

# MQE: Economic Inference from Data:

## Module 6: Regression Discontinuity

Claire Duquennois

6/9/2020

# Regression discontinuity research designs

Introduced in other fields as far back as the 1960s.

Gain popularity in economics in the past 20 years or so as economists:

- ▶ increasingly focus on causal inference
- ▶ large administrative datasets became more widely available

When correctly applied, RD designs are very transparent in how they achieve causal identification which makes them very appealing.

# Regression discontinuity research designs

RD designs leverage the researchers knowledge of a rule or policy that determines treatment.

Identification comes from how some rules are applied in a fairly arbitrary way.

This arbitrary application generates randomness we can exploit to estimate causal effects.

# The Set Up

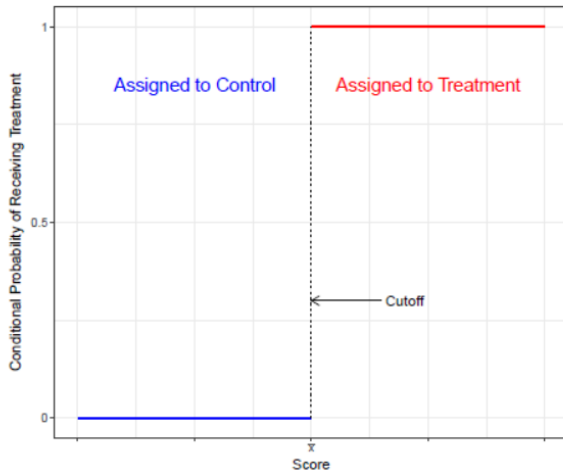
Suppose that we want to estimate the effect of some binary treatment  $D_i$  on an outcome  $Y_i$ . Using the potential outcomes framework:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

Suppose  $D_i$  is completely (or partially) determined by whether some predictor  $R_i$  lies above or below a certain threshold,  $c$ .

## RD Set up:

Figure 2.1: Conditional Probability of Receiving Treatment in Sharp RD Design



# The Set Up

The “running variable”  $R_i$  need not be randomly assigned:

- ▶  $R_i$  is related to  $Y_i$
- ▶ absent treatment, this relationship is smooth ( $Y_i$  does not jump discontinuously as  $R_i$  changes).

$\Rightarrow$  any discontinuous change in  $Y_i$  as  $R_i$  crosses  $c$  can thus be interpreted as a causal effect of  $D_i$ .

## Where to find RD setups?

Often from administrative situations in which units are assigned a program, treatment or award based on a numerical index being above or below a certain threshold.

# Where to find RD setups?

## Examples:

- ▶ A politician may be elected if and only if the differential between the vote share that she receives and the vote share that her opponent receives exceeds 0.
- ▶ A student may be assigned to summer school if and only if his performance on a combination of tests falls below a certain threshold.
- ▶ A toxic waste site may receive cleanup funds if and only if it's hazard rating falls above a certain level.



# Why does RD work?

The idea:

- ▶ Units whose indices  $R$  lie directly below the threshold  $c$  are considered to be comparable to individuals or units whose indices  $R$  lie directly above the threshold  $c$ .
- ▶ We can estimate the treatment effect by taking a difference in mean outcomes for units directly above the threshold and units directly below the threshold.

## Key RD Assumption:

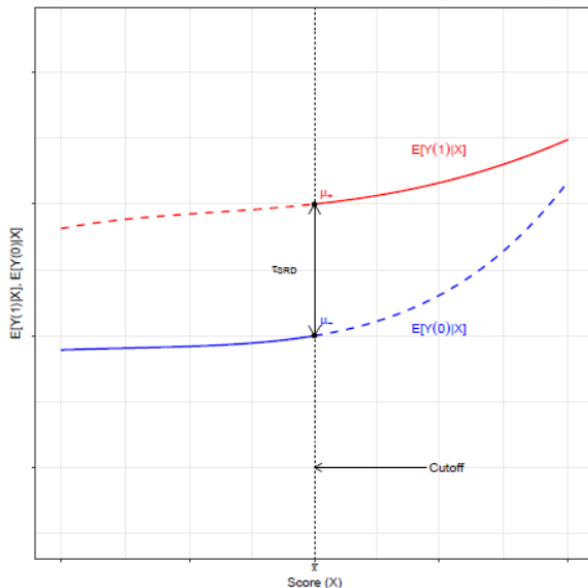
**The continuity assumption:**

**$E[Y_i(0)|R_i = r]$  and  $E[Y_i(1)|R_i = r]$  are continuous in  $r$ .**

(Absent treatment, there would be no discontinuity in outcomes.)

# RD designs: The continuity assumption

Figure 2.2: RD Treatment Effect in Sharp RD Design



# RD Types

RD designs come in two flavors:

- ▶ Sharp:
  - ▶ the probability that  $D = 1$  changes from 0 to 1 as the running variable crosses  $c$ .
  - ▶ no one with  $R < c$  gets treated and everyone with  $R \geq c$  gets treated (or vice versa)
- ▶ Fuzzy:
  - ▶ the probability of treatment changes by some nonzero amount as the running variable crosses the threshold  $c$ ,
  - ▶ the change in probability is less than 100 percentage points  
 $\Rightarrow D_i$  is no longer a deterministic function of  $R_i$ .

I start by discussing the general elements common to all RD designs.  
I will then cover each of these specifically.

# RD Graphs

Key to a good RD project is the graphical analysis (statistical results really take a back seat).

A discontinuous change in treatment and outcomes (if  $\tau \neq 0$ ) should be visible as  $R_i$  crosses the threshold  $c$ .

Failure to show visually perceptible breaks at  $c$  challenges the credibility of the approach (regardless of the regression results).

A break that is visually perceptible will almost surely be statistically significant.

# RD Graphs

A simple scatterplot of your data is unlikely to reveal the patterns you wish to illustrate.

The key RD graphs are histogram-type plot that presents the average value of:

- ▶ treatment status
- ▶ the outcome
- ▶ covariates
- ▶ the density of the running variable

at evenly spaced values of the running variable.

# RD Graphs

Generating these plots requires choosing two key parameters:

- ▶ the binwidth,  $h$ ,
- ▶ the number of bins shown to the left and right of the threshold value,  $K_0$  and  $K_1$ .

Once these choices are made, construct:

- ▶  $K_0$  evenly spaced bins of width  $h$  below the threshold value
- ▶  $K_1$  evenly spaced bins of width  $h$  above the threshold value
- ▶ (avoid having any bin crossing the threshold value  $c$ ).

## RD Treatment status graph

RD papers often include a graph that plots treatment by the running variable.

We expect to see a visually perceptible discontinuity in the probability of treatment as  $R$  crosses the threshold.

After constructing the bins described above, calculate  $\bar{D}_k$ , the average treatment level in the bin

$$\bar{D}_k = \frac{1}{N_k} \sum_{i=1}^N D_i * \mathbb{1}(b_k < R_i \leq b_{k+1}).$$

Plot these values against the midpoint of each of the bins.



## Outcome Graphs

The main course of an RD paper is a plot of the outcome by the running variable.

If  $\tau \neq 0$ , we expect to see a discontinuity here too.

Plotting this graph requires calculating,  $\bar{Y}_k$ , the average outcome in each bin

$$\bar{Y}_k = \frac{1}{N_k} \sum_{i=1}^N Y_i * \mathbb{1}(b_k < R_i \leq b_{k+1})$$

and plotting these values against the midpoint of each of the bins.

# Outcome Graphs

A visual break at  $c$

- ▶  $\Rightarrow$  crossing  $c$  has a significant effect on the outcome,
- ▶  $\Rightarrow$  the treatment has a significant effect on the outcome. This graph is the equivalent of the reduced form in an IV analysis.

## Outcome Graphs: Robustness

You should also look for other discontinuities of similar (or greater) magnitude at other values of  $R$ .

If there discontinuities for no clear reason, the research design is called into question - effectively we have detected a violation of Assumption 1 (smoothness in expected potential outcomes).

## Robustness Graphs: Covariates

It is common to plot covariates that may be related to the outcome but should not be affected by the treatment.

As above, we calculate  $\bar{X}_k$  where

$$\bar{X}_k = \frac{1}{N_k} \sum_{i=1}^N X_i * \mathbb{1}(b_k < R_i \leq b_{k+1})$$

is plotted against the midpoint of each bin.

If the research design is valid there should not be any discontinuity in  $\bar{X}_k$  as the running variable crosses the threshold  $c$ .

This is equivalent to a balance check across the threshold- you are checking that the treated and un-treated groups are similar.

## Robustness Graphs: Density of the Running Variable

It is also common to plot the density of the running variable.

For each bin, you calculate

$$N_k = \sum_{i=1}^N \mathbb{1}(b_k < R_i \leq b_{k+1})$$

and plot these against the midpoint of the bin.

A concern in RD is that individuals may “game” the assignment rule: they manipulate their  $R_i$  to place themselves just above (below)  $c$ .

This sorting would create selection bias and bias our estimate of  $\tau$ .

If units are manipulating  $R_i$  around  $c$ , we will see a discontinuity in the distribution of  $R_i$  as it crosses  $c$ .

If the distribution of  $R_i$  is smooth as it crosses  $c$ , then it's unlikely that individuals are gaming the assignment mechanism.

# Robustness Graphs: Density of the Running Variable

Example:

- ▶ consider a scholarship if test scores fall above a certain threshold  $c$ . Shrewd students could retake the test many times until they pass the threshold.
- ▶ If a researcher uses an individuals maximum test score as the running variable, motivated individuals who retake the test many times are more likely to fall just above the threshold, then just below it.
- ▶ this group of observations is selected and no longer directly comparable to the observations that fall directly below the threshold.

# RD Estimation

RD estimations can be done quite easily in a regression framework.

There are a couple things to keep in mind when it comes to RD estimation strategies:

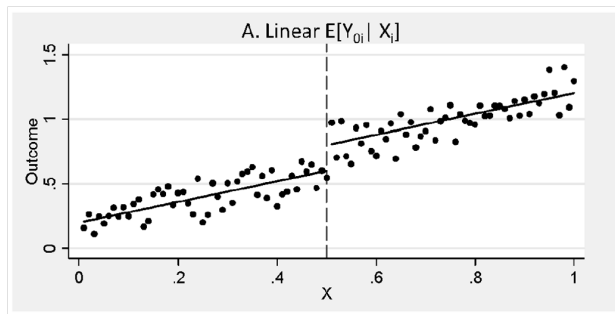
- ▶ the choice of functional form
- ▶ the choice of bandwidth.

## RD Estimation: Functional form

To identify the treatment effect, you need a good model of the underlying relationship between the running variable and the outcome.



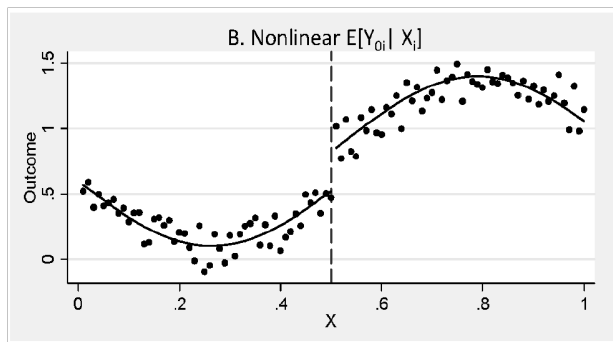
## RD Estimation: Functional form



Easy if the underlying relationship is linear:

- ▶ a linear regression fits the data well
- ▶ the treatment effect is the “jump” between the intercepts at the threshold value.

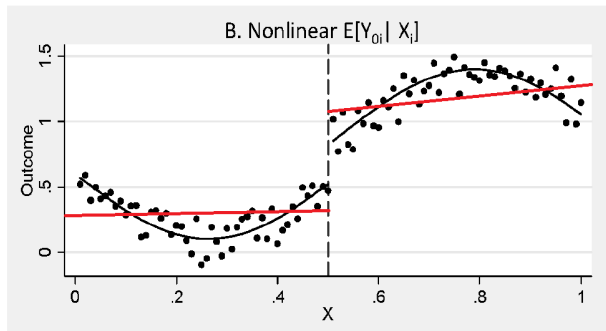
## RD Estimation: Functional form



If not linear, need higher order polynomial terms.

**What if I fit a linear function?**

## RD Estimation: Functional form

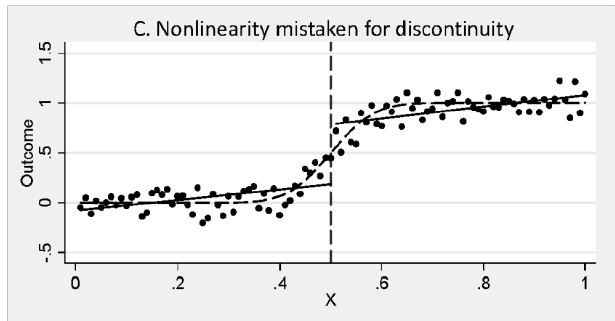


If not linear, need higher order polynomial terms.

### **What if I fit a linear function?**

If we do not fit the data correctly, the “jump” at the the threshold will not correctly estimate the treatment effect.

## RD Estimation: Functional form



There is no discontinuity at the threshold.

If I force a linear regression to this data I will detect a discontinuity:  
I find a treatment effect where there is none.

## RD Estimation: Robustness to functional form

Good graphs are important: they will alert you to these problems.

Run specifications with higher order polynomials and check that your estimated treatment effect is not sensitive to their inclusion (or exclusion).

## RD Estimation: Choosing the bandwidth

The bandwidth: how close to the threshold an observation's running variable must be to be in the regression sample.

**What bandwidth,  $h$ , should you use?**

There is an econometric literature on this.

In practice this choice is more art than science (ie fairly arbitrary).

## RD Estimation: Choosing the bandwidth

The choice is a tradeoff:

- ▶ With a large bandwidth, you will have a larger sample which can shrink your standard errors and improve the precision of your results.
- ▶ Two main drawbacks to a large bandwidth:
  - ▶ With a very large bandwidth, it becomes more difficult to argue that observations on either side of the threshold are similar.
  - ▶ Making sure you select the right functional form will be more important if you are working with a wide bandwidth. Example: see figure C.
- ▶ Good practice: check that results are not sensitive to the choice of bandwidth.

## RD Estimation

The specifics of the estimation will depend on whether we are dealing with a sharp or fuzzy RD.



Sharp RD

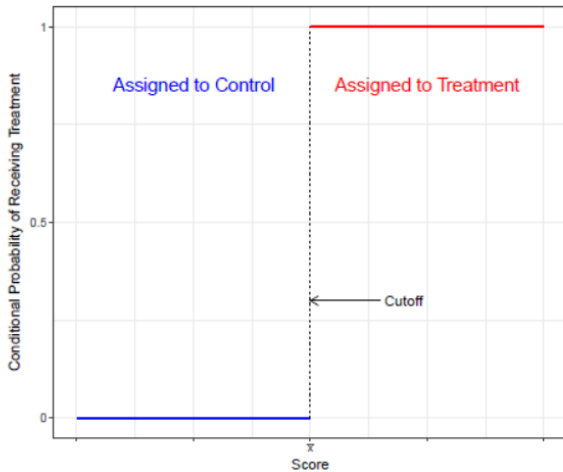
## Sharp RD Set up:

The probability that  $D = 1$  changes from 0 to 1 as the running variable crosses  $c$ .

- ▶ no one with  $R_i < c$  gets treated
- ▶ everyone with  $R_i \geq c$  gets treated
- ▶  $\Rightarrow D_i$  is a deterministic function of  $R_i$  :  $D_i = 1$  if  $(R_i \geq c)$ .
- ▶ (or vice versa)

## Sharp RD Set up:

Figure 2.1: Conditional Probability of Receiving Treatment in Sharp RD Design



## Sharp RD Set up:

To estimate the causal effect of  $D_i$  on some outcome  $Y_i$ , we simply take the difference in mean outcome on either side of  $c$ :

$$\lim_{r \rightarrow c} E[Y_i | R_i = r] - \lim_{r \leftarrow c} E[Y_i | R_i = r] = \lim_{r \rightarrow c} E[Y_i(1) | R_i = r] - \lim_{r \leftarrow c} E[Y_i(0) | R_i = r]$$

This estimates  $\tau_{SRD}$ , the causal effect for individuals with  $R_i = c$ :

$$\tau_{SRD} = E[Y_i(1) - Y_i(0) | R_i = c]$$

## Sharp RD Assumption:

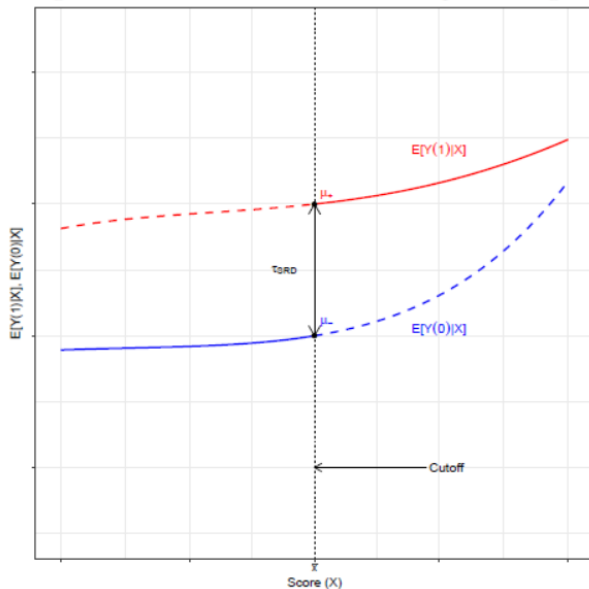
If the continuity assumption holds:

$$\tau_{SRD} = \lim_{r \rightarrow c} E[Y_i | R_i = r] - \lim_{r \leftarrow c} E[Y_i | R_i = r]$$

We can estimate  $\tau_{SRD}$  as the difference between the two regression functions estimated in the **neighborhood** of  $c$ .

## Sharp RD designs:

Figure 2.2: RD Treatment Effect in Sharp RD Design



## Sharp RD Simulation

You are the superintendent of a large school district.

Last year you made participation in small reading groups **mandatory** for all students whose 3rd grade reading score was 75 points or less.

Your data includes the 3rd and 4th grade reading scores for all 5000 students in your school district.

**How did these reading groups affected student performance on their 4th grade reading tests?**

# Sharp RD Simulation

```
set.seed(7000)

sharp<-rnorm(5000, mean=80, sd=5)
sharp<-as.data.frame(sharp)

names(sharp)<-c("read3")
sharp$error<-rnorm(5000, mean=0, sd=5)
sharp$pe3<-rnorm(5000, mean=90, sd=4)
sharp$height<-rnorm(5000, mean=130, sd=15)
```



# Sharp RD Simulation

```
sharp$treated<-0
sharp$treated[sharp$read3<=75]<-1

tau=10
#the DGP
sharp$read4<-(-6)+0.8*sharp$read3+tau*sharp$treated+sharp$error

#selecting observations that fall in our bandwidth
sharp<-sharp[sharp$read3<78 & sharp$read3>72,]
```

## Sharp RD: Treatment status graph

- ▶ Step 1: construct the histogram bins
- ▶ Step 2: calculate  $\bar{D}_k$ , the average treatment level in the bin

$$\bar{D}_k = \frac{1}{N_k} \sum_{i=1}^N D_i * \mathbb{1}(b_k < R_i \leq b_{k+1}).$$

In a sharp RD,  $\bar{D}_k$  should be either 0 or 1.

# Sharp RD: Treatment status graph Simulation

```
#I will break up the data into 60 bins (30 above and 30 below the threshold)
cuts<-c(72,72.1,72.2,72.3,72.4,72.5,72.6,72.7,72.8,72.9,73,
       73.1,73.2,73.3,73.4,73.5,73.6,73.7,73.8,73.9,74,
       74.1,74.2,74.3,74.4,74.5,74.6,74.7,74.8,74.9,75,
       75.1,75.2,75.3,75.4,75.5,75.6,75.7,75.8,75.9,76,
       76.1,76.2,76.3,76.4,76.5,76.6,76.7,76.8,76.9,77,
       77.1,77.2,77.3,77.4,77.5,77.6,77.7,77.8,77.9,78)
midpoints<-cuts[2:61]-0.05

sharp$bins <- cut(sharp$read3,
                  breaks=cuts,
                  include.lowest=TRUE,
                  right=FALSE,
                  labels=midpoints)

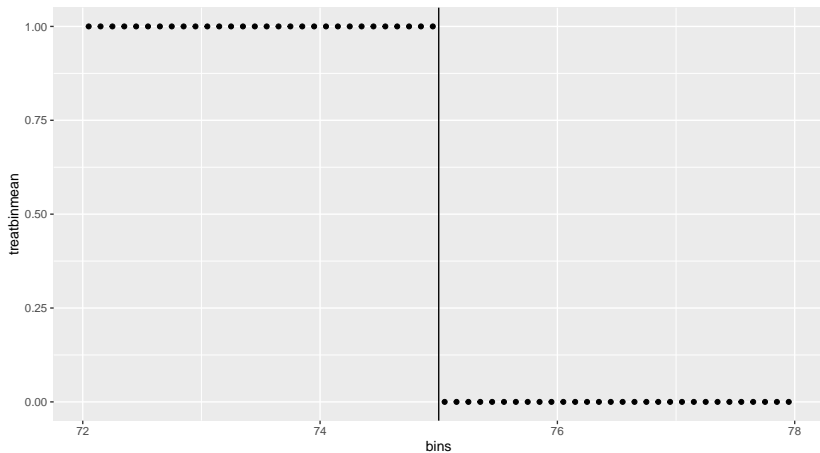
sharp_mean<-sharp %>%
  group_by(bins) %>%
  dplyr::summarize(outbinmean = mean(read4, na.rm=TRUE),
                  treatbinmean=mean(treated, na.rm=TRUE),
                  pebinmean=mean(pe3, na.rm=TRUE),
                  heightbinmean=mean(height, na.rm=TRUE), numb=n())

sharp_mean$bins<-as.numeric(as.character(sharp_mean$bins))

plot1shp<-ggplot(sharp_mean, aes(x=bins, y=treatbinmean))+
  geom_point()+
  geom_vline(xintercept = 75)
```

# Sharp RD: Treatment status graph Simulation

plot1shp



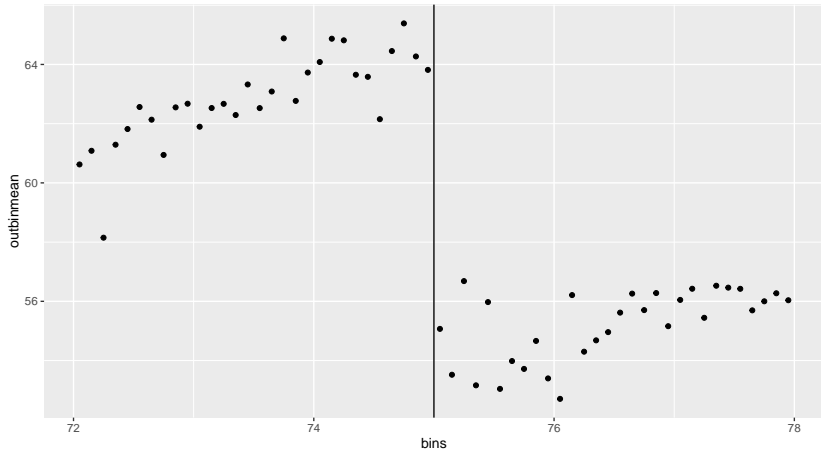
# Outcome Graphs: Sharp

In the earlier chunk, we calculated,  $\bar{Y}_k$ , the average outcome in each bin

$$\bar{Y}_k = \frac{1}{N_k} \sum_{i=1}^N Y_i * \mathbb{1}(b_k < R_i \leq b_{k+1})$$

```
plot2shp<-ggplot(sharp_mean, aes(x=bins, y=outbinmean))+  
  geom_point()+  
  geom_vline(xintercept = 75)
```

# Outcome Graphs: Sharp



# Robustness Graphs: Covariates sharp

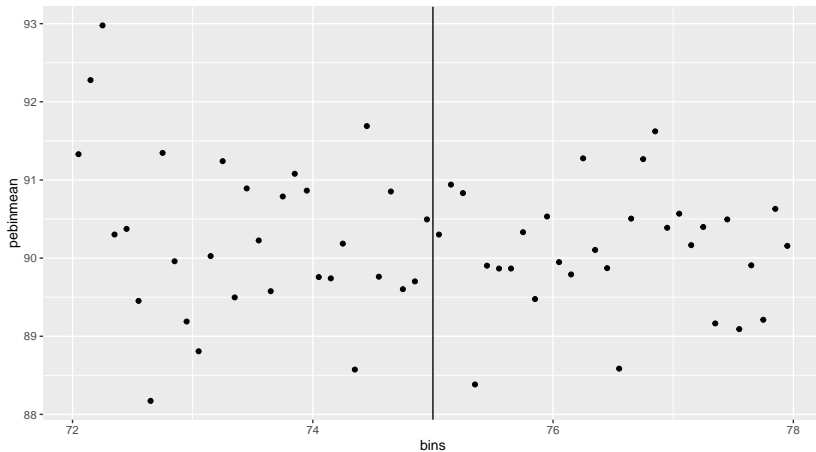
In the earlier chunk, we calculated  $\bar{X}_k$  where

$$\bar{X}_k = \frac{1}{N_k} \sum_{i=1}^N X_i * \mathbb{1}(b_k < R_i \leq b_{k+1})$$

```
plot3shp<-ggplot(sharp_mean, aes(x=bins, y=pebinmean))+  
  geom_point()+  
  geom_vline(xintercept = 75)
```

# Robustness Graphs: Covariates sharp

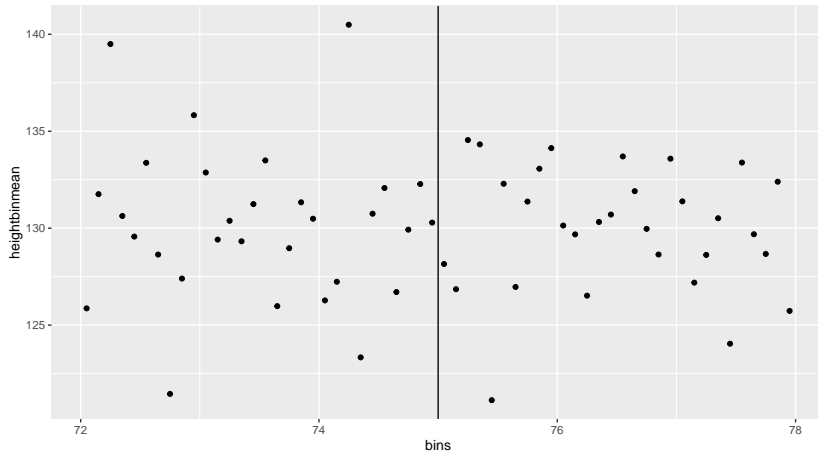
plot3shp





# Robustness Graphs: Covariates sharp

```
plot4shp<-ggplot(sharp_mean, aes(x=bins, y=heightbinmean))+  
  geom_point()+  
  geom_vline(xintercept = 75)  
plot4shp
```



# Robustness Graphs: Density of the Running Variable (Sharp)

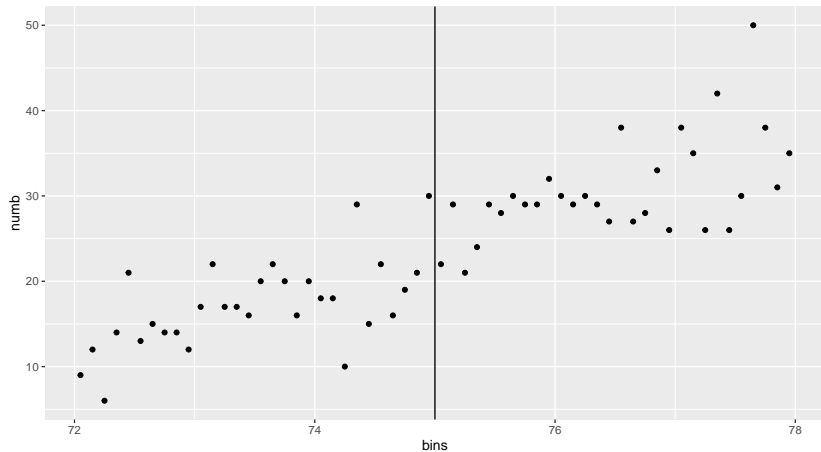
In an earlier chunk we calculated

$$N_k = \sum_{i=1}^N \mathbb{1}(b_k < R_i \leq b_{k+1})$$

```
plot5shp<-ggplot(sharp_mean, aes(x=bins, y=numb))+  
  geom_point()+  
  geom_vline(xintercept = 75)
```

# Robustness Graphs: Density of the Running Variable (Sharp)

plot5shp



## RD Estimation: Sharp

Idea: Fit a linear regression on either side of the threshold point for the samples with  $R_i \in (c - h, c)$  and  $R_i \in [c, c + h)$ .

Estimate some version (you can include covariates) of the following specification,

$$Y_i = \alpha + \tau D_i + \beta(R_i - c) + \gamma(R_i - c) * D_i + u_i \text{ for } c - h \leq R_i < c + h.$$

With this estimation strategy,  $\hat{\tau}_{SRD}$  will estimate the treatment effect for units right at the threshold.

# RD Estimation: Sharp

```
sharp$runminc<-sharp$read3-75  
shpestim<-felm(read4~treated+runminc+treated*runminc, sharp)  
  
stargazer(shpestim, type="latex", header=FALSE)
```

Table 1:

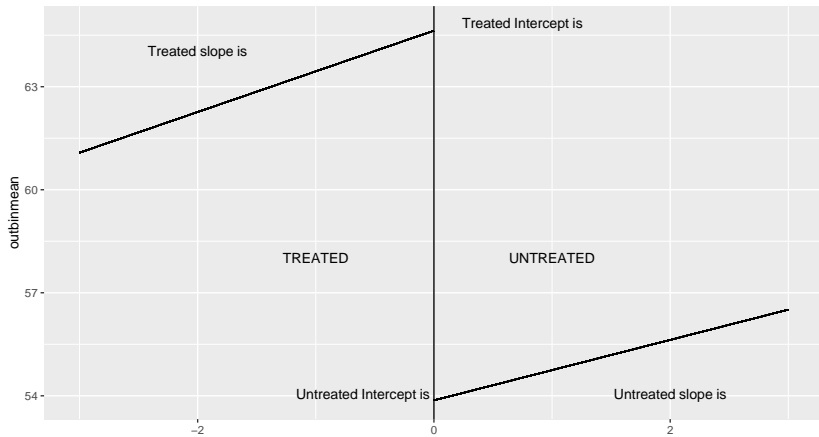
<i>Dependent variable:</i>	
	read4
treated	10.764*** (0.555)
runminc	0.880*** (0.197)
treated:runminc	0.304 (0.333)
Constant	53.868*** (0.356)
Observations	1,436
R <sup>2</sup>	0.358
Adjusted R <sup>2</sup>	0.356
Residual Std. Error	5.128 (df = 1432)
Note:	* p<0.1; ** p<0.05; *** p<0.01

## RD Estimation: Sharp

$$Y_i = \alpha + \tau D_i + \beta(R_i - c) + \gamma(R_i - c) * D_i + u_i \text{ for } c - h \leq R_i < c + h.$$

How do these coefficients translate to the RD graph?

⇒ Top Hat



# RD Estimation: Sharp

It is helpful to see how the coefficients estimated above translate to the RD graph.

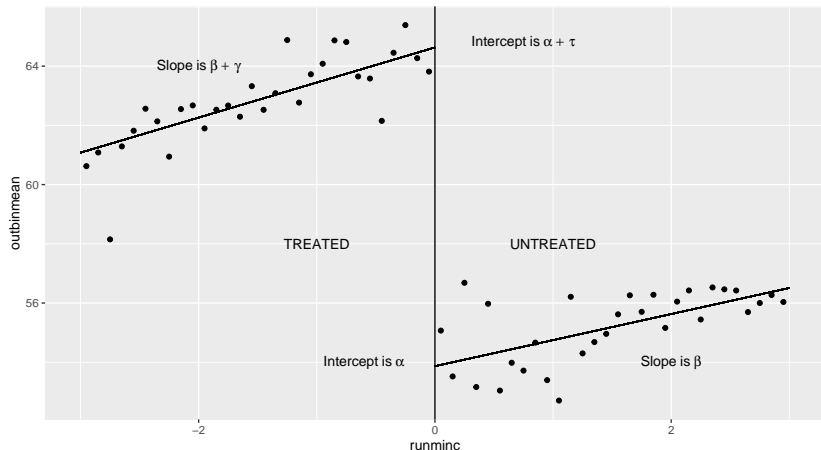
```
sharp_mean$runminc<-sharp_mean$bins-75

plot6shp<-ggplot(sharp_mean, aes(x=runminc, y=outbinmean))+
  geom_point()+
  geom_vline(xintercept = 0)+
  geom_segment(aes(x = 0, xend = 3,
                  y = shpestim$coefficients[1],
                  yend = shpestim$coefficients[1]
                  +3*shpestim$coefficients[3]))+
  geom_segment(aes(x = -3, xend = 0,
                  y = shpestim$coefficients[1]
                  + shpestim$coefficients[2]
                  +(-3*( shpestim$coefficients[3]+ shpestim$coefficients[4])),
                  yend = shpestim$coefficients[1]
                  + shpestim$coefficients[2]))+
  #adding some labeling for course notes:
  annotate("text", x = 0.75, y = 64.8,
          label = "Intercept-is-alpha-+-tau" ,parse = TRUE)+
  annotate("text", x = -0.6, y = 54,
          label = "Intercept-is-alpha" ,parse = TRUE)+
  annotate("text", x = -2, y = 64,
          label = "Slope-is-beta-+-gamma" ,parse = TRUE)+
  annotate("text", x = 2, y = 54,
          label = "Slope-is-beta" ,parse = TRUE)+
  annotate("text", x = -1, y = 58,
          label = "TREATED" ,parse = TRUE)+
  annotate("text", x = 1, y = 58,
          label = "UNTREATED" ,parse = TRUE)
```

# RD Estimation: Sharp

$$Y_i = \alpha + \tau D_i + \beta(R_i - c) + \gamma(R_i - c) * D_i + u_i \text{ for } c - h \leq R_i < c + h.$$

plot6shp





Fuzzy RD

## Fuzzy RD Set Up

Potentially more common than the sharp RD set ups.

$D_i$  is no longer a deterministic function of  $R_i$ .

The probability of treatment changes by some nonzero amount as the running variable crosses the threshold  $c$ .

The change in probability is less than 100 percentage points.

$$0 < \lim_{r \rightarrow c} P(D_i = 1 | R_i = r) - \lim_{r \leftarrow c} P(D_i = 1 | R_i = r) < 1$$

# Fuzzy RD Set Up

There are now two causal effects to be estimated:

- ▶ the effect of crossing the threshold on the probability of treatment (which is 1 in the sharp RD)
- ▶ the effect of crossing the threshold on the outcome.

Formally, the fuzzy RD estimator is

$$\tau_{FRD} = \frac{\lim_{r \rightarrow c} E[Y_i | R_i = r] - \lim_{r \leftarrow c} E[Y_i | R_i = r]}{\lim_{r \rightarrow c} E[D_i | R_i = r] - \lim_{r \leftarrow c} E[D_i | R_i = r]}.$$

**What should this remind you of?**

# Fuzzy RD IV

RD analog of an IV estimator:

- ▶ the instrument is an indicator for whether  $R_i$  lies directly above(below)  $c$ .
- ▶ similar to how the IV estimator allowed us to recover the LATE estimate in an RCT.

-you (randomly) fall above (below)  $c$  which changes the probability of treatment. We leverage this change in probability as an instrument to estimate the effect of treatment on the outcome.

## Fuzzy RD IV

In a fuzzy RD:

- ▶ the ITT group (say who have  $R_i \leq c$  for example)
- ▶ the control group ( $R_i > c$ ).

Because we are in fuzzy land:

- ▶ being in the ITT group does not necessarily mean you get treated (there are never takers)
- ▶ some in the control get treated (always takers).
- ▶ but there is a group of observations, the compliers, whose treatment status changes if they go from control to ITT (ie if they were moved across the threshold).

## Fuzzy RD IV

If I just compare the outcomes of those that are above and below the threshold ( $\lim_{r \rightarrow c} E[Y_i | R_i = r] - \lim_{r \leftarrow c} E[Y_i | R_i = r]$ ) the effect is “diluted”. **Why?**

## Fuzzy RD IV

If I just compare the outcomes of those that are above and below the threshold ( $\lim_{r \rightarrow c} E[Y_i | R_i = r] - \lim_{r \leftarrow c} E[Y_i | R_i = r]$ ) the effect is “diluted”. **Why?**

- ▶ for many observations, crossing the threshold has no effect (since they are always or never takers).

## Fuzzy RD IV

To get the LATE, scale treatment estimates by the change in probability of being treated ( $\lim_{r \rightarrow c} E[D_i | R_i = r] - \lim_{r \leftarrow c} E[D_i | R_i = r]$ ).

$\Rightarrow$  the fuzzy RD design measures the average treatment effect for RD compliers at the threshold (the LATE),

$$\tau_{FRD} = E[Y_i(1) - Y_i(0) | \text{unit } i \text{ is a complier and } R_i = c].$$



## Fuzzy RD Simulation:

You are the superintendent of a large school district.

Last year you **strongly encouraged** students to participate in small reading groups if their 3rd grade reading score fell below 75 points.

Your data includes the 3rd and 4th grade reading scores for all 5000 students in your school district.

**How did these reading groups affected student performance on their 4th grade reading tests?**

# Fuzzy RD Simulation:

```
set.seed(2000)

fuzzy<-rnorm(5000, mean=80, sd=5)
fuzzy<-as.data.frame(fuzzy)

names(fuzzy)<-c("read3")
fuzzy$error<-rnorm(5000, mean=0, sd=5)
fuzzy$pe3<-rnorm(5000, mean=90, sd=4)
fuzzy$height<-rnorm(5000, mean=130, sd=15)

fuzzy$lowprob<-rbinom(5000,1,0.3)
fuzzy$highprob<-rbinom(5000,1,0.8)
fuzzy$treated<-NA
fuzzy$treated[fuzzy$read3>75]<-fuzzy$lowprob[fuzzy$read3>75]
fuzzy$treated[fuzzy$read3<=75]<-fuzzy$highprob[fuzzy$read3<=75]

tau=10

#the DGP
fuzzy$read4<-(-6)+0.8*fuzzy$read3+tau*fuzzy$treated+fuzzy$error

fuzzy<-fuzzy[fuzzy$read3<78 & fuzzy$read3>72,]
```

## Fuzzy RD: Treatment status graph

We calculate  $\bar{D}_k$ , the average treatment level in the bin

$$\bar{D}_k = \frac{1}{N_k} \sum_{i=1}^N D_i * \mathbb{1}(b_k < R_i \leq b_{k+1}).$$

In a fuzzy RD,  $\bar{D}_k$  can take on many possible values.

This plot should show that there is a visual discontinuity in the probability of getting treated at the threshold  $c$ .

This graph is equivalent to the first stage in an IV analysis: It shows that we have found a tool that generates some random variation we can leverage to estimate unbiased treatment effects.

# Fuzzy RD: Treatment status graph Simulation

```
#I will break up the data into 60 bins (30 above and 30 below the threshold)
```

```
cuts<-c(72,72.1,72.2,72.3,72.4,72.5,72.6,72.7,72.8,72.9,73,  
       73.1,73.2,73.3,73.4,73.5,73.6,73.7,73.8,73.9,74,  
       74.1,74.2,74.3,74.4,74.5,74.6,74.7,74.8,74.9,75,  
       75.1,75.2,75.3,75.4,75.5,75.6,75.7,75.8,75.9,76,  
       76.1,76.2,76.3,76.4,76.5,76.6,76.7,76.8,76.9,77,  
       77.1,77.2,77.3,77.4,77.5,77.6,77.7,77.8,77.9,78)  
midpoints<-cuts[2:61]-0.05
```

```
fuzzy$bins <- cut(fuzzy$read3,  
                 breaks=cuts,  
                 include.lowest=TRUE,  
                 right=FALSE,  
                 labels=midpoints)
```

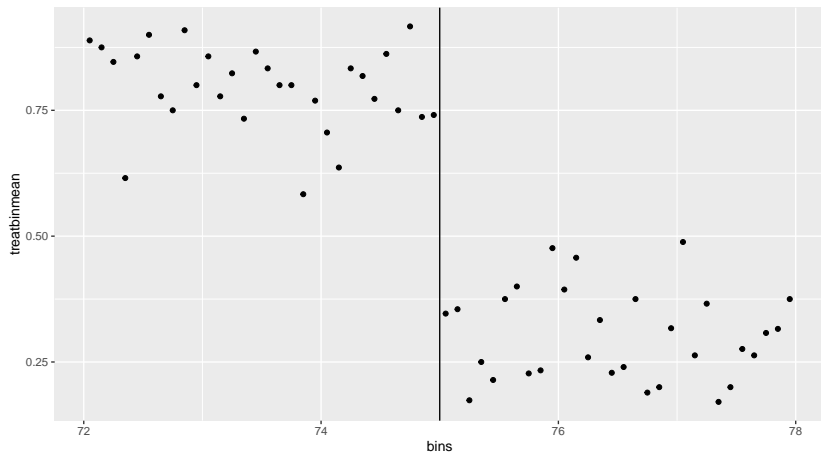
```
fuzzy_mean<-fuzzy %>%  
  group_by(bins) %>%  
  dplyr::summarize(outbinmean = mean(read4, na.rm=TRUE),  
                  treatbinmean=mean(treated, na.rm=TRUE),  
                  pebinmean=mean(pe3, na.rm=TRUE),  
                  heightbinmean=mean(height, na.rm=TRUE), numb=n())
```

```
fuzzy_mean$bins<-as.numeric(as.character(fuzzy_mean$bins))
```

```
plot1fuz<-ggplot(fuzzy_mean, aes(x=bins, y=treatbinmean))+  
  geom_point()+  
  geom_vline(xintercept = 75)
```

# Fuzzy RD: Treatment status graph Simulation

plot1fuz



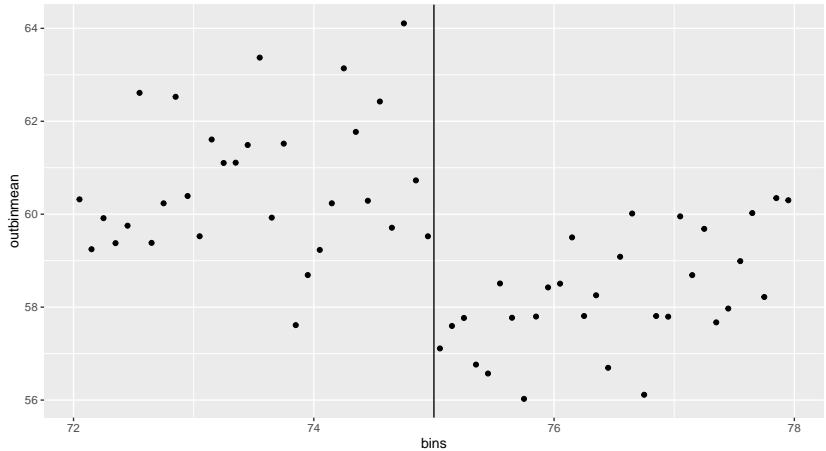
## Outcome Graphs: Fuzzy

In the earlier chunk, we calculated,  $\bar{Y}_k$ , the average outcome in each bin

$$\bar{Y}_k = \frac{1}{N_k} \sum_{i=1}^N Y_i * \mathbb{1}(b_k < R_i \leq b_{k+1})$$

```
plot2fuz<-ggplot(fuzzy_mean, aes(x=bins, y=outbinmean))+  
  geom_point()+  
  geom_vline(xintercept = 75)
```

# Outcome Graphs: Fuzzy



# Robustness Graphs: Covariates fuzzy

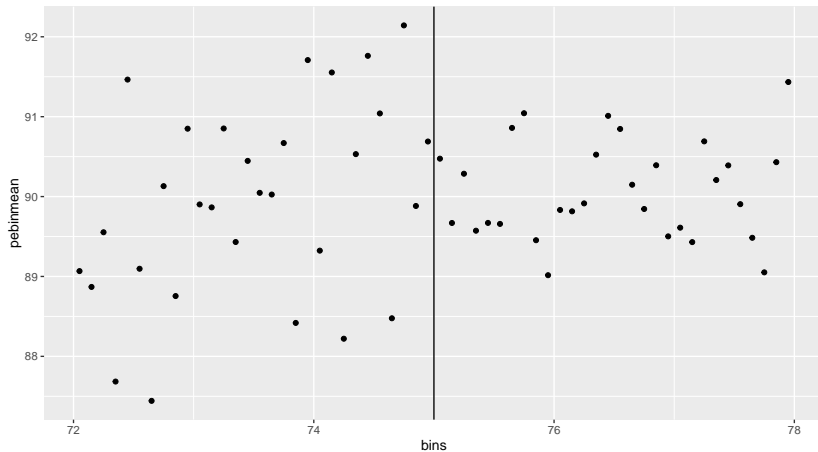
In the earlier chunk, we calculated  $\bar{X}_k$  where

$$\bar{X}_k = \frac{1}{N_k} \sum_{i=1}^N X_i * \mathbb{1}(b_k < R_i \leq b_{k+1})$$

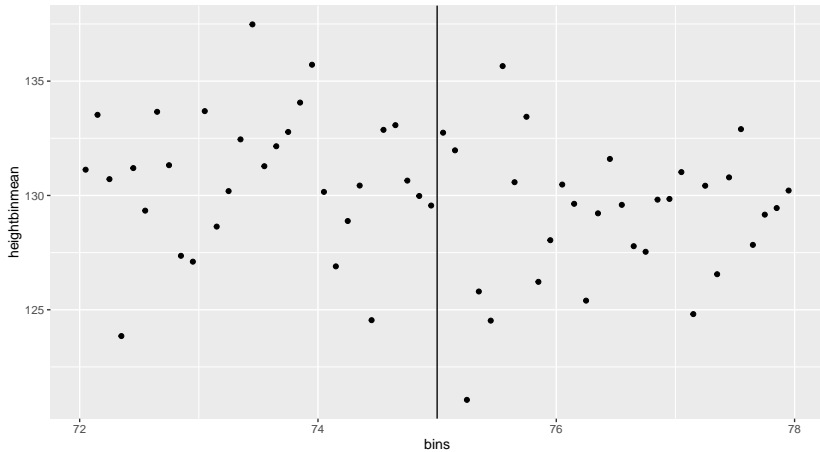
```
plot3fuz<-ggplot(fuzzy_mean, aes(x=bins, y=pebinmean))+  
  geom_point()+  
  geom_vline(xintercept = 75)
```



## Robustness Graphs: Covariates fuzzy



## Robustness Graphs: Covariates fuzzy



# Robustness Graphs: Density of the Running Variable (Fuzzy)

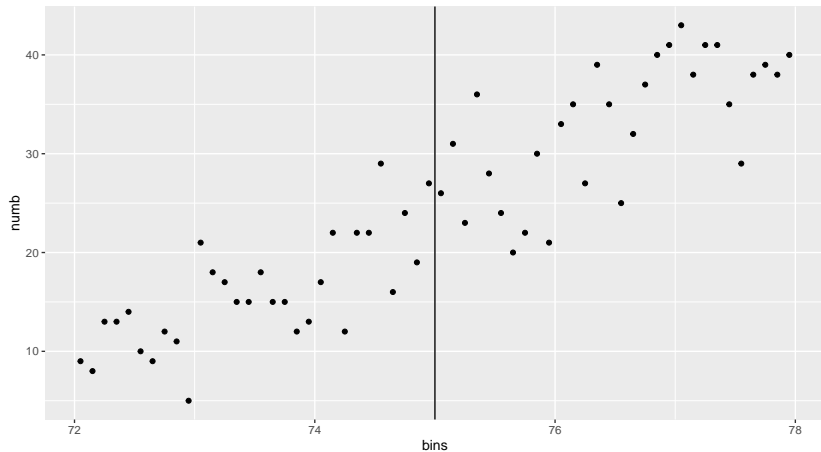
In an earlier chunk we calculated

$$N_k = \sum_{i=1}^N \mathbb{1}(b_k < R_i \leq b_{k+1})$$

```
plot5fuz<-ggplot(fuzzy_mean, aes(x=bins, y=numb))+  
  geom_point()+  
  geom_vline(xintercept = 75)
```

# Robustness Graphs: Density of the Running Variable (Fuzzy)

```
plot5fuz
```



## RD Estimation: Fuzzy

In the fuzzy RD design, we have two effects to estimate:

- ▶ the effect of crossing the threshold on the treatment (the “first stage”)
- ▶ the effect of crossing the threshold on the outcome (the “reduced form”).

We apply the same methodology as in earlier IV's to estimate the effect of crossing the threshold on  $Y_i$  and the effect of crossing the threshold on  $D_i$ .

## RD Estimation: Fuzzy

For the sample with  $c - h \leq R_i < c + h$  we run some version (you can include covariates) of the following “reduced form” IV regression

$$Y_i = \pi_0 + \pi_1 Z_i + \pi_2 (R_i - c) + \pi_3 (R_i - c) * Z_i + u_i$$

where the first stage is given by

$$D_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 (R_i - c) + \gamma_3 (R_i - c) * Z_i + v_i$$

where  $Z_i = \mathbb{1}(R_i > c)$ .

## RD Estimation: Fuzzy

The fuzzy RD estimator is then

$$\hat{\tau}_{FRD} = \frac{\hat{\pi}_1}{\hat{\gamma}_1}.$$

The FRD estimator is the ratio of the reduced form and the first stage estimates (aka the IV estimator).

i.e. the effect of crossing the discontinuity threshold on the outcome, scaled by the effect of crossing the discontinuity threshold on the probability of treatment.

# RD Estimation: Fuzzy

```
fuzzy$runminc<-fuzzy$read3-75

#first stage
fuzzy$ittgroup<-0
fuzzy$ittgroup[fuzzy$read3<=75]<-1

fuzfs<-felm(treated~ittgroup+runminc+ittgroup*runminc,fuzzy)

#reduced form
fuzrf<-felm(read4~ittgroup+runminc+ittgroup*runminc,fuzzy)

fuzzy$interedog<-fuzzy$treated*fuzzy$runminc
fuzzy$interinst<-fuzzy$ittgroup*fuzzy$runminc
#IV
fuziv<-felm(read4~runminc|0|(treated|interedog~ittgroup+runminc+interinst),fuzzy)

## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```



# RD Estimation: Fuzzy

```
stargazer(fuzfs, fuzrf, fuziv, type="latex", header=FALSE)
```

Table 2:

	<i>Dependent variable:</i>		
	treated	read4	
	(1)	(2)	(3)
ittgroup	0.464*** (0.048)	4.253*** (0.722)	
runminc	-0.004 (0.017)	0.821*** (0.251)	1.057*** (0.332)
ittgroup:runminc	-0.013 (0.029)	-0.479 (0.441)	
'treated(fit)'			9.201*** (1.162)
'interedog(fit)'			-0.685 (0.631)
Constant	0.308*** (0.030)	57.010*** (0.460)	54.182*** (0.640)
Observations	1,460	1,460	1,460
R <sup>2</sup>	0.213	0.037	0.454
Adjusted R <sup>2</sup>	0.211	0.035	0.453
Residual Std. Error (df = 1456)	0.443	6.728	5.065

Note:

\* p<0.1; \*\* p<0.05; \*\*\* p<0.01

## RD Estimation: Fuzzy

Note: you will not be able to plot the “scaled” treatment effect as the graphs are limited to the graphical equivalent of the first stage and the reduced form estimates.

## RD and External Validity

Notice the conditioning statements in our treatment estimators.

$$\tau_{SRD} = E[Y_i(1) - Y_i(0) | R_i = c]$$

$$\tau_{FRD} = E[Y_i(1) - Y_i(0) | \text{unit } i \text{ is a complier and } R_i = c].$$

# RD and External Validity

- ▶ We focus on observations that are in the neighborhood of the threshold. **Why?**

## RD and External Validity

- ▶ We focus on observations that are in the neighborhood of the threshold. **Why?**
  - ▶ implicitly we are assuming that an observation falling just above or just below the threshold is effectively random
  - ▶  $\Rightarrow$  we can measure a LATE that is valid around  $c$
  - ▶ observations far below and far above the threshold likely differ from each other in observable and unobservable ways.
- ▶ With fuzzy RD's, we further condition by estimating effects on compliers

# RD and External Validity

Because of these conditions, RD estimates are:

- ▶ inherently localized since the effects are estimated for a sub population where their  $R_i$  is in the neighborhood of  $c$ .
- ▶ have a relatively high degree of internal validity,
- ▶ it is important to think about their external validity:
  - ▶ treatment effects could be quite different for observation where  $R_i$  is quite different from  $c$ .