

# Topic 4: Regression Discontinuity

*ECON 5783 – University of Arkansas*

Prof. Kyle Butts

Fall 2024

# Regression Discontinuity Design (RDD)

## *Example 1*

Lee (2008, JOE) studies the “incumbency advantage”, the hypothesis that being a serving elected official improves future election outcomes

# Regression Discontinuity Design (RDD)

## *Example 1*

Lee (2008, JOE) studies the “incumbency advantage”, the hypothesis that being a serving elected official improves future election outcomes

The problem, of course, is that candidates who won their election usually are different than those that do not

- e.g. are more charming, in a more one-sided district, have a better resume

# Regression Discontinuity Design (RDD)

## *Example 1*

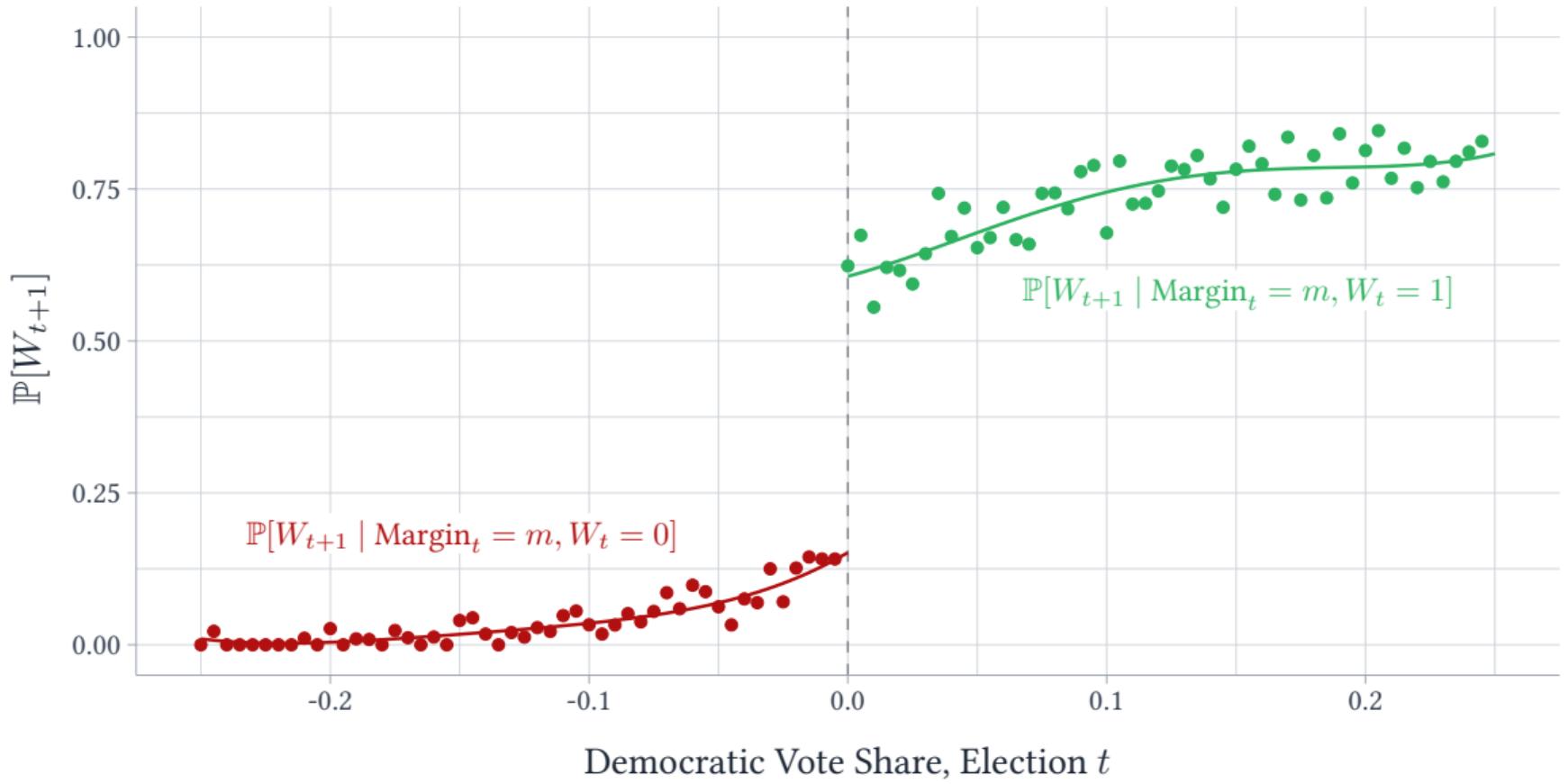
Lee (2008, JOE) studies the “incumbency advantage”, the hypothesis that being a serving elected official improves future election outcomes

The problem, of course, is that candidates who won their election usually are different than those that do not

- e.g. are more charming, in a more one-sided district, have a better resume

Lee’s idea is to compare candidates who narrowly lost to those that narrowly won

- By having similar vote percentage, Lee hopes, to be comparing candidates with similar unobservables



# Regression Discontinuity Design (RDD)

## *Example 1*

There is a clear jump from candidates that narrowly lost to candidates that narrowly won

- The main concern is that candidates that just lost look different in terms of unobservables to those that just won

# Regression Discontinuity Design (RDD)

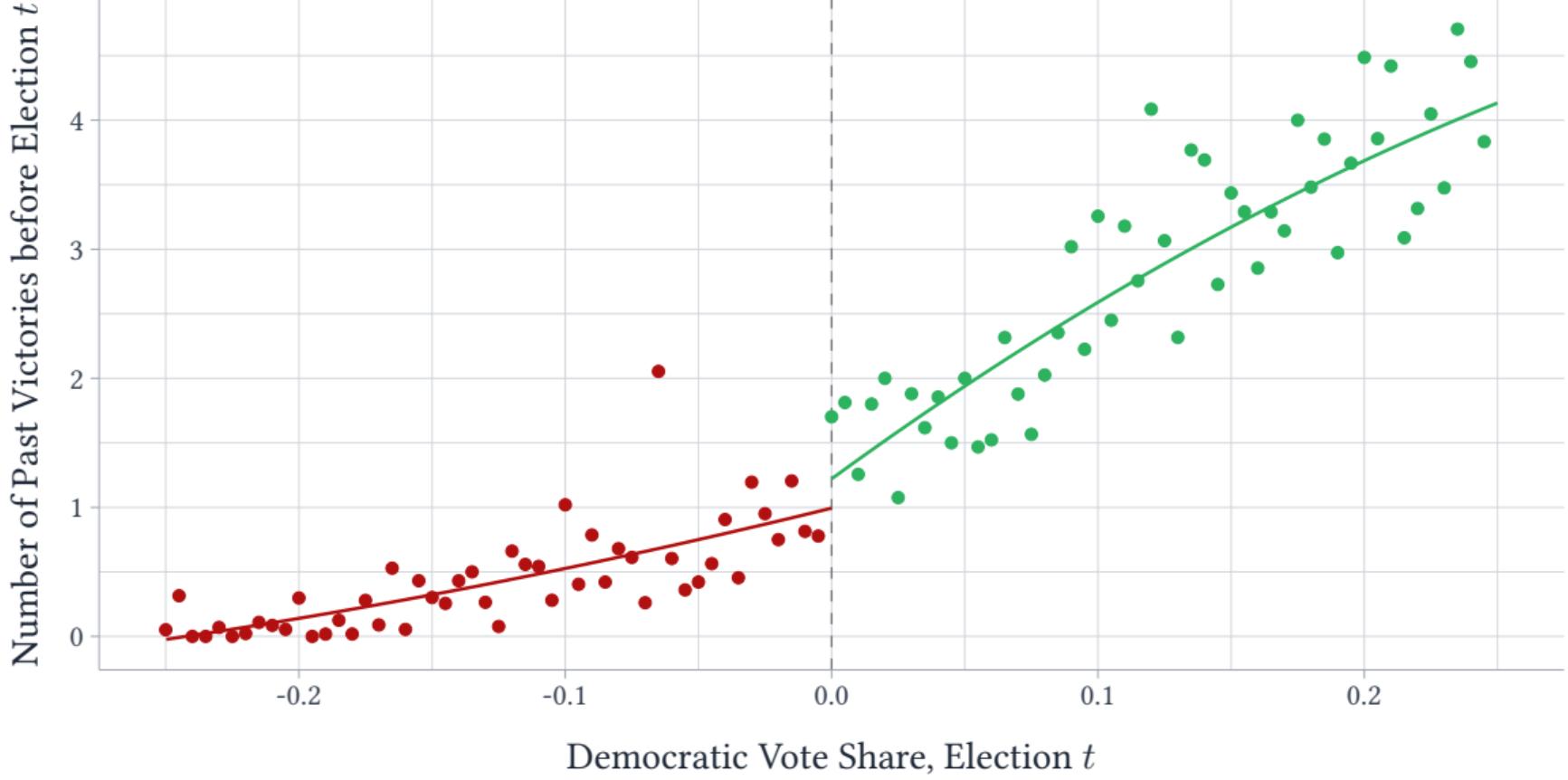
## *Example 1*

There is a clear jump from candidates that narrowly lost to candidates that narrowly won

- The main concern is that candidates that just lost look different in terms of unobservables to those that just won

To try and alleviate this concern, Lee checks if other observed variables ‘jump’ at the cutoff

- A jump in other ‘pre-determined’ variables at the cutoff would be problematic to our story



# Regression Discontinuity Design (RDD) Terminology

In the RDD literature, we will have a '**running' variable** ('score' variable),  $X_i$ , and a '**cutoff**' value  $c$

- Units with  $X_i < c$  do not receive the 'policy' and units with  $X_i \geq c$  do

The treatment variable is defined as  $D_i = \mathbb{1}[X_i \geq c]$

# Regression Discontinuity Design (RDD)

## *Example 2*

Bleemer and Mehta (2022, AEJ Applied) study the returns to being an economics major by leveraging a requirement of a 2.8 GPA threshold in Econ 1 and 2 at UCSC:

- Comparing students just above a 2.8 GPA to those just below helps address selection into economics major
- A potential concern is that highly-motivated students near the threshold might ask for higher grade, extra credit, etc.

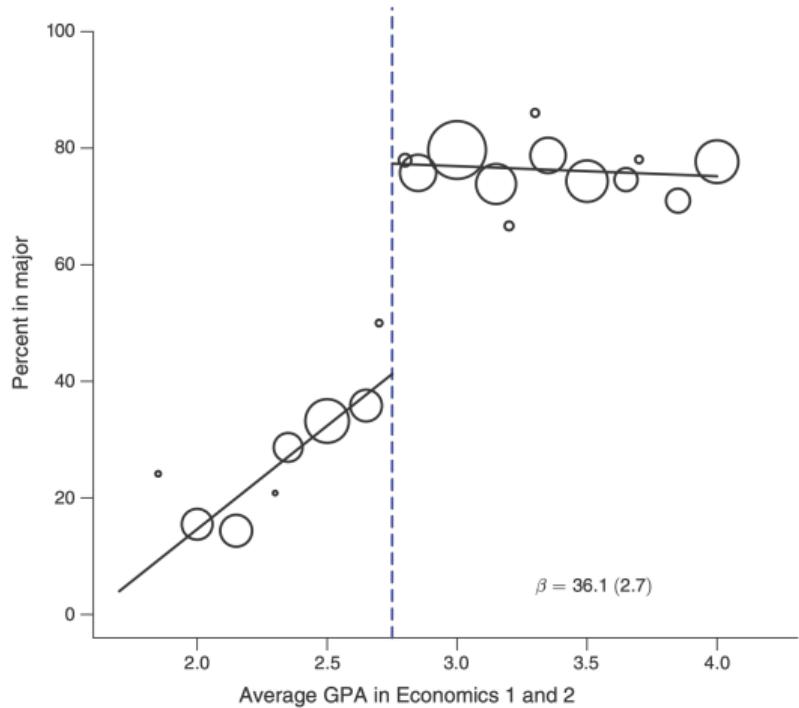


FIGURE 1. THE EFFECT OF THE UCSC ECONOMICS GPA THRESHOLD ON MAJORING IN ECONOMICS

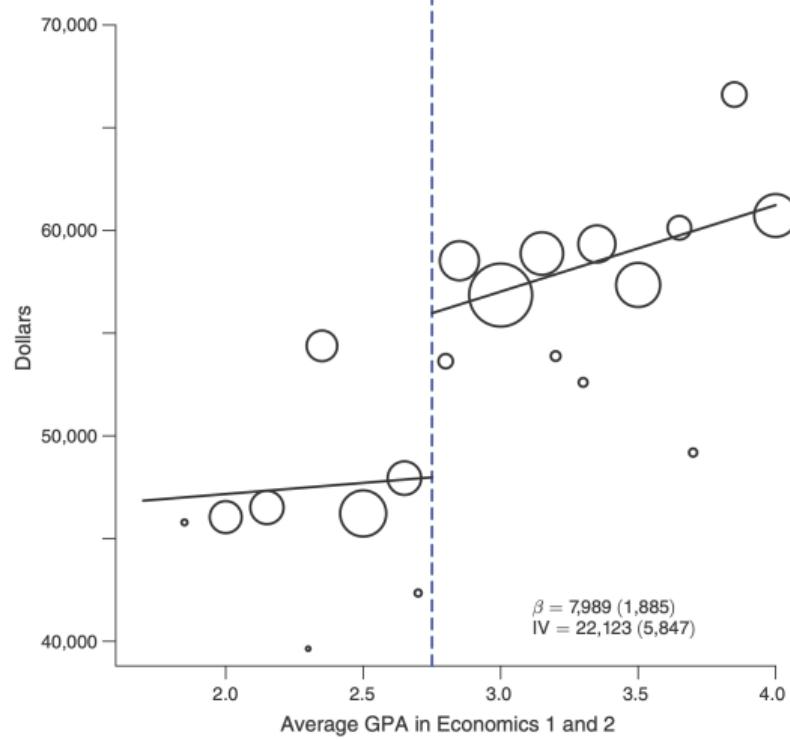


FIGURE 2. THE EFFECT OF THE UCSC ECONOMICS GPA THRESHOLD ON ANNUAL WAGES

# Regression Discontinuity Design (RDD)

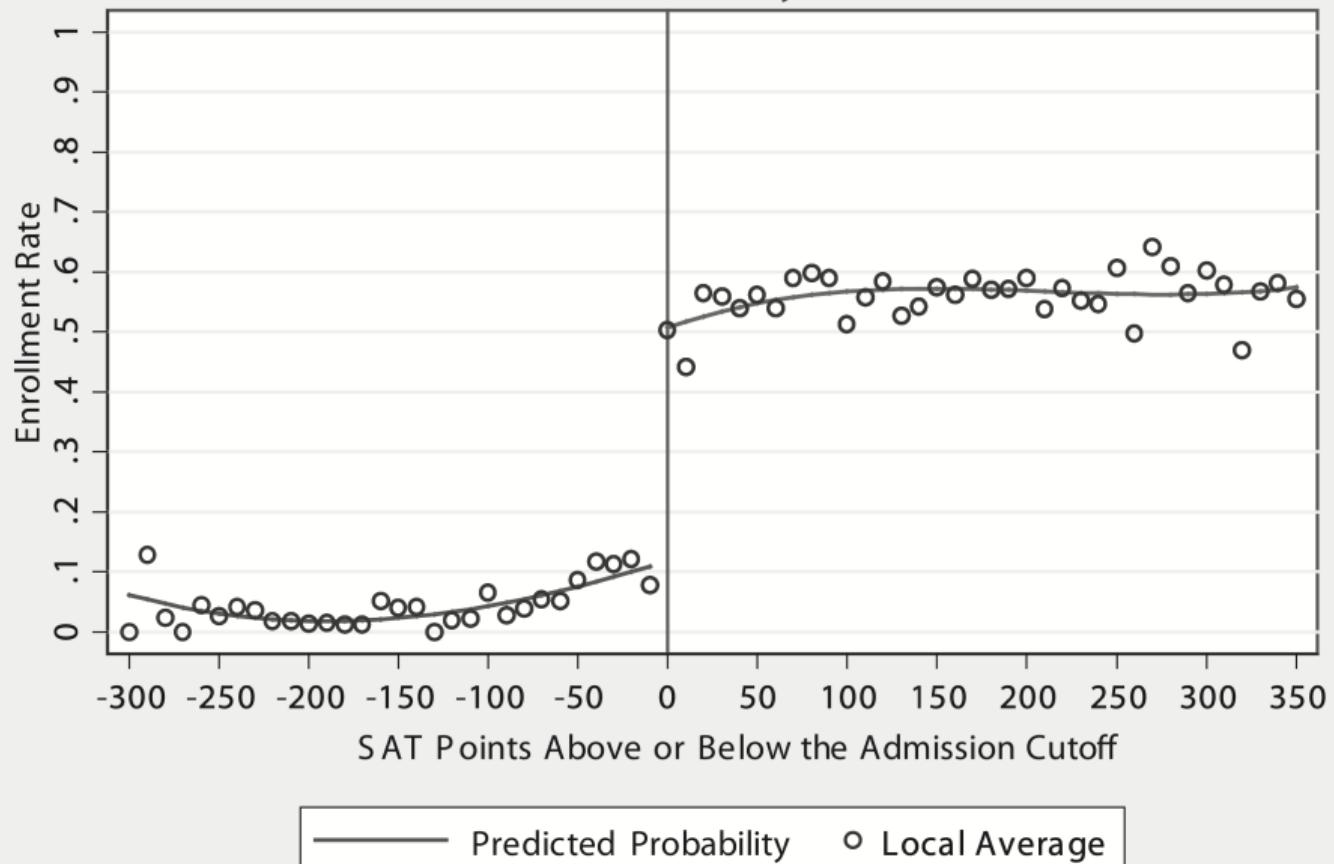
## *Example 3*

Hoekstra (2009, RESTAT) estimates the impact of attending a 'flagship university' on future earnings

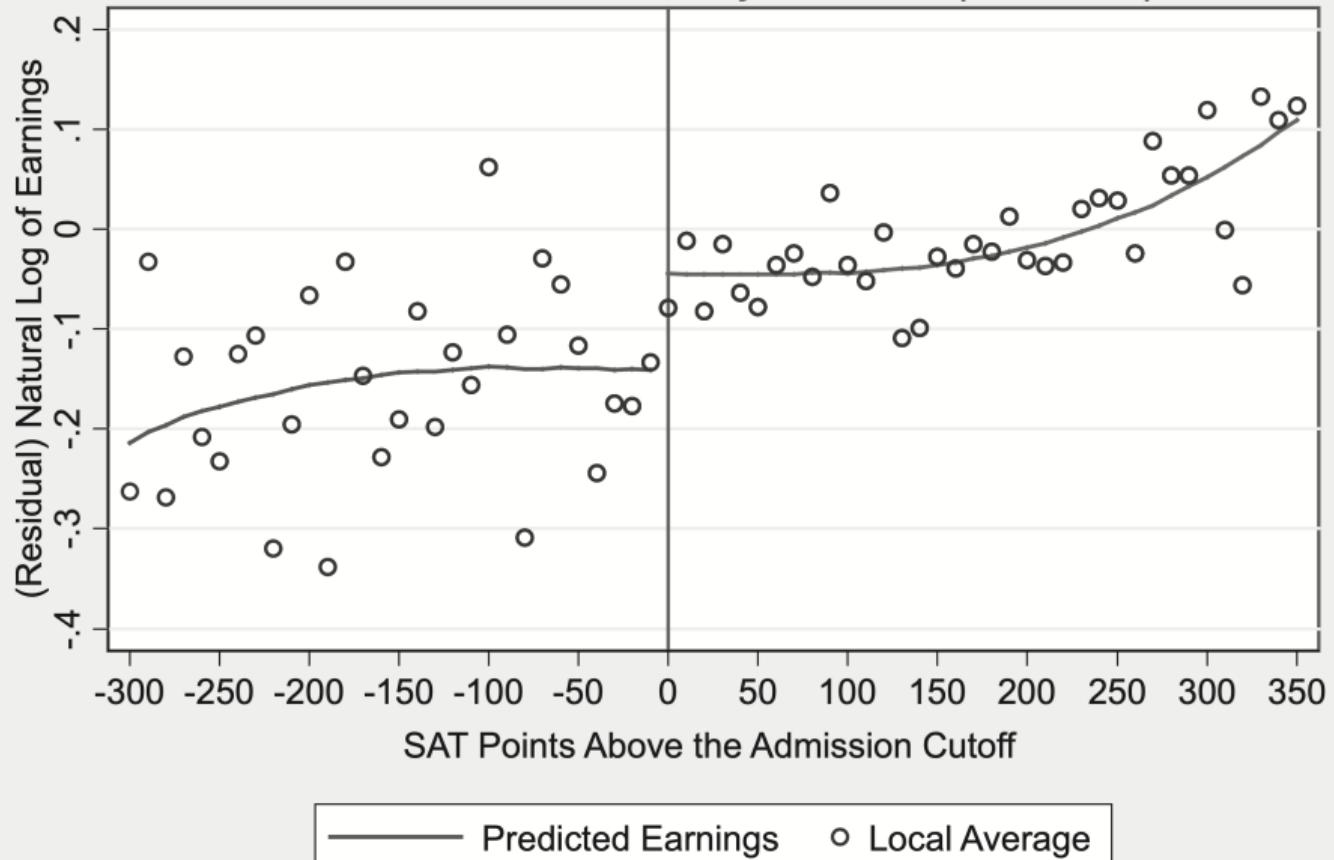
- There was an internal (secret) SAT cutoff that had a large increase in the probability of acceptance

That this cutoff was secret helps with concerns over retaking SAT, e.g.

Estimated Discontinuity = 0.388 ( $t=10.57$ )



Estimated Discontinuity = 0.095 ( $z = 3.01$ )



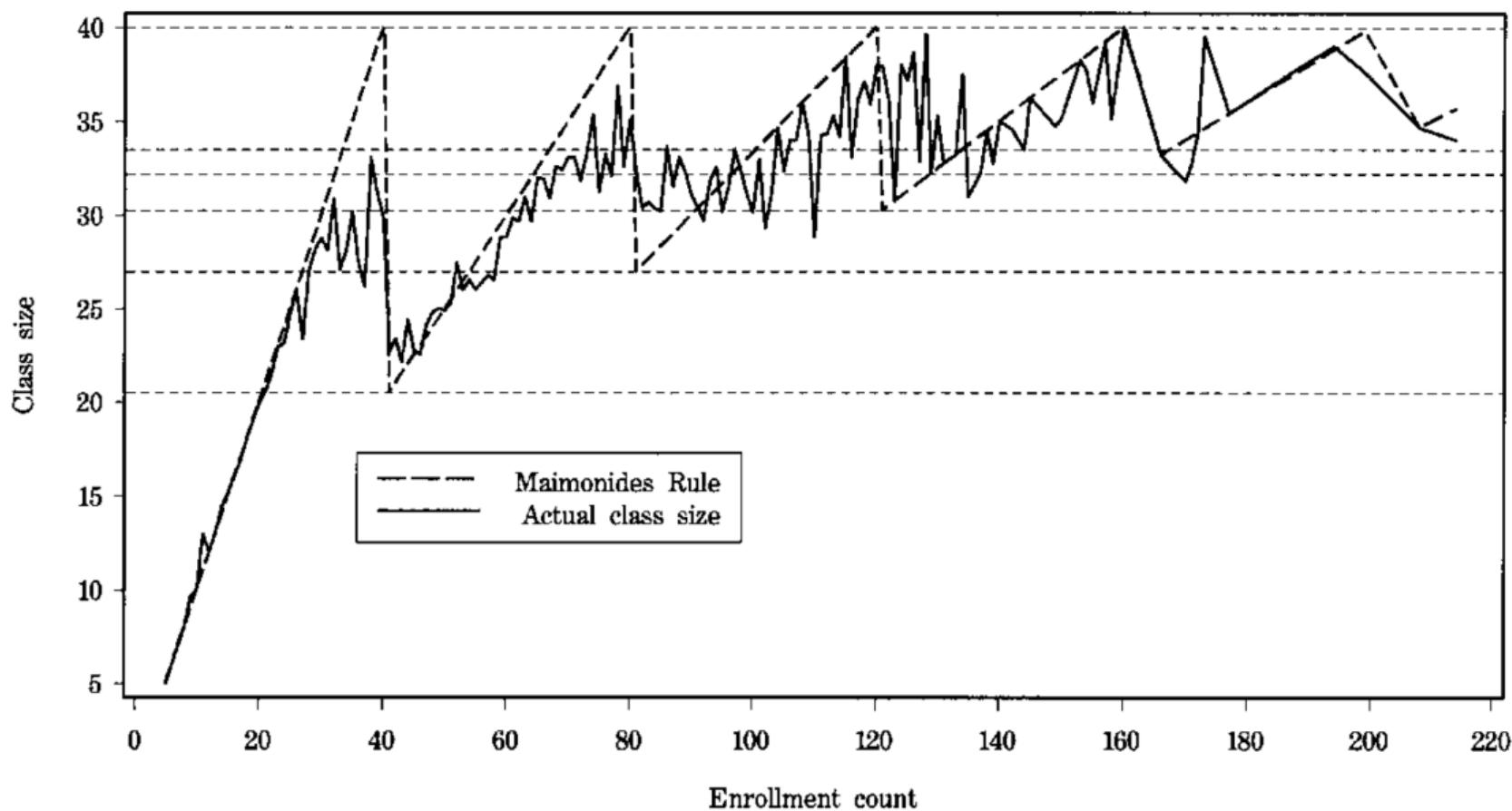
# Regression Discontinuity Design (RDD)

## *Example 4*

Angrist and Lavy (1999, QJE) study the effect of class size on kid's learning. They use a rule in Israel that requires all classes to have 40 or fewer students

- This creates sharp drops in class size at 41, 81, 121, 161 students

a. Fifth Grade



# Regression Discontinuity Design (RDD)

## *Example 5*

Anderson and Magruder (2011, Economics Journal) study the impact of restaurant Yelp ratings on restaurant business outcomes

- Yelp takes the average rating (e.g. 3.27) and shows the nearest number of stars rounding/up or down. E.g. 3.24 rounds down to 3 and 3.25 rounds up to 3.5

# Regression Discontinuity Design (RDD)

## *Example 5*

Anderson and Magruder (2011, Economics Journal) study the impact of restaurant Yelp ratings on restaurant business outcomes

- Yelp takes the average rating (e.g. 3.27) and shows the nearest number of stars rounding/up or down. E.g. 3.24 rounds down to 3 and 3.25 rounds up to 3.5

Their argument is that the score, average rating, is a noisy measure of the true quality and so restaurants look the same on either side of the cutoff

- Their main concern is 'manipulation' of running variable (we'll come back to this)

# Regression Discontinuity Design (RDD)

## *Example 6*

Turner et. al. (2014, ECTA) study the impacts of land-use regulations on the value of land

- They compare homes on one side of a zoning border to homes on the others
- The assumption is that the neighborhood doesn't 'abruptly' change when crossing the zoning boundary

This is an example of a spatial RDD which we'll discuss more about later

## Difficulties with RDD

We can not do our usual strategy of comparing treated and untreated individuals with the same  $X_i$

- Either everyone is treated or no one is

But, there is one point where we *kind of* have treated and untreated units: the cutoff  $c$

- This is our intuition for using 'just above' versus 'just below' the cutoff.

# Roadmap

Formalizing RDD Identification Arguments

Spatial RDD

# Formalizing RDD (Take 1)

*"Noisy" running variable*

Let's use a canonical example of students taking a test ( $X_i$ ) and students above a certain cutoff,  $c$ , are treated (e.g. put in honor's class).

Students will vary in terms of their expected score,  $\sigma_i = \mathbb{E}[X_i]$ . We think that future outcomes  $Y_i$  vary systematically based on  $\sigma_i$

- Comparing students above and below the cutoff will be biased

## “Noisy” running variable creates an experiment

For random reasons (a weird question, skipped breakfast, etc.) students score is given by  $X_i = \sigma_i + \varepsilon_i$ , where  $\varepsilon$  is random noise.

- For students with  $\sigma_i$  close to  $c$ ,  $\varepsilon_i$  will make them above or below the cutoff

## “Noisy” running variable creates an experiment

For random reasons (a weird question, skipped breakfast, etc.) students score is given by  $X_i = \sigma_i + \varepsilon_i$ , where  $\varepsilon$  is random noise.

- For students with  $\sigma_i$  close to  $c$ ,  $\varepsilon_i$  will make them above or below the cutoff

This means for  $\sigma_i$  close to  $c$ , we have a *natural experiment*

- Of course, we don't observe  $\sigma_i$ , so instead we take students within some range  $c \pm h$ .
- If  $h$  is “small enough”, we can do a difference in means between students below and above the cutoff

$Y_i(1)$  and  $Y_i(0)$

2  
1  
0  
-1  
-2

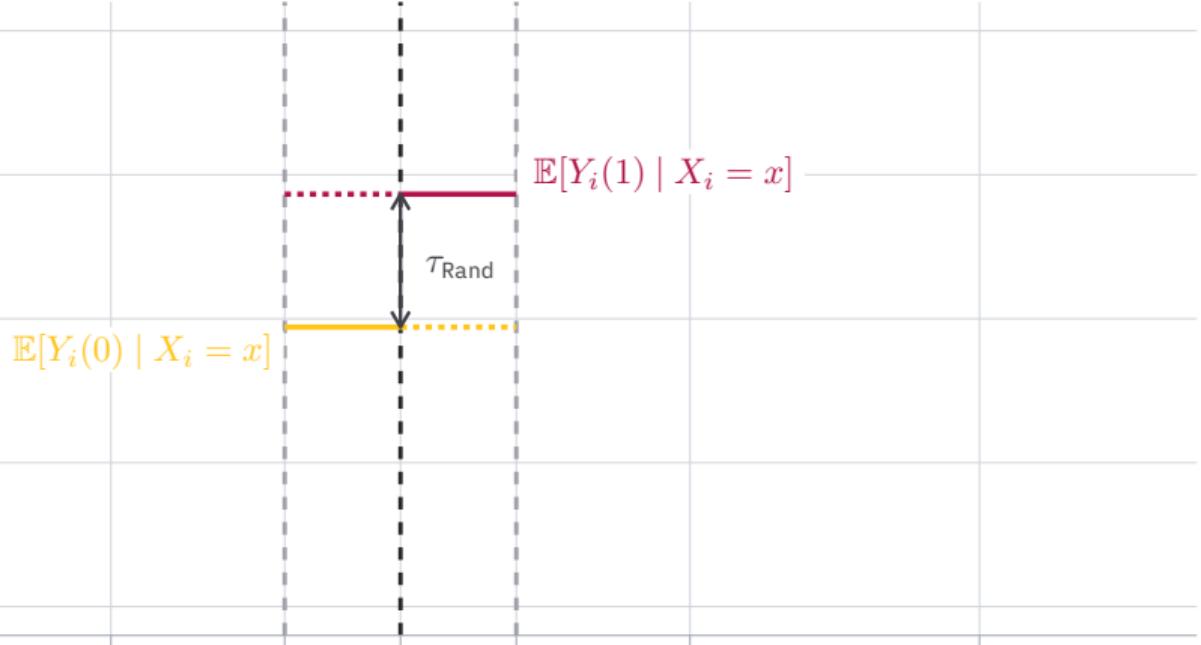
$\mathbb{E}[Y_i(0) | X_i = x]$

$\mathbb{E}[Y_i(1) | X_i = x]$

$c - h \quad c \quad c + h$

(Test) Score,  $X_i$

$\tau_{\text{Rand}}$



## “Local” treatment effect

In essence, we throw out all the data outside of  $[c - h, c + h]$  and compare students with  $\sigma_i \approx c$

This means we only learn about what the impact of treatment is for students with  $\sigma_i \approx c$

- E.g. if the cutoff is an SAT score of 1950, then we estimate the impact of treatment for students with around that score

## “Local” treatment effect

That is, we compare the following:

$$\begin{aligned} & \mathbb{E}[Y_i \mid c < \sigma_i < c + h] - \mathbb{E}[Y_i \mid c - h < \sigma_i < c] \\ &= \mathbb{E}[Y_i(1) \mid c < \sigma_i < c + h] - \mathbb{E}[Y_i(0) \mid c - h < \sigma_i < c] \end{aligned}$$

## “Local” treatment effect

That is, we compare the following:

$$\begin{aligned} & \mathbb{E}[Y_i \mid c < \sigma_i < c + h] - \mathbb{E}[Y_i \mid c - h < \sigma_i < c] \\ &= \mathbb{E}[Y_i(1) \mid c < \sigma_i < c + h] - \mathbb{E}[Y_i(0) \mid c - h < \sigma_i < c] \\ &\approx \mathbb{E}[Y_i(1) \mid \sigma_i = c] - \mathbb{E}[Y_i(0) \mid \sigma_i = c] \end{aligned}$$

That is, we estimate the CATE for people with  $\sigma_i \approx c$

## How much ‘noise’ is there in the running variable?

How do we know how large to make the cut-off  $h$ ?

- The gold-star answer is to have some application-specific understanding to know how much noise there is and take  $h$  to be half that noise

Otherwise, we are left to trying to determine a ‘reasonable’  $h$

## How much ‘noise’ is there in the running variable?

There is a trade-off between using a smaller or larger  $h$

- On the one hand, a smaller  $h$  makes it more likely that units in  $(c - h, c)$  look similar to units in  $(c, c + h)$
- On the other, a larger  $h$  uses more observations for estimation. A larger  $h$  runs a risk of including units that differ systematically from the  $\sigma_i = c$  units

When data on attributes of units are available, then we can use those to help with determining  $h$

## Estimation of $h$ using extra covariates

Cattaneo, Frandsen and Titiunik (2015, Journal of Causal Inference) recommend a procedure (available in `rdlocrand`) where:

- Start with a very small  $h_1$  and test if the mean of  $X_i$  are the same in  $(c - h_1, c)$  and  $(c, c + h_1)$
- If you fail to reject the null of no difference in means, then expand to  $h_2$
- Continue until you reject the null

The idea being, select the largest  $h$  where units 'look the same' on both sides of the cutoff.

## rdlocrand basic syntax

```
library(rdlocrand)

# Estimate effect and get p-values
rdrandinf(
  Y = df$y, R = df$score, cutoff = 0, wl = 0.025, wr = 0.025
)

# Estimate optimal h
rdwinselect(
  R = df$score, X = cbind(df$x1, df$x2, df$x3), obsmin = 10, wobs = 5
)
```

## Local randomization argument

This approach has a nice intuition: local to the cutoff we have a quasi-experiment due to noise in the running variable

Despite this, this approach is less popular than the next approach we discuss

- In part, sometimes we don't believe there is noise in the running variable
- The plots, like in Lee (2008), are the bread and butter of the RDD
  - These are not "difference-in-means" but instead rely on trying to identify a "jump" in smoothed lines

# Formalizing RDD (Take 2)

## *Continuity of outcomes*

The second approach we will discuss is the older (and more commonly used) approach to RDD estimation

- Instead of assuming the score is randomly assigned, this method will rely on a ‘continuity assumption’ of potential outcomes

# Formalizing RDD (Take 2)

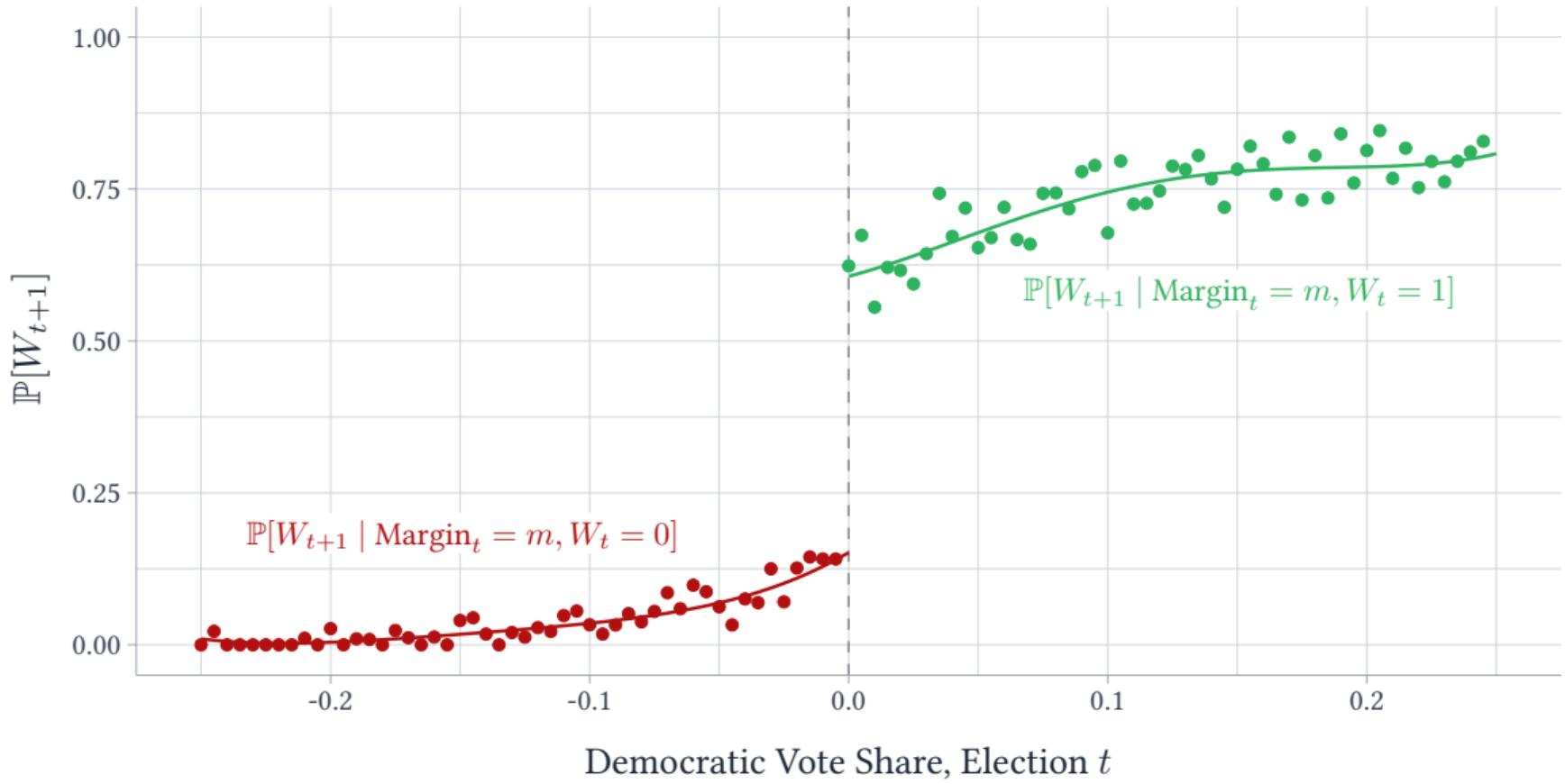
## *Continuity of outcomes*

The second approach we will discuss is the older (and more commonly used) approach to RDD estimation

- Instead of assuming the score is randomly assigned, this method will rely on a ‘continuity assumption’ of potential outcomes

The **continuity assumption** says that both the treated and untreated potential outcomes evolve ‘smoothly’ and do not have an abrupt ‘jump’ (discontinuity) at the cutoff

- This is what our mind ‘tells us to do’ when we see these RDD plots



# Formalizing continuity

The goal of our treatment effect estimator is to identify the following:

$$\mathbb{E}[Y_i(0) \mid X_i = c] \text{ and } \mathbb{E}[Y_i(1) \mid X_i = c]$$

We don't typically observe anyone with  $X_i$  exactly equal to 0 (assuming continuous  $X_i$ )

- So, necessarily we need to extrapolate to the cutoff using observations away from the cutoff

# Formalizing continuity

Similar to the selection on observables, define:

$$\mu_d(x) = \mathbb{E}[Y_i(d) | X_i = x]$$

to be the conditional expectation of  $Y_i(0)/Y_i(1)$  conditional on the running variable being equal to  $x$

We will fit  $\mu_0(x)$  using observations with  $X_i < c$  and  $\mu_1(x)$  using  $X_i > c$

- E.g. fit a linear model of  $X_i$  with observations in  $(c - h, c) / (c, c + h)$

## Extrapolation as limits

We are able to learn about the relationship between  $Y_i$  and  $X_i$  away from the cutoff. We are going to need to take our model and **extrapolate** it to the cutoff

- Imagine taking averages of  $Y_i$  from  $(c, c + h)$  to estimate  $\mathbb{E}[Y_i(1) | X_i = c]$ . Assuming infinite data, as  $h \rightarrow 0$  our average should get closer to closer to the true CEF.

More formally, under 'continuity' we have

$$\mathbb{E}[Y_i(1) | X_i = c] = \lim_{s \downarrow c} \mathbb{E}[Y_i(1) | X_i = x]$$

- Note here, we are taking the limit from above to only use obs. with  $Y_i = Y_i(1)$

## RDD Estimand

The **regression discontinuity** estimand is formed as follows:

$$\begin{aligned}\tau_{\text{RD}} &= \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = c] \\ &= \lim_{s \downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{s \uparrow c} \mathbb{E}[Y_i \mid X_i = x]\end{aligned}$$

- In words, take the difference between the right-hand limit of  $\mu_1(x)$  and the left-hand limit of  $\mu_0(x)$

## Why ‘continuity’?

Since we are fitting a model using observations away from the cutoff  $c$ , we are extrapolating the estimated  $\mu_0(x)$  from the range of  $X_i$  we used for estimation to the cutoff  $X_i = c$

- For this to work, we need the  $\mu_d(x)$  to be continuous in a neighborhood around  $c$

## Why might continuity fail?

One of the main concerns people have with an RDD empirical application is that units are 'sorting' near the cutoff:

- E.g. if there is an SAT cutoff for a scholarship, some kinds of students who were close to making the threshold might retake the SAT multiple times

If students who retook look different than those that don't,  $\implies$  at a jump in  $Y_i(0)$  at the cutoff

- $c + \varepsilon$  is 'contaminated' by the retakers

## Why might continuity fail?

In general, the main intuition is that we want there to be no 'sorting' around the threshold

- If we think units' characteristics (observable and unobservable) are smooth over the cutoff, then it's reasonable to assume the potential outcomes are smooth too

## Why might continuity fail?

In general, the main intuition is that we want there to be no 'sorting' around the threshold

- If we think units' characteristics (observable and unobservable) are smooth over the cutoff, then it's reasonable to assume the potential outcomes are smooth too

For the covariates you do observe, can show that these do not jump at the cutoff

- Hopefully, the unobservables do not as well

## Using observations ‘near’ the cutoff

Returning to the notion of continuity, to estimate  $\mu_d(x)$  at the cutoff  $X_i = c$ , we want to use observations very close to the cutoff

- So that we extrapolate as little as possible

In finite samples, there might be very few observations near the cutoff

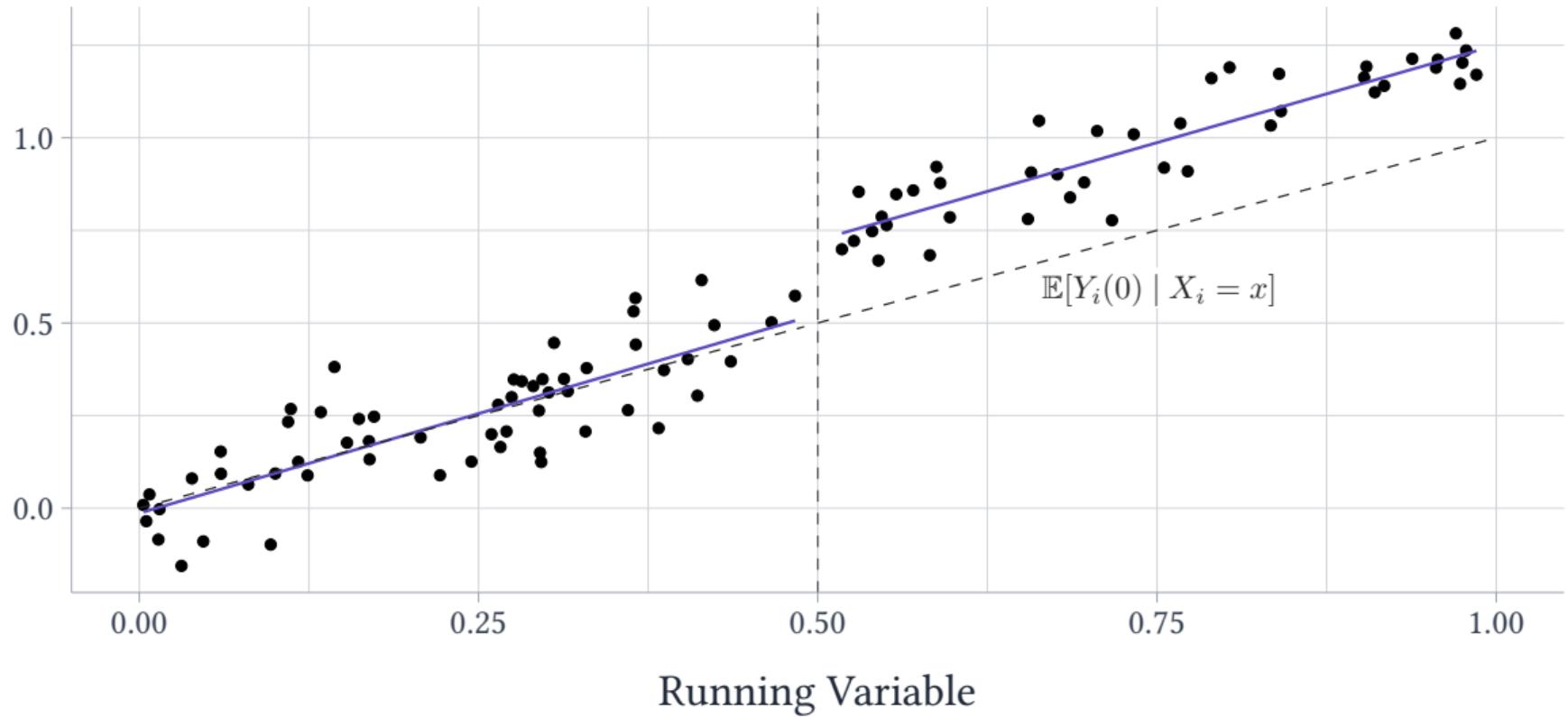
- In these settings, noise in the data can make it hard to learn information about the data-generating process

## Estimation of RDD

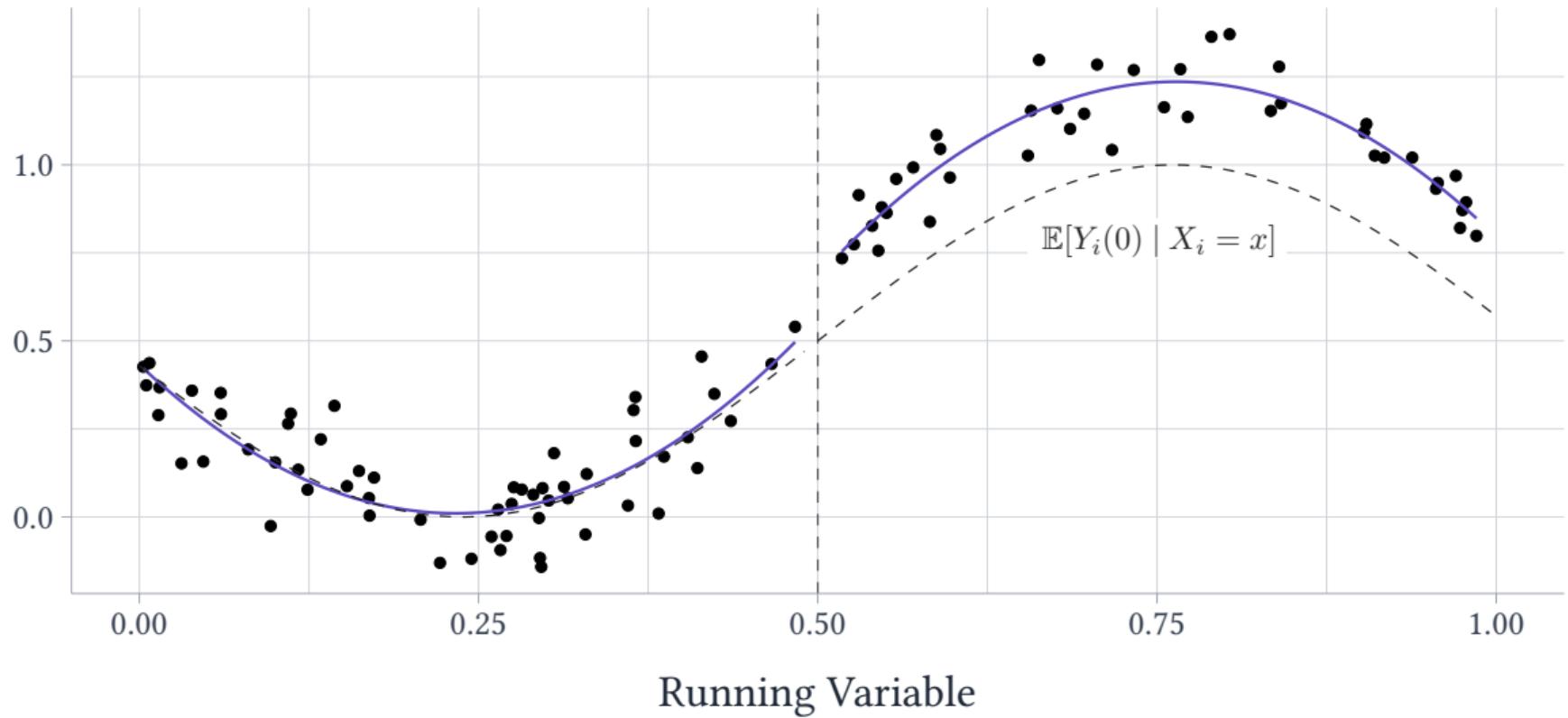
There are *many* different approaches to estimation of the RD coefficient, but there are a few common questions:

1. How large of a window around the cutoff should we use?
  - Smaller window relies on less ‘extrapolation’ but is ‘noisier’ (bias-variance trade-off)
2. How should we fit the model of  $\mathbb{E}[Y_i(d) \mid X_i = x]$ 
  - Simpler models are more robust to extrapolation, but may get the functional form wrong

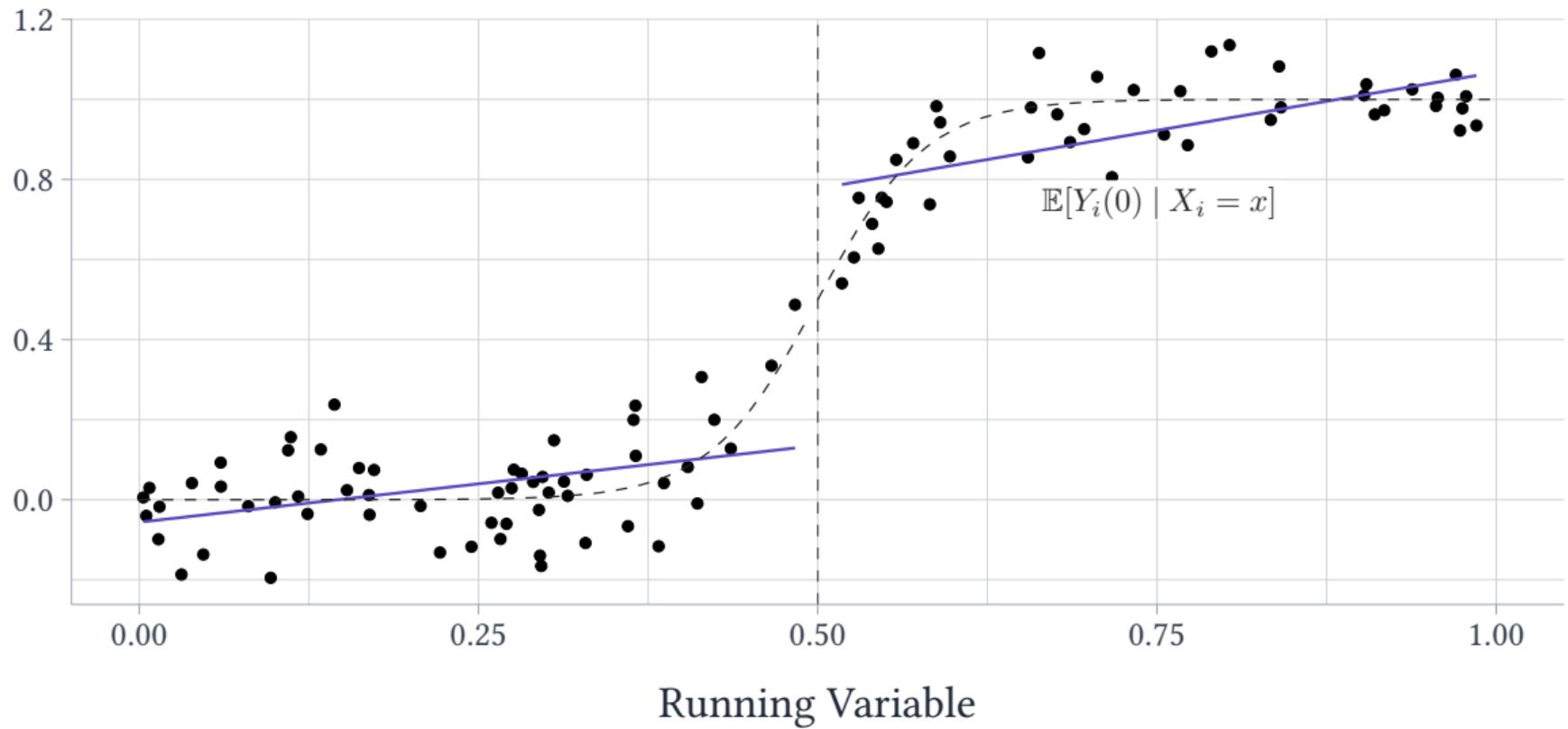
Linear  $\mathbb{E}[Y_i(0) \mid X_i]$



## Non-Linear $\mathbb{E}[Y_i(0) \mid X_i]$



## Non-Linear mistaken for linear



## Difference-in-means estimation

For now, take the window to be  $(c - h, c + h)$ ; we will return to the choice of  $h$  later in the slides

The simplest estimator is the *locally constant estimator*

$$\mu_0(x) = \hat{\mathbb{E}}[Y_i \mid X_i \in (c - h, c)] \text{ and } \mu_1(x) = \hat{\mathbb{E}}[Y_i \mid X_i \in (c - h, c)]$$

- This is our difference-in-means estimator

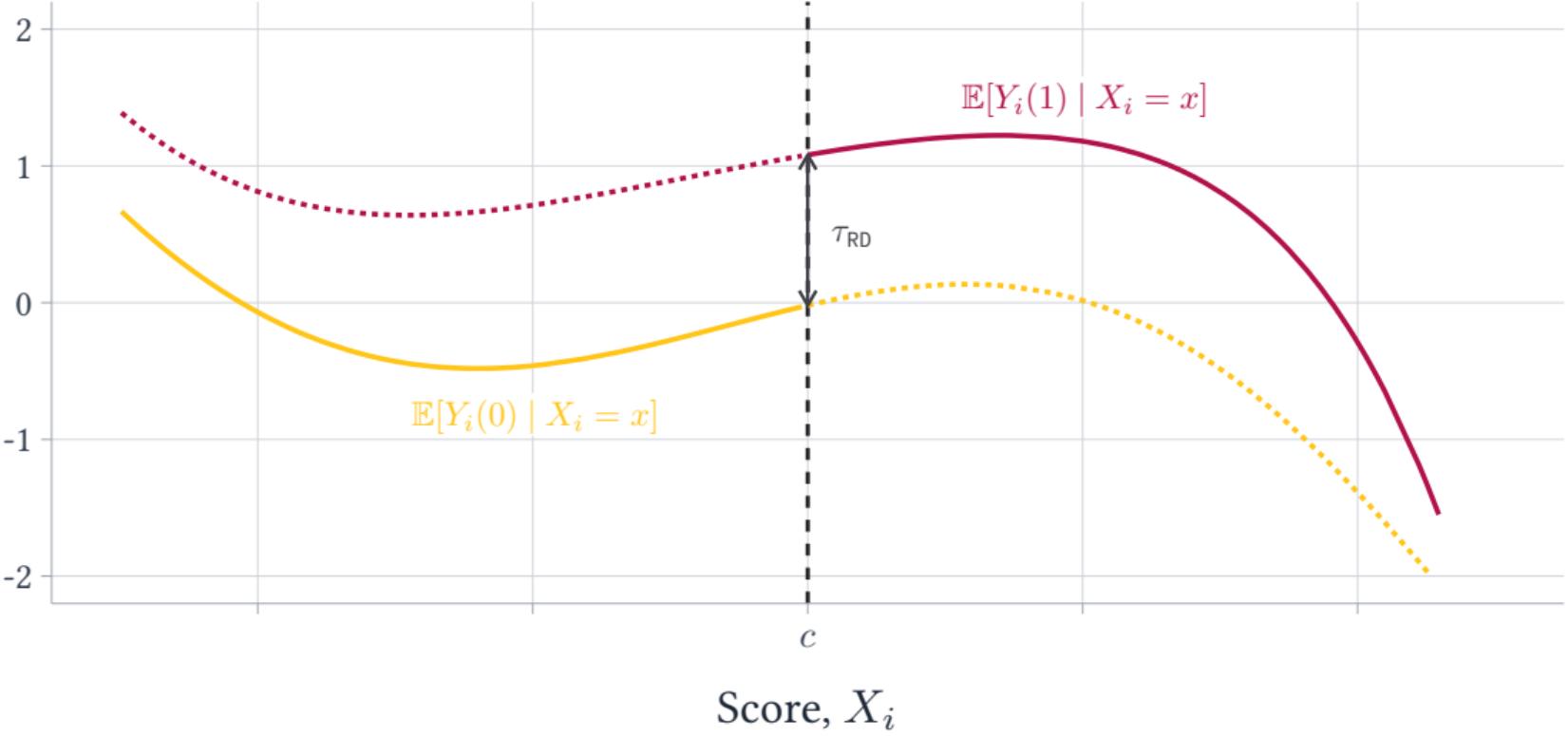
# Estimation via Regression

This can be estimated with the simple regression:

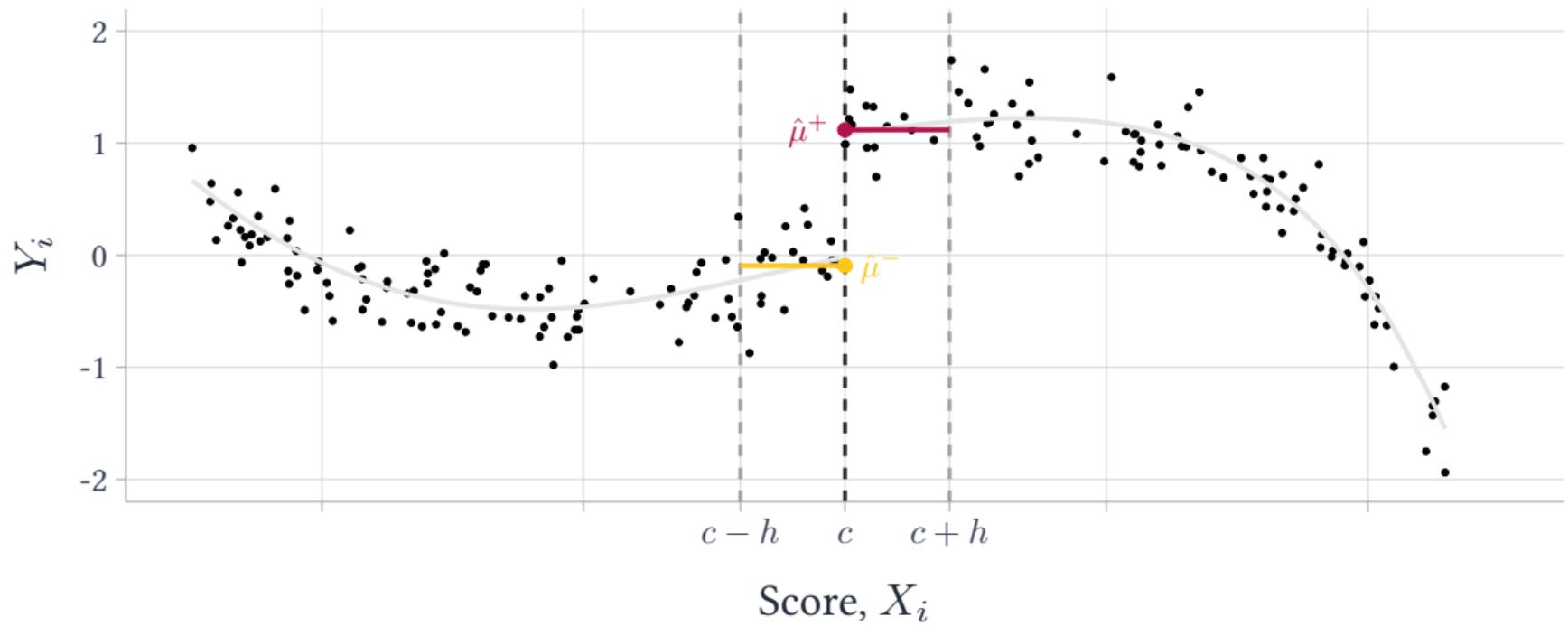
$$Y_i = \alpha_0 + \alpha_1 D_i + u_i$$

on the subsample with  $X_i \in (c - h, c + h)$

$Y_i(1)$  and  $Y_i(0)$



$$\hat{\tau}_{\text{RD}} = \hat{\mu}^+ - \hat{\mu}^-$$



## Locally-linear estimation

Now, let's use a linear-model  $\mu_d(x) = \alpha_d + \beta_d X_i$

- Note we let the slope vary for  $Y_i(0)$  and  $Y_i(1)$

Take our estimated models and form the regression adjustment estimator as:

$$\hat{\tau} = (\hat{\alpha}_1 + \hat{\beta}_1 c) - (\hat{\alpha}_0 + \hat{\beta}_0 c)$$

- This looks like our regression adjustment estimator!

## Estimation via Regression

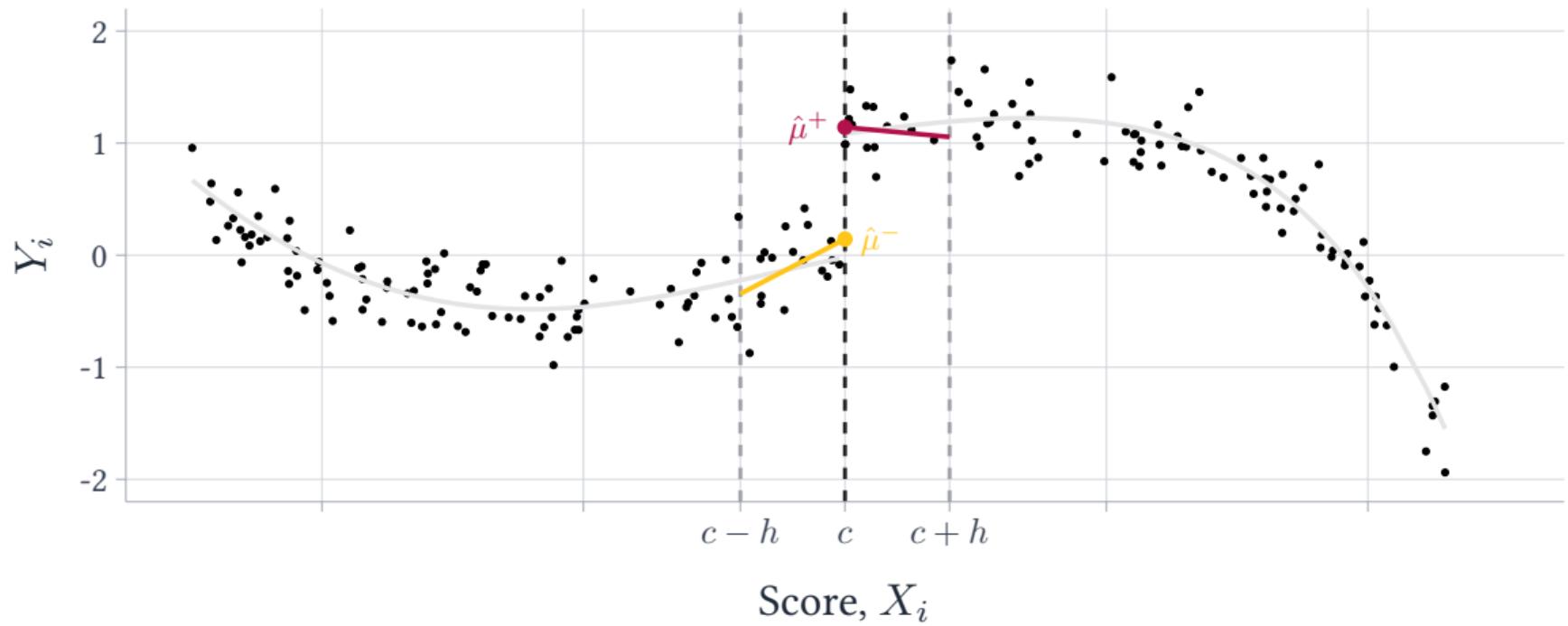
The locally linear estimator can be estimated with an interacted regression (like with regression adjustment):

$$Y_i = \alpha_0 + \alpha_1 D_i + \beta_0(X_i - c) + \beta_1 D_i(X_i - c) + u_i$$

on the subsample with  $X_i \in (c - h, c + h)$

- Note we recenter,  $X_i - c$ , so that  $\hat{\alpha}_1$  is the RDD estimate (e.g. "margin of victory" instead of "vote-share" in Lee, 2008)

$$\hat{\tau}_{\text{RD}} = \hat{\mu}^+ - \hat{\mu}^-$$



## Why linear if our data looks wiggly?

There is a lingering question, if we think our data is very 'wiggly' why are we assuming linearity in  $X_i$ ?

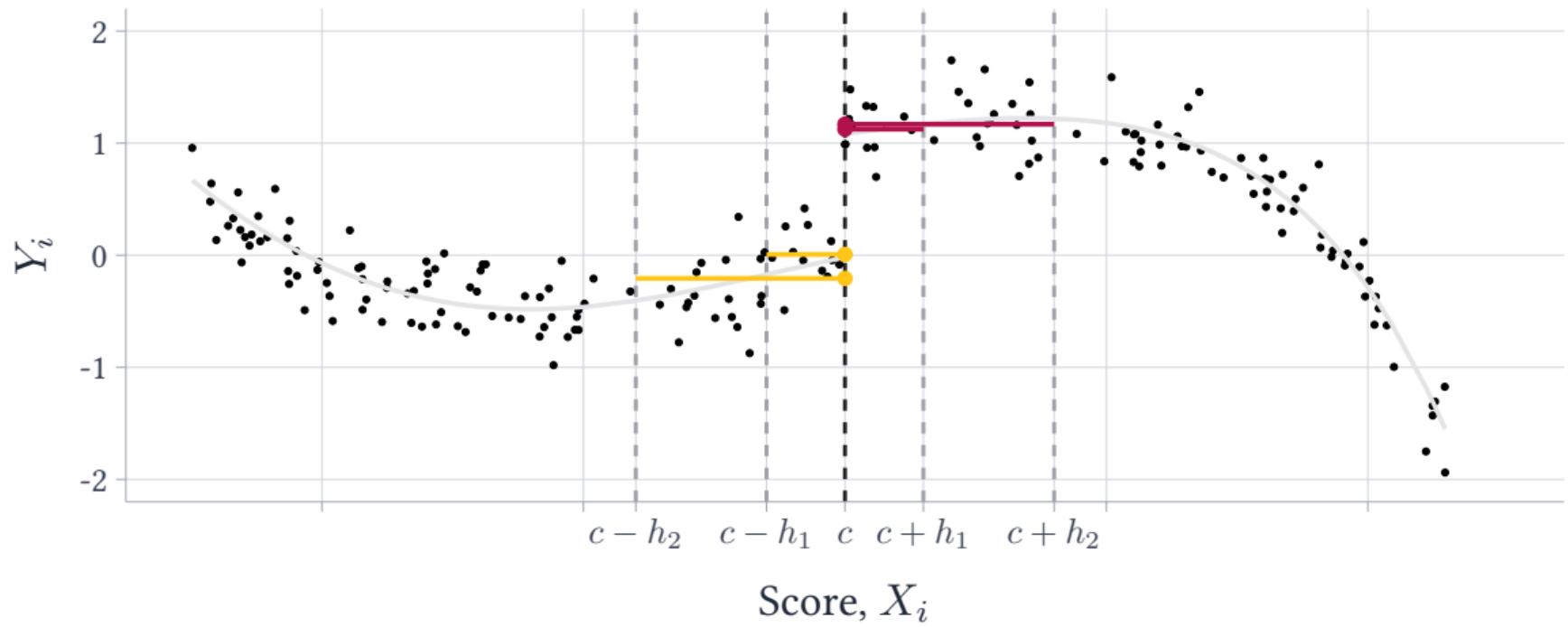
- Remember we are using only observations *local* to the cutoff, so (at least asymptotically) wiggly functions are approximately linear on small bandwidths
- Essentially we are using the logic of the Taylor expansion that we can approximate a function locally using a linear function

## Different Bandwidths

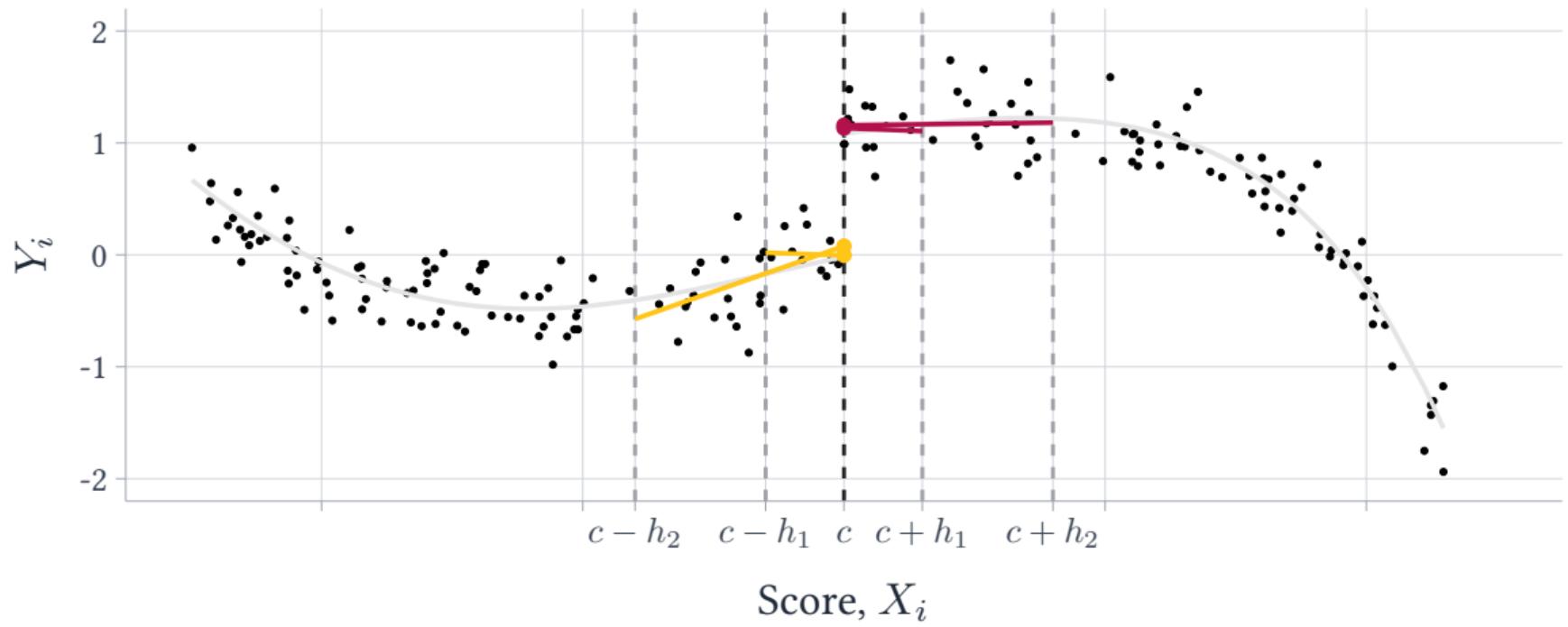
Our estimator, implicitly depends on our choice of bandwidth  $h$

- Larger  $h$  uses more observations so should help with precision of our estimate
- but relies more on functional form for extrapolation (Taylor approximation only holds locally)

$$\hat{\tau}_{\text{RD}} = \hat{\mu}^+ - \hat{\mu}^-$$



$$\hat{\tau}_{\text{RD}} = \hat{\mu}^+ - \hat{\mu}^-$$



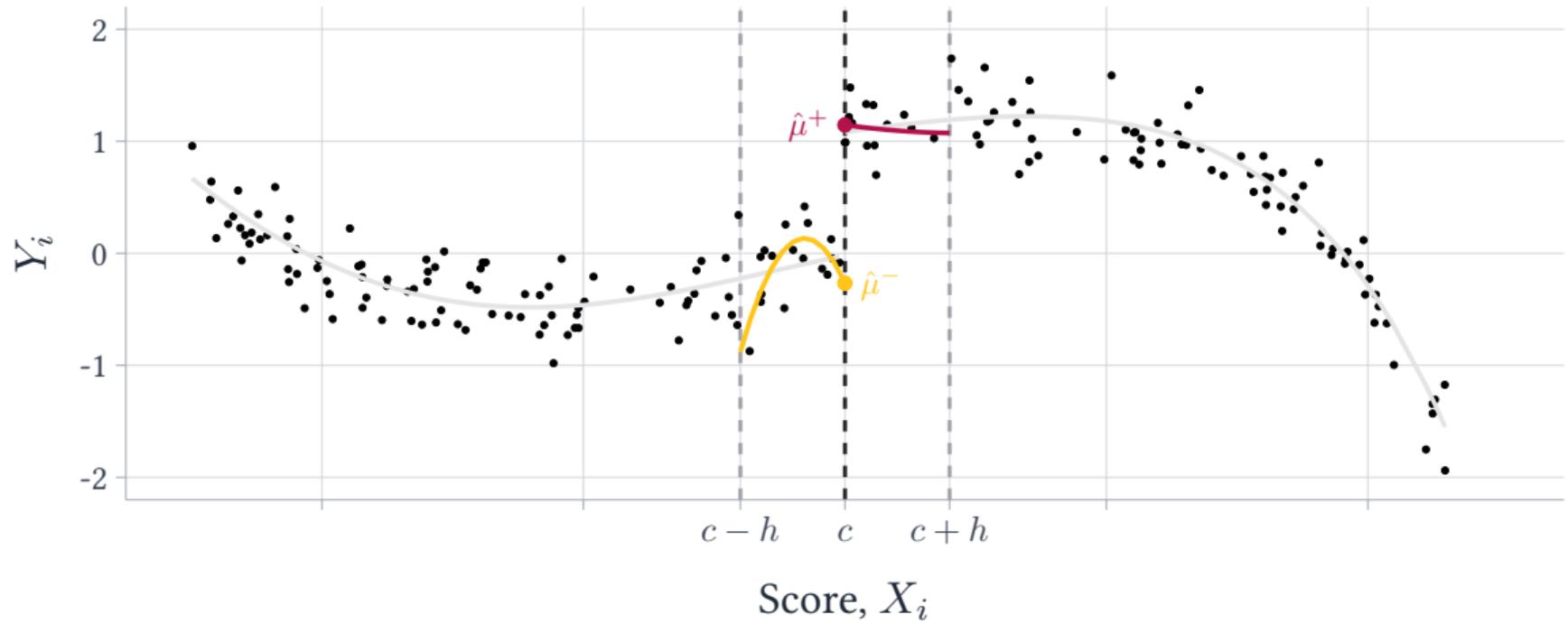
# Local polynomial estimation

We can extend our logic to higher-order polynomials:

$$Y_i = \alpha_0 + \alpha_1 D_i \sum_{p=1}^{\rho} \beta_{0,p} (X_i - c)^p + \sum_{p=1}^{\rho} \beta_{1,p} (X_i - c)^p + u_i$$

Still,  $\hat{\alpha}_1$  is our RDD estimator

$$\hat{\tau}_{\text{RD}} = \hat{\mu}^+ - \hat{\mu}^-$$



## RDD By Hand

The most ‘straight forward’ way to estimate the RDD is to do two regressions:

1. Regress  $Y_i$  on a  $k$ -th order polynomial of  $X_i - c$  for  $c - h < X_i < c$
2. Regress  $Y_i$  on a  $k$ -th order polynomial of  $X_i - c$  for  $c < X_i < c + h$

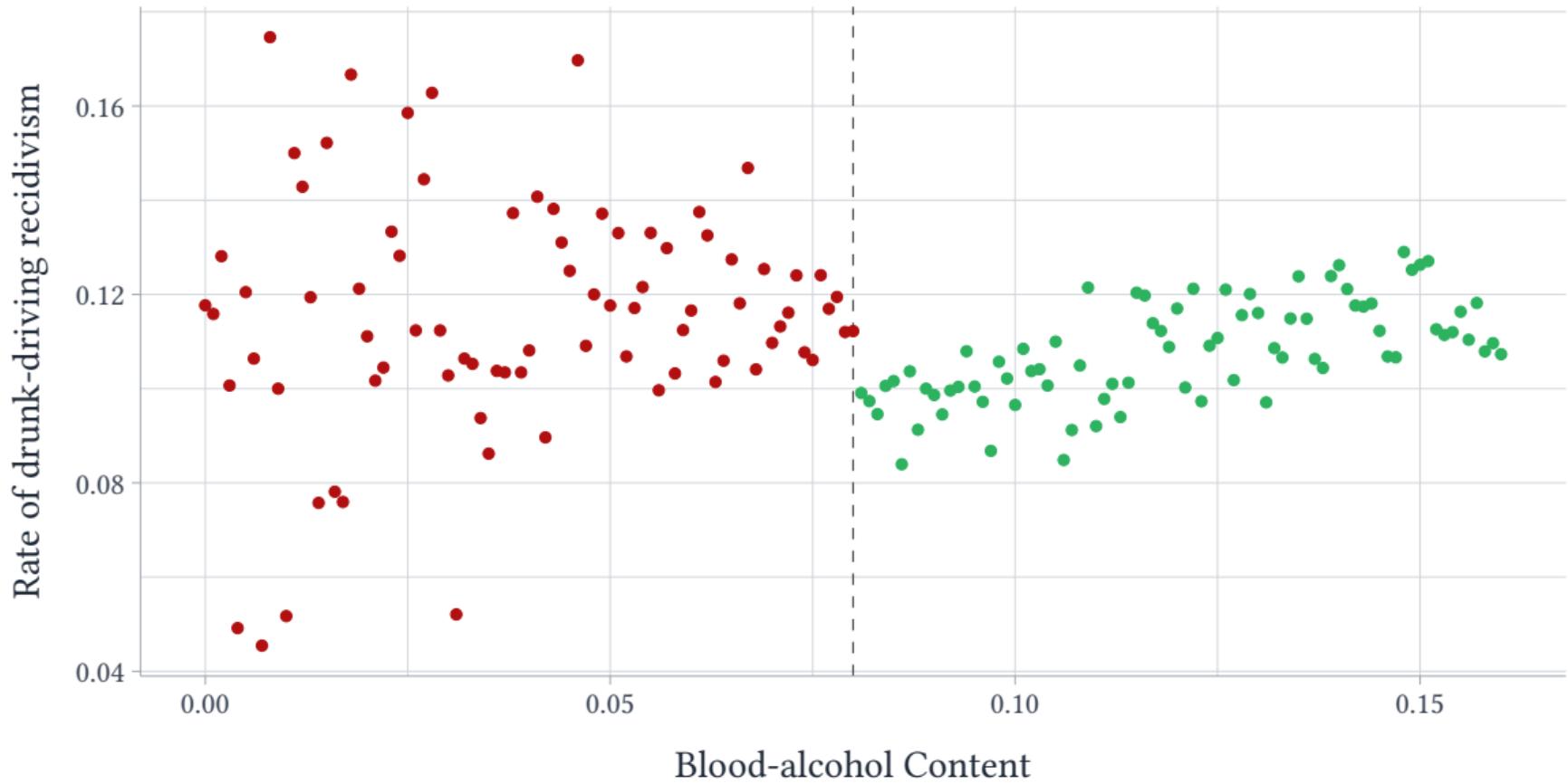
Predict  $Y$  at  $X_i = 0$  for both models. These are your estimates for  $\hat{\mu}^-$  and  $\hat{\mu}^+$  respectively. Then our RD estimate can be formed as

$$\hat{\tau}_{RD} = \hat{\mu}^+ - \hat{\mu}^-$$

## Example: Punishment and Deterrence: Evidence from Drunk Driving

Hansen (2015, AER) considers the impact of getting a DUI (driving while drunk) has on future drinking behavior

- Uses the 'legal limit' of a blood-alcohol content of 0.08 as the RDD 'cutoff'



## RDD by hand: Randomization-based estimation

```
feols(  
  recidivism ~ 1 + over_limit,  
  data = subset(hansen, bac1 >= 0.07 & bac1 <= 0.09)  
)
```

## RDD by hand: Randomization-based estimation

```
OLS estimation, Dep. Var.: recidivism
Observations: 18,897
Standard-errors: IID

            Estimate Std. Error   t value   Pr(>|t|) 
(Intercept)  0.114784  0.003253 35.28149 < 2.2e-16 ***
over_limitTRUE -0.017903  0.004472 -4.00349 6.2652e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## RDD by hand: Locally-linear estimation

```
feols(  
    recidivism ~ 1 + over_limit +  
    I(bac1 - 0.08) + I(bac1 - 0.08) * over_limit,  
    data = subset(hansen, bac1 >= 0.04 & bac1 <= 0.12)  
)
```

## RDD by hand: Locally-linear estimation

OLS estimation, Dep. Var.: recidivism

Observations: 72,880

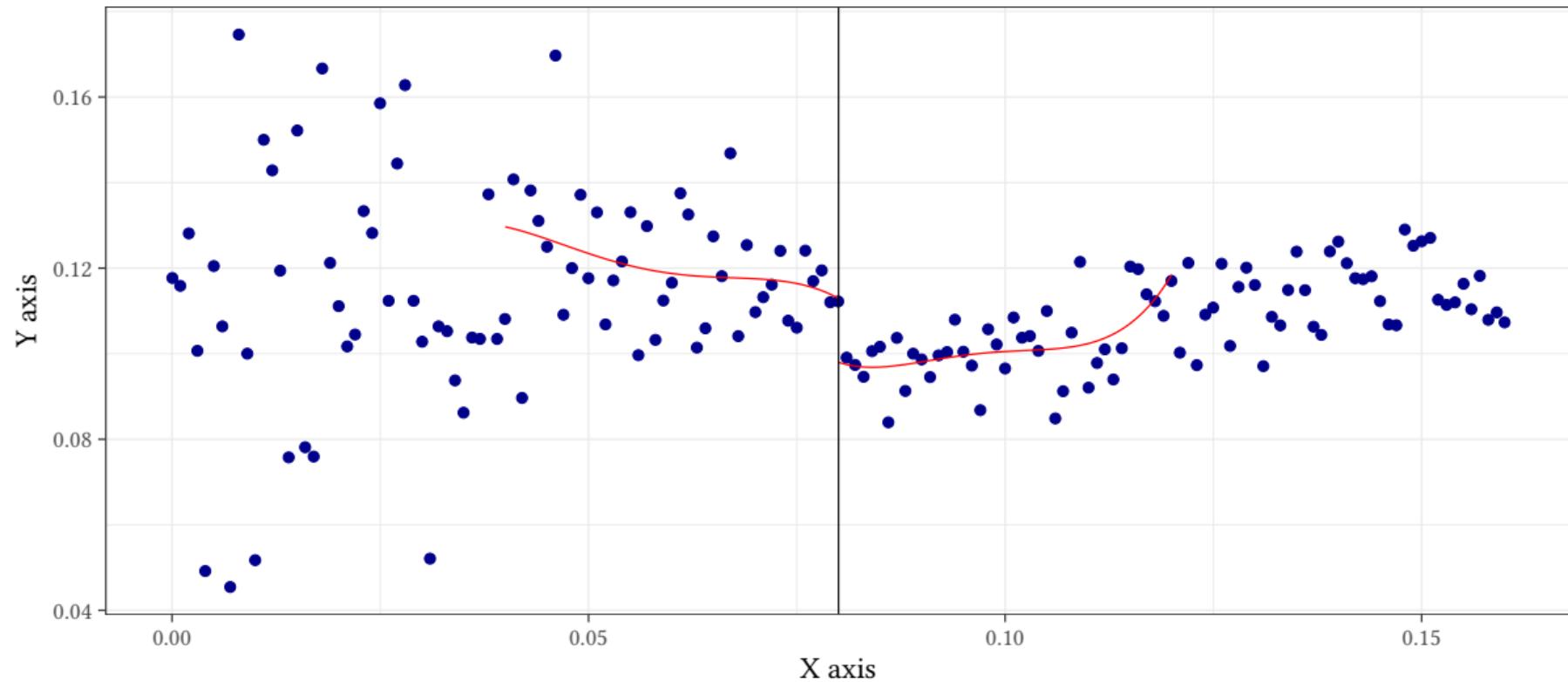
Standard-errors: IID

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.113956	0.003718	30.65215	< 2.2e-16 ***
over_limitTRUE	-0.020381	0.004694	-4.34212	1.4131e-05 ***
I(bac1 - 0.08)	-0.312061	0.226113	-1.38011	1.6756e-01
over_limitTRUE:I(bac1 - 0.08)	0.707713	0.254253	2.78350	5.3790e-03 **
---				
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
	0.1 ' '	1		

## RDD plot using rdplot from rdrobust

```
library(rdrobust)
rdplot(y = hansen$recidivism, x = hansen$bac1, c = 0.08, h = 0.4)
```

RD Plot



## RDD using rdrobust

```
est <- rdrobust(  
  y = hansen$recidivism, x = hansen$bac1, c = 0.08  
)  
summary(est)
```

## RDD using rdrobust

... Details ...

Method	Coef.	Std. Err.	z	P> z	[ 95% C.I. ]
Conventional	-0.018	0.006	-3.083	0.002	[-0.030 , -0.007]
Robust	-	-	-2.452	0.014	[-0.031 , -0.003]

## Sorting based on the running variable

As we discussed before, our main threat to the continuity assumption is that units are 'sorting' onto either side of the cutoff based on their observable characteristics

There are two standard 'gut checks' that folks will want to see

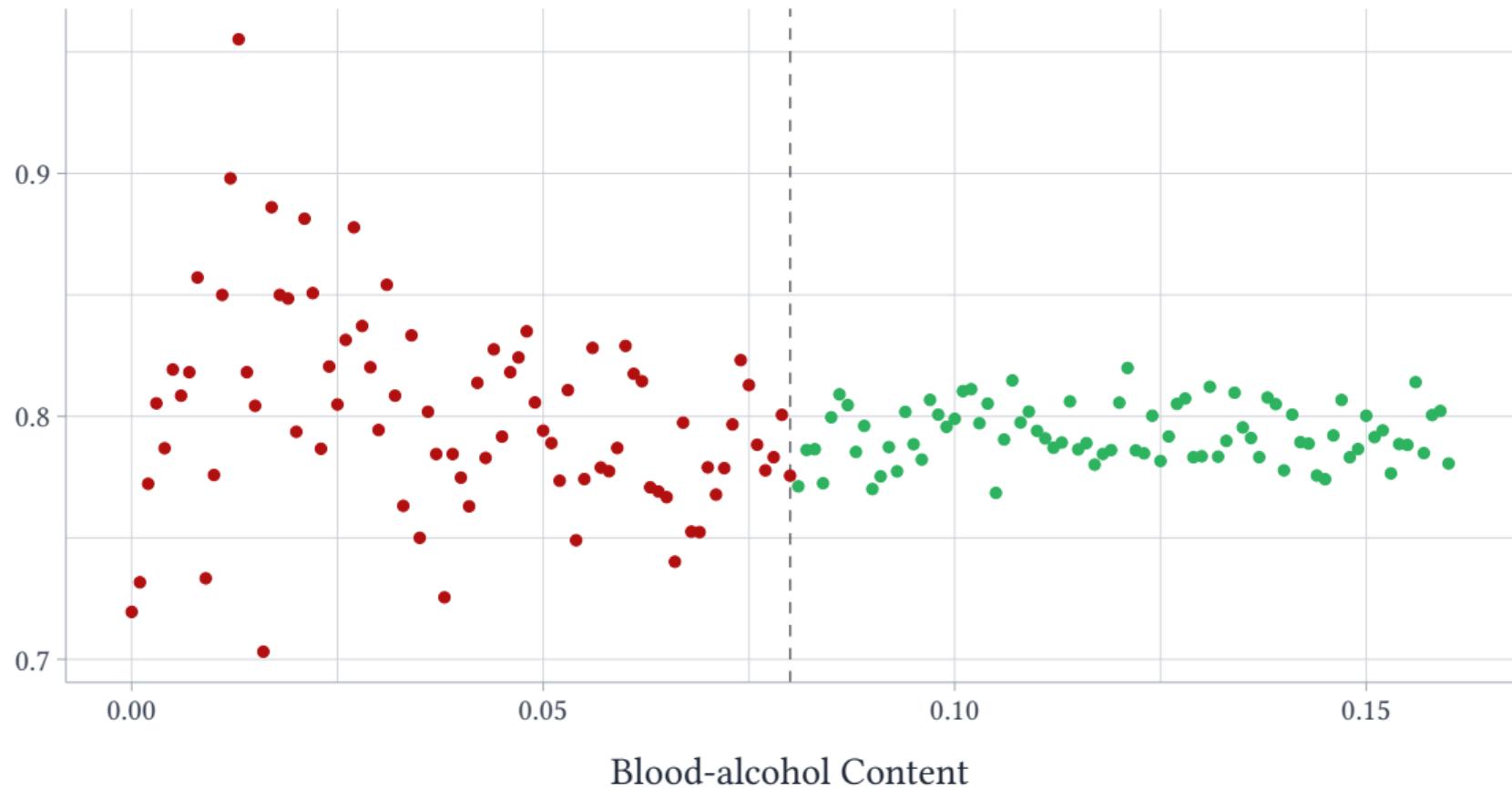
## Check 1: Balance

First, if we have any ‘pre-determined’ covariates, we want see that there are no jumps in the average of covariates at the cutoff

- Rerun RDD estimator using  $X_i$  as the outcome variable. Test  $\hat{\tau}_{RD} = 0$

Use covariates that you would likely see a jump in if there was sorting (e.g. a proxy for ‘motivation’ for SAT example)

Percent of drivers that are male



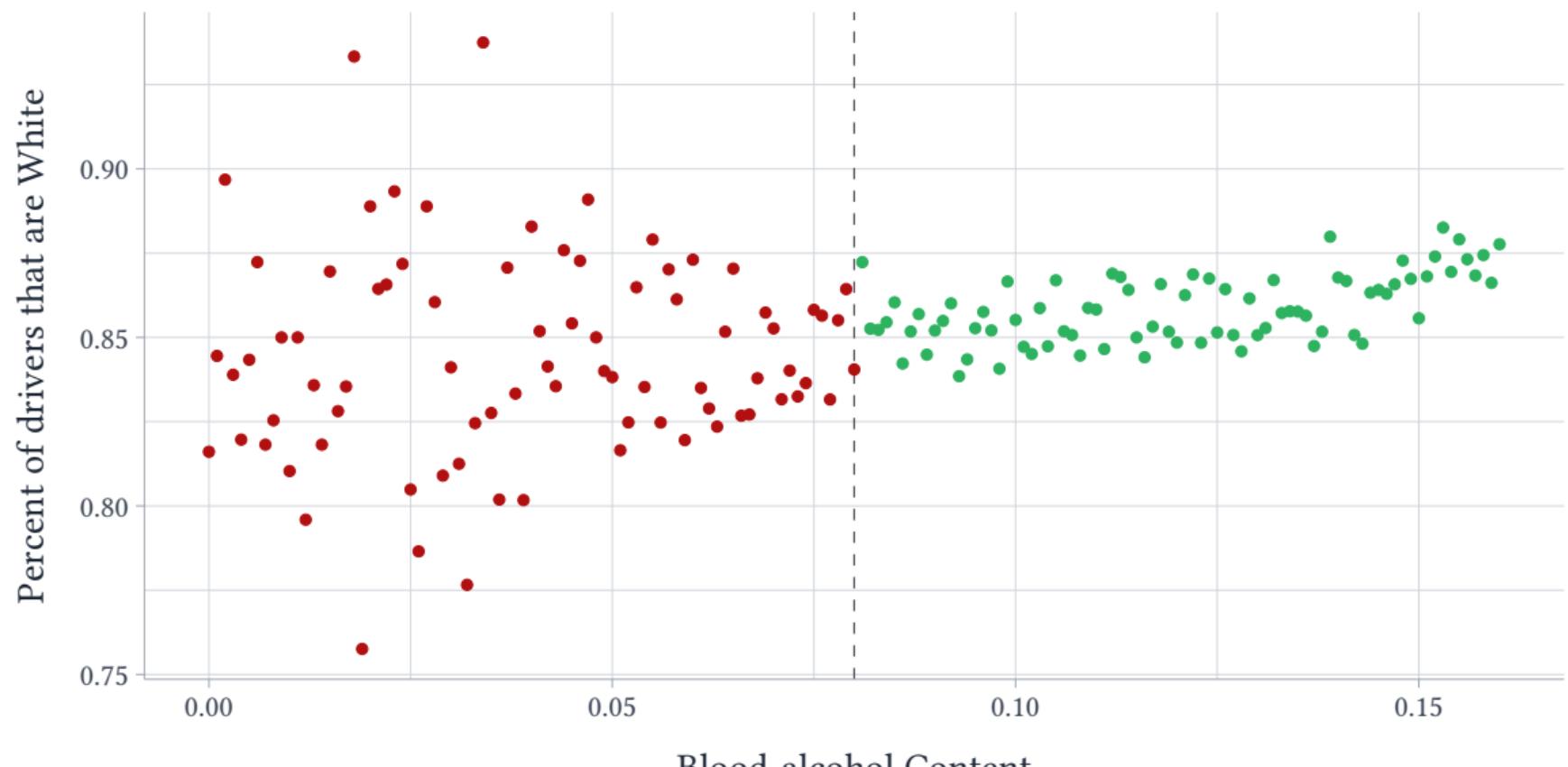
Percent of drivers that are White

0.00 0.05 0.10 0.15

Blood-alcohol Content

0.75 0.80 0.85 0.90

A vertical dashed line is drawn at a Blood-alcohol Content of approximately 0.08.

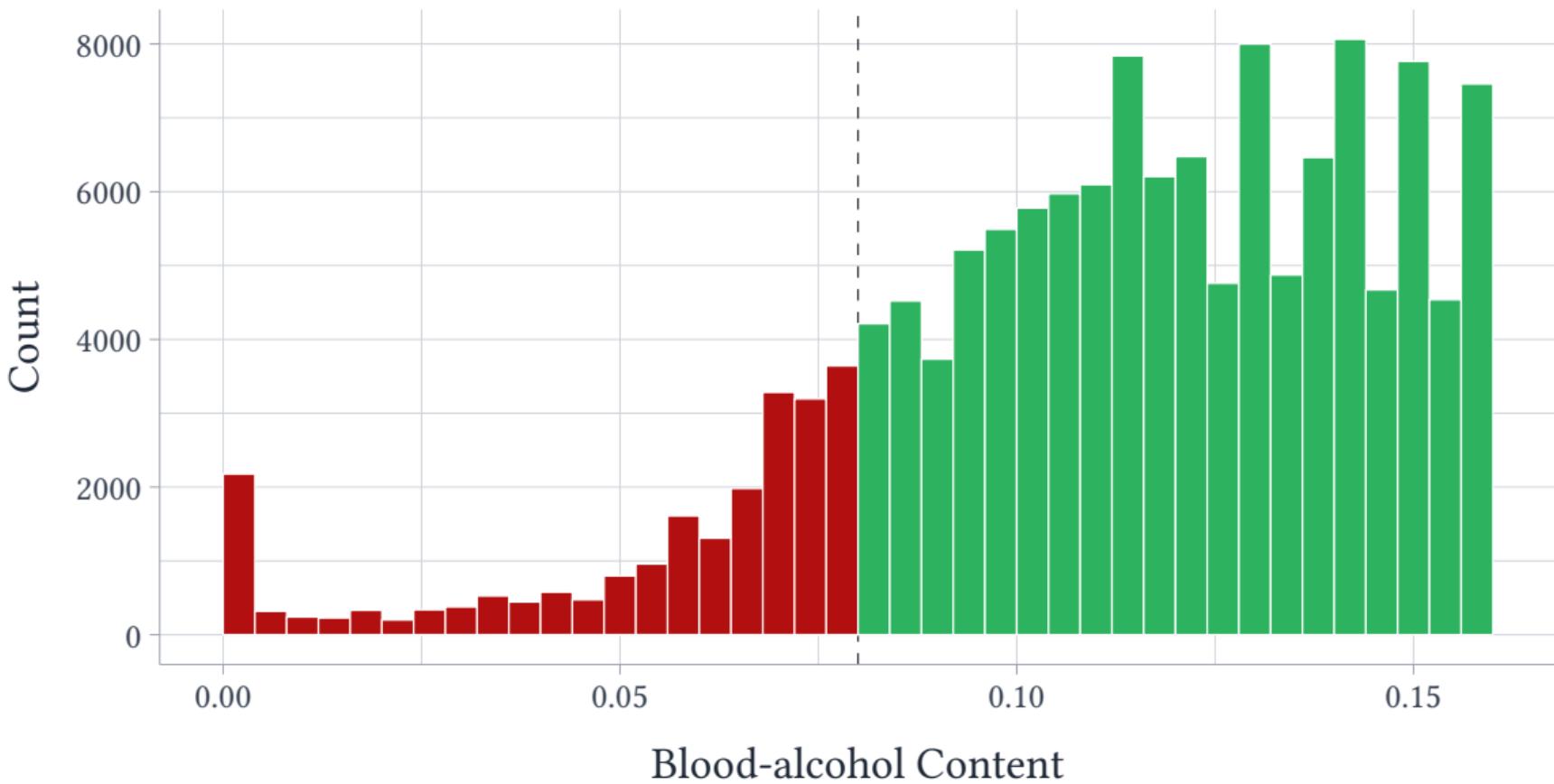


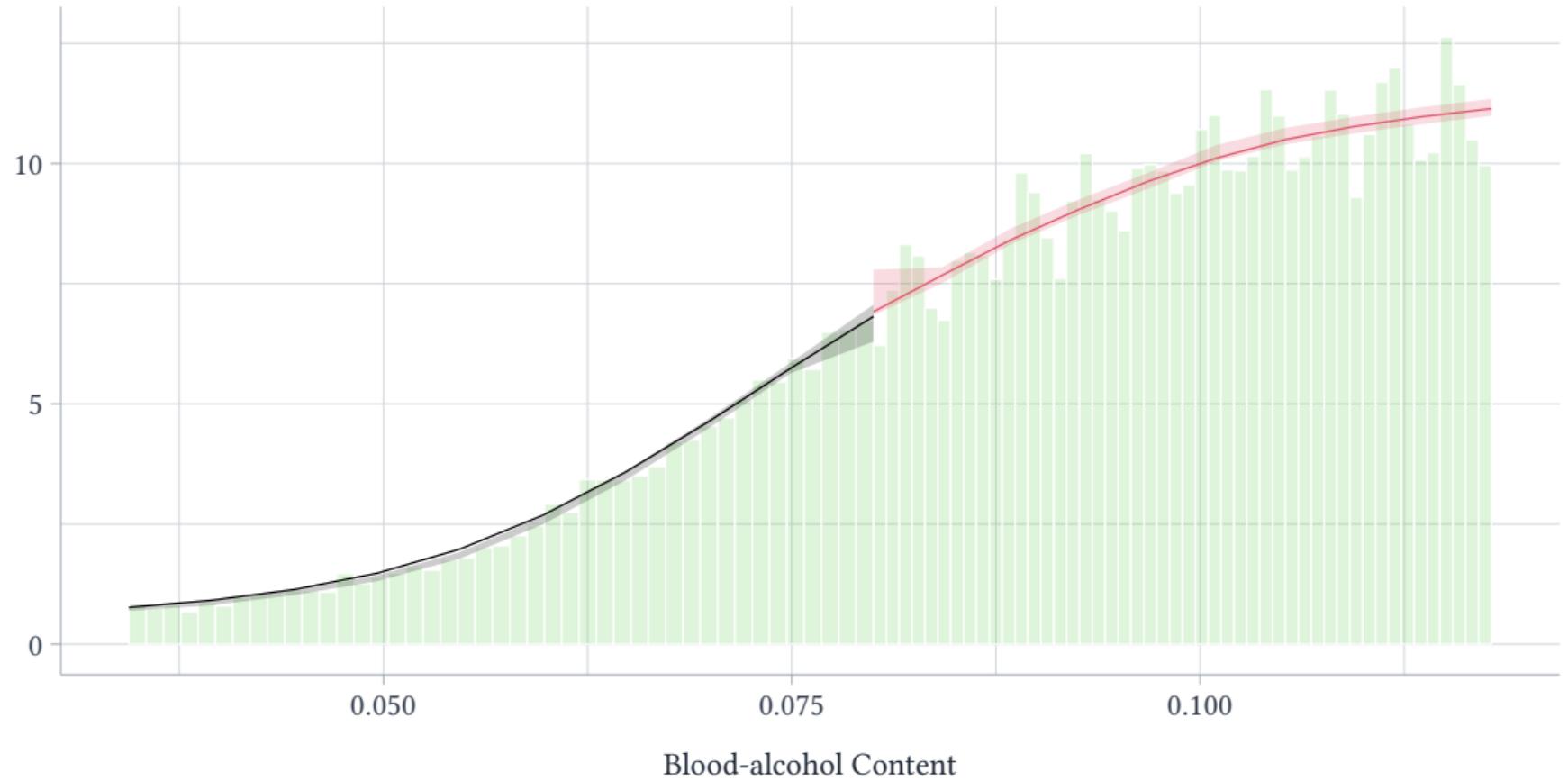
## Check 2: McCrary Density

If you don't have extra covariates, you can check for 'continuous density' a la McCrary (2006, JOE)

The idea is that if we think there is no sorting, then we would expect to see no strong 'jump' in the histogram of the running variable

- A jump at the discontinuity suggests that agents are sorting onto one side of the border (assuming the density would be smooth in the absence of the cutoff policy)



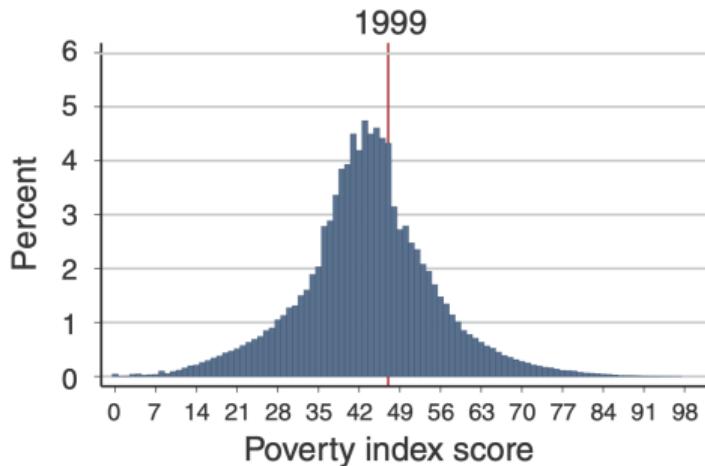
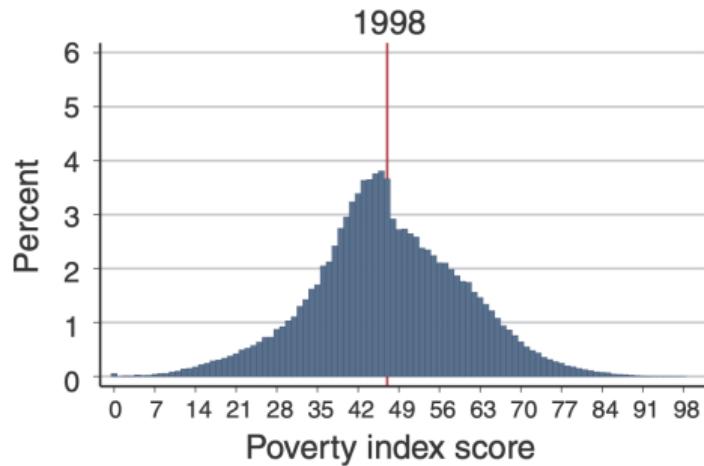
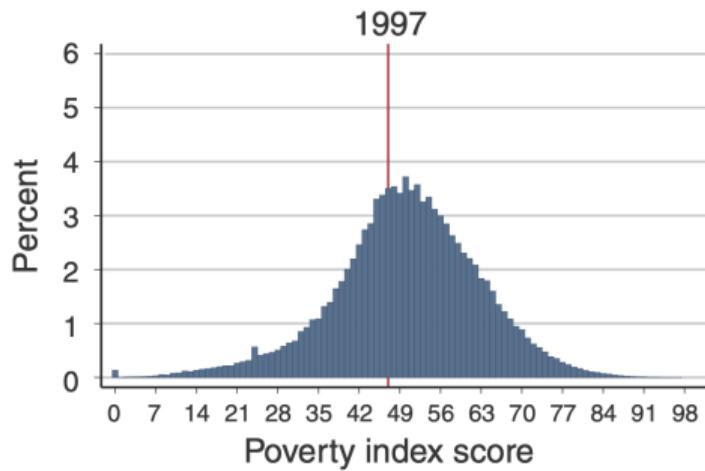
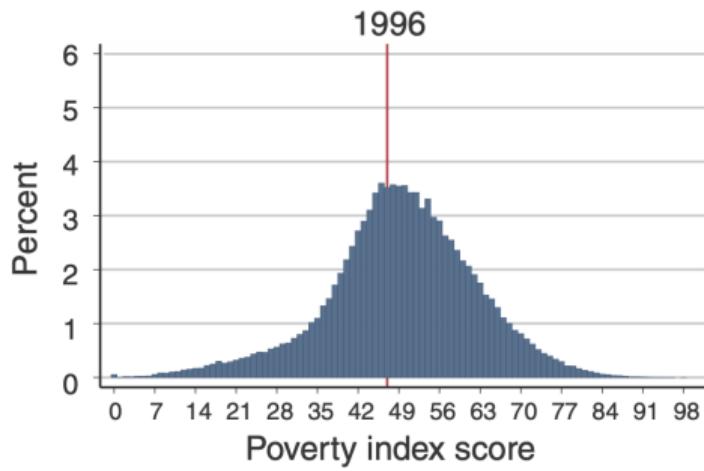


## Example

*Camacho and Conover (2011, AEJ: EP)*

Camacho and Conover (2011, AEJ: EP) discuss a real example of 'manipulation'

- Towns in Colombia receive social programs if their poverty index score is below a cutoff
- The poverty index algorithm becomes public in 1997



## Example

*Camacho and Conover (2011, AEJ: EP)*

The authors show in their paper that the towns that corruptly gamed their score looked different than those that did not

- Those that gamed their scores had more political competition

In this setting, if we saw a big jump at the poverty index cutoff, then we can't know if it's from the social programs or from the higher-level of political competition

## Optimal Bandwidth selection

As we discussed before, there is a tension with selection of bandwidth:

- Smaller bandwidths use data closer to the cutoff and rely less on functional form
- Larger bandwidths lets you use more data to estimate RD

All identification relies on the bandwidth getting smaller as the sample size grows

- But we are left wondering what the 'optimal bandwidth' is given a finite sample

## Optimal Bandwidth selection

Calonico, Cattaneo and Farrell (2018, Economics Journal) use statistical methods to identify the 'optimal' bandwidth. The optimal bandwidth is based on the following:

- Of course, the bandwidth should shrink as the sample size grows
- If the curvature of  $\mathbb{E}[Y_i | X_i = x]$  is more wiggly, the bandwidth should be smaller
- If the variance of  $Y_i$  near the cutoff is larger, then you should use a larger bandwidth

## Optimal Bandwidth selection

Calonico, Cattaneo and Farrell (2018, Economics Journal) use statistical methods to identify the ‘optimal’ bandwidth. The optimal bandwidth is based on the following:

- Of course, the bandwidth should shrink as the sample size grows
- If the curvature of  $\mathbb{E}[Y_i | X_i = x]$  is more wiggly, the bandwidth should be smaller
- If the variance of  $Y_i$  near the cutoff is larger, then you should use a larger bandwidth

The upshot is this is done automatically for you with `rdrobust`

- See Cattaneo et. al.’s review article for more details

# Roadmap

Formalizing RDD Identification Arguments

Spatial RDD

# Spatial RDD

## *Example*

Dell (2010, ECTA) considers the impact of historical colonizer institutions on long-run impacts:

- In now Peru and Bolivia, there was a set of Spanish mines that had a catchment area called the 'Mita'

The boundary was drawn somewhat arbitrarily around the mines, so either side of the Mita boundary should have the same 'latent opportunity' (in 2000) except for the fact of the Mita boundary

- This argument is that  $\mathbb{E}[Y(0)]$  would be continuous right around the border

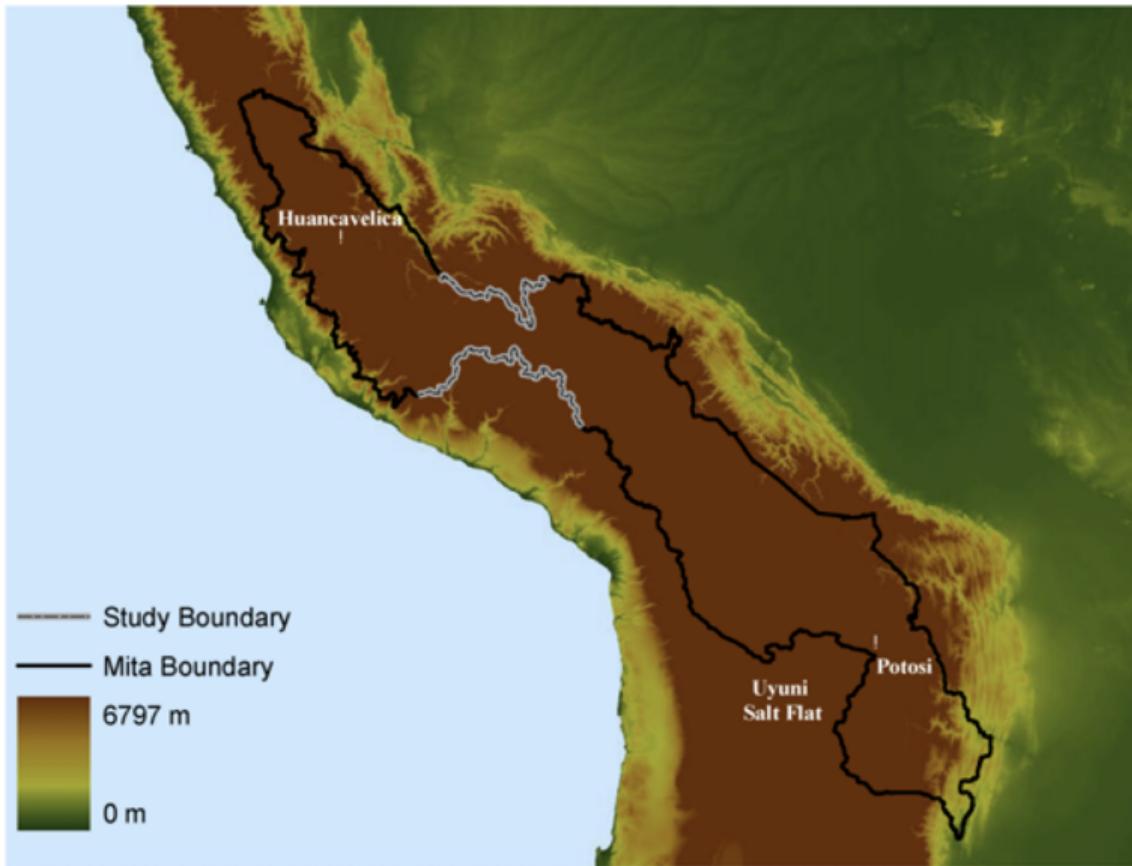
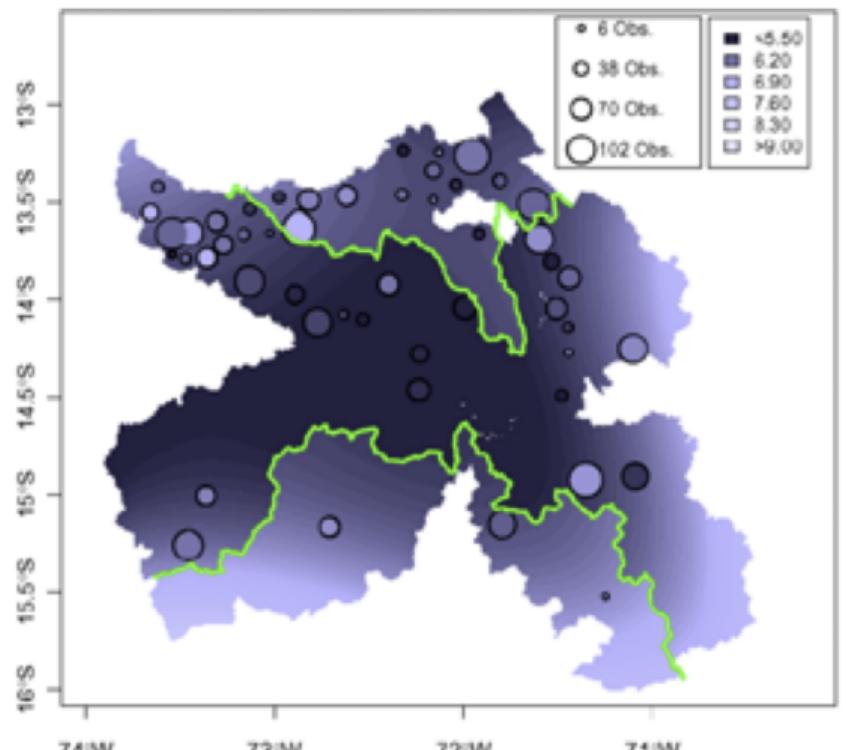
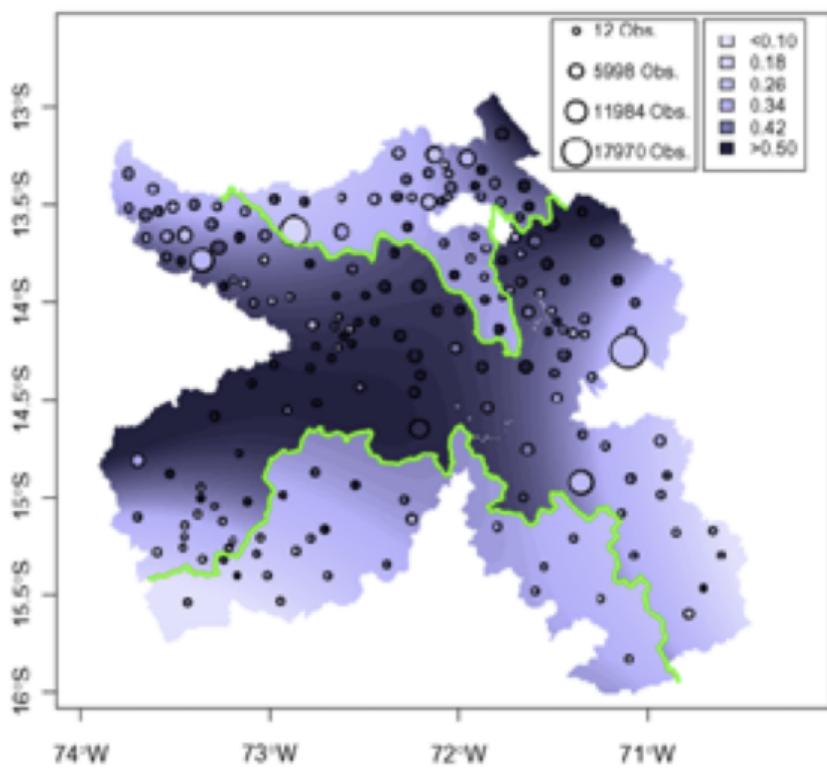


FIGURE 1.—The *mita* boundary is in black and the study boundary in light gray. Districts falling inside the contiguous area formed by the *mita* boundary contributed to the *mita*. Elevation is shown in the background.



(a) Consumption (2001)



(b) Stunting (2005)

## Spatial RDD Setup

Units have a location in geographic space  $(x_{1,i}, x_{2,i})$  and there is some border consisting of lines in  $X_1/X_2$  space

- The borders will separate units into  $D_i = 0$  and  $D_i = 1$  space.

## Spatial RDD Setup

Units have a location in geographic space  $(x_{1,i}, x_{2,i})$  and there is some border consisting of lines in  $X_1/X_2$  space

- The borders will separate units into  $D_i = 0$  and  $D_i = 1$  space.

Since we are in two-dimensions, it's not immediately clear how to think about "approaching the border". For now, let's use the Euclidean distance to the closest point on the border

- Positive distances are within  $D_i = 1$  area; negative distances are outside  $D_i = 1$  areas

## Spatial RDD Idea

Let  $\text{dist}_i$  be the distance to the closest border with a positive sign if on the  $D_i = 1$  area and negative sign if in the  $D_i = 0$  area

- The cutoff  $c$  is therefore  $\text{dist}_i = 0$

The RDD identification assumption is the same as before viewing  $\text{dist}_i$  as the single running variable

- We assume  $\mathbb{E}[Y_i | \text{dist}_i = d]$  does not jump at the boundary

## Discontinuity in Dell (2010)

In the context of Dell (2010), the assumption is that in the absence of the Mita boundaries, the outcomes

- Most of the boundary was at the base of the Andes mountain, so this assumption is implausible there
- Dell argues that the part of the boundary she uses was a bit 'randomly drawn' in the sense that it was not placed on some obvious cutoff

To bolster her argument, she presents some evidence that the boundary was not

## Discontinuities are much more likely with spatial boundaries

Turner et. al. (2014, ECTA) used data at the city/town-level for their analysis, so their borders primarily were at the town level

- They need to make the strong assumption that on either side of a town-border did not differ in ways that would impact home prices in a way that is correlated with zoning laws

This seems unlikely because, for example, towns with more restrictive zoning laws tend to have higher school budgets

## Multiple Treatments at Boundaries

We have already discussed one concern regarding sorting onto sides of the border. Another common problem is that multiple treatments might turn on at a border

- E.g. many policy differences across town borders

## Multiple Treatments at Boundaries

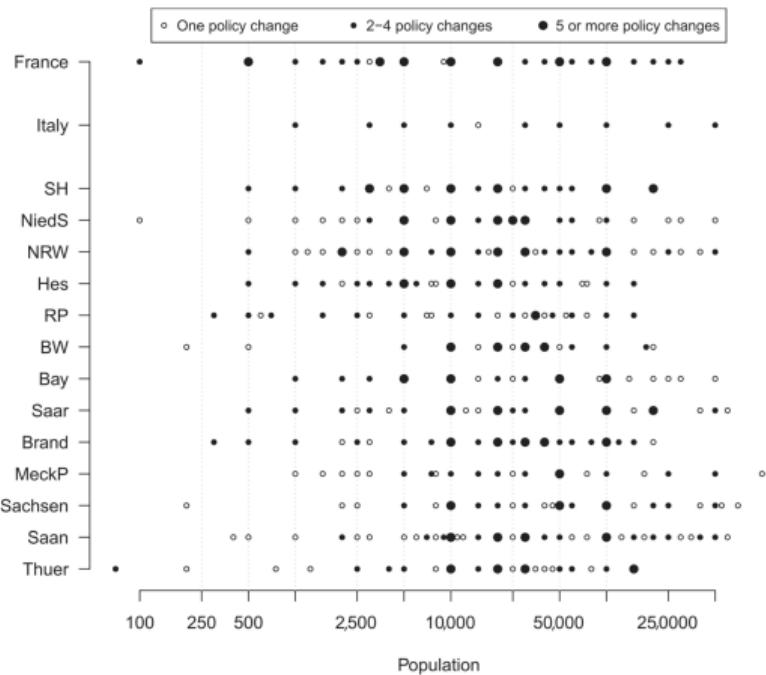
Grembi et. al. (2018, AJPS) look at a common occurrence of 'population thresholds' for policies to take place

- Since policymakers chose arbitrary numbers, they usually pick nice round numbers
- Also common is some multiple of the federal poverty line

This creates a bundle of different treatments that turn on at a threshold

- How do you disentangle effects from different policies?

**FIGURE 1 Population Thresholds at Which Municipal Policies Change: France, Italy, and German States**



*Note:* Each dot indicates a population threshold at which a policy changes; solid dots indicate more than one policy changing at the same threshold.

From Grembi et. al. (2018, AJPS)

Many 'round' population thresholds have multiple policy changes

## Solution Idea #1: Find better cutoffs

The first idea is to find better borders where we more confidently think there aren't other policy changes

E.g. Kulka, Sood, and Chiumenti (2023, WP) look at the impact of zoning changes on home prices like in Turner et. al.

- They have zoning microdata and only compare zoning borders that fall within town and school-districts
  - Compare homes within the same town and school district on opposite side of zoning borders

## Solution Idea #2: Compare before and after a policy is put into place

In some cases, you can observe units before vs. after a new policy is put into place

If all other policies are in place and all sorting has already taken place, we can estimate an RD before and an RD after

- The RD before the new policy is put into place tells you about the jump in  $Y_i(0)$  that already exists (from sorting and the treatment effect of other policies)

## Solution Idea #2: Compare before and after a policy is put into place

In some cases, you can observe units before vs. after a new policy is put into place

If all other policies are in place and all sorting has already taken place, we can estimate an RD before and an RD after

- The RD before the new policy is put into place tells you about the jump in  $Y_i(0)$  that already exists (from sorting and the treatment effect of other policies)
- The RD after the new policy still has the previous jump, but also a new jump from the new policy

The 'difference-in-jumps' estimates the effect of the new policy

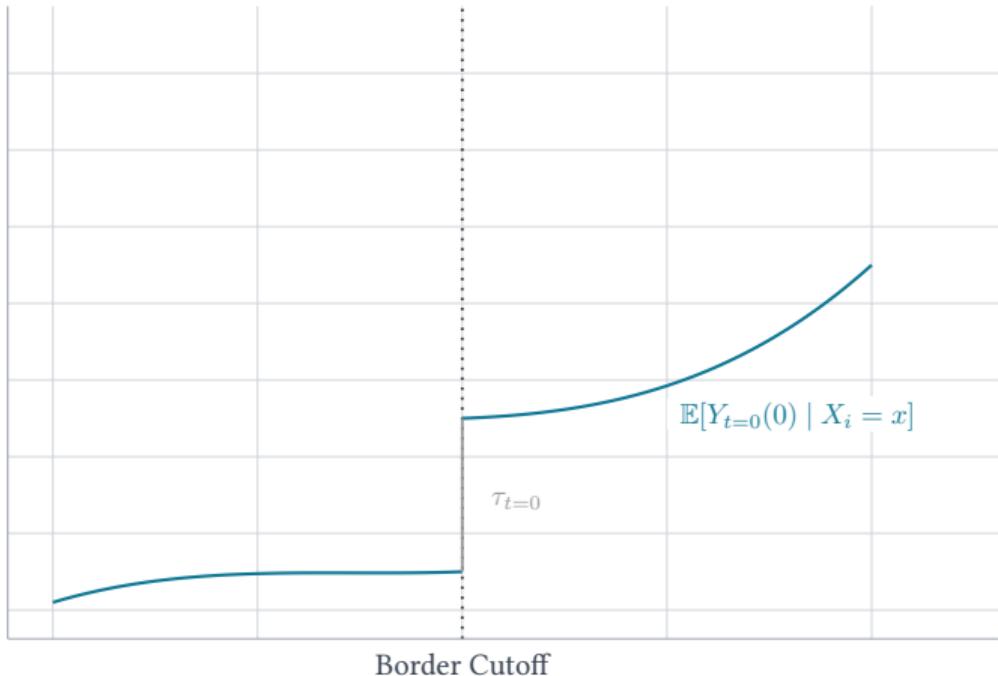
## Difference in discontinuities

This method was first proposed by Grembi et. al. (2016, AEJ Applied) and they coined it 'difference in discontinuities' (we will see why later)

This relies on two assumptions:

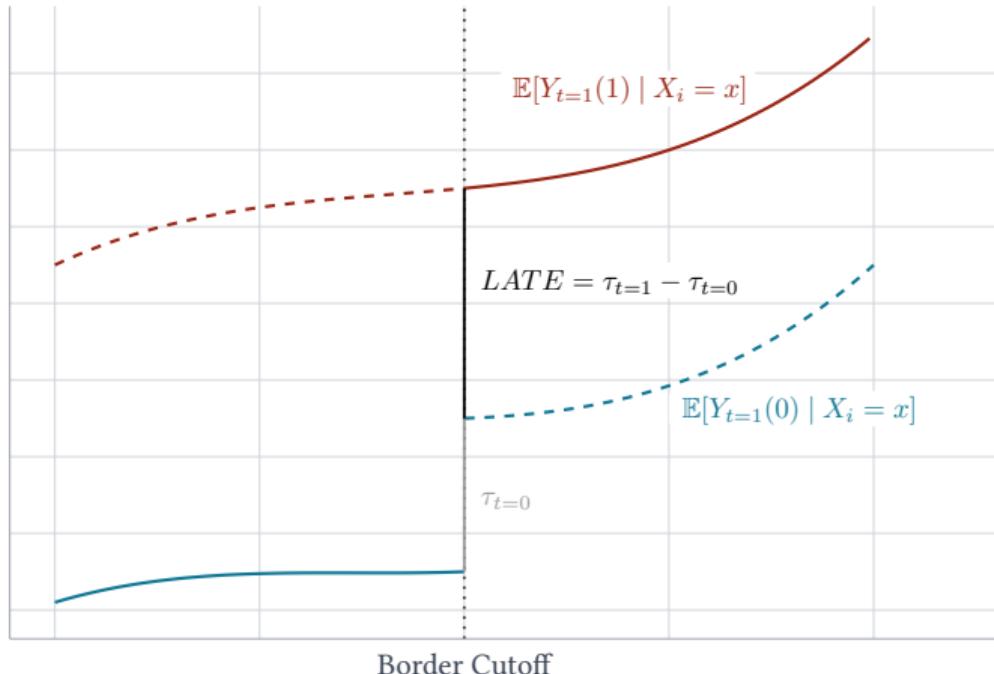
1. The jump in  $Y_i(0)$  stays constant over time, so that we can estimate it using the pre-period
  - There is no *new* effect and no *new* sorting from *previous* treatments
2. The  $Y_i(1)$  curve will also have a jump of the same magnitude as  $Y_i(0)$ 
  - There is no *additional* sorting occurring from the *new* policy

## RDD at $t = 0$



Pre-treatment RD informs us  
of effects of other policy  
changes and non-treatment  
based-sorting

## RDD at $t = 1$



Difference of post RD and pre RD is the treatment effect of interest

## Implementing Diff-in-disc

Butts (2021, AEL) (that's me) shows an easy way to estimate the difference-in-discontinuities:

- Instead of doing two RDs and taking the difference, do an RD on first-differenced data

i.e. use `rdrobust` using `df$y1 - df$y0` as the outcome variable