Week 9: Regression Discontinuity Designs

PUBL0050 Causal Inference

Julia de Romémont

Term 2 2023-24

UCL Departement of Political Science

This Lecture

Intuition

Sharp Regression Discontinuity Design

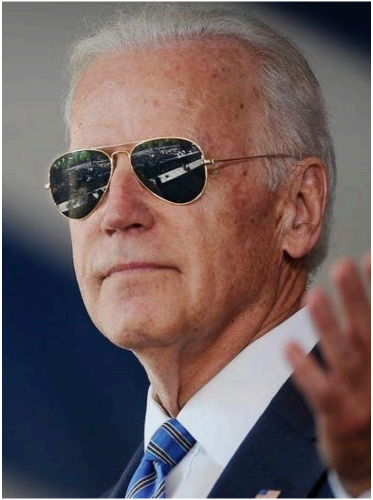
RDD Estimation

RDD Validation

Fuzzy Regression Discontinuity Design

Intuition

Joe



Bernie

"Sanders'...success frightens Democrats who worry that the socialism label remains a potent pejorative among the swing voters they'll need to defeat Trump" – *The Atlantic*, Feb 2020.

Running example

What happens when extremists win primaries?

What are the consequences of nominating an extremist candidate in a primary election for electoral outcomes? Hall (2015) studies primary elections for the US House between 1980 and 2010 where the contest was between an extremist candidate and a moderate candidate. Extremism is determined by receving donations from extreme interest groups. The outcomes of these races are used to compare the electoral outcomes of moderates and challengers in subsequent general elections.

- Outcome ($Y_{i,p,t}$): Party vote share in district i in the general election at time t
- Treatment ($D_{i,p,t}$): 1 if the party's primary winner in district i is an "extremist"
- Running variable ($X_{i,p,t}$): Extremist candidate's vote-share winning/losing margin in the primary in district i

Naive DIGM

```
vote_share_extreme <- mean(hall$vote_share_general[hall$extreme == 1])
vote_share_moderate <- mean(hall$vote_share_general[hall$extreme == 0])

vote_share_extreme - vote_share_moderate
```

```
## [1] -0.02736295
```

Why can't we interpret this causally?

[1] As *always*.

Possible solutions?

- **Randomize who wins the election**
  - $\rightarrow$ obvious implementation issues
- **Condition on observed** differences between extremists and moderates
  - $\rightarrow$ OVB remains an issue
- Find an **instrument** that increases the probability of an extremist winning the primary, but that has no effect otherwise
  - $\rightarrow$ difficult to come up with a good instrument
- Use variation over time in party vote shares using a **difference-in-differences** analysis
  - $\rightarrow$ possibly too infrequent repeated measurement (i.e. elections) with very different contexts

Regression Discontinuity Designs (RDD)

Compare vote share of parties in districts where extremists **narrowly won** their primary races to districts where extremists **narrowly lost**

$\rightarrow$ Assume that winning is **as good as random** in **close** races

In a nutshell

- Each unit has a score on a running variable which determines treatment
- The cutoff is the value of the running variable at which treatment is assigned
  - assigned when unit running variable score is above a known cutoff
  - not assigned when unit running variable score is below cutoff
- There is a discontinuous change in probability of receiving the treatment at the cutoff
  - Abrupt change in treatment probability can be used to learn about the **local** causal effect of the treatment on an outcome of interest

Discontinuities to overcome selection bias

RDD core intuition

Units with scores barely below the cutoff can be used as counterfactuals for units with scores barely above it.

- RDD is widely used in rule-based settings, where it is clear how and when $D_i = 1$ is asssigned:
  - Elections
  - Administrative programmes
  - Geographic boundaries
- The design is reliant on us **knowing about** and **having access to** a running variable that determines the treatment status.

Sharp Regression Discontinuity Design

Set-up

Imagine that our binary treatment variable, D_i, is completely determined by the value of an explanatory variable, X_i, according to:

$$D_i = 1 \times (X_i > c) \qquad \text{so} \qquad D_i = \begin{cases} D_i = 1 & \text{if } X_i \geq c \\ D_i = 0 & \text{if } X_i < c \end{cases}$$

where

- X_i is known as the “forcing” or “running” variable, and may be correlated with the outcomes (Y_i) and potential outcomes (Y_{1i}, Y_{0i})
- c is a fixed cutoff point

Implications

- D_i is a **deterministic function** of X_i

  $\rightarrow$ when we know X_i, we know D_i

- D_i is a **discontinuous function** of X_i

  $\rightarrow$ no matter how close to c we are, $D_i = 0$ until $X_i \geq c$

Examples of X_i, D_i, and c

- **Eggers (2015)**
  - Y_i – turnout (aggregate)
  - D_i – proportional representation in a French town
  - X_i – population of the town
  - c – 3500
- **de Kadt (2017)**
  - Y_i – turnout after 1994 (individual)
  - D_i – voting in South Africa in 1994
  - X_i – age in 1994
  - c – 18
- **Hall (2015)**
  - Y_i – party vote share in general election
  - D_i – primary election won by an extremist
  - X_i – margin of victory in the primary election
  - c – 0

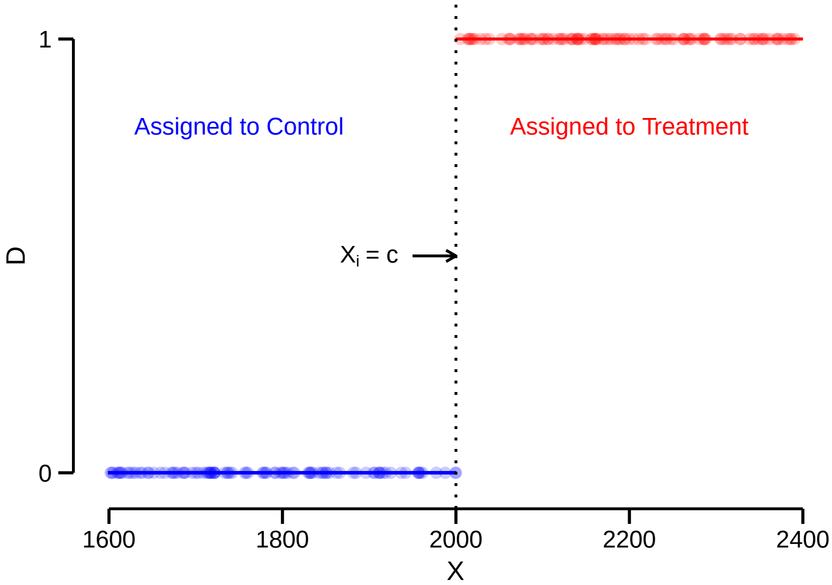
Graphical illustration

Do scholarships increase earnings?

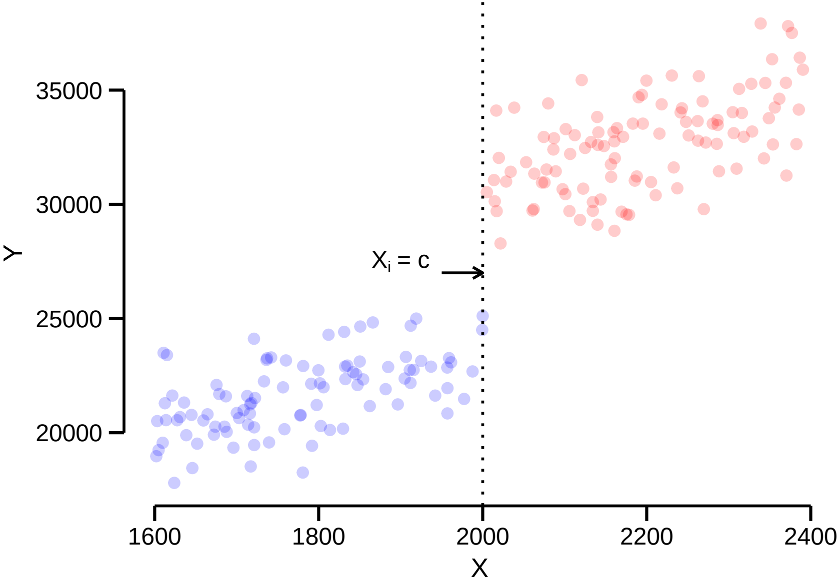
Thistlethwaite and Campbell (1960) study the effects of college scholarships on employment outcomes for students later in life. They study the allocation of "merit awards", which were given out to students based on a score, and anyone with a score above some cutoff received the merit award, whereas everyone below that cutoff did not.

- Outcome (Y_i): Adult earnings ($)
- Treatment (D_i): Receipt of a merit award
- Running variable (X_i): Score on a standardized test
- Cutoff (c): Scores of 2000 on more on X_i result in a merit award.

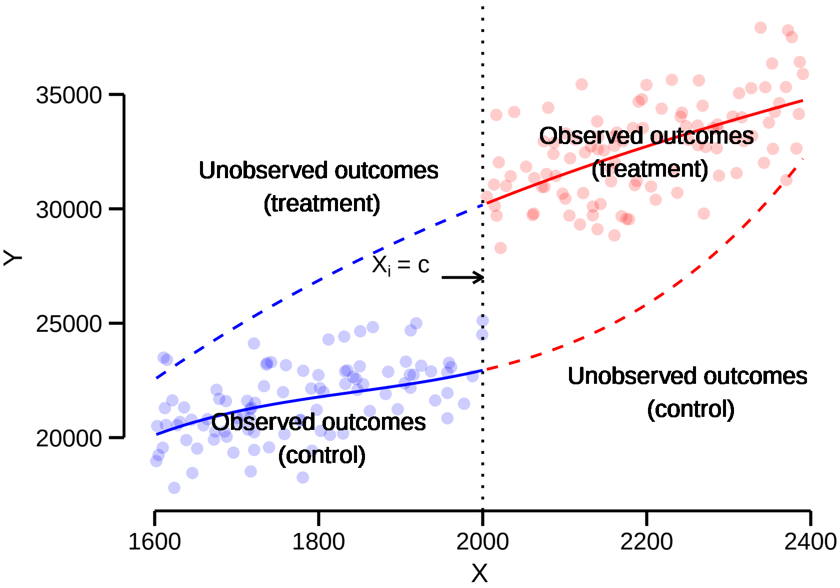
Graphical illustration (X_i and D_i)



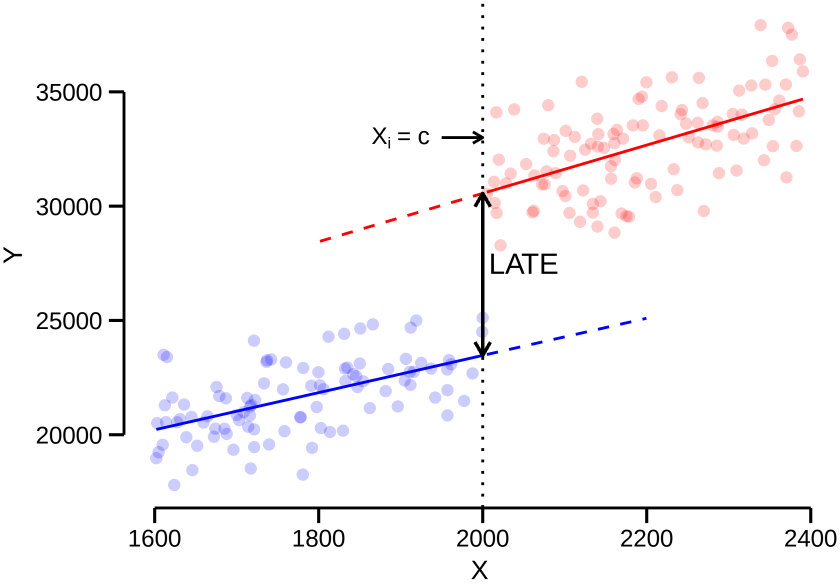
Graphical illustration (X_i and Y_i)



Graphical illustration (X_i, Y_{1i} and Y_{0i})



Graphical illustration (τ_{ATE} at c)



Identification with sharp RDD

- We want to be able to estimate the difference between $D_i = 1$ and $D_i = 0$ at the threshold c.
- Can we estimate this?

$$\begin{aligned}\tau_{\mathsf{LATE}} &= E[Y_{1i}|X_i = c] - E[Y_{0i}|X_i = c] \\ &= E[Y_i|X_i = c, D_i = 1] - E[Y_i|X_i = c, D_i = 0]\end{aligned}$$

Identification with sharp RDD

Identification Assumption

The potential outcomes $E[Y_{1i}|X_i, D_i]$ and $E[Y_{0i}|X_i, D_i]$ are continuous in X around c.

Identification Result

The treatment effect at the threshold c is identified by:

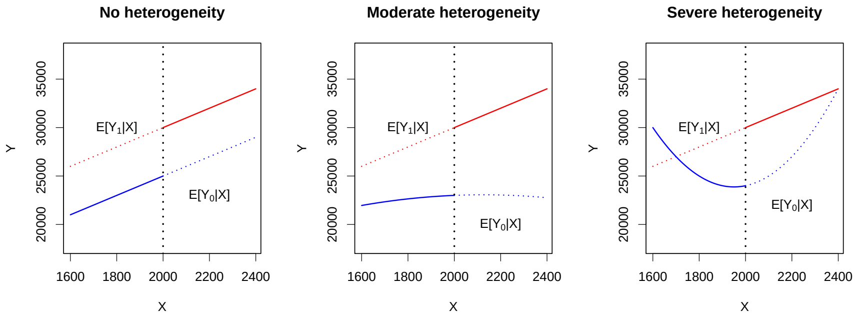
$$\begin{aligned} \tau_{\text{LATE}} &= E[Y_{1i} - Y_{0i}|X = c] \\ &= E[Y_{1i}|X = c] - E[Y_{0i}|X = c] \end{aligned}$$

But we can't observe both of these! However, **if the potential outcomes are continuous around** c, then we can estimate these values through the limits of the observed outcomes from above and below c:

$$\hat{\tau}_{\text{LATE}} = \lim_{X \downarrow c} E[Y_i|X = c, D_i = 1] - \lim_{X \uparrow c} E[Y_i|X = c, D_i = 0]$$

- We extrapolate a small amount to infer potential outcomes at c
- Without further assumptions, the LATE only identifies the ATE at c

Local nature of the RD effect



Implications

- RDDs estimate **L**ocal **A**verage **T**reatment **E**ffects that are the average causal effect for units **exactly at the cutoff**
- *Only* when treatment effects are homogeneous, meaning that every unit is affected by treatment in the same way, then $\tau_{LATE} = \tau_{ATE}$
  - This is very often an unconvincing assumption to make as units far from the cutoff are likely *very* different

RDD Estimation

- Recode the running variable to deviations from c: $\tilde{X}_i = X_i - c$
  - $\tilde{X}_i = 0$ if $X_i = c$
  - $\tilde{X}_i > 0$ if $X_i > c$ and so $D_i = 1$
  - $\tilde{X}_i < 0$ if $X_i < c$ and so $D_i = 0$
- Decide on a regression model for $E[Y_i|X_i, D_i]$
  - Linear, same slope for $E[Y_{0i}|X_i]$ and $E[Y_{1i}|X_i]$
  - Linear, different slopes
  - Polynomial
  - Local linear
- Produce an RD plot, visualising the discontinuity
- Statistical inference via regression standard errors

Estimation

- Consider the following model where $\tilde{X}_i = X - c$:

$$E[Y_i|D_i, X_i] = \alpha + \beta\tilde{X}_i + \tau D_i$$

- τ identifies the LATE in this model, i.e the difference between $E[Y_i|X_i = c, D_i = 1]$ and $E[Y_i|X_i = c, D_i = 0]$
- Why?

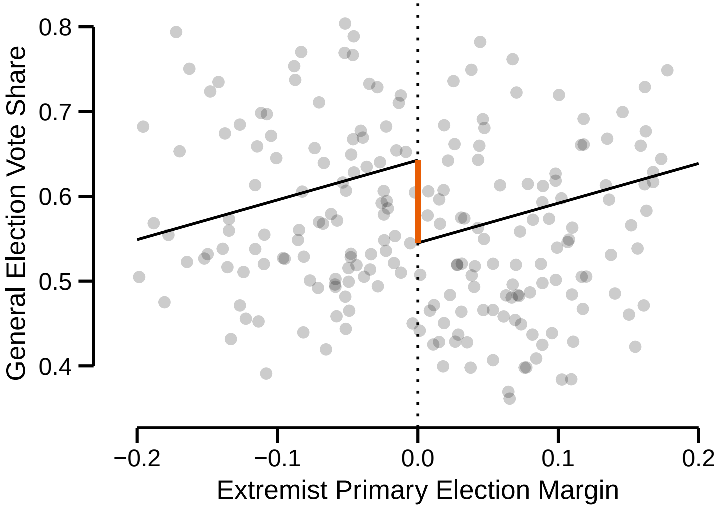
Estimation in R

```r
same_slope_model <- lm(vote_share_general ~ extreme + running_variable,
                       data = hall_subset)

different_slope_model <- lm(vote_share_general ~ extreme * running_variable,
                            data = hall_subset)

polynomial_model <- lm(vote_share_general ~ extreme * running_variable +
                                            extreme*I(running_variable^2) +
                                            extreme*I(running_variable^3),
                       data = hall_subset)
```

- The functional form of the running variable can be specified even more flexibly as LOESS or other smooth terms

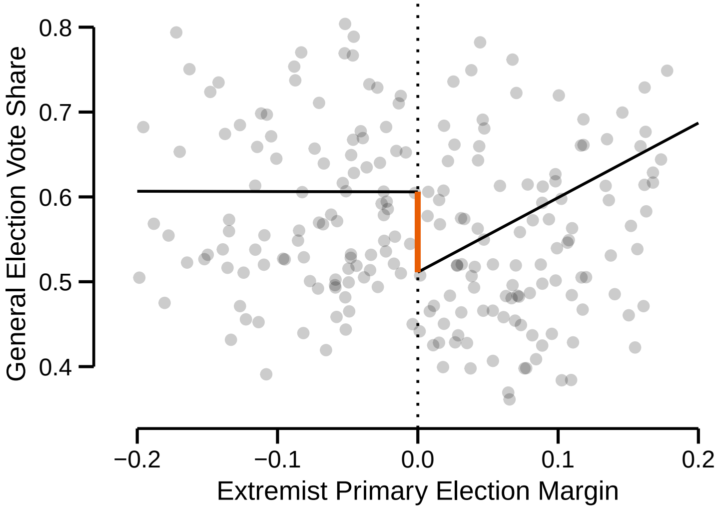
Linear model, same slopes

$$E[Y|\tilde{X}_i, D_i] = \alpha + \beta_1 \tilde{X}_i + \tau D_i$$



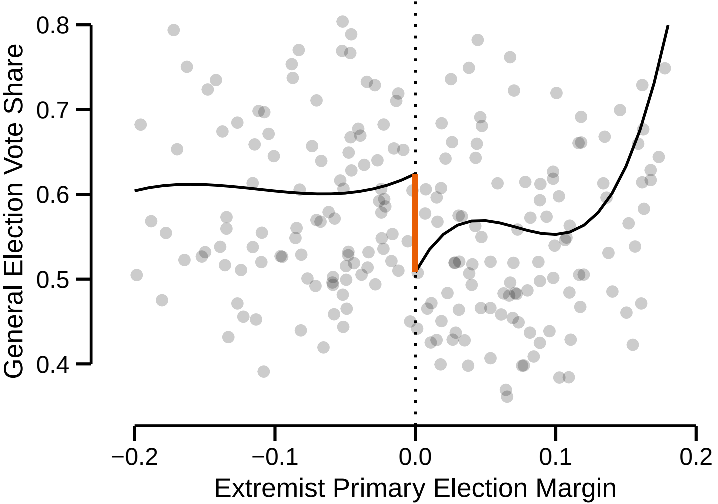
Linear model, different slopes

$$E[Y|\tilde{X}_i, D_i] = \alpha + \beta_{01}\tilde{X}_i + \beta_1(\tilde{X}_i D_i) + \tau D_i$$



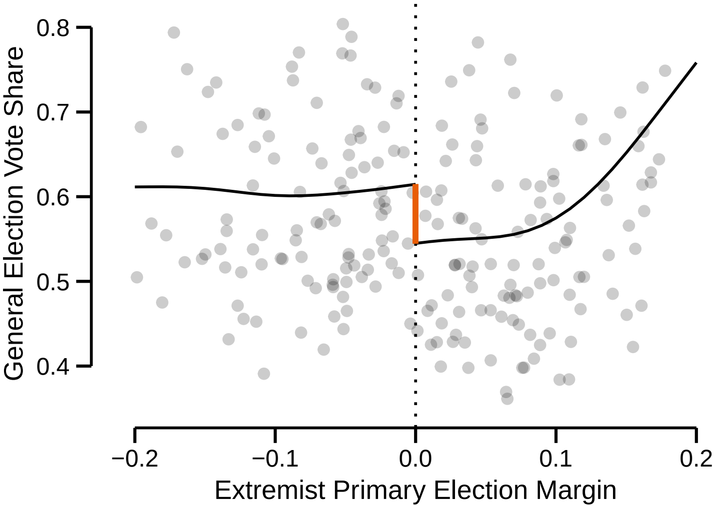
Polynomial model

$$E[Y|\tilde{X}_i, D_i] = \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \beta_{03}\tilde{X}_i^3 + \beta_1(\tilde{X}_i D_i) + \beta_2(\tilde{X}_i^2 D_i) + \beta_4(\tilde{X}_i^3 D_i) + \tau D_i$$



Nonlinear model I

$$E[Y|\tilde{X}_i, D_i] \quad = \quad \alpha + f(\tilde{X}_i) + \tau D_i$$



Non-linear model II

- The plot in the previous slide is based on a nonparametric regression model[2]
- You will use `RDestimate()` from the package `rdd` which uses local linear nonparametric regression
  - You don't need to understand the specifics, but the gist of it is: it calculates several regressions 'locally' within rolling windows of values the running variable

```
# install.packages(rdd)
library(rdd)
rd_model <- RDestimate(vote_share_general ~ running_variable,
                       cutpoint = 0, data=hall_subset)
```

```
...
## Estimates:
##          Bandwidth  Observations  Estimate  Std. Error  z value  Pr(>|z|)
## LATE     0.2        233           -0.07522  0.03180     -2.365   0.01801   *
...
```

[2] Specifically a generalised additive model, but you can immediately forget about this.

Comparing models

	Vote Share in GE				
	Same slope	Different slope	Polynomial	Non-linear	Local-linear
Intercept	0.643***	0.606***	0.624***	0.631***	0.609***
	(0.019)	(0.024)	(0.053)	(0.027)	(0.023)
Extremist Candidate	-0.098**	-0.095**	-0.116	-0.070	-0.075*
	(0.034)	(0.034)	(0.074)	(0.048)	(0.031)
R2	0.035	0.060	0.102	0.069	0.037
Num.Obs.	233	233	233	233	233

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

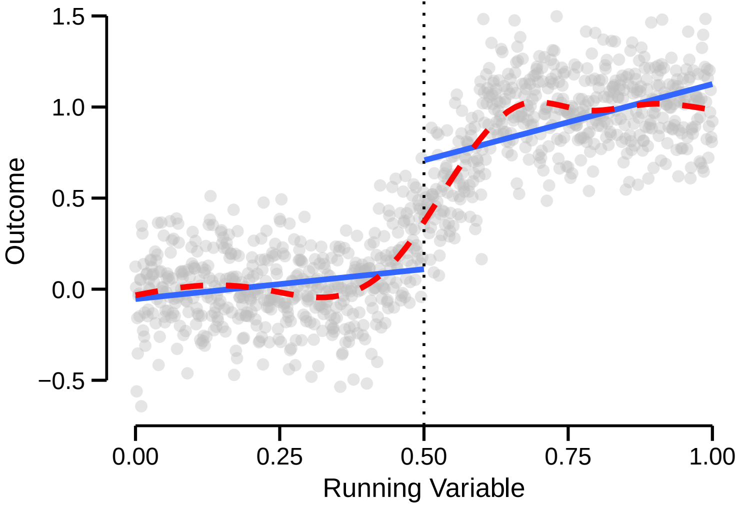
Interpretation

When an extremist wins a "coin-flip" US primary election, the party's GE vote share decreases, on average, by 7-12 percentage points.

What to do?

Non-linearity mistaken for discontinuity

The functional form for $\tilde{X}$ can be consequential for inferences about $\hat{\tau}_{\text{LATE}}$:



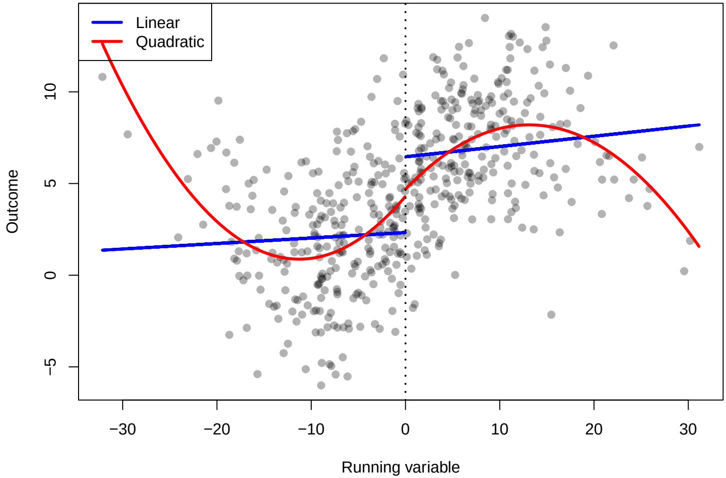
Bandwidth selection

- One way to reduce this type of **model dependence** is to focus only on observations that are close to the cutoff.
- In practice, this means only keeping observations with:

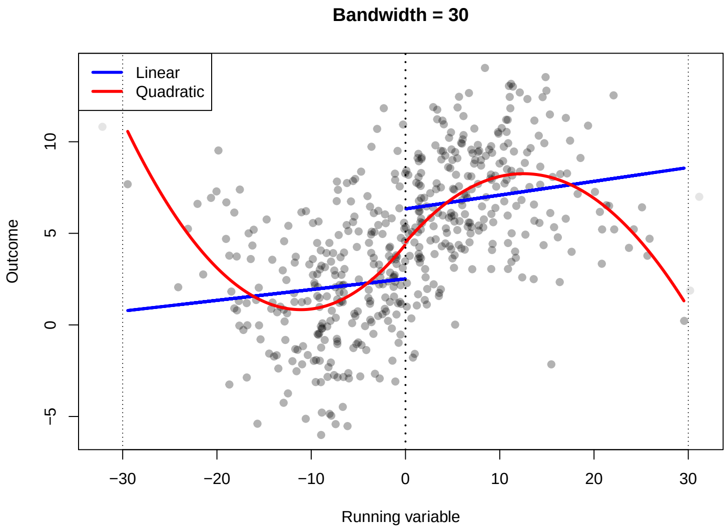
$$c - h \leq X_i \leq c + h$$

  where h is a positive value determining the window or bandwith size.
- The bandwidth controls the size of the neighbourhood around the cutoff that is used to calculate the discontinuity.
- h directly affects the properties of the estimation process and empirical findings can be sensitive to the particular value that one chooses for h.

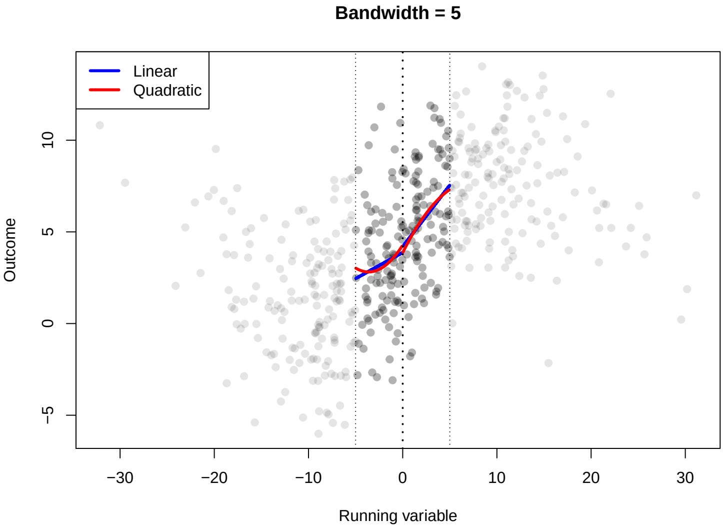
Bandwidth selection



Bandwidth selection



Bandwidth selection



Implications of bandwidth selection

- Comparing average outcomes in a small neighbourhood to the right and left of the cutoff leads to:
  - Estimates of LATE that are less dependent on the functional form specification for $\tilde{X}$
  - Decreases the bias that comes from misspecification
  - Leads to a smaller sample size, thus increasing the variance
- In picking h we face a **bias-variance** trade-off:
  - Smaller values of $h \rightarrow$ less bias in $\hat{\tau}_{\text{LATE}}$
  - Smaller values of $h \rightarrow$ greater variance in $\hat{\tau}_{\text{LATE}}$ $(SE(\hat{\tau}_{\text{LATE}}) \uparrow)$
- The choice of h is important for the estimates of $\hat{\tau}_{\text{LATE}}$.

How do we pick h?

Two approaches:

1. "Optimal" bandwidth selection
   - Use algorithmic bandwidth selection methods
   - Most common $\rightarrow$ Imbens-Kalyanaraman procedure
   - Choose h to balance bias-variance tradeoff
   - h is chosen to minimise the expected mean-square error of the RD estimator
2. Reporting results from multiple bandwidths
   - In practice, it is common to show that how much (if at all) the estimate of $\hat{\tau}_{\text{LATE}}$ changes as we vary the bandwidth

Optimal bandwidth in practice

```r
library(rdd)
optimal_bandwidth <- IKbandwidth(X = hall$running_variable,
                                 Y = hall$vote_share_general,
                                 cutpoint = 0)
optimal_bandwidth
```

```
## [1] 0.08507813
```

```r
rd_est <- RDestimate(vote_share_general ~ running_variable,
                     cutpoint = 0,
                     bw = optimal_bandwidth,
                     data = hall)
rd_est
```

```
##
## Call:
## RDestimate(formula = vote_share_general ~ running_variable, data = hall,
##     cutpoint = 0, bw = optimal_bandwidth)
##
## Coefficients:
##      LATE    Half-BW  Double-BW
##  -0.09506   -0.08582   -0.08792
```

Different bandwidths in practice

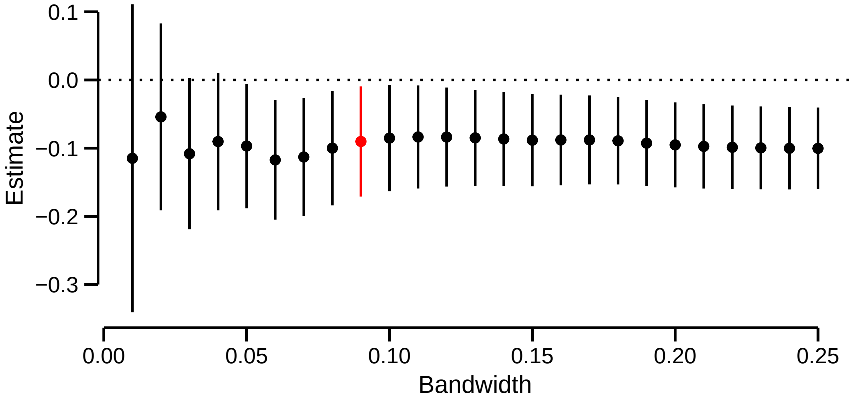
```
bandwidths <- c(seq(0.01,0.25,0.01))
rd_est <- RDestimate(vote_share_general ~ running_variable,
                     cutpoint=0,
                     bw=bandwidths,
                     data=hall)
rd_est
```

```
##
## Call:
## RDestimate(formula = vote_share_general ~ running_variable, data = hall,
##     cutpoint = 0, bw = bandwidths)
##
## Coefficients:
##  [1]  -0.11490  -0.05413  -0.10819  -0.09029  -0.09689  -0.11730  -0.11299
##  [8]  -0.09999  -0.09023  -0.08525  -0.08358  -0.08372  -0.08490  -0.08655
## [15]  -0.08835  -0.08802  -0.08790  -0.08926  -0.09271  -0.09522  -0.09742
## [22]  -0.09872  -0.09959  -0.10020  -0.10031
```

Code for figure

```
results <- data.frame(bw = bandwidths,
                      est = rd_est$est,
                      se = rd_est$se,
                      lo = rd_est$est - 1.96*rd_est$se,
                      hi = rd_est$est + 1.96*rd_est$se,
                      opt = bandwidths==round(optimal_bandwidth,2))

ggplot(results, aes(x= bw, y= est, color=opt)) +
  geom_point() +
  geom_linerange(aes(ymin=lo,ymax=hi)) +
  geom_hline(yintercept = 0, linetype="dotted") +
  scale_x_continuous("Bandwidth",breaks = seq(0,0.25,.05)) +
  scale_color_manual(values = c("black","red")) +
  ylab("Estimate") +
  theme_clean() +
  lemon::coord_capped_cart(bottom="both",left="both") +
  theme(plot.background = element_rect(color=NA),
        panel.grid.major.y = element_blank(),
        legend.position = "none",
        axis.ticks.length = unit(2,"mm"))
```

Different bandwidths for extremist candidates



Interpretation

The effect is consistently negative and significantly different from zero in most cases and therefore can be said to be **robust to different bandwidths**.

RDD Validation

Falsification checks

- **Balance checks**
  - Are covariates discontinuous at the threshold?
- **Placebo thresholds**
  - Do we estimate significant treatment effects at “placebo” thresholds, c^*?
- **Sorting**
  - Are units able to “sort” around the threshold?

Remember

If treatment is indeed as if randomly assigned around c, then treated and control units around c will be the same (in expectation) with respect to both observed and unobserved covariates.

We cannot check balance of unobserved covariates, but we can assess balance on **observed** covariates Z_i via...[3]

- Visual inspection
  - Plot $E[Z_i|\tilde{X}_i, D_i]$
  - The relationship between $\tilde{X}_i$ and Z_i should be smooth around c
- RD model for covariates
  - Estimate $E[Z_i|\tilde{X}_i, D_i] = \alpha + \beta_{01}\tilde{X}_i + \beta_1(\tilde{X}_i D_i) + \tau_Z D_i$
  - This should yield $\tau_Z = 0$ if Z_i is balanced at the threshold

[3] Z does not refer to an instrument here, but just other covariates that are not the runnig variable X.

Balance checks in practice

- Are women more or less likely to win the primary (i.e. be in the treatment group) at the cutoff?

```
rd_est_female <- RDestimate(winner_female ~ running_variable,
                            cutpoint=0, bw=optimal_bandwidth, data=hall)
```

- Are more experienced candidates more or less likely to win the primary at the cutoff?

```
rd_est_qual <- RDestimate(qual ~ running_variable,
                          cutpoint=0, bw=optimal_bandwidth, data=hall)
```

- Are candidates with more donations more or less likely to win at the cutoff?

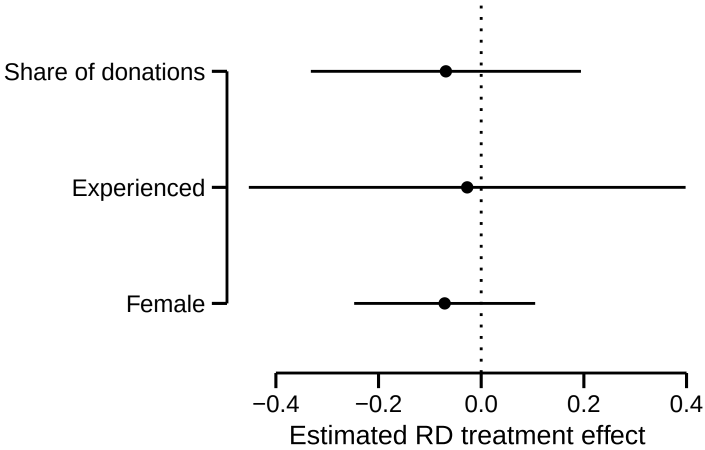
```
rd_est_pac <- RDestimate(prim_pac_share ~ running_variable,
                         cutpoint=0,bw=optimal_bandwidth, data=hall)
```

A "yes" to any of these questions suggests that the identification assumption does not hold.

Code for figure

```
female <- data.frame(pe = rd_est_female$est[1],
                     hi = rd_est_female$est[1] + rd_est_female$se[1]*1.96,
                     lo = rd_est_female$est[1] - rd_est_female$se[1]*1.96)
qual <- data.frame(pe = rd_est_qual$est[1],
                   hi = rd_est_qual$est[1] + rd_est_qual$se[1]*1.96,
                   lo = rd_est_qual$est[1] - rd_est_qual$se[1]*1.96)
pac_share <- data.frame(pe = rd_est_pac$est[1],
                        hi = rd_est_pac$est[1] + rd_est_pac$se[1]*1.96,
                        lo = rd_est_pac$est[1] - rd_est_pac$se[1]*1.96)

out <- rbind(female, qual, pac_share)
out <- cbind(out, "var" = c("Female", "Experienced", "Share of donations"))

 ggplot(out, aes(x= pe,  y= var)) +
  geom_point() +
  geom_linerange(aes(xmin=lo,xmax=hi)) +
  geom_vline(xintercept = 0, linetype="dotted") +
  scale_x_continuous("Estimated RD treatment effect", breaks = seq(-.4,.4,.2)) +
  scale_y_discrete("", limits=c("Female","Experienced","Share of donations")) +
  theme_clean() +
  lemon::coord_capped_cart(bottom="both",left="both") +
  theme(plot.background = element_rect(color=NA),
        panel.grid.major.y = element_blank(),
        legend.position = "none",
        axis.ticks.length = unit(2,"mm"))
```

Balance checks for extremist candidates



Interpretation

There does not seem to be systematic differences in terms of **observable** characteristics between the treatment and control groups.

Placebo thresholds

- We can also check whether the discontinuity only appears where it "should" appear, and that it is zero at other values of the cutoff.
- If we have a placebo value $c^* \neq c$, then define $\tilde{X}_i^* = X_i - c^*$ and estimate:

$$E[Y_i|\tilde{X}_i^*, D_i] = \alpha + \beta_{01}\tilde{X}_i^* + \beta_1(\tilde{X}_i^* D_i) + \tau^* D_i$$

  or more flexible specifications thereof.

Implication

If our RDD is valid, we should find no significant treatment effects τ^* for any $c^* \neq c$.

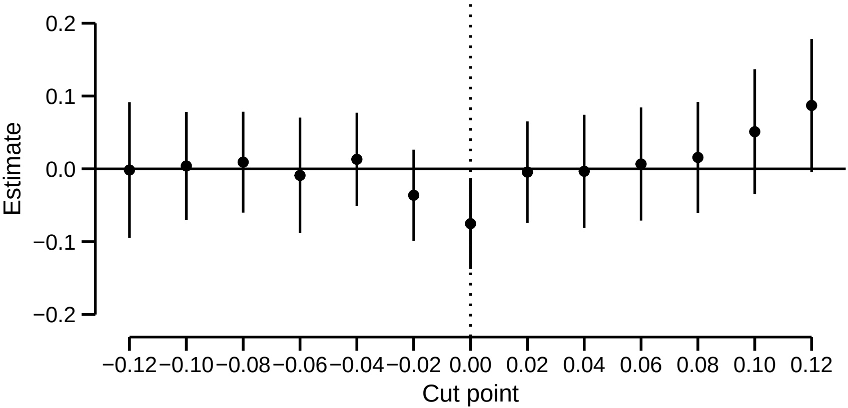
Placebo tests in practice

```
# create empty data
out <- data.frame(cstar = NA, est = NA, se = NA)

# placebo cut-offs
cuts <- seq(-.12,.12,0.02)

# loop and store results
for (cstar in cuts){
    rd_est_placebo <- RDestimate(vote_share_general ~ running_variable,
                                 cutpoint=cstar, bw=optimal_bandwidth,
                                 data=hall)
    res <- data.frame(cstar = cstar,
                      est = rd_est_placebo$est[1],
                      se = rd_est_placebo$se[1])
    out <- rbind(out,res)
}

# confidence intervals
out$lo <- out$est - 1.96*out$se
out$hi <- out$est + 1.96*out$se
```

Code for figure

```r
ggplot(out, aes(x= cstar, y= est)) +
  geom_point() +
  geom_linerange(aes(ymin=lo,ymax=hi)) +
  geom_hline(yintercept = 0) +
  geom_vline(xintercept = 0, linetype="dotted") +
  scale_x_continuous("Cut point",breaks = seq(-.12,.12,0.02)) +
  scale_y_continuous("Estimate",breaks = seq(-.2,.2,.1),limits = c(-.21,.21)) +
  theme_clean() +
  lemon::coord_capped_cart(bottom="both",left="both") +
  theme(plot.background = element_rect(color=NA),
        panel.grid.major.y = element_blank(),
        legend.position = "none",
        axis.ticks.length = unit(2,"mm"))
```

Placebo test for extremist candidates



Interpretation

We only find a statistically significant, negative, effect at the **actual** cutoff, strengthening confidence in the validity of the RDD.

Sorting

- The RDD is based on the assumption that there is continuity in the potential outcomes at the threshold.
- One way this assumption might be violated is if units can control their values of the running variable.
- If this happens, this implies sorting of units into treatment

Examples of sorting

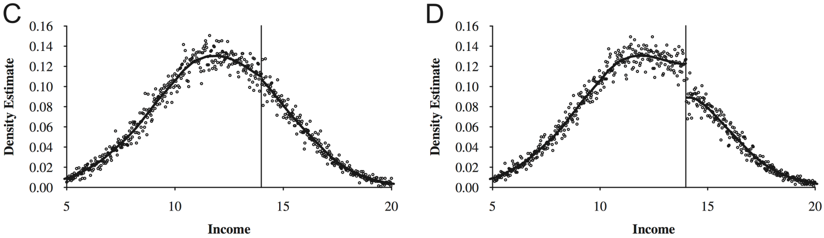
- Population thresholds: Administrators might misreport population in town/district if particular benefits are received at certain thresholds (Eggers et al., 2018)
- Earnings thresholds: Individuals may reduce their earnings if benefits are granted to those below a certain income **(e.g. McCrary, 2008)**
- Geographic thresholds: Businesses might locate in different areas if benefits are allocated differentially across localities **(e.g. Keele and Titunik, 2015)**

Implication

Sorting $\rightarrow$ "as if random" assumption violated $\rightarrow$ selection bias.

Investigating sorting in practice

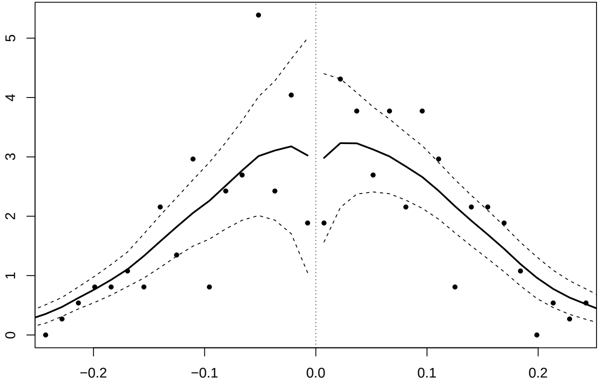
McCrary (2008) proposes a test to detect sorting in $\tilde{X}_i$:



- Look for evidence of discontinuous jumps in the **running variable** at c
- Null hypothesis $\rightarrow$ no sorting (small p-values suggest evidence of sorting)
- `DCdensity(running_variable)` in R

```r
DCdensity(hall$running_variable)
```

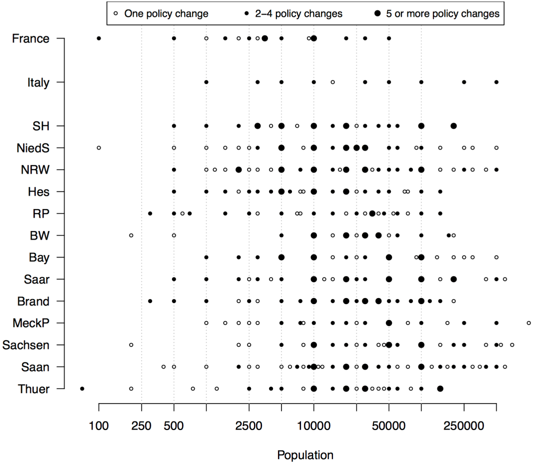
```
## [1] 0.9563002
```



Compound treatments

- RDD assumes that **the only** thing that is determined by X_i at the cutoff is the probability of receiving the treatment.
- It is often the case that there are **multiple** changes at a given cutoff, and so we can only estimate a **compound treatment effect**
- Eggers (2015) uses the fact that French towns with ≥ 3500 people hold PR elections while towns with < 3500 hold majoritarian elections.
  - Outcome (Y_i): Turnout in municipality
  - Treatment (D_i): PR election system
  - Running variable (X_i): Population of municipality
  - Cutoff (c): 3500
- Key question: Is the electoral system the only thing that changes at 3500?

Compound treatments (Eggers et. al., 2018)



Fuzzy Regression Discontinuity Design

Fuzzy RDD

- Thresholds/cutoffs may not perfectly determine treatment status, but might still create discontinuities in the **probability of treatment exposure**. E.g.
  - Incentives to participate in a program may change discontinuously at a threshold…
  - … but the incentives are not powerful enough to move all units from nonparticipation to participation
- We can think of the cutoff as **assigning** units to a treatment condition, where only some units will **comply** with the treatment.
- We can use **discontinuities** to produce **instrumental variable** estimators of the treatment (close to the discontinuity).

Assumptions in Fuzzy RDD

1. **First stage**
   - There should be a discontinuity in treatment probability at the cutoff
   - Empirically: check RD plots with running variable on X and treatment probability on Y
2. ***Local* independence of the instrument**
   - The treatment assignment should be as good as random around the cutoff
   - Empirically: check RD balance plots of covariates
3. **Monotonicity**
   - No units should be discouraged from taking the treatment at the cutoff
4. **Exclusion restriction**
   - Crossing the cutoff should only affect the outcome through a unit's treatment values, not through any other channel
   - No compound effects

Fuzzy RD example

Does education decrease anti-immigrant views?

Although low-levels of education are powerful predictors of anti-immigrant sentiment, it is difficult to establish a causal relationship between education and attitudes towards immigrants. Cavaillé and Marshall (2018) use an RDD to address this question by exploiting changes to the length of mandatory education in five countries (Denmark, France, UK, Netherlands, and Sweden).

- Outcome (Y_i): Index of anti-immigrant attitudes
- Treatment (D_i): 1 if respondent was affected by the reform
- Running variable (X_i): Birth year of the respondent minus year the birth year of those first affected by the policy

Schooling and immigration attitudes

TABLE 1. Compulsory Education Reforms

Country	Date of reform passing	Year reform came into effect	Change in minimum school leaving age	Change in years of compulsory education	Year of birth of first affected cohort
Denmark	1958	1958	14 to 15	7 to 8	1944
France	1959	1967	14 to 16	8 to 10	1953
Great Britain	1944	1947	14 to 15	9 to 10	1933
Great Britain†	1962	1972	15 to 16	10 to 11	1957
Netherlands	1975	1974	15 to 16	9 to 10	1959
Sweden	1962	1965	14 to 15	8 to 9	1951

Notes: †The second reform in Great Britain was first passed in 1962, but not implemented until a 1972 statutory instrument. The British reforms did not affect Northern Ireland.

- Here, treatment is determined by age:

$$D_{i,c} = \begin{cases} 1 & \text{if birth year}_{i,c} \geq \text{birth year of first effected}_c \\ 0 & \text{if birth year}_{i,c} < \text{birth year of first effected}_c \end{cases}$$

- But many students would have stayed in school longer even in the absence of a reform. We therefore have some non-compliance (**always-takers**).

Fuzzy RD estimation

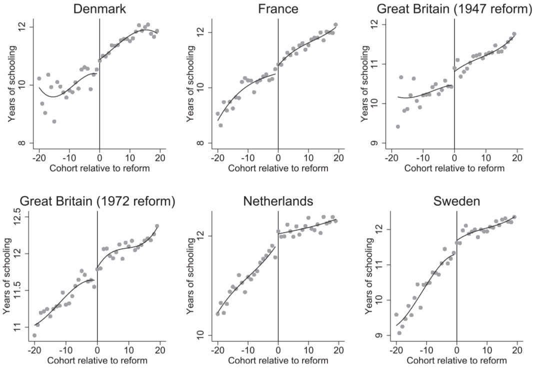
- Restrict data to small window above and below the cutoff ($\pm h$)
- Code the instrument using the running variable $Z_i = 1\{X_i > c\}$)
- Fit 2SLS

$$Y_i = \alpha + \beta_1 \tilde{X}_i + \beta_2 \hat{D}_i \tilde{X}_i + \tau \hat{D}_i$$

  where $\hat{D}_i$ is instrumented by Z_i and $\tilde{X}_i = X_i - c$
- We can, as before, add more flexible specifications for $\tilde{X}_i$
- We would normally also plot and estimate both the first- and second-stage discontinuities

First stage

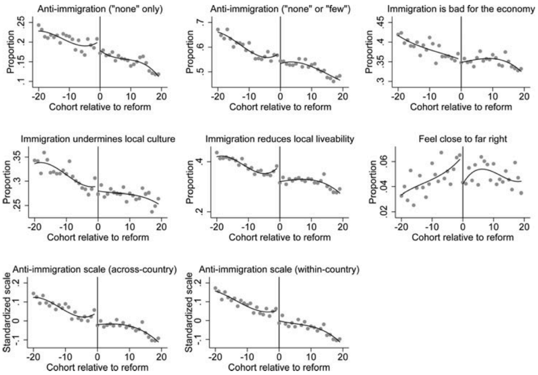
FIGURE 1. Years of Completed Schooling among Cohorts around Compulsory Schooling Reforms, by Reform (Third-Order Polynomials Either Side of the Reform)



Implication

Among respondents who where within 20 years of the affected cohort, reforms are associated with an increase in secondary schooling by 0.29 years on average.

FIGURE 3. Anti-immigration Attitudes among Cohorts around Compulsory Schooling Reforms, Pooled across Reforms (Third-Order Polynomials Either Side of the Reform).



Implication

On many indexes, reform-affected students are less opposed to immigration.

LATE

	Years of completed schooling (1)	Anti-immigration ("none" only) (2)	Anti-immigration ("none" or "few") (3)	Immigration is bad for the economy (4)	Immigration undermines local culture (5)	Immigration reduces local livability (6)	Feel close to far-right (7)	Anti-immigration scale (across) (8)	Anti-immigration scale (within) (9)
Panel H: Fuzzy RD (instrumental variables) estimates—all reforms pooled									
Years of completed schooling		−0.087*	−0.083+	−0.031	−0.024	−0.180**	−0.065*	−0.214**	−0.183**
		(0.038)	(0.049)	(0.044)	(0.042)	(0.058)	(0.031)	(0.077)	(0.079)
Bandwidth		9	7	10	9	6	9	7	8
Observations		24,278	19,281	26,789	24,278	16,740	14,910	19,281	21,777
Outcome mean		0.18	0.55	0.36	0.29	0.34	0.05	0.00	0.01
Years of completed schooling mean		11.32	11.32	11.32	11.32	11.32	11.32	11.32	11.32
First stage *F* statistic		22.0	20.6	22.6	22.1	19.0	12.9	19.3	21.1

Note that the LATE estimated in a fuzzy RD is "local" in two ways:

- Local to the threshold
- Local to the compliers

Fuzzy RDD in R 'by hand'

```
# Restrict data to +/- 7 years round cutoff
schoolreform2 <- schoolreform[schoolreform$running %in% c(-7:7),]

# First stage
firststage <- lm(eduyrs ~ treatment + running,
                 data = schoolreform2)

# Predict D hat
schoolreform2$Dhat <- predict(firststage)

# Second stage
secondstage <- lm(anti_immigr_index ~ Dhat + running,
                  data = schoolreform2)
```

- Note that the running variable **has** to be included in both stages, as it is a control, **not the instrument**!

Fuzzy RDD with `rdd`

- The endogenous variable (i.e. the one that is being instrumented) should be included in the `formula` argument *behind* the running variable

```
fuzzyrdd <- RDestimate(anti_immigr_index ~ running + eduyrs,
                       data = schoolreform,
                       bw = 8)
```

	Linear OLS	Local Linear
First stage	0.28*** (0.062)	0.32*** (0.058)
LATE	-0.17* (0.072)	-0.20*** (0.06)
Num. Obs.	19,261	19,261

The ever-looming question of validity

- Internal validity
  - RDD is a transparent approach to inference which requires less stringent assumptions that IV (at least in the Sharp RDD case)
  - Many of the key identifying assumptions are empirically verifiable
  - RDD has been shown to do a very good job at recovering known experimental benchmarks (Cook & Wong, 2008)
- External validity
  - Sharp RDD only identifies the ATE *at the point of the discontinuity*
  - Fuzzy RDD only identifies the ATE *at the point of the discontinuity*, **amongst compliers**
  - Generalizability depends on how weird the units are at the cutoff, and how weird the compliers are

Data requirements for RDD

- Unlike some other methods we have studied, RDD designs do not require either extensive covariates or repeated data on the same units over time.
- The main requirement is to find a discontinuity!
- The key is to find a running variable that
  1. ...leads to a discontinuous jump in the probability of treatment
  2. ...is not possible for units to manipulate
- This often requires a great deal of substantive knowledge of different settings.