

More Methods

EC420 MSU Online

Justin Kirkpatrick

Last updated June 10, 2021

Lectures:

(1) Motivation

Difference in Differences

(2) Introducing Difference in Differences

(3) DiD Estimation

(4) DiD Assumptions

(5) DiD in R

Regression Discontinuity

(6) Introducing Regression Discontinuity

(7) RD Assumptions and Estimation

(8) RD Examples

(9) More Regression Discontinuity

Motivation

[top](#)

This section is about

- Difference in differences
- Regression Discontinuity

These econometric methods

- Estimate the *ATE* or *SATE*
- Have a causal interpretation
- Depend on assumptions

In fact, Difference in Differences and Regression Discontinuity use tools you already have

- Estimated with OLS
- Assume MLR.1 - MLR.6

But they add some assumptions

- Statements about what is randomly assigned
- Restrictions on the counterfactual values

It turns out, we can make some minimal assumptions that are, in some contexts, easily justified, that lead to an unbiased estimate of the SATE.

Assumptions

Econometrics is filled with assumptions

- Some are very innocuous, like assuming the data \mathbf{X} is full rank.
- Some, it turns out, are frequently not met, like MLR.4
- Some can be relaxed (like MLR.5 on homoskedasticity)

Good applied econometric work

- Makes minimal assumptions necessary
- States assumptions clearly

Assumptions

A physicist, a chemist, and an economist are adrift at sea with only a can of food and no way to open it. They decide to each devise and share a plan for opening the can of food. The physicist proposes a mechanism to leverage the waves at sea to open the can. The chemist proposes a saltwater-based solution that will decompose the can's seal to open it. Finally, when the economist is called on to present his plan, he stands up in the life raft and starts with "assume we have a can opener..."

- We make assumptions to simplify the problem and make it *tractable*
- In economics, we make this clear and transparent, allowing for robust discussion about those assumptions.
 - In a way, it provides a forum for discussion the accuracy of a model

Introducing Difference in Differences

[top](#)

Difference-in-differences

In the absence of random assignment, treatment and control groups are likely to differ for many reasons

MM, Ch. 5

Our goal is to get a valid counterfactual for $E[Y_0|D = 1]$

So that we have solved our selection bias problem

But we might not have a good instrument

Or we might worry that there are other unobserved things about the group being treated

DiD is another tool for your econometric toolbox

Diff-in-diff is closely related to our discussion on fixed effects

$$y_{it} = \beta_0 + \beta_1 x_{it} + \phi_i + \delta_t + u_{it}$$

- Where ϕ_i is a vector of effects for each i (except the base or omitted level)
- And where δ_t is a vector of effects for each time period t (except for a base or omitted level)

ϕ_i controls for persistent individual-specific effects

- "persistent" means "constant over time" within i
- δ_t does the same, but for effects in a time period t that are constant over all i

Of course, we must have panel data for DID

- We need multiple i and multiple t

We have to have some "treated" i , and some "untreated" i

- The "treatment group"
- The "control group"
 - This group helps us build our counterfactual
 - And our counterfactual helps us fix selection bias
 - And fixing selection bias gets us to an *ATE*!

The intuition:

Let's start with a simple difference. We observe some before and after data, where τ is the time that the treatment starts:

$$y_t = \beta_0 + \beta_1 D_t + u$$

and

$$D_t = 1(t > \tau)$$

Since β_1 is the difference in $y_{t < \tau}$ and $y_{t > \tau}$, then:

$$E[y|D = 1] = E[y|t > \tau]$$

$$E[y|D = 0] = E[y|t \leq \tau]$$

Which means

$$\beta_1 = E[y|D = 1] - E[y|D = 0]$$

plus:

$$\beta_0 = E[y|D = 0]$$

The simple difference is a naive estimator of the treatment effect of interest.

Even if we take a bunch of before-after observations: $y_{it} = \beta_0 + \beta_1 D_{it} + u_{it}$, we might still have a problem.

The problem here is that *treatment is not necessarily as good as randomly assigned*.

The biggest threat is that there is something *unobserved* that *varies between treated units* that might be correlated with the start of treatment, τ .

An example

Let's say τ is the time that a minimum wage increases in a state, and y is total hours worked in the city

- Interesting because we want to see the effect of minimum wage on employment
- **But** when states decide to increase their minimum wage, it's *may be* because they think the national economy is on the rise
- So treatment is correlated with *better economy* and that can drive employment up.
- Treatment time is endogenous!

Some more examples:

τ may be the time in which a superfund site is cleaned up, and y is the average home value around the superfund site

- Economic rationale: If cleaning up a highly polluted site increases the value of homes (by more than the cost of cleanup), we are increasing welfare by cleaning up pollution.
- **But** it may be the house prices are on the way up when there are good economic times, and government has more revenue during good economic times for cleanup. Or people who move into this neighborhood might be better at lobbying for cleanup.
- Treatment time is endogenous!

Yet another example:

τ may be the start of an incentive program for solar panel installation in a state, and y is the number of solar panel installations in a given month.

- If this incentive program is effective, we would expect to see an increase in the number of solar panels installed after τ
- **But** it might be that the program was enacted because policymakers thought solar would be becoming more popular everywhere
- Treatment time is endogenous!

The common theme is that treatment time may be endogenous, so β_1 is biased.

- $E[u|D] \neq 0$
- Those time periods (months, years) that are treated have different (Y_{t0}, Y_{t1})
- Here, potential outcomes are over **time**.

Counfounder:

We might say:

- | There is an unobserved, time-variant trend confounding our estimate

Confounder is something that is biasing our estimate by "looking like" our treatment variable D .

National trends **confound** our estimate when they coincide with treatment time τ .

When we have panel data, we have multiple i 's with treatment:

$$y_{1t} = \beta_0 + \beta_1 D_{1t} + u_{1t}$$

$$y_{2t} = \beta_0 + \beta_1 D_{2t} + u_{2t}$$

Unfortunately, we can't just fixed-effect away the problem. We would worry that:

$$y_{it} = \beta_0 + \beta_1 D_t + \phi_i + u$$

is still potentially biased because *some other time-variant effect possibly correlated to the treatment time* confounds β_1 .

- "There were going to be lots of solar panels installed anyways", etc.
- National trend in employment going up

The control group

- Observed i that are not treated
- They provide the counterfactual
- They are affected by the same unobserved time trend
- We should be able to "subtract off" that time trend
- Let's use a new dummy variable for members of this group:
 - $TREAT_i = 0$ for those in the control group
 - $TREAT_i = 1$ for those in the treatment group

Since treatment is time-variant, let's re-name D_t :

- $POST_t = 0$ in the pre-treatment period
- $POST_t = 1$ in the post-treatment period.

A treated t (time) for i (individual/unit) is $POST_t \times TREAT_i == 1$

First, take the before-after differences for the i that get treated:

$$\begin{aligned}\delta_{TREAT} &= E[y|POST == 1, TREAT == 1] \\ &\quad - E[y|POST == 0, TREAT == 1]\end{aligned}$$

δ_{TREAT} is what we had before just looking at treated units before-after.

As usual, we could get this from a regression of Y on $TREAT$ and $POST$

Second, take the before-after differences for the i that did not get treated:

$$\begin{aligned}\delta_{CONTROL} = & E[y|POST == 1, TREAT == 0] \\ & - E[y|POST == 0, TREAT == 0]\end{aligned}$$

If we assume that *this* $\delta_{CONTROL}$ is "what would have happened to the treated in the absence of treatment", then it provides a good counterfactual.

Finally, take the *difference* in the *differences*:

$$\delta_{DID} = \delta_{TREAT} - \delta_{CONTROL}$$

That's it!

Plugging in those two equations and using \bar{y} for $E[y]$:

$$\begin{aligned}\hat{\delta}_{DID} = & \left(\bar{y}_{TREAT=1, POST=1} - \bar{y}_{TREAT=1, POST=0} \right) \\ & - \left(\bar{y}_{TREAT=0, POST=1} - \bar{y}_{TREAT=0, POST=0} \right)\end{aligned}$$

We know that we can get an unbiased and consistent estimate of

$$E[y|TREAT = 1, POST = 1] - E[y|TREAT = 1, POST = 0]$$

using a regression **only** on $TREAT_i == 1$:

$$y_{it} = \beta_0 + \beta_2 POST_t + u \quad (\text{using only treated } i)$$

Similarly, we could do the same for those observations where $TREAT_i == 0$

But there is an easier way to do it. In fact, we can do it in one well-specified regression we can estimate with OLS.

DiD Estimation

top

We could regress:

$$y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \underbrace{\beta_3 TREAT_i * POST_t}_{\delta_{DID}} + u$$

We can see this by calculating:

- $E[Y|TREAT = 1, POST = 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$
- $E[Y|TREAT = 1, POST = 0] = \beta_0 + \beta_1$
- $E[Y|TREAT = 0, POST = 1] = \beta_0 + \beta_2$
- $E[Y|TREAT = 0, POST = 0] = \beta_0$

$$\begin{aligned} E[y|TREAT = 1, POST = 1] - E[y|TREAT = 1, POST = 0] &= \\ (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1) &= \beta_2 + \beta_3 \end{aligned}$$

and

$$\begin{aligned} E[y|TREAT = 0, POST = 1] - E[y|TREAT = 0, POST = 0] &= \\ (\beta_0 + \beta_2) - (\beta_0) &= \beta_2 \end{aligned}$$

In the regression:

$$y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \underbrace{\beta_3 TREAT_i * POST_t}_{\delta_{DID}} + u$$

δ_{DID} is the difference between the before-after differences. It is the unexpected "jump" that occurs in the treatment i at time of treatment $t = \tau$ relative to those that don't get treatment.

We can see this visually, and see the counterfactual.

DiD

Assumptions

top

Assumptions:

We made an assumption when we were talking about controls and counterfactuals. Specifically, we made the assumption that the *unobserved time trend* would affect both the **treatment** and the **control** the same.

Another way of saying this is (from MM):

Absent any policy differences (treatment), the **trend** in the control group is what we would have seen in the treatment group.

Or, from Wooldridge:

The average trends would be the same for the control and treatment in the absence of treatment

This is the **parallel trends assumption**

In the absence of treatment, the trends in both groups would have been the same.

We do not get to observe this

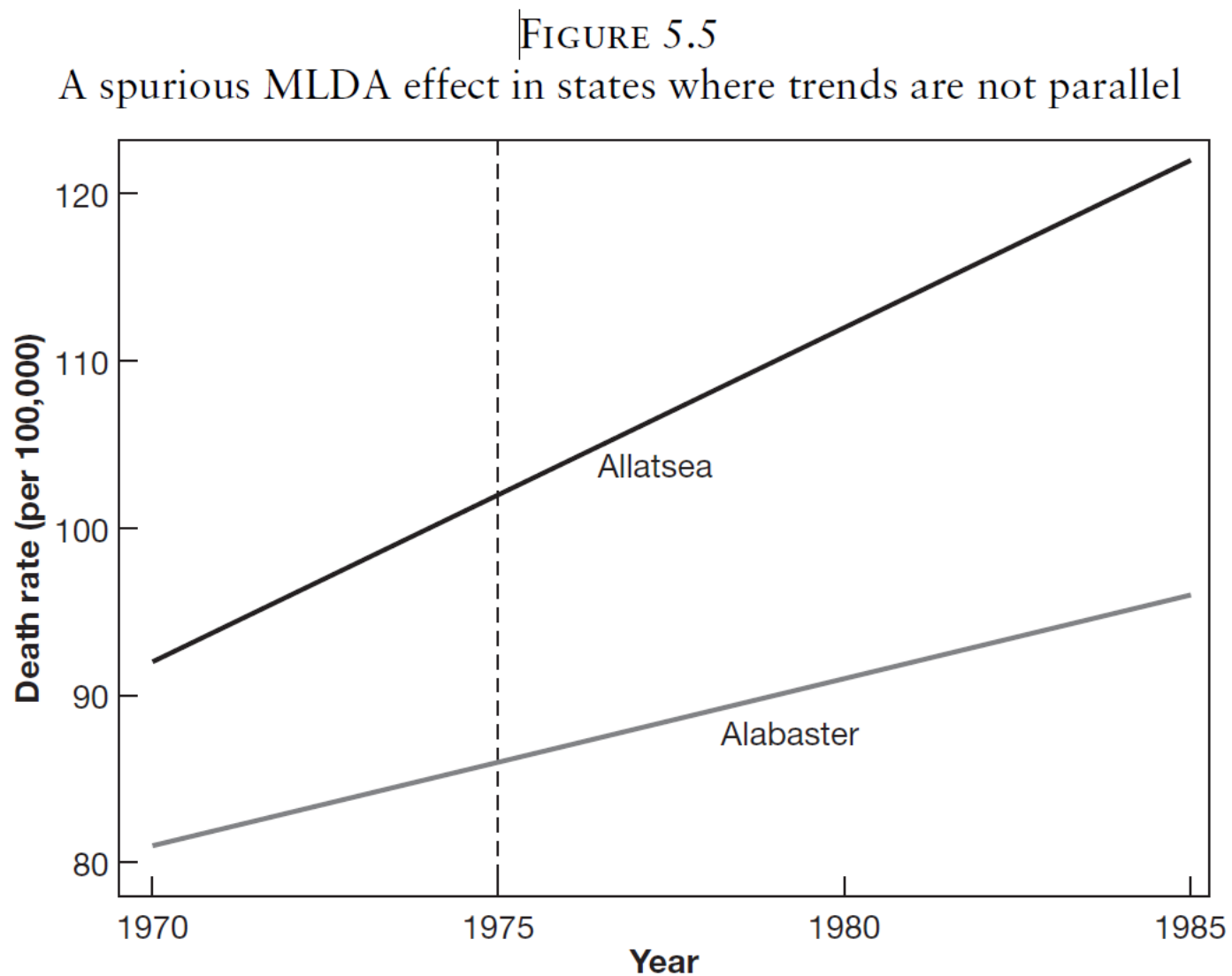
We never see the treatment group's untreated trend after treatment.

- That's why this is an assumption.

But we do see the pre-treatment trend

Which may help us make a convincing argument

- But can never guarantee that we our parallel trend assumption holds



Restating the parallel trends assumption a different way:

We have to assume that there are no time-variant, i -specific effects. We can have constant i -specific fixed effects, we can have t -specific fixed effects (trends), but we can't have something that changes over time (e.g. changes right at the treatment time τ) and *only* happens to the treated units.

The parallel trends assumption is our *identifying assumption*

If it does not hold, we are not identifying the ATE as we cannot write it in terms of population parameters we can estimate from the data.

If it does hold, then we do identify the ATE and it is δ_{DiD} .

More flexible DID

We don't have to limit our fixed effects to $TREAT_i$. We can be more flexible:

$$y_{it} = \beta_0 + \beta_1 POST_t + \beta_2 TREAT_i + \beta_3 POST_t * TREAT_i + \phi_i + u_{it}$$

Here, β_2 wouldn't be identified because it's just a combination of ϕ_i 's (remember our *full rank* assumption). Dropping it out results in a **DID** estimator that lets every treated (and control) group have its own intercept instead of one common one for all treated i . β_3 is the **DID** estimator, has the same interpretation, and is identified.

Time-varying treatment

What if we have a treatment that isn't the same time period for all i ?

$$y_{it} = \beta_0 + \beta_1 D_t + \beta_2 TREAT_i + \beta_3 TREAT_i * D_t + u_{it}$$

This wouldn't work because there isn't a common D_t , and if we used D_{it} , we'd be estimating β_3 , which is already it -specific!

But:

$$y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 TREAT_i * D_t + \delta_t + u_{it}$$

Where δ_t flexibly controls for a common time trend. This would result in β_2 being the *DID* estimate.

Finally, a very flexible:

$$y_{it} = \beta_0 + \beta_1 D_{it} + \delta_t + \phi_i + u_{it}$$

Where $D_{it} = 1$ during a post-treatment period in a treated i .

i -specific trends

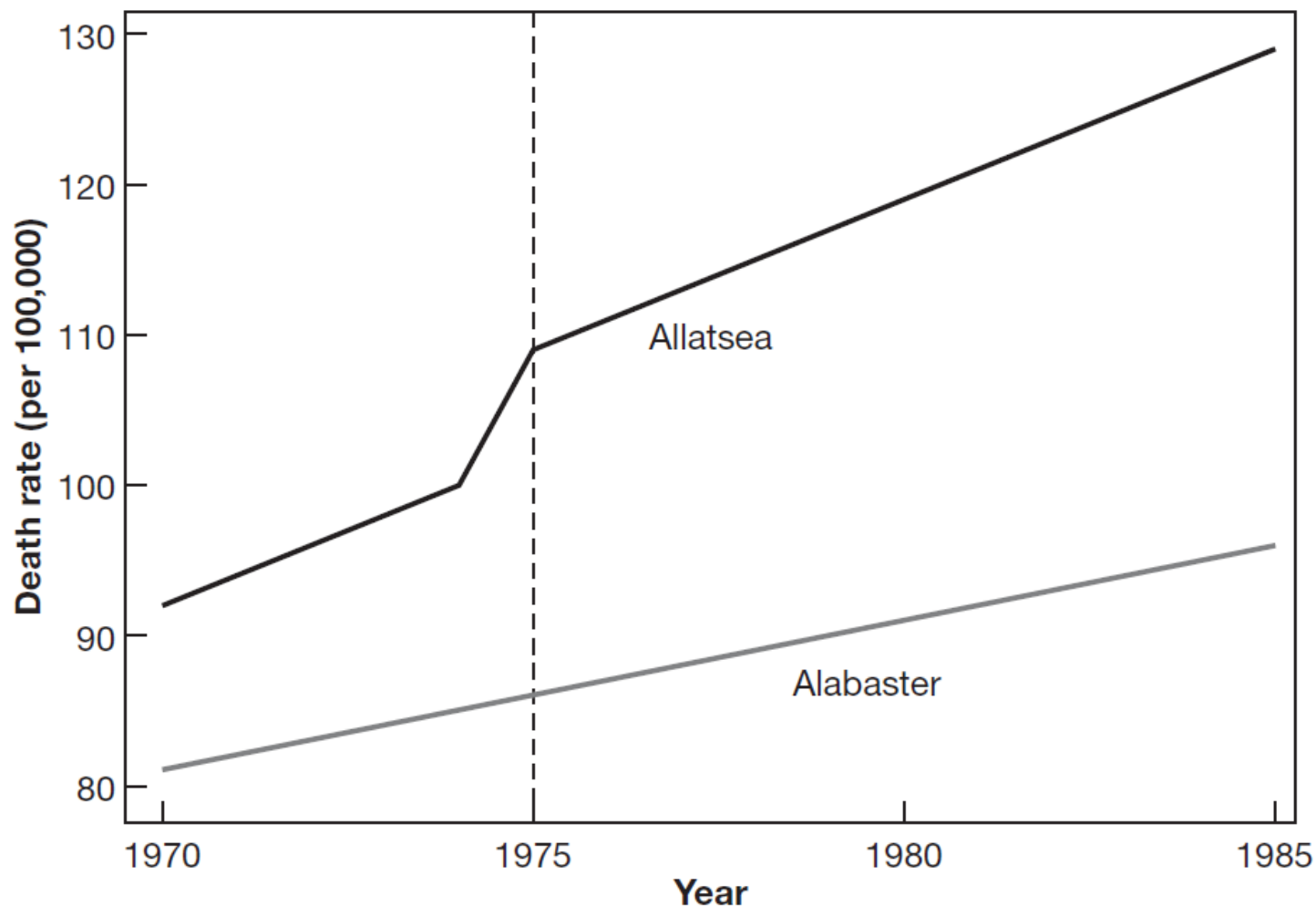
We can relax the parallel trends assumption a little bit by assuming that we can estimate different time trends (using an interaction) and that these trends would *continue to hold even in the absence of treatment*. This is a "parallel trend in trends" assumption.

$$y_{it} = \beta_0 + \beta_1 D_{it} + \phi_i + \underbrace{\sum_{t=1}^T \delta_i * t}_{i\text{-specific time trend}} + u_{it}$$

So, we can have non-parallel trends, but we **do** have to assume that those trends **continue** in the absence of treatment.

FIGURE 5.6

A real MLDA effect, visible even though trends are not parallel



There is no "test" for the parallel trends assumption

- The assumption is a statement about *what the treated unit would have done in the absence of treatment*
- We can *look* at the pre-treatment trends, but they do not guarantee us post-treatment trends
 - If no pre-treatment parallel trends, really hard to say that post-treatment would be parallel.

We can try a few things

- A balance comparison like before when we talked about selection on observables
- We could plot all of the treated units over time, centering them on the treatment date
 - E.g. plot of 10,9,8,... years before treatment with colors by treated/untreated
- Placebo test (randomly assign treatment and see if significant. It shouldn't be).
- Add leads of the treatment variable:

$$y_{it} = \beta_0 + \beta_1 TREATMENT_i + \beta_2 POST_t + \beta_3 TREATMENT_i \times POST_t + \underbrace{\beta_4 TREATMENT_i \times POST_{t+1}}_{\text{Lead of 1}} + u_{it}$$

Time-varying treatment

It turns out that time-varying treatment is more complicated than we thought. This is because of two things

- The treatment effect may be different over time
 - E.g. in the first year after treatment vs. in the 5th year after treatment
 - "Event study" finds the effect relative to the treatment year
- When treatments are applied at different times, the "untreated" group being used to estimate the difference is different over time
 - Never-treated
 - Not-yet-treated

Work by many econometricians like Goodman-Bacon, Callaway and Sant'anna, Chaisemartin and D'Haultfoeuille, Sun and Abrahams, and more are working to establish difference-in-differences estimates that acknowledge this.

DiD in R

top

We will use an example from **Scott Cunningham's excellent Causal Inference Mixtape**.

- I highly encourage you to read the DiD chapter in this free book, and the rest of the text as well if you're interested in econometrics with causal inference. The book also dives into the Goodman-Bacon DiD decomposition, which addresses some of the varied timing problems we discussed at the end of the last video.
- For now, we'll stick to the traditional DiD example, but note that there are issues with the estimate due to differential timing.

This example looks at the effect of the *castle doctrine* on homicide rates by states. The *castle doctrine* basically permits citizens to use deadly force within their home without a duty to retreat. Critics have indicated that this leads to more gun violence, while proponents claim it makes people safer and allows for self-defense. We are econometricians, not politicians, so we are interested in the empirical effects.

This is a good opportunity for a difference-in-differences because we think there are some state-specific characteristics that might lead to both higher homicide rates *and* selection into passing a castle doctrine.

Before we begin

You'll need the `haven` packages installed to read `.dta` files. We won't do the full paper recreation done on the Causal Inference website. Instead, we'll generate our interacted (DiD) treatment variable and only use the two-way fixed effects. The book uses far more controls, but the same underlying two-way fixed effects.

The data has a field called `effyear` that shows what year the castle doctrine was enacted in each state. Those with `NA` for `effyear` are never-adopters. So, we need to make a variable that is `1` if `year > effyear` **and** the state is a treatment state (enacted the castle doctrine). We can do that as below:

```
library(haven) # you need the haven package to read .dta files
castle = read_dta('https://raw.githubusercontent.com/scunning1975/mixtape/master/castle.dta')

castle$TREATMENT = !is.na(castle$effyear)
castle$INTX = castle$TREATMENT==T & castle$year>=castle$effyear
```

So, `castle$INTX` is our interaction of `TREATMENT` and `POST`. The coefficient will be the DiD coefficient.

```
myDID = lm(l_homicide ~ INTX + unemployrt + income + police + as.factor(year) + as.factor(state))
coeftest(myDID, vcov = vcovHC, 'HC1')
```

```
##
## z test of coefficients:
##
##
```

	Estimate	Std. Error	z value	Pr(> z)	
## (Intercept)	1.5686e+00	3.1805e-01	4.9320	8.139e-07	***
## INTXTRUE	9.0679e-02	3.5627e-02	2.5452	0.0109203	*
## unemployrt	-7.7754e-03	1.1497e-02	-0.6763	0.4988352	
## income	5.1356e-06	5.0716e-06	1.0126	0.3112398	
## police	5.2316e-04	4.4073e-04	1.1870	0.2352218	
## as.factor(year)2001	3.1064e-02	5.1337e-02	0.6051	0.5451105	
## as.factor(year)2002	1.9001e-02	5.1660e-02	0.3678	0.7130218	
## as.factor(year)2003	6.7740e-02	5.2723e-02	1.2848	0.1988509	
## as.factor(year)2004	5.8649e-02	4.7427e-02	1.2366	0.2162367	
## as.factor(year)2005	7.2961e-02	4.7480e-02	1.5367	0.1243701	
## as.factor(year)2006	6.2021e-02	4.3789e-02	1.4164	0.1566691	
## as.factor(year)2007	4.8966e-02	4.9130e-02	0.9967	0.3189241	
## as.factor(year)2008	1.9367e-02	5.5733e-02	0.3475	0.7282187	
## as.factor(year)2009	-3.4763e-02	8.0289e-02	-0.4330	0.6650351	
## as.factor(year)2010	-8.2367e-02	7.8454e-02	-1.0499	0.2937703	
## as.factor(state)Alaska	-4.4131e-01	1.3358e-01	-3.3037	0.0009541	***
## as.factor(state)Arizona	-2.3742e-02	5.7070e-02	-0.4160	0.6773937	
## as.factor(state)Arkansas	-7.9500e-02	6.4654e-02	-1.2296	0.2188415	
## as.factor(state)California	-1.6096e-01	8.9771e-02	-1.7930	0.0729665	.
## as.factor(state)Colorado	-7.8393e-01	9.9331e-02	-7.8921	2.972e-15	***
## as.factor(state)Connecticut	-8.7247e-01	1.2179e-01	-7.1639	7.841e-13	***
## as.factor(state)Delaware	-6.4259e-01	1.3857e-01	-4.6372	3.531e-06	***
## as.factor(state)Florida	-3.0502e-01	6.1623e-02	-4.9497	7.431e-07	***

The result

Using this simple DiD specification, and assuming parallel trends (see [Causal Inference The Mixtape](#) book for a long discussion on parallel trends here), we get a results of $\beta^{DiD} = .0906$. Since the outcome variable is logged homicide rate, this results suggests an 9.06% increase in the homicide rate when castle doctrine is enacted.

Of course, there are many things omitted to be controlled for, and one could argue that the assumptions do not actually hold. But our task -- learning DiD -- is done.

For other interesting examples, see Video 5.1 at <https://mattmasten.github.io/bootcamp/>

Introducing Regression Discontinuity

[top](#)

Let's return to the endogeneity issue:

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_i + u_i$$

And we have the usual endogeneity problem that: D_i may be correlated with u_i , even once we condition on exogenous x_i .

Let's set up an example

- y_i is semesters of college attended
- D_i is "getting a scholarship"
- x_i is parental income
- u_i is potentially endogenous error

We might think that *ability*

1. Affects semesters of college attended
2. Unobserved (and thus in u_i)
3. Correlated with getting a scholarship

$$\Rightarrow (Y_{i0}, Y_{i1}) \not\perp D_i | x_i$$

Let's return to the endogeneity issue:

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_i + u_i$$

What if we had a scholarship that was awarded to all students with a 1200 SAT score or above?

- $D_i = 1(SAT_i > 1200)$

Does this solve our endogeneity problem?

- Is SAT_i , the student's SAT score, exogenous?
- What do we learn about scholarships in general from understanding the effect of *this* particular scholarship? We'll return to this in a little bit.

Does this solve our endogeneity problem?

- Is SAT_i , the student's SAT score, exogenous?

What if I compared a student with an 800 to a student with a 1600
(and with the exact same x_i)

- Would you be concerned about that unobserved *ability* in u_i ?

What about comparing a 900 student to a 1500 student?

What about comparing a 1000 student to a 1400 student?

What about comparing a 1100 student to a 1300 student?

What about comparing a 1190 student to a 1210 student?

What about comparing a 1199 student to a 1201 student?

The intuition behind RD is as follows:

If we compare students within a *small enough window* around a threshold, then treatment D_i is **as good as randomly assigned**

Whatever we were worried about in the error term that was correlated with D_i (and was correlated with SAT score) would be unlikely to be correlated with such a tiny, tiny difference in SAT scores. 1199 and 1201 is not much of a difference.

→ *as good as randomly assigned* lets us treat the treatment as **exogenous**

The *unobserved* endogenous problem in u_i is no longer a problem!

Let's make this assumption formal, and see how to leverage this intuition

RD Assumptions and Estimations

What do we need for an RD?

I. A **threshold**

- And it has to be exogenous
- Usually from a policy
- Arbitrary policies are...great!



What do we need for an RD?

II. A **running variable**

- We need some variable that crosses the threshold
 - It cannot be binary
- It has to be *observed*
- It has to determine the treatment of interest [†]

"Although treatment isn't randomly assigned, we know where it comes from"

— Mastering Metrics, Ch. 4

[†] A fuzzy RD relaxes this assumption

The **running variable** in our example is SAT score

- Crosses the threshold
- Non-binary
- Observed
- Determines scholarship (treatment) discontinuously
 - A "jump"

What do we need for an RD?

III. A **window**

- We need somewhere to "draw the line"
- In our SAT example, some of us would have been OK with 1150 and 1250.
 - Some would want 1190 to 1210
 - Some would say 1195 to 1205

A **threshold**

- Set exogenously

A **running variable**

- Determines treatment
- Continuous
- Observable

A **window**

- How close is close enough

The identifying assumption:

Within a small enough **window**, the treatment (as determined by the **running variable** and the **threshold**) is as good as randomly assigned

Specification of an RD

- Let's call the running variable x_r
- Let's call the threshold τ

Our RD equation would be:

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_{i,r} + \beta_3 x_{exogenous} + u_i$$

- y_i is outcome
- $D_i = 1(x_{i,r} > \tau)$
- τ is the threshold
- $x_{i,r}$ is the *running variable*
- $x_{i,exogenous}$ is an exogenous covariate (a control)
- u_i is the error

Our RD equation would be:

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_{i,r} + \beta_3 x_{exogenous} + u_i$$

and

$$D_i = 1(x_{i,r} > \tau)$$

Note that we have a continuous running variable $x_{i,r}$ with a linear coefficient

- We can still control for the effect of the running variable
- That's why we need the τ threshold where treatment "jumps"

And we can still control for other things

- $x_{exogenous}$

Our RD equation would be:

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_{i,r} + \beta_3 x_{exogenous} + u_i$$

But where is the window?

- We *could* just use a sample within the window
- Define the window as being (a,b)
 - R: `myData[myData$xir >=a & myData$xir <=b,]`
- Another option is to use *local linear regression*, which can be implemented as a *weighted regression*

A quick aside to establish weighted regression:

We know that OLS minimizes the sum of the squared error term u_i

$$\sum (y_i - \beta_0 - \beta_1 x_1)^2 = \sum (u_i)^2$$

Weighted regression just puts weights on that sum:

$$\sum w_i (u_i)^2$$

and

$$\sum w_i = 1$$

Weights can set the estimation sample

If we said that $w_i = 0$ if i is not "in the window", and constant otherwise, we would just be setting the estimating sample.

Weights can be continuous

What if we set w_i to be some function of *how far away from the threshold* an observation $x_{i,r}$ is?

- If $x_{i,r}$ is far away from τ , then it gets a low weight
- If $x_{i,r}$ is practically equal to τ , then it gets a high weight

$$\tilde{w}_i = \frac{1}{|x_{i,r} - \tau|}$$

Of course, we might have to re-scale all the w_i 's to make sure they add to 1:

$$w_i = \frac{\tilde{w}_i}{\sum_{j=1}^N \tilde{w}_j}$$

This "local linear regression" because it puts more weight on those observations where $x_{i,r}$ is very close to τ

- You still have to determine the form of the weights
- We call the function that generates weights the *kernel*
- Since we are not making an assumption about how the distance away from τ enters the main regression, this is **non-parametric**.

Other kernels include:

- Rectangular Kernel
 - Puts equal weight on all observations between a and b
- Triangle Kernel
 - Puts weight on those closer to the center

Applying weights to OLS is easy:

```
myOLS = lm(Y ~ X1 + D + X2, data=myData, weights = 1/abs(myData$X1 - tau))
```

"Local Linear Regression" will also usually allow for a different slope on each side of the cutoff τ

$$y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 D \cdot X_1 + \beta_4 X_2 + U$$

```
myOLS = lm(Y ~ X1 + D + X1*D + X2, data = myData, weights = 1/abs(myData$X1-tau))
```

RD Examples

top

Example 1

Maimonides Rule

- 12th century rabbinic scholar who determined that 40 students was the max. class size.
- Anything larger had to be cut into two 20/21-person classes
- Angrist and Lavy (1999) used threshold to look at effect of class size on student outcomes
 - What is the endogeneity problem between student outcomes and *assignment* of class sizes?
 - What is the threshold and is it exogenously set?
 - What is running variable?
 - Evaluate whether or not we have an as-good-as-random treatment assignment

Example 2

Superfund cleanup

- All superfund sites have a risk score
- Congress determines how far down the
- Greenstone and Gallagher (2008) and Gamper-Rabindran and Timmins (2011) used threshold to look at effect of cleanup on housing prices
 - What is the endogeneity problem between home values and *assignment* of superfund cleanup?
 - What is the threshold and is it exogenously set?
 - What is the running variable?
 - Evaluate whether or not we have an as-good-as-random treatment assignment

More Regression Discontinuity

[top](#)

We have to be very careful with our *running variable* as it crosses the threshold

- We assume that it is "smooth" over the threshold
- This means we can estimate its effect on the outcome y separate from the effect of crossing the threshold
 - Remember, crossing the threshold triggers treatment

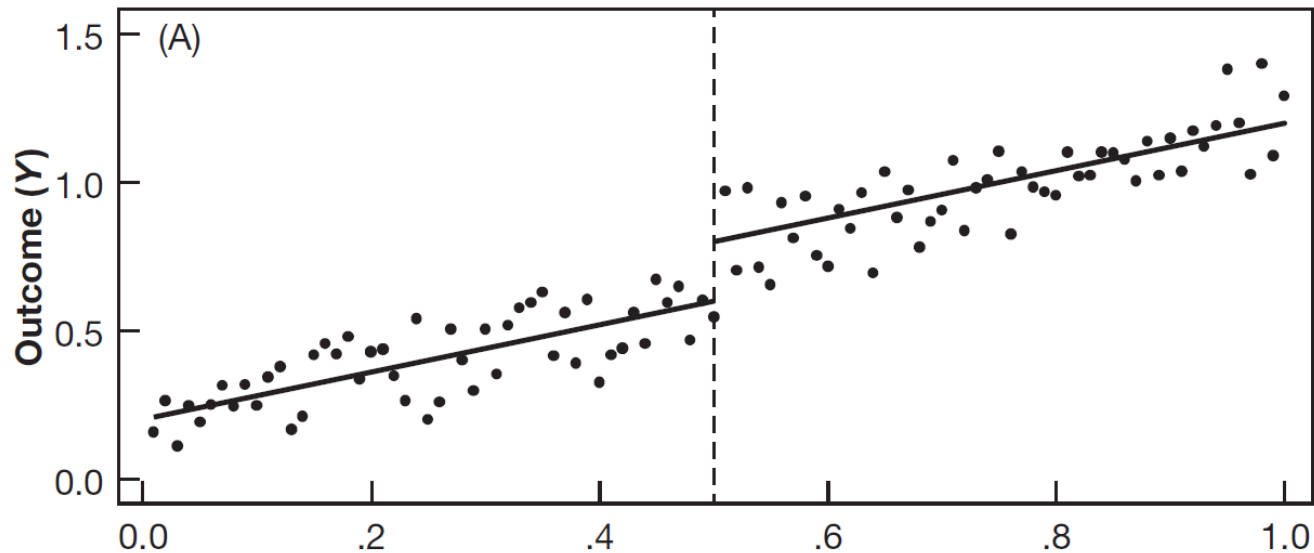
We control *flexibly* for the relationship between x , the running variable; and y , the outcome.

- How?

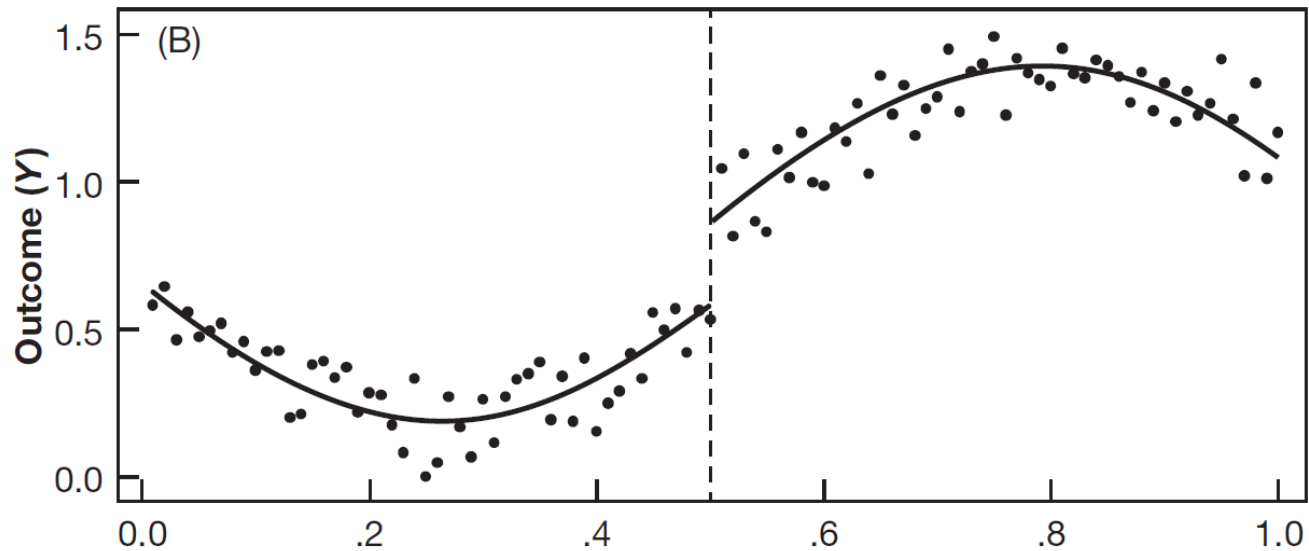
A linear control

Assumes it's a constant marginal effect (a straight line)

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_{i,r} + u_i$$

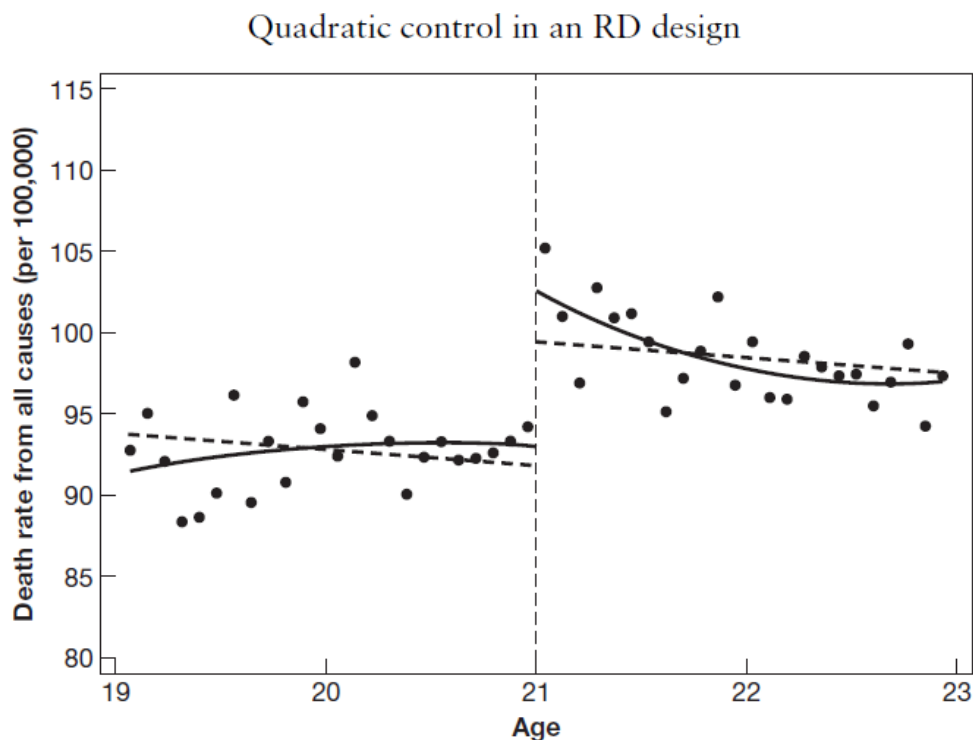


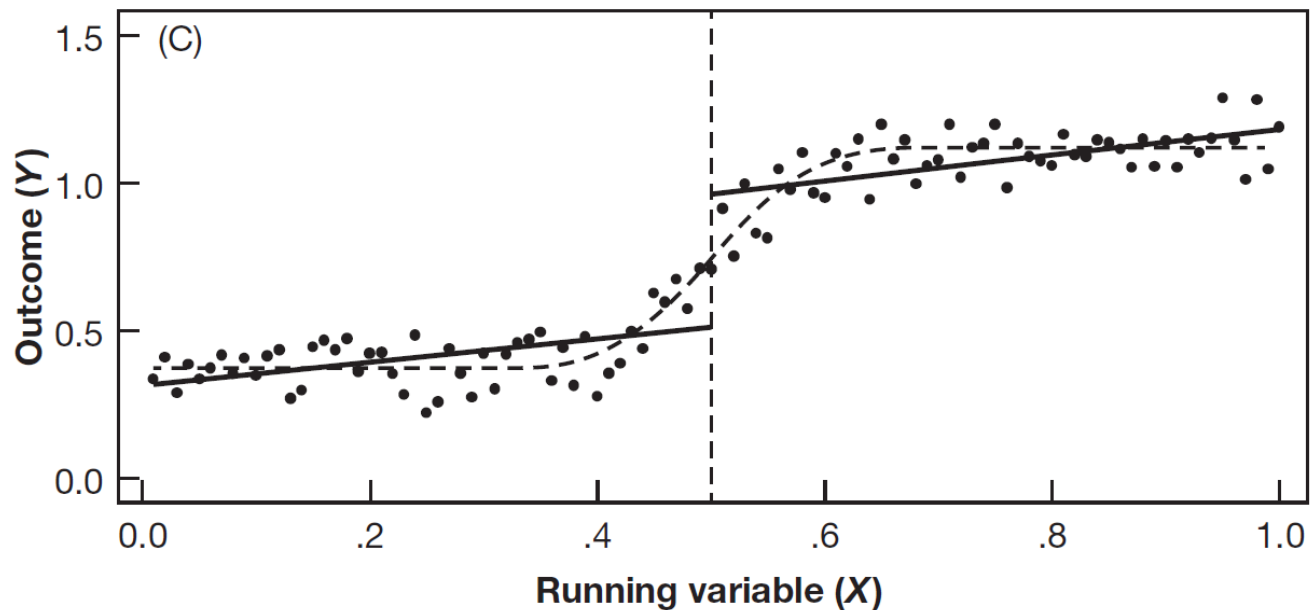
$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_{i,r} + \beta_3 x_{i,r}^2 + \beta_4 x_{i,r}^3 + u_i$$



$$y_i = \beta_0 + \beta_1 D_i + \beta_2(x_{i,r} - \tau) + \beta_3(x_{i,r} - \tau) * D_i + u_i$$

Here, we let the *slope* of the coefficient on the running variable change when it crosses the threshold. To the left of the threshold, $\frac{dy}{dx_{i,r}} = \beta_2$, and to the right, it is $\beta_2 + \beta_3$ (dashed line).





This one is not so good - the dashed line fits a possible non-linear trend.

- Note that there is an upward trend just before the threshold
- That trend is more consistent with a polynomial trend, and the linear trend + threshold gives a false effect.

The implicit assumption in our RD "window"

is that everything else *around* the threshold τ varies *smoothly*

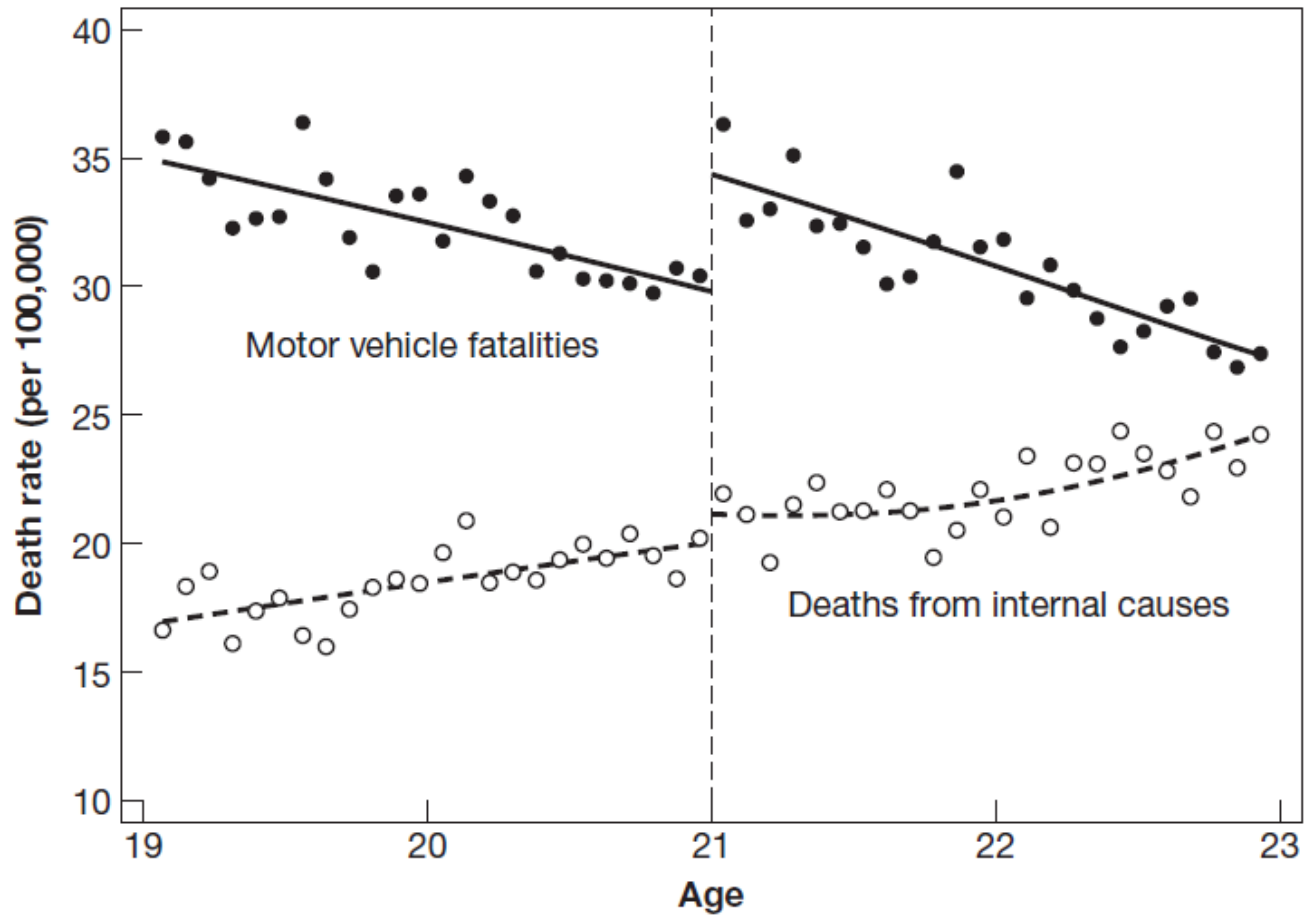
- Then, we can control for it parametrically.

Which means that *other variables* not affected by the treatment should vary smoothly over the threshold

- After all, if some unrelated variable is "jumping" at the threshold, then we shouldn't have much confidence in that threshold being exogenous
 - And this would ruin our assumption.

(Z)ero effects on outcomes that should be unchanged by treatment raise our confidence in the causal effects we are after

(-) Mastering Metrics



What if the threshold doesn't perfectly determine treatment, but the other assumptions hold?

In this case, we would have:

- Something exogenous
- That has an affect on treatment
- But doesn't affect outcome except through treatment

It's an instrument!

Back to our original example, except maybe we don't have perfect awarding of scholarships to $SAT > 1200$

First Stage:

$$SCHOLARSHIP_i = \alpha_0 + \alpha_1 SAT_i + \alpha_2 1(SAT_i > 1200) + \alpha_3 x_i + v_i$$

This is just like our [original](#) equation, but with $SCHOLARSHIP_i$ on the LHS.

Second Stage:

$$y_i = \beta_0 + \beta_1 \widehat{SCHOLARSHIP}_i + \beta_2 SAT_i + \beta_3 x_i + u_i$$

β_1 is our treatment of interest.

Note that SAT_i is still in the second stage **but** the threshold dummy $1(SAT_i > 1200)$ is *not*.

This just follows the IV method, but gets the first-stage exogeneity from the RD specification! The *crossing of the threshold* is the exogenous event.

In practice...

We rarely get a threshold that is perfectly "sharp".

- Age of drinking is pretty sharp
- But many other things don't have a perfect relationship

So fuzzy is more common in the literature

- And remember: when you have an instrument that is almost perfectly correlated with the endogenous variable, IV turns into OLS.
 - If the first stage Z has a perfect prediction of D , then $\hat{D} = D = Z$

Hopefully, the RD *intuition* is clear

- We find a source of exogeneity
- We make our identifying assumptions about it explicit
 - Which allows others to poke at them. Scientific method and such
- And we implement it with OLS
 - Or, in the case of weighted regression, something WOLS

RD is only appropriate in certain circumstances

- It's not always the right tool in the toolbox
- Many seemingly random spatial boundaries (county lines, etc.) actually have a lot of variation in other unobservables when you cross them
 - School district lines
- Use wisely, and be prepared to *state and justify assumptions*.