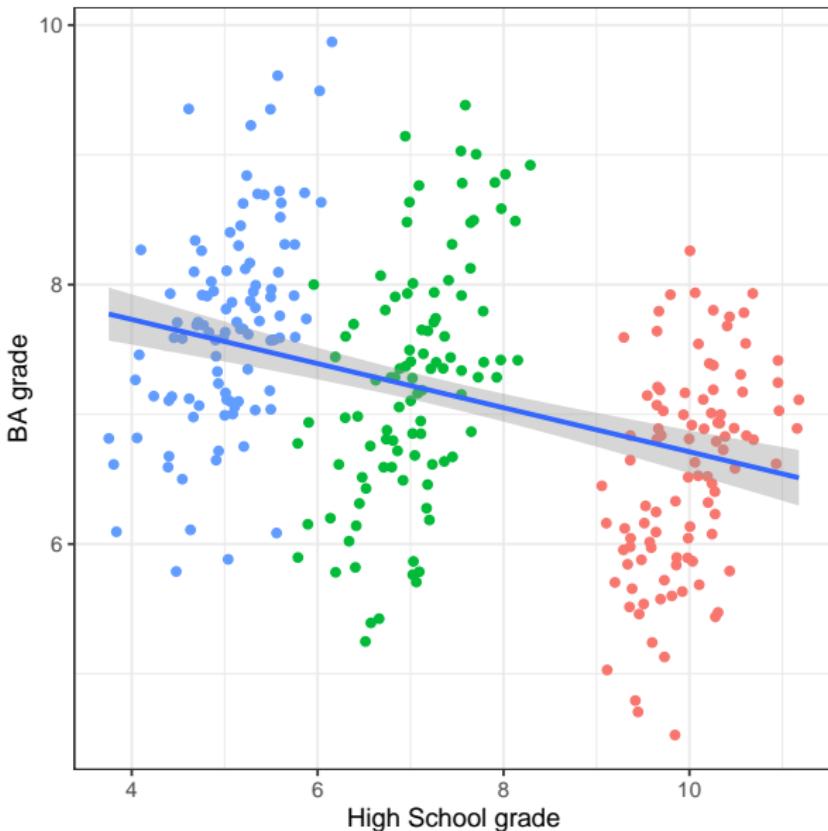
Fixed effects



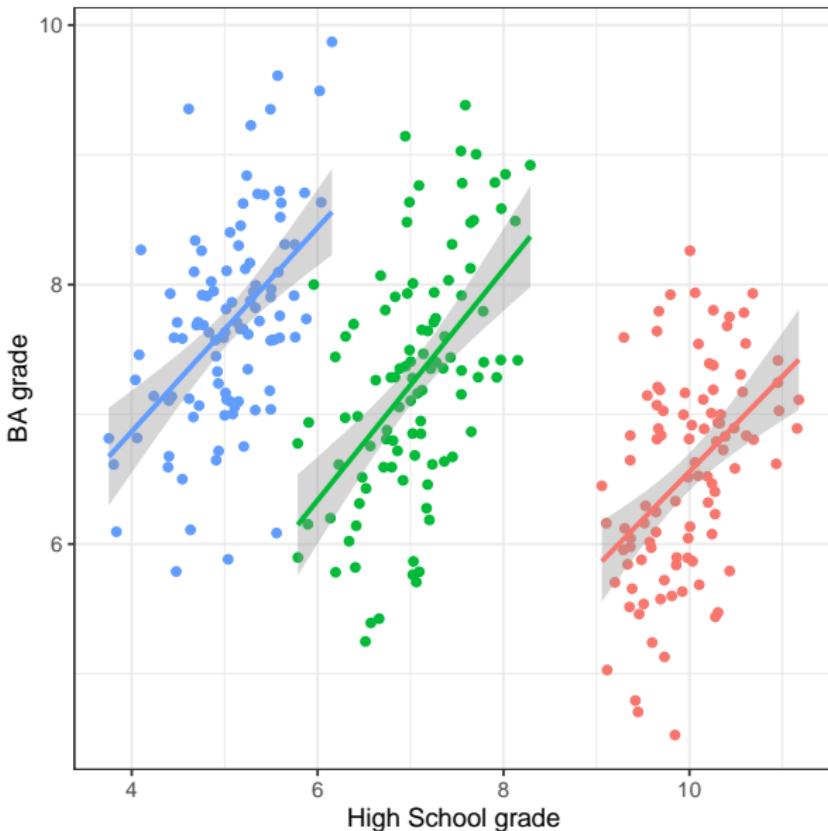
Fixed effects



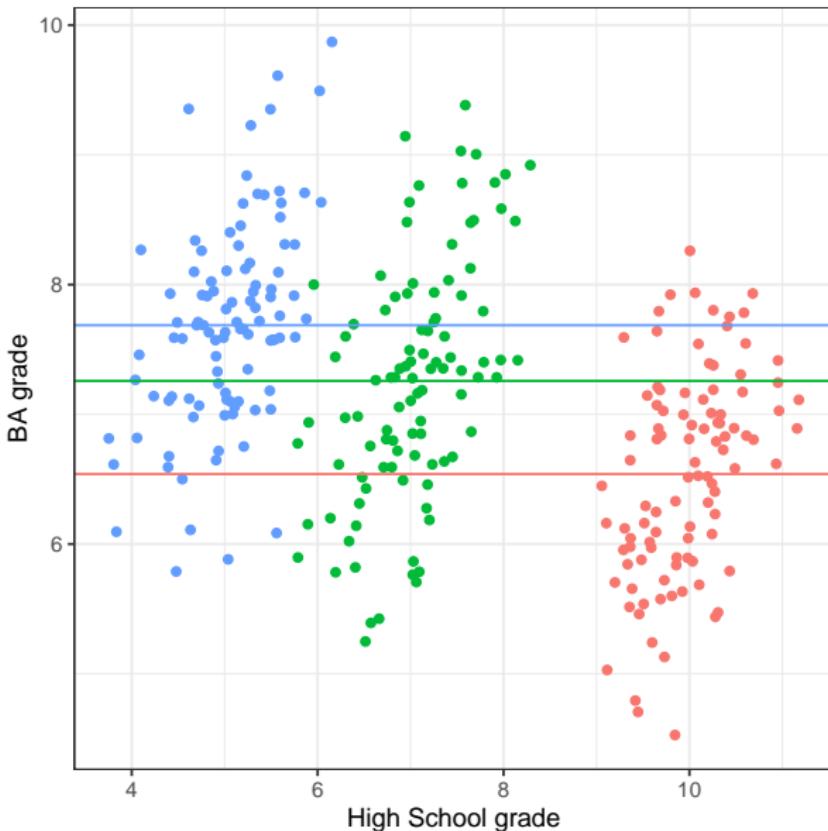
Fixed effects



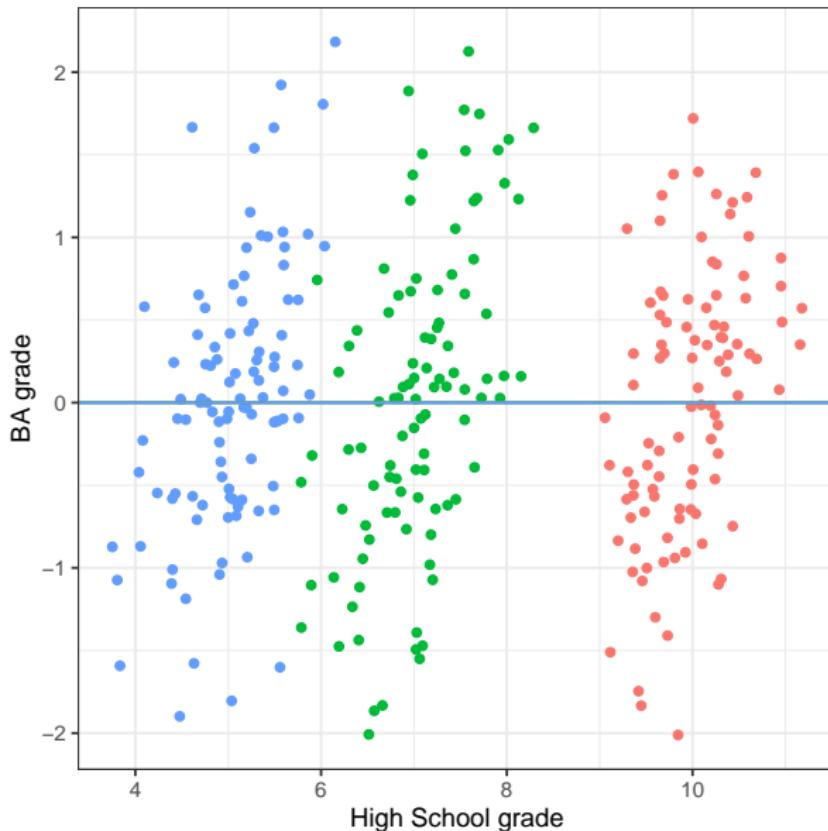
Fixed effects



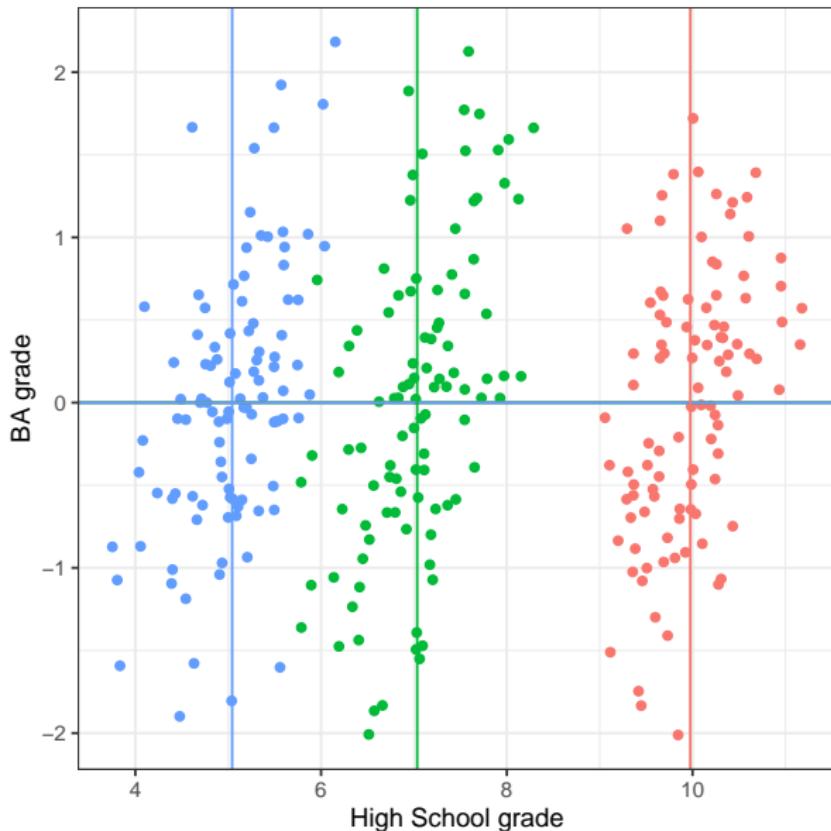
Fixed effects



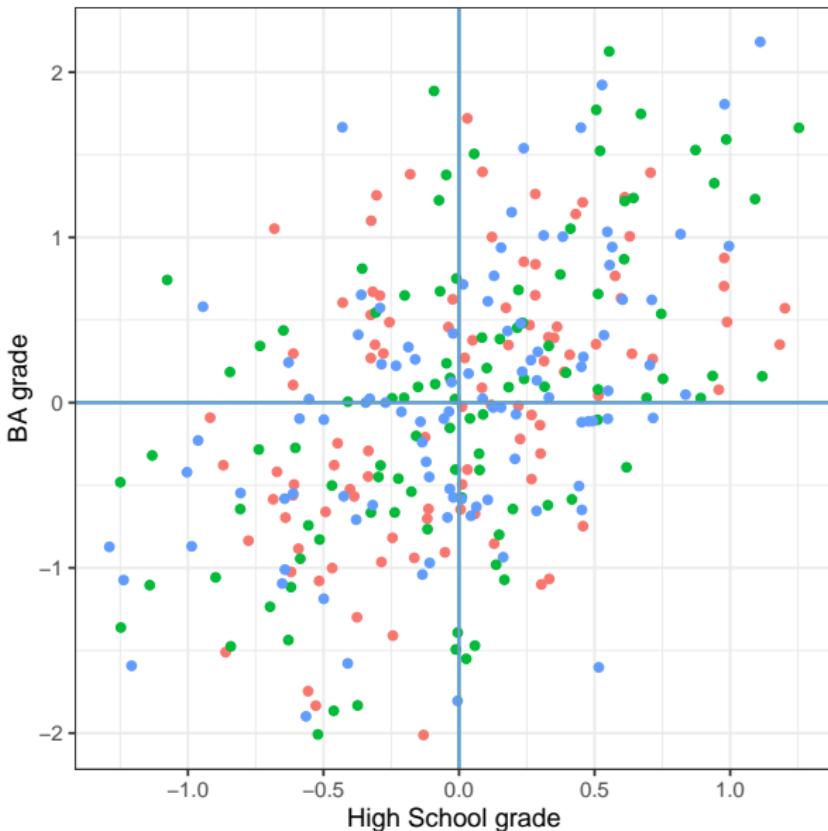
Fixed effects



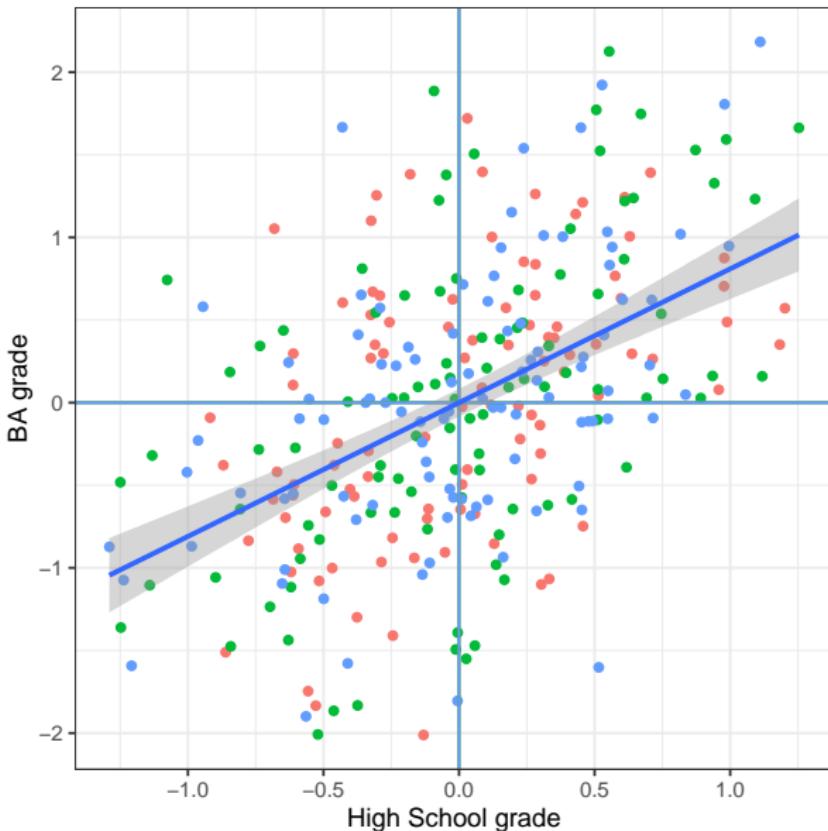
Fixed effects



Fixed effects



Fixed effects



Fixed effects and regression

```
# Simulate data: 300 students from three high schools,  
# normally distributed grades  
df = data.frame(high_school = rep(c("A", "B", "C"), each = 100)) %>%  
  group_by(high_school) %>%  
  mutate(hs_grade = rnorm(100, 7, 0.5)) %>%  
  ungroup()  
  
# But school A inflated grades a lot, and C was notoriously difficult  
df$hs_grade[df$high_school == "A"] = df$hs_grade[df$high_school == "A"] + 3  
df$hs_grade[df$high_school == "C"] = df$hs_grade[df$high_school == "C"] - 2  
  
# First-year BA grades are a function of the HS grade +/- the school inflation  
df$ba_grade = 2 + 0.75 * df$hs_grade + rnorm(300, 0, 0.75)  
df$ba_grade[df$high_school == "A"] = df$ba_grade[df$high_school == "A"] - 3  
df$ba_grade[df$high_school == "C"] = df$ba_grade[df$high_school == "C"] + 2
```

Our true causal model

Fixed effects and regression

Call:

```
lm(formula = ba_grade ~ hs_grade, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.42897	-0.57050	-0.05923	0.66227	2.76385

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.70808	0.19412	44.86 < 0.0000000000000002	***
hs_grade	-0.22059	0.02544	-8.67 0.0000000000000283	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9061 on 298 degrees of freedom

Multiple R-squared: 0.2014, Adjusted R-squared: 0.1988

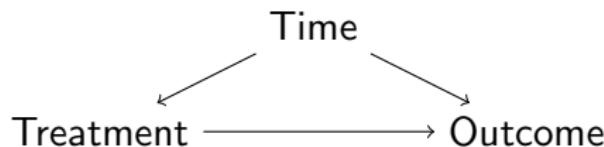
F-statistic: 75.17 on 1 and 298 DF, p-value: 0.00000000000002826

Fixed effects and regression

```
Call:  
lm(formula = ba_grade ~ hs_grade + factor(high_school), data = df)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-2.0124 -0.5464  0.0042  0.4984  2.2539  
  
Coefficients:  
              Estimate Std. Error t value          Pr(>|t|)  
(Intercept) -1.68390   0.98549 -1.709          0.0886 .  
hs_grade      0.80803   0.09871  8.186 0.00000000000000811 ***  
factor(high_school)B 3.23797   0.30984 10.451 < 0.0000000000000002 ***  
factor(high_school)C 5.26454   0.49388 10.659 < 0.0000000000000002 ***  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 0.7719 on 296 degrees of freedom  
Multiple R-squared:  0.4243, Adjusted R-squared:  0.4185  
F-statistic: 72.73 on 3 and 296 DF,  p-value: < 0.0000000000000022
```

Difference-in-differences

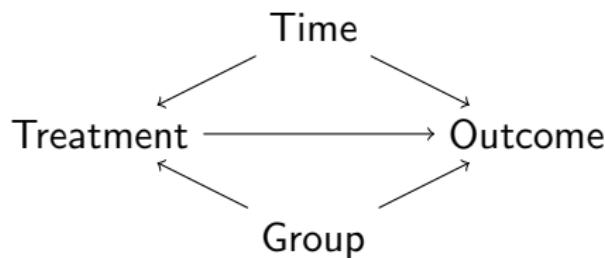
- Treatments usually occur at a particular moment in time, e.g.:
 - Minimum wage increase
 - Terrorist attack
 - Influx of refugees
 - ...
- In those cases, if we have before & after observations, we have something like this:



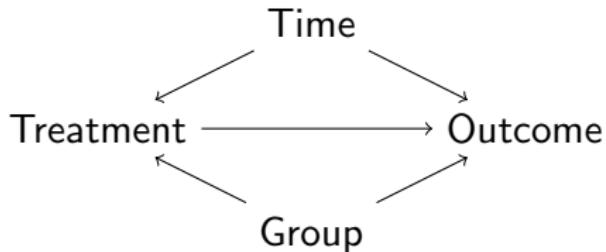
- The **problem** is that **all** the variation goes through time, so if we close that back door, we're left with nothing

Difference-in-differences

- So one strategy we can use is to bring additional group that is *not treated* and for which we also have before/after observations
 - Minimum wage increase: maybe those earning above MW?
 - Influx of refugees: other countries? regions far from the border?
 - Terrorist attack: do we have a control (untreated) group?
 - ...

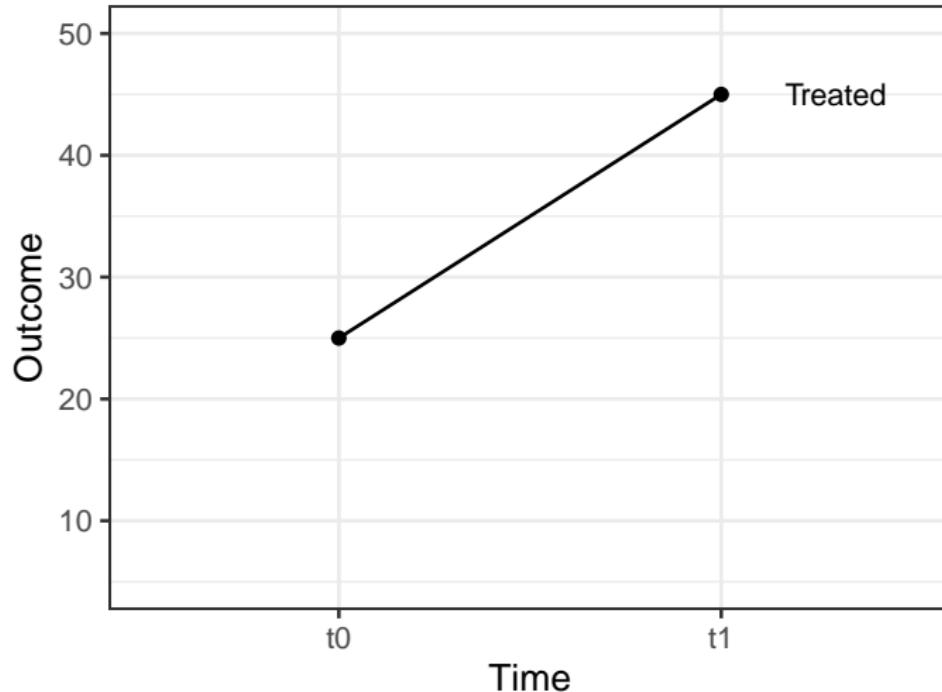


Difference-in-differences

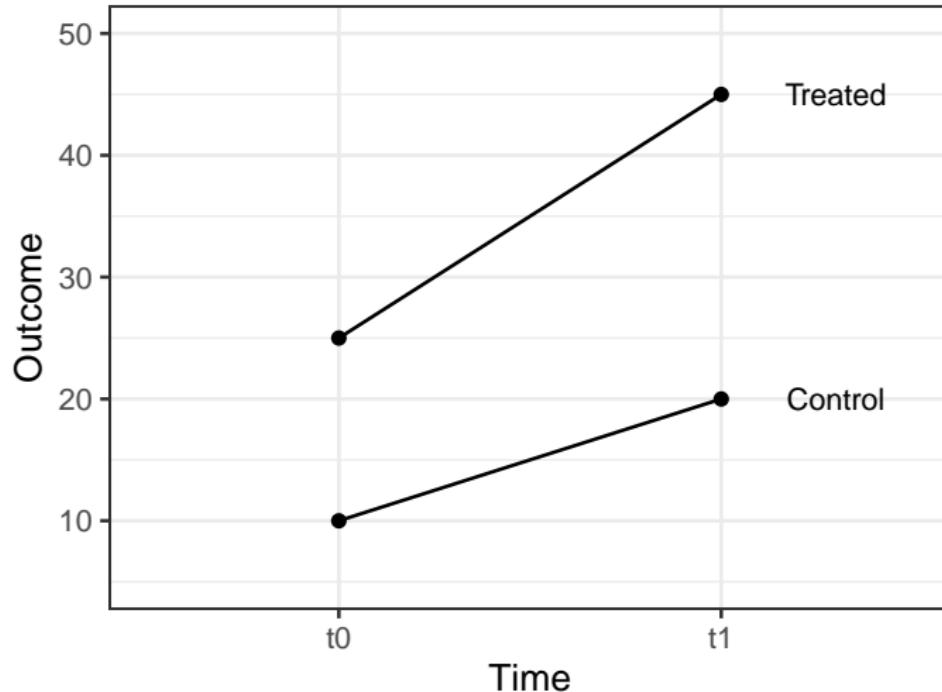


- We can compare changes across time *within* the treated and control groups (closing the back door through group)
- Compare within-group variation between treated and control (since time affects both ‘within-variations’ the same way, we are closing the other back door through time)

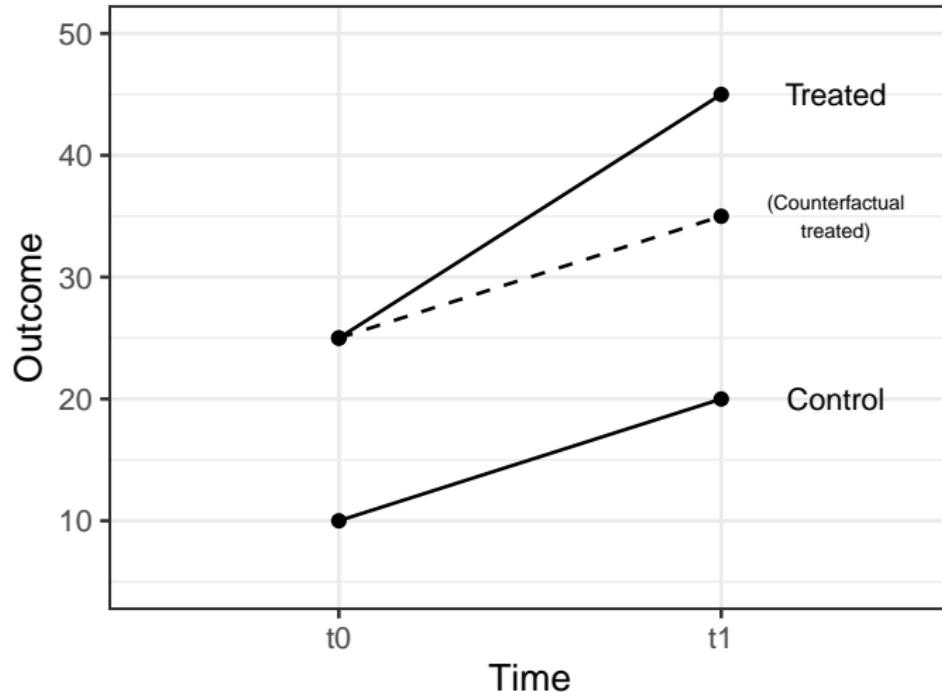
Difference-in-differences



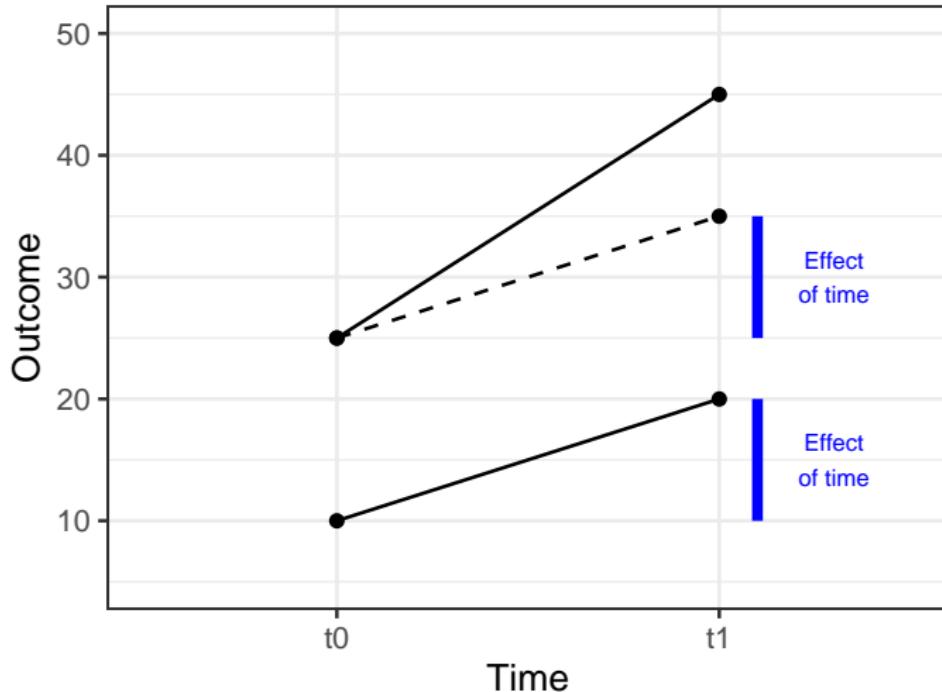
Difference-in-differences



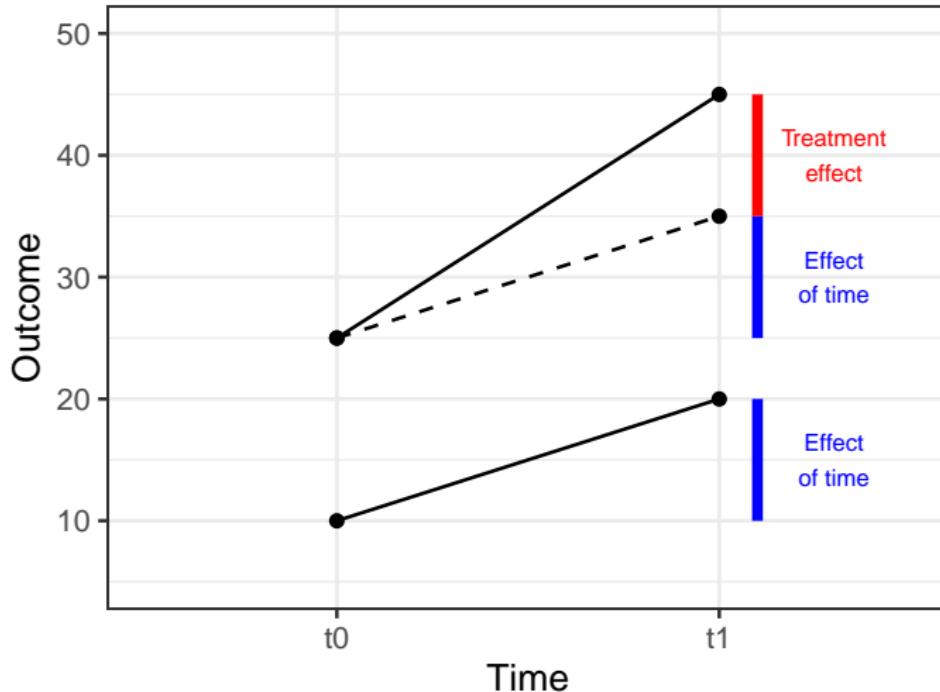
Difference-in-differences



Difference-in-differences



Difference-in-differences



Difference-in-differences: Cholera in London



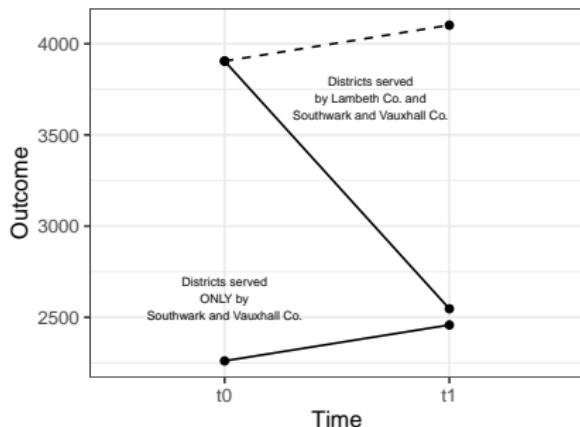
- Probably first use of DiD and natural experiments?
- Lambeth Company moved water intake upriver in 1852, Southwark & Vauxhall Company still got it from downstream

Difference-in-differences: Cholera in London

TABLE XII.

2261	2458
3905	2547
162	37

Southwk. & Vauxhall.
 Both Companies.
 Lambeth Company.



Sub-Districts.	Deaths from Cholera in 1849.	Deaths from Cholera in 1854.	Water Supply.
St. Saviour, Southwark .	283	371	
St. Olave .	157	161	
St. John, Horsleydown .	192	148	
St. James, Bermondsey .	949	363	
St. Mary Magdalene .	259	244	
Leather Market .	226	237	
Rotherhithe* .	352	282	
Wandsworth .	97	59	
Battersea .	111	171	
Putney .	8	9	
Camberwell .	935	240	
Peckham .	92	174	
Christchurch, Southwark	256	113	
Kent Road .	267	174	
Borough Road .	312	270	
London Road .	257	93	
Trinity, Newington .	318	210	
St. Peter, Walworth .	446	388	
St. Mary, Newington .	143	92	
Waterloo Road (1st)	193	58	
Waterloo Road (3rd)	243	117	
Lambeth Church (1st)	915	49	
Lambeth Church (2nd)	544	193	
Kennington (1st)	187	303	
Kennington (2nd)	153	142	
Brixton .	81	48	
Clapham .	114	165	
St. George, Camberwell	176	133	
Norwood .	9	10	
Streatham .	154	15	Lambeth Company only.
Dulwich .	1	—	
Sydenham .	5	12	
First 12 sub-districts .	2261	2458	Southwk. & Vauxhall.
Next 16 sub-districts .	3905	2547	Both Companies.
Remaining 20 sub-districts .			

Difference-in-differences

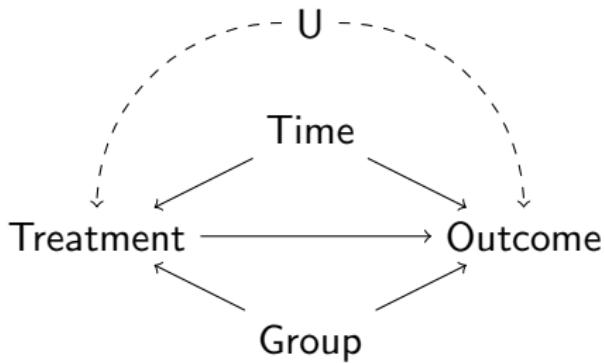
- You can estimate this effect just with group means
- But it is often easier to use regression, also because you can include controls:

$$Y_{it} = \beta_0 + \beta_1 Treated_i + \beta_2 After_t + \beta_3 (Treated_i \times After_t) + \beta^\top x_i + \epsilon_{it} \quad (1)$$

- But why do we need all this?

Difference-in-differences

- Because DiD identification **depends** on the assumption that the **control group is a good counterfactual to the treated group**

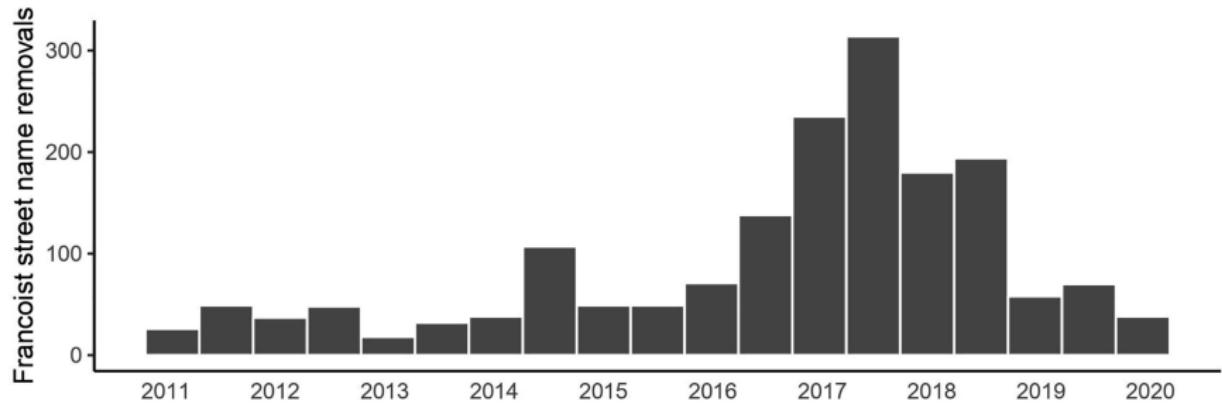


- One way to test this is checking if the **parallel trends assumption** holds (we need data further back in time)

DiD example

- What is the effect of symbolic TJ policies?
- journals.sagepub.com/doi/full/10.1177/20531680211058550

DiD example



DiD example

Francoist names	Removed Francoist names, 2016–2018?	
In June 2016?	No	Yes
No	6455 (100%)	0 (0%)
	1184 (72%)	454 (28%)

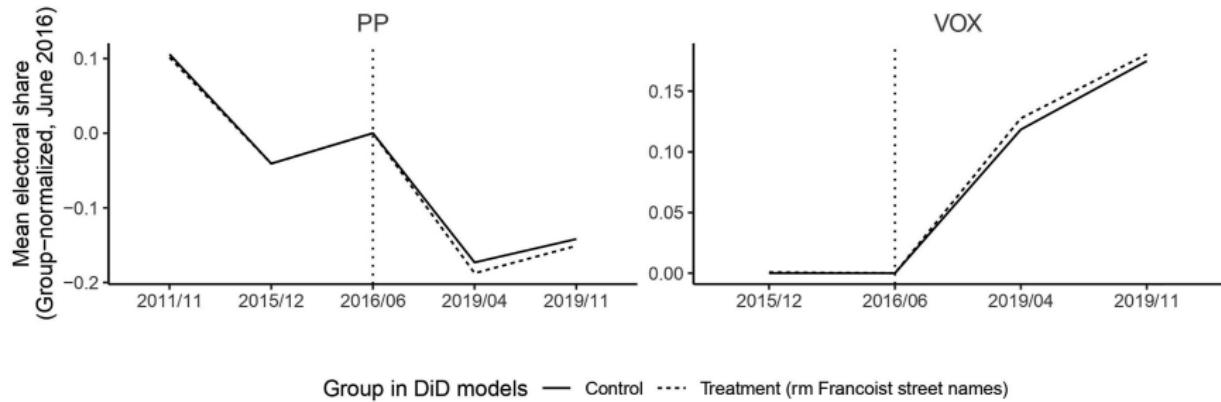
Note: Row percentages. Changes in 2016–2018 refer to the period between 30/06/2016 and 31/12/2018.

DiD example

Table 2. Mean electoral share in sample.

Party	June 2016			April 2019			
	<i>Control</i>	<i>Treated</i>	Δ	<i>Control</i>	<i>Treated</i>	Δ	$\Delta_{2019} - \Delta_{2016}$
Vox	0.21	0.21	0	12.54	13.28	0.74	0.74
PP	41.22	46.77	5.55	23.83	27.68	3.85	-1.7
PSOE	29.13	28.01	-1.12	33.38	32.03	-1.35	-0.23

DiD example



Regression discontinuity

The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi

DANIEL N. POSNER *University of California, Los Angeles*

This paper explores the conditions under which cultural cleavages become politically salient. It does so by taking advantage of the natural experiment afforded by the division of the Chewa and Tumbuka peoples by the border between Zambia and Malawi. I document that, while the objective cultural differences between Chewas and Tumbukas on both sides of the border are identical, the political salience of the division between these communities is altogether different. I argue that this difference stems from the different sizes of the Chewa and Tumbuka communities in each country relative to each country's national political arena. In Malawi, Chewas and Tumbukas are each large groups vis-à-vis the country as a whole and, thus, serve as viable bases for political coalition-building. In Zambia, Chewas and Tumbukas are small relative to the country as a whole and, thus, not useful to mobilize as bases of political support. The analysis suggests that the political salience of a cultural cleavage depends not on the nature of the cleavage itself (since it is identical in both countries) but on the sizes of the groups it defines and whether or not they will be useful vehicles for political competition.

Regression discontinuity

FIGURE 1. Research Sites

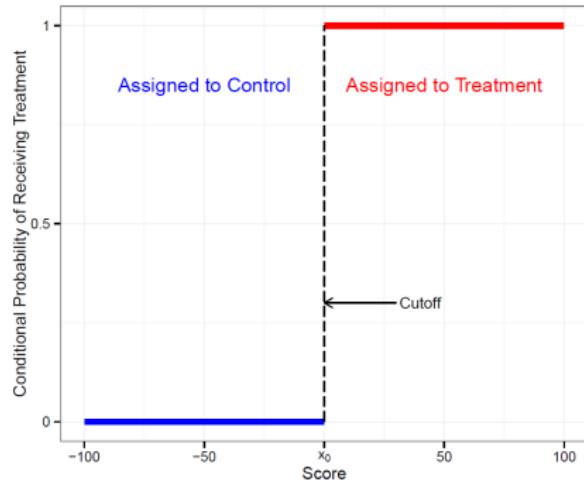


Regression discontinuity

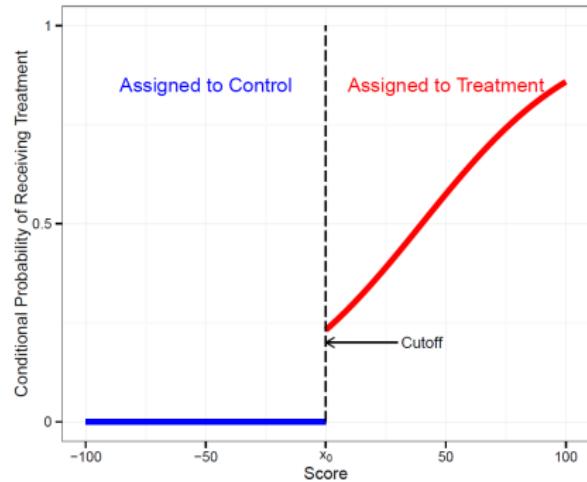
- RDD works well when assignment into treatment depends on a cutoff along a **running variable**
 - Do incumbent politicians have an electoral advantage? (vote share)
 - What is the effect of being drafted into the military? (birth year)
 - Effect of national policies in ethnic identification in Africa? (distance to colonial borders)
- This is the source of the exogenous variation (or if you will, the natural experiment):
 - Although many variables confound the relationships between X and Y , nothing should be too different *around the cutoff* between treatment and control groups (local randomization assumption)
 - Sometimes you look at different *bandwidths* to check this

RDD

Figure 1: Conditional Probability of Receiving Treatment in Sharp vs. Fuzzy RD Designs



(a) Sharp RD

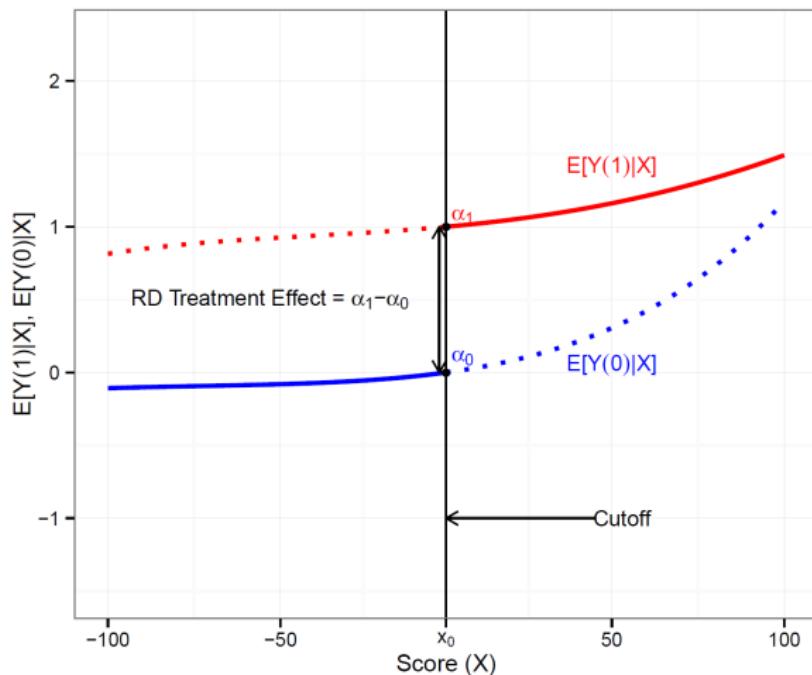


(b) Fuzzy RD (One-Sided)

<https://bookdown.org/paul/applied-causal-analysis/rddbasics2.html>

RDD

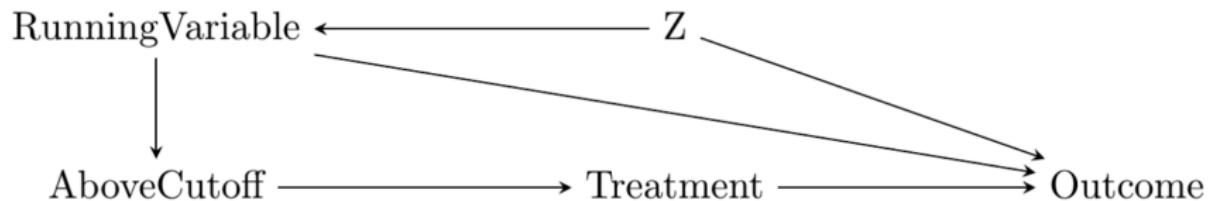
Figure 2: RD Treatment Effect in Sharp RD Design



<https://bookdown.org/paul/applied-causal-analysis/rddbasics3.html>

RDD

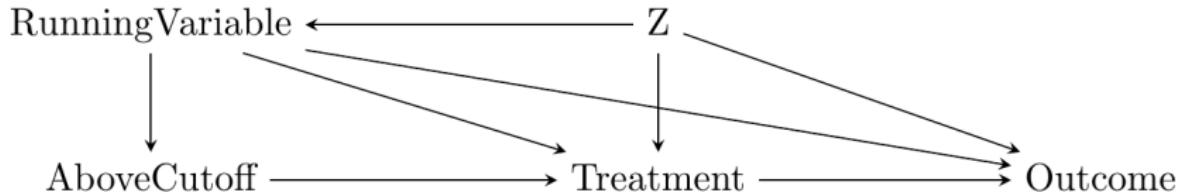
- Underlying assumption: other confounders also vary along the running variable, but are independent to the *jump*



Huntington-Klein, *The Effect*, p.508

RDD

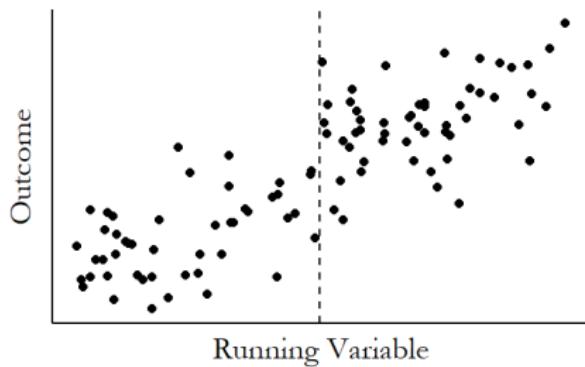
- Underlying assumption: other confounders also vary along the running variable, but are independent to the *jump*
- Even if in a *fuzzy* design



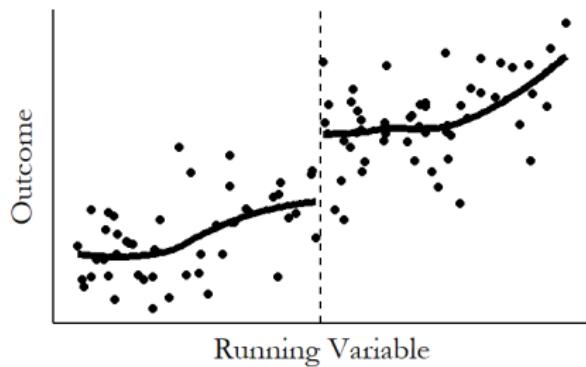
Huntington-Klein, *The Effect*, p.508

RDD implementation

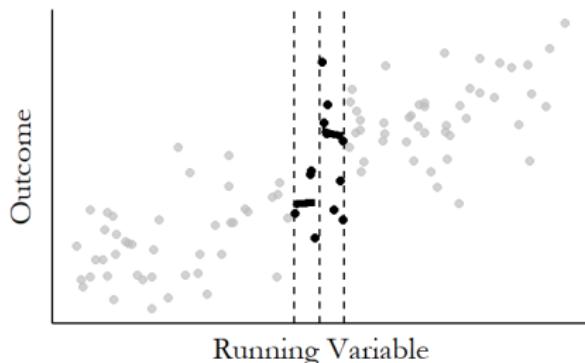
(a) Raw Data



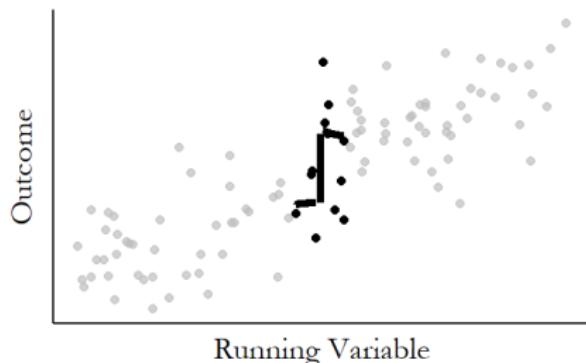
(b) Predict Values Near the Cutoff



(c) Pick a Bandwidth

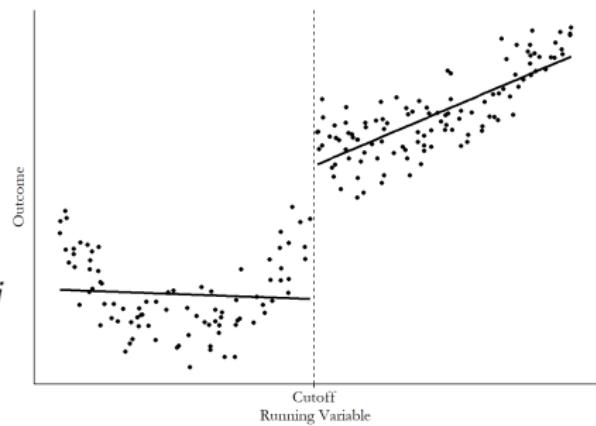


(d) Estimate Jump at the Cutoff



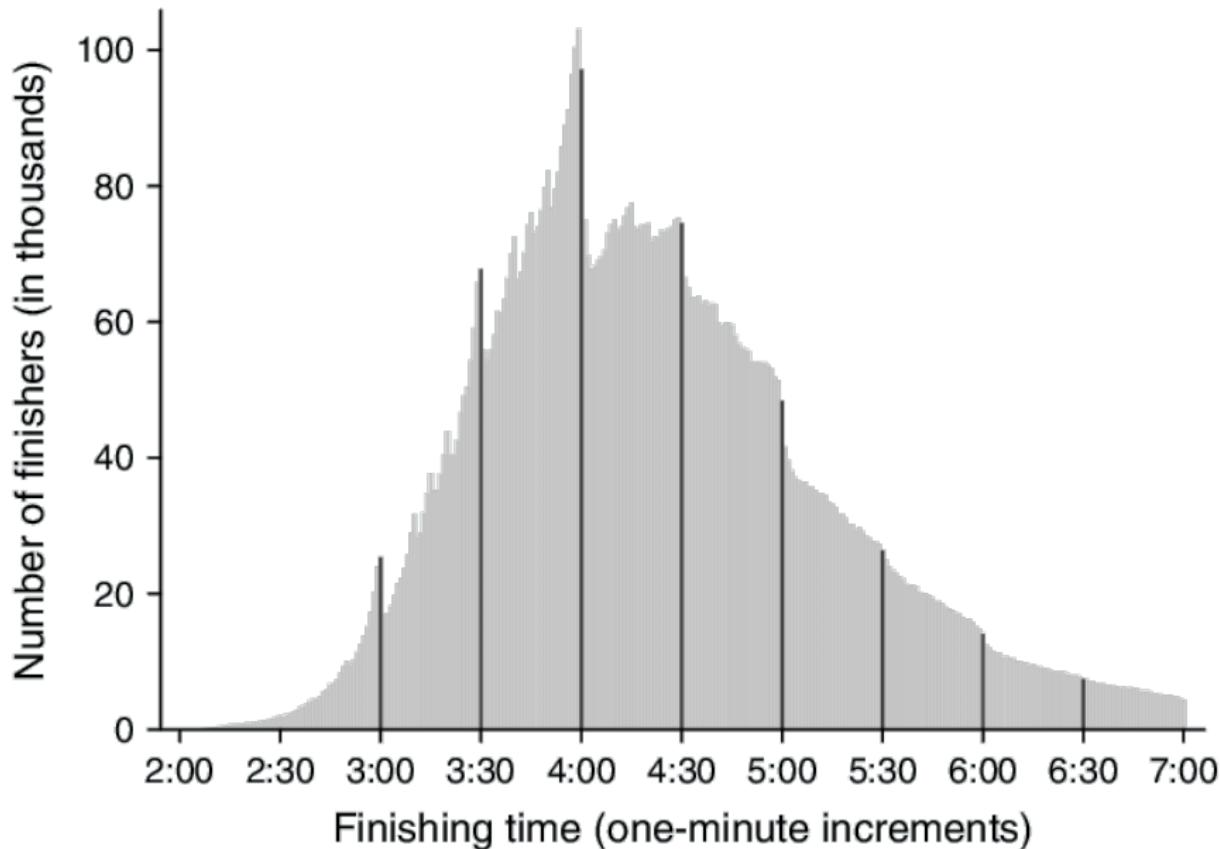
RDD and regression

$$Y = \beta_0 + \beta_1 Distance + \\ \beta_2 Treated + \\ \beta_3 (Treated \times Distance) + \beta^T x_i \quad (2)$$



- But there's no need to use linear regression, other methods available as well

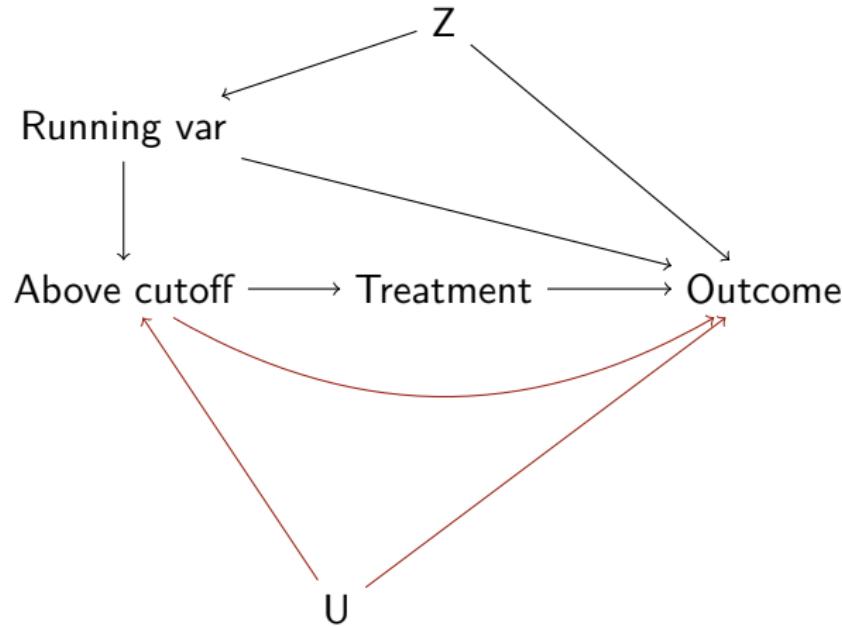
Threats to RDD: precise sorting



Threats to RDD: cutoff ← outcome



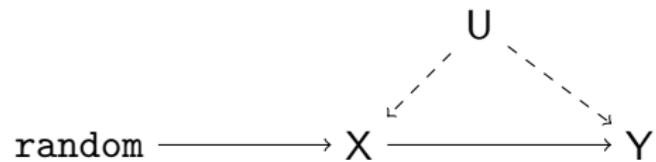
Threats to RDD



Some extensions & combinations

- DiD with multiple treatment periods (units being treated at different times)
- Matched DiD
- Difference-in-discontinuities

Instrumental variables



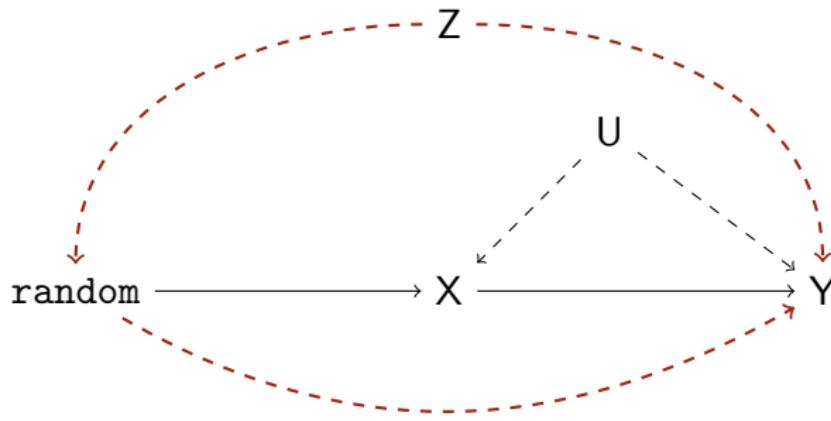
Instrumental variables

- Find an exogenous source of variation in the treatment variable
- Isolate that variation and use it to identify the causal effect

Instrumental variables

- Find an exogenous source of variation in the treatment variable
- Isolate that variation and use it to identify the causal effect
- Assumptions:
- **Relevance:** the instrument explains at least some part of the treatment variable
- **Validity or exclusion restriction:** no back door paths between the instrument and the outcome

Instrumental variables



IV threats

Economic Shocks and Civil Conflict: An Instrumental Variables Approach

Edward Miguel

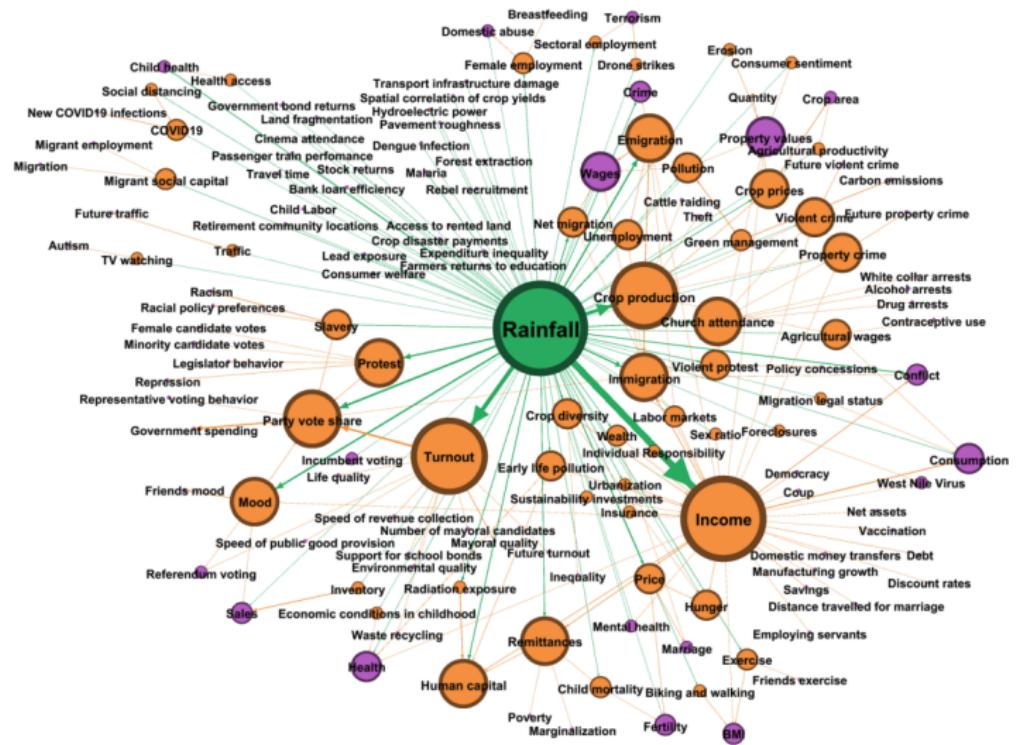
University of California, Berkeley and National Bureau of Economic Research

Shanker Satyanath and Ernest Sergenti

New York University

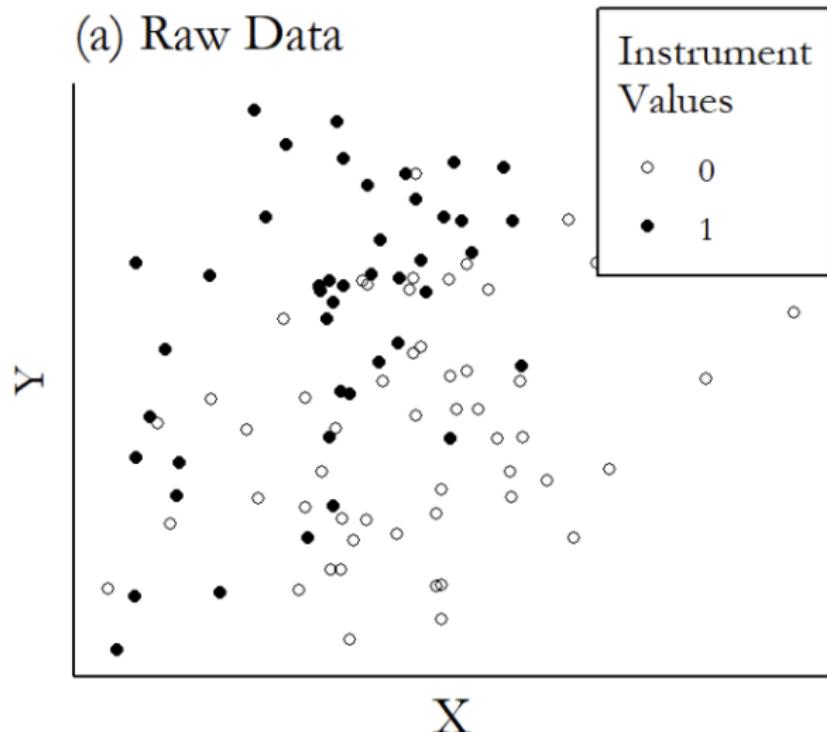
Estimating the impact of economic conditions on the likelihood of civil conflict is difficult because of endogeneity and omitted variable bias. We use rainfall variation as an instrumental variable for economic growth in 41 African countries during 1981–99. Growth is strongly negatively related to civil conflict: a negative growth shock of five percentage points increases the likelihood of conflict by one-half the following year. We attempt to rule out other channels through which rainfall may affect conflict. Surprisingly, the impact of growth shocks

IV threats



Jonathan Mellon (2022) Rain, Rain, Go Away: 192 Potential Exclusion-Restriction

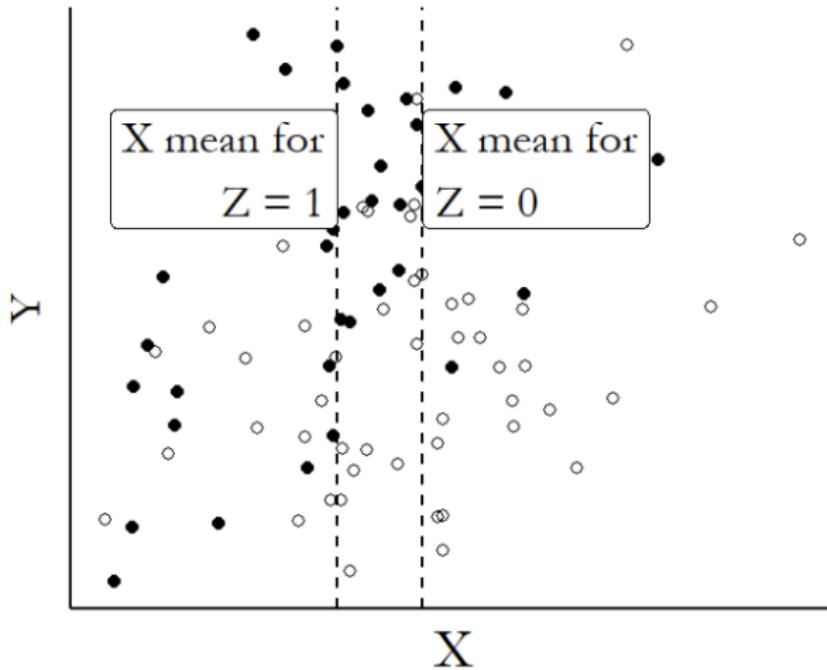
How does IV work?



(Huntington-Klein, *The Effect*, p 472)

How does IV work?

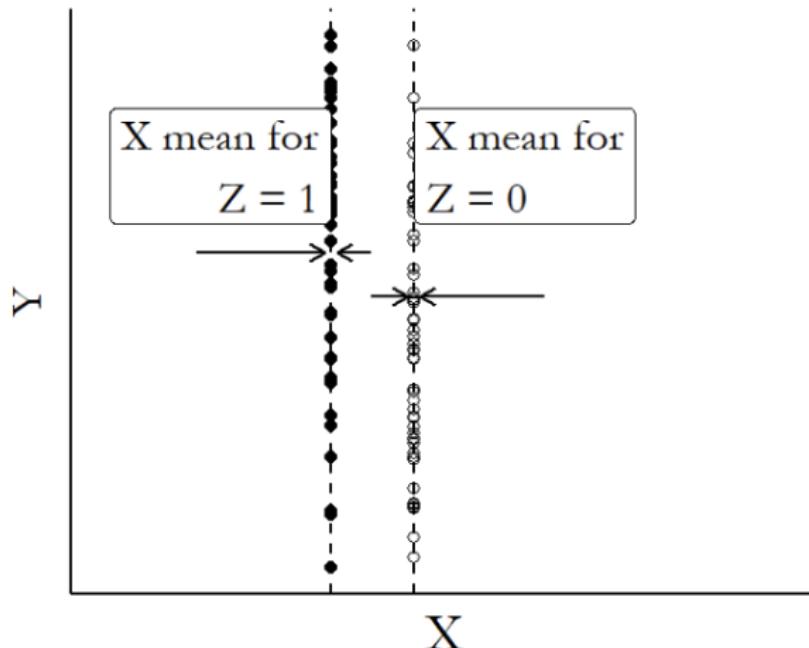
(b) Predict X with Z



(Huntington-Klein, *The Effect*, p 472)

How does IV work?

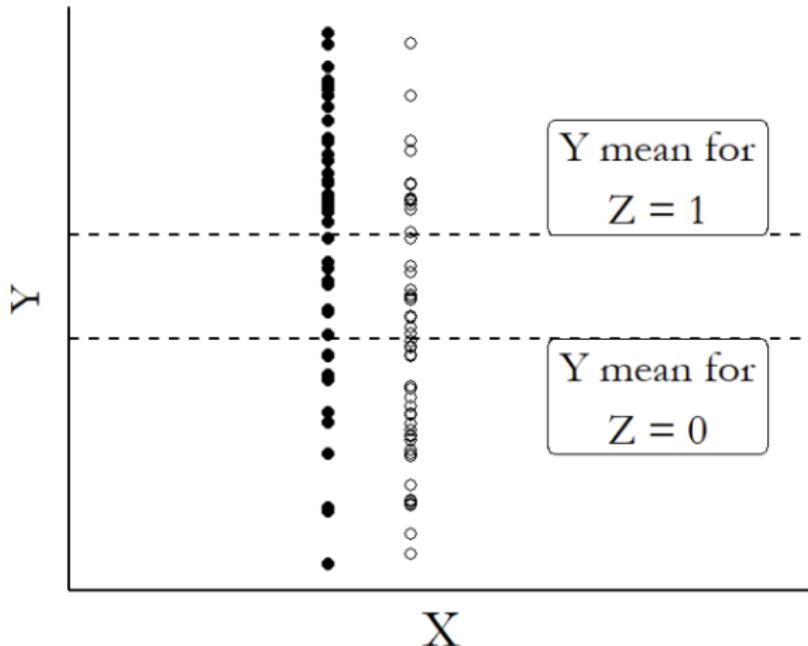
(c) Only Use Predicted Variation



(Huntington-Klein, *The Effect*, p 472)

How does IV work?

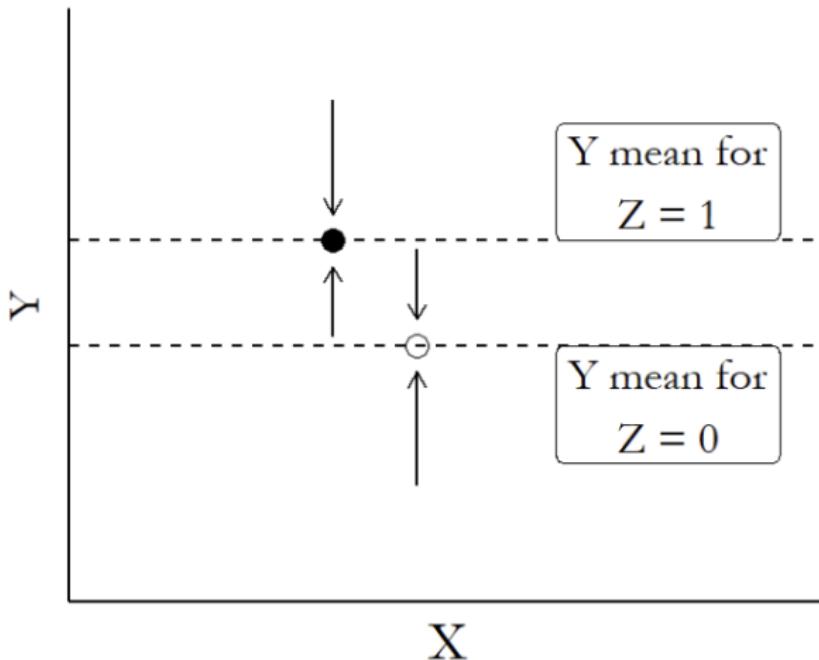
(d) Predict Y with Z



(Huntington-Klein, *The Effect*, p 472)

How does IV work?

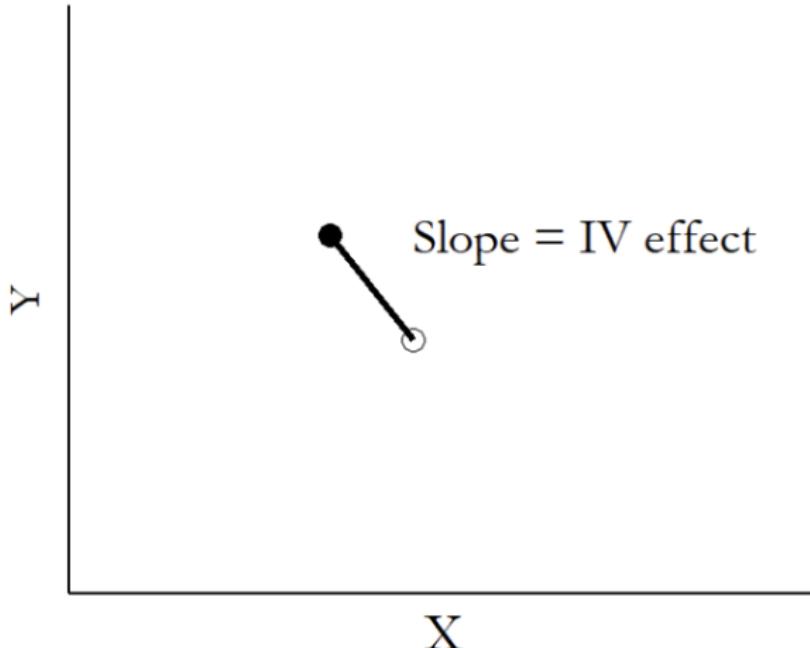
(e) Only Use Predicted Variation



(Huntington-Klein, *The Effect*, p 472)

How does IV work?

(f) Relate Predicted Y to Predicted X



(Huntington-Klein, *The Effect*, p 472)

How does IV work?

- Usually: two-stage least squares, or **2SLS**
1. Run a 'first-stage' regression to predict the treamtnet with the instrument
 2. Use the predicted values to predict the outcome in the 'second-stage'

Alternative approaches to IV: build your own

Roads to Rule, Roads to Rebel: Relational State Capacity and Conflict in Africa

Carl Müller-Crepon¹ , Philipp Hunziker²,
and Lars-Erik Cederman³

Journal of Conflict Resolution
2021, Vol. 65(2-3) 563-590

© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0022002720963674
[journals.sagepub.com/home/jcr](https://jcr.sagepub.com/home/jcr)



Abstract

Weak state capacity is one of the most important explanations of civil conflict. Yet, current conceptualizations of state capacity typically focus only on the state while ignoring the relational nature of armed conflict. We argue that opportunities for conflict arise where relational state capacity is low, that is, where the state has less control over its subjects than its potential challengers. This occurs in ethnic groups that are poorly accessible from the state capital, but are internally highly interconnected. To test this argument, we digitize detailed African road maps and convert them into a road atlas akin to Google Maps. We measure the accessibility and internal connectedness of groups via travel times obtained from this atlas and simulate road networks for an instrumental variable design. Our findings suggest that

Alternative approaches to IV: build your own

Instrumental Variable Approach

We complement our robustness checks with an instrumental variable (IV) strategy that addresses potential omitted variable biases not captured by the previous tests. In particular, there might be hitherto unmeasured group-level characteristics that have affected colonial road building and recent conflict. To address such endogeneity as well as potential systematic measurement bias in the Michelin maps, our IV approach exploits variation from road networks simulated on the basis of countries' population distribution. Our IV approach improves identification by isolating the component of RSC that is due to the spatial population distribution within a country. While population distributions are less malleable than road networks, populations are not randomly distributed. We must therefore rely on the assumption that the population distributions that produce our simulated road networks are conditionally exogenous to conflict. We address potential violations of this assumption below.

Alternative approaches to IV: build your own

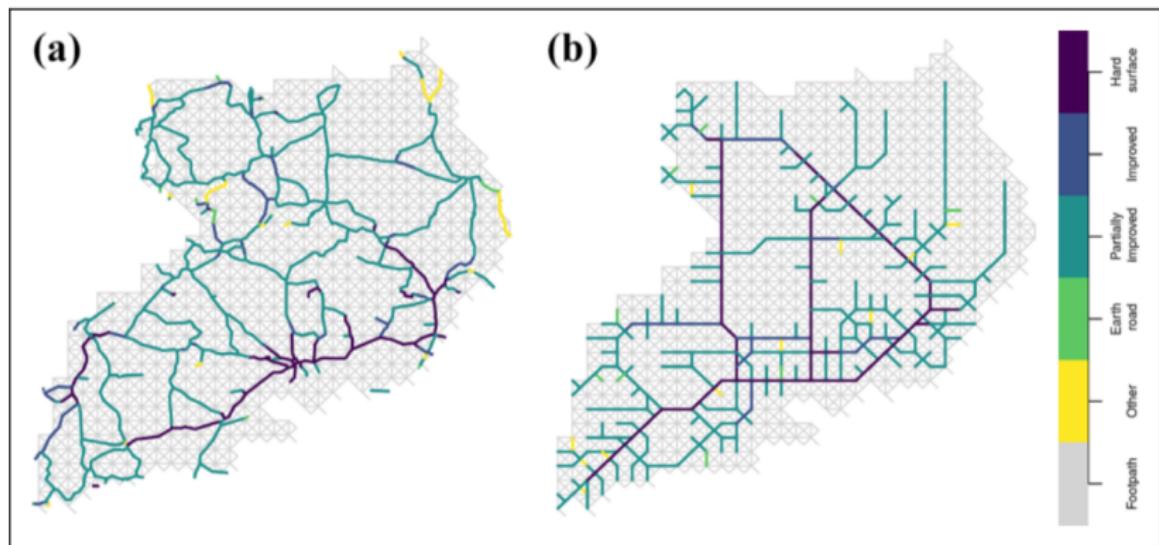


Figure 3. Observed and simulated road network in Uganda, 1966. (a) Observed network.
(b) Simulated network.