Causal Inference: Regression Discontinuity Designs

Ken Stiller

20th June 2023

Overview

Strategies for estimating effects of treatments so far:

- Randomize treatment and take the DIGM
- Identify and control for confounding variables such that the CIA holds
- Identify a plausible counterfactual using time and unit variation in panel data
- Identify an instrumental variable and use two-stage-least-squares to estimate average treatment effect for compliers.

Today: take advantage of special situations where treatment is applied based on a cutoff

Discontinuities

"Natura non facit saltus." (Nature makes no jump.)
- Gottfried Leibniz

But humans sometimes use rules that create jumps:

- Students pass with a 4.0 or better
- French municipalities use PR elections if population is 3,500 (now 1,000) or higher
- A candidate is elected if she receives more votes than any other candidate
- Journalists report a recession if the economy shrinks for two consecutive quarters

Regression Discontinuity Design (RDD)

- RDD is a fairly old idea (Thistlethwaite and Campbell, 1960) but this design experienced a renaissance in recent years
- Assignment to treatment and control is not random, but we know the assignment rule influencing how people are assigned or selected in to treatment
- Widely applicable in a rule based world (administrative programs, elections, etc.)
- High internal validity: In their validation study Cook and Wong (2008) identify RDD as one of the few observational study designs that can accurately reproduce experimental benchmarks

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Assignment Mechanism

- Trying to do causal inference by modeling $f(Y_0|D)$ is not appealing because Y_0 is unobserved whenever $D = 1$
- Often, better strategy is to study the assignment mechanism:

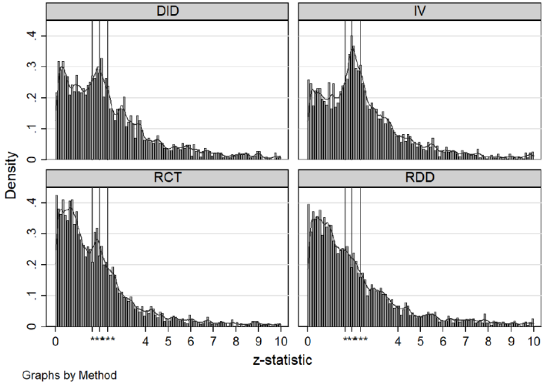
Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units get which treatment. Examples include:

- random assignment - sometimes not doable
- selection on observables - usually unknown
- selection on unobservables

- Most statistical models (in causal inference) attain identification of treatment effects by restricting the assignment mechanism
- RD uses assignment based on rules

Why Assignment Matters!



This figure displays histograms of test-coefficients for $z \in [0, 10]$. Coefficients are partitioned by identification method: difference-in-differences (DID), instrumental variables (IV), randomized control trial (RCT) and regression discontinuity design (RDD). Bins are 0.1 wide. Reference lines are displayed at conventional significance levels. We have also superimposed an Epanechnikov kernel. Test statistics have been de-rounded.

Source: Brodeur, Cook, and Heyes AER (2020).

Outline

Introduction

Identification

Sharp RDD: Estimation

Falsification Checks

Hypothetical Illustration Example: Sharp RDD

- Thistlethwaite and Campbell (1960) study the effects of college scholarships on later students' achievements
- Scholarships are given on the basis of whether or not a student's test score exceeds some *arbitrary* cut-off c
  - Treatment D is scholarship
  - Forcing variable X is SAT score with cutoff c
  - Outcome Y is subsequent earnings
  - Y_0 denotes potential earnings without the scholarship
  - Y_1 denotes potential earnings with the scholarship
- Y_1 and Y_0 are correlated with X: on average, students with higher SAT scores obtain higher earnings

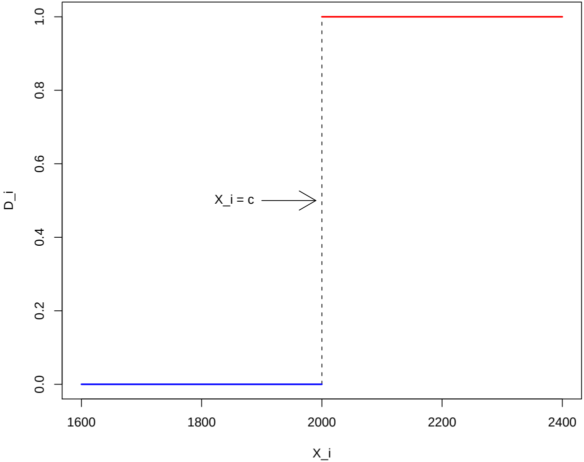
Sharp Regression Discontinuity Design

- Imagine a binary treatment D that is completely determined by the value of a predictor X_i being on either side of a fixed cutoff point c:

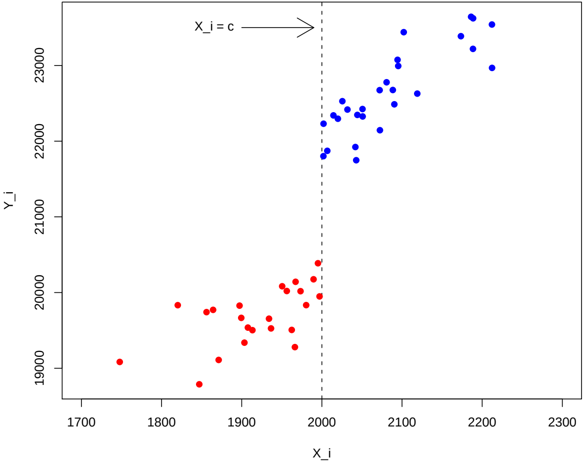
$$D_i = 1\{X_i > c\} \text{ so } D_i = \begin{cases} D_i = 1 & \text{if } X_i > c \\ D_i = 0 & \text{if } X_i < c \end{cases}$$

- X_i, called the forcing variable, may be correlated with the outcomes Y so comparing treated and untreated units does not provide causal estimates
  - (e.g. students with higher SAT scores obtain higher earnings even without the scholarship)
- Design arises often from administrative decisions, where the allocation of units to a program is partly limited for reasons of resource constraints, and sharp rules rather than discretion by administrators is used for allocation

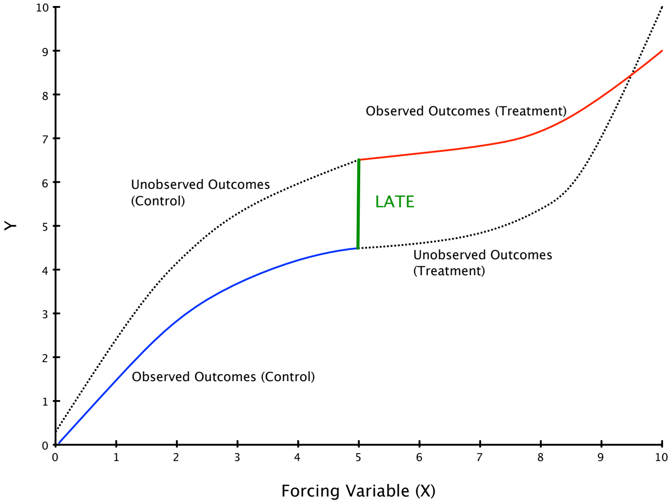
Sharp RDD: Graphical Illustration



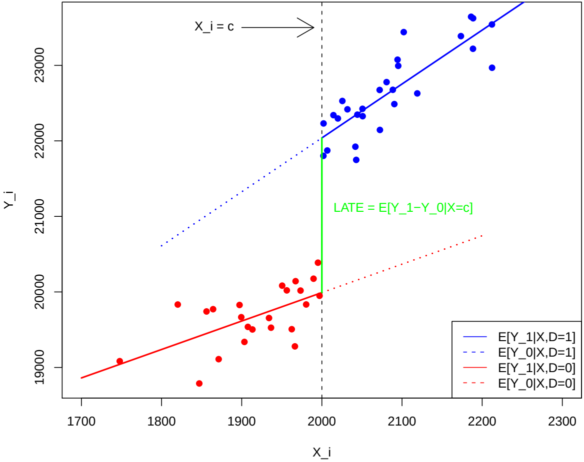
Sharp RDD: Graphical Illustration



Potential Outcomes



Sharp RDD: Graphical Illustration



Sharp RDD: Identification

Identification Assumption

$E[Y_1|X, D]$ and $E[Y_0|X, D]$ are continuous in X around the threshold $X = c$ (individuals have imprecise control over X around the threshold)

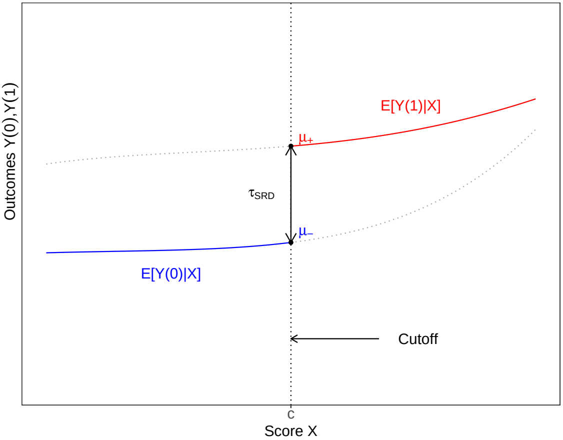
Identification Result

The treatment effect is identified at the threshold as:

$$\begin{aligned} \alpha_{SRDD} &= E[Y_1 - Y_0|X = c] \\ &= E[Y_1|X = c] - E[Y_0|X = c] \\ &= \lim_{x \downarrow c} E[Y_1|X = c] - \lim_{x \uparrow c} E[Y_0|X = c] \end{aligned}$$

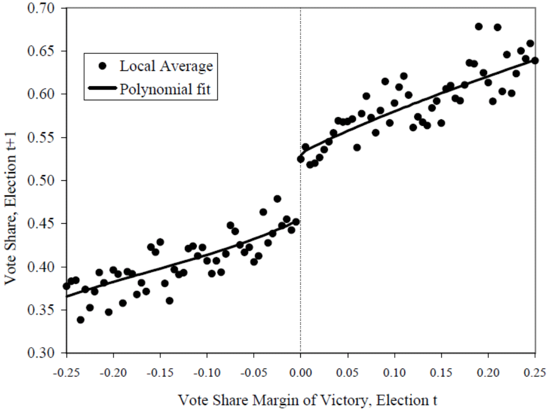
Without further assumptions α_{SRDD} is only valid at the threshold.

$$\tau_{\texttt{SRD}} = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]}_{\text{Unobservable}} = \underbrace{\lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x]}_{\text{Estimable}} - \underbrace{\lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]}_{\text{Estimable}}$$



Sharp RDD: Incumbency Advantage

Figure IVa: Democrat Party's Vote Share in Election t+1, by Margin of Victory in Election t: local averages and parametric fit



What is the role of peer effects in voting?

- What affects the likelihood of voting? Education? Income? Your upbringing? Peers?
- How would you estimate the role of peer effects on the likelihood of voting?
- What could be the sources of selection bias?

Daahlgard et al (BJOPS) study couples that moved in right before vs right after an election. Why is this smart? What is the cutoff? What is the running variable?

What is the role of peer effects in voting? (Daahlgard et al, 2021)

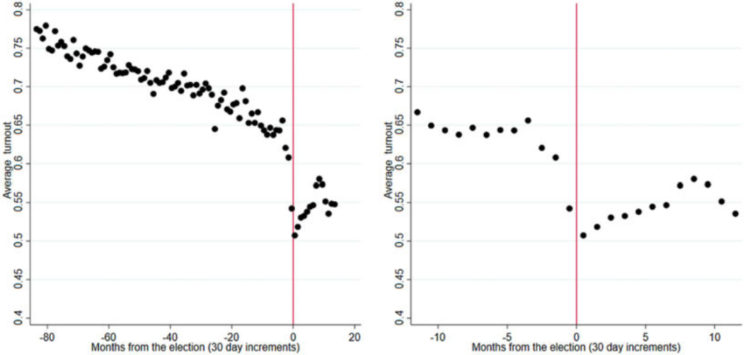


Figure 1. Turnout by month of cohabitation

Note: The figure shows average turnout binned by 30-day windows around election day represented by the vertical line. The households are placed in bins based on when they moved together. Households on the left side of the vertical line formed before the election and households on the right side formed after.

Outline

Introduction

Identification

Sharp RDD: Estimation

Falsification Checks

Estimate $\alpha_{SRDD} = E[Y_1|X = c] - E[Y_0|X = c]$

1. Trim the sample to a reasonable window around the threshold c (discontinuity sample):
   - $c - h \leq X_i \leq c + h$, were h is some positive value that determines the size of the window
   - h may be determined by one of the methods (e.g. Imbens and Kalyanaraman 2012, refinement in CCT 2020) discussed later
2. Recode running variable to deviations from threshold: $\tilde{X} = X - c$.
   - $\tilde{X} = 0$ if $X = c$
   - $\tilde{X} > 0$ if $X > c$ and thus $D = 1$
   - $\tilde{X} < 0$ if $X < c$ and thus $D = 0$
3. Decide on a model for $E[Y|X]$
   - linear, different slopes for $E[Y_0|X]$ and $E[Y_1|X]$
   - non-linear, different slopes
   - start with visual inspection (scatter plot with kernel/lowess)
   - usually local polynomial of order 1 are used

Estimate $\alpha_{SRDD} = E[Y_1|X = c] - E[Y_0|X = c]$

1. Trim the sample to a reasonable window around the threshold c (discontinuity sample):
   - $c - h \leq X_i \leq c + h$, were h is some positive value that determines the size of the window
   - h may be determined by one of the methods (e.g. Imbens and Kalyanaraman 2012, refinement in CCT 2020) discussed later
2. Recode running variable to deviations from threshold: $\tilde{X} = X - c$.
   - $\tilde{X} = 0$ if $X = c$
   - $\tilde{X} > 0$ if $X > c$ and thus $D = 1$
   - $\tilde{X} < 0$ if $X < c$ and thus $D = 0$
3. Decide on a model for $E[Y|X]$
   - linear, different slopes for $E[Y_0|X]$ and $E[Y_1|X]$
   - non-linear, different slopes
   - start with visual inspection (scatter plot with kernel/lowess)
   - usually local polynomial of order 1 are used

Estimate $\alpha_{SRDD} = E[Y_1|X = c] - E[Y_0|X = c]$

1. Trim the sample to a reasonable window around the threshold c (discontinuity sample):
   - $c - h \leq X_i \leq c + h$, were h is some positive value that determines the size of the window
   - h may be determined by one of the methods (e.g. Imbens and Kalyanaraman 2012, refinement in CCT 2020) discussed later
2. Recode running variable to deviations from threshold: $\tilde{X} = X - c$.
   - $\tilde{X} = 0$ if $X = c$
   - $\tilde{X} > 0$ if $X > c$ and thus $D = 1$
   - $\tilde{X} < 0$ if $X < c$ and thus $D = 0$
3. Decide on a model for $E[Y|X]$
   - linear, different slopes for $E[Y_0|X]$ and $E[Y_1|X]$
   - non-linear, different slopes
   - start with visual inspection (scatter plot with kernel/lowess)
   - usually local polynomial of order 1 are used

Different models

- Same slope at both sides of the cut-off: Run a regression of Y on D and the margin $\tilde{X} = X - c$.

$$Y = \gamma + \underbrace{\alpha}_{Treat\ Effect} D + \underbrace{\beta \overbrace{\tilde{X}}^{Forcing\ Variable}}_{Control\ Function} + \varepsilon$$

- Different slopes at different sides of the cut-off:

$$Y = \gamma + \underbrace{\alpha}_{Treat\ Effect} D + \underbrace{\beta_0 \tilde{X} + \beta_1 (\tilde{X} \cdot D)}_{Control\ Function} + \varepsilon$$

- Polynomial function:

$$Y = \gamma_0 + \underbrace{\alpha_0}_{Treat\ Effect} D + \underbrace{\gamma_1 \tilde{X} + \gamma_2 \tilde{X}^2 + \gamma_3 \tilde{X}^3 + \alpha_1 (\tilde{X} \cdot D) + \alpha_2 (\tilde{X}^2 \cdot D) + \alpha_3 (\tilde{X}^3 \cdot D)}_{Control\ Function} +$$

Different models

- Same slope at both sides of the cut-off: Run a regression of Y on D and the margin $\tilde{X} = X - c$.

$$Y = \gamma + \underbrace{\alpha}_{Treat\ Effect} D + \underbrace{\beta \overbrace{\tilde{X}}^{Forcing\ Variable}}_{Control\ Function} + \varepsilon$$

- Different slopes at different sides of the cut-off:

$$Y = \gamma + \underbrace{\alpha}_{Treat\ Effect} D + \underbrace{\beta_0 \tilde{X} + \beta_1 (\tilde{X} \cdot D)}_{Control\ Function} + \varepsilon$$

- Polynomial function:

$$Y = \gamma_0 + \underbrace{\alpha_0}_{Treat\ Effect} D + \underbrace{\gamma_1 \tilde{X} + \gamma_2 \tilde{X}^2 + \gamma_3 \tilde{X}^3 + \alpha_1 (\tilde{X} \cdot D) + \alpha_2 (\tilde{X}^2 \cdot D) + \alpha_3 (\tilde{X}^3 \cdot D)}_{Control\ Function} +$$

Different models

- Same slope at both sides of the cut-off: Run a regression of Y on D and the margin $\tilde{X} = X - c$.

$$Y = \gamma + \underbrace{\alpha}_{Treat\ Effect} D + \underbrace{\beta \overbrace{\tilde{X}}^{Forcing\ Variable}}_{Control\ Function} + \varepsilon$$

- Different slopes at different sides of the cut-off:

$$Y = \gamma + \underbrace{\alpha}_{Treat\ Effect} D + \underbrace{\beta_0 \tilde{X} + \beta_1 (\tilde{X} \cdot D)}_{Control\ Function} + \varepsilon$$

- Polynomial function:

$$Y = \gamma_0 + \underbrace{\alpha_0}_{Treat\ Effect} D + \underbrace{\gamma_1 \tilde{X} + \gamma_2 \tilde{X}^2 + \gamma_3 \tilde{X}^3 + \alpha_1 (\tilde{X} \cdot D) + \alpha_2 (\tilde{X}^2 \cdot D) + \alpha_3 (\tilde{X}^3 \cdot D)}_{Control\ Function} +$$

Sharp RDD Estimation: Differential Slopes

- $E[Y_0|X]$ and $E[Y_1|X]$ are distinct linear functions of X, so the average effect of the treatment $E[Y_1 - Y_0|X]$ varies with X:

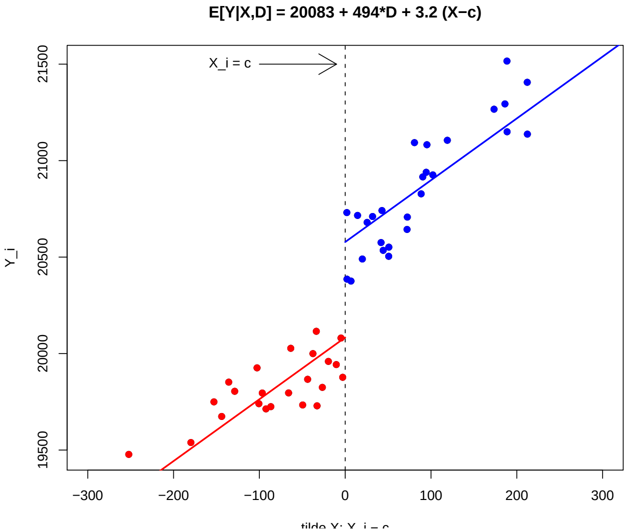
$$E[Y_0|X] = \mu_0 + \beta_0 X, \qquad E[Y_1|X] = \mu_1 + \beta_1 X$$

- So $\alpha(X) = E[Y_1 - Y_0|X] = (\mu_1 - \mu_0) + (\beta_1 - \beta_0)X$ we have

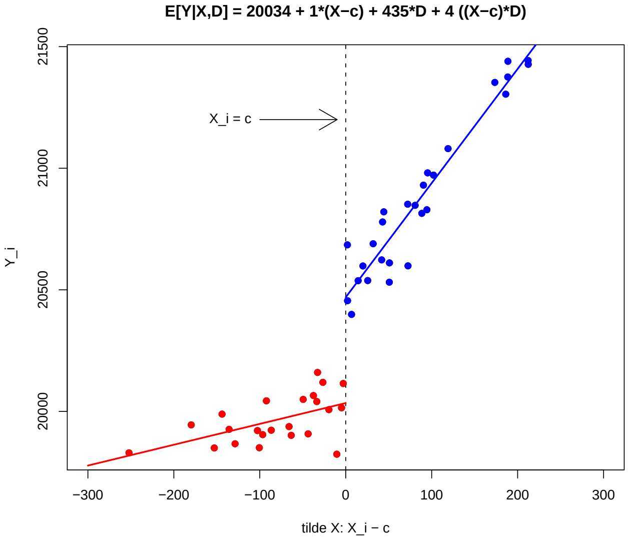
$$\begin{aligned} E[Y|X, D] =& D \cdot E[Y_1|X] + (1 - D) \cdot E[Y_0|X] \\ =& \mu_1 D + \beta_1(X \cdot D) + \mu_0(1 - D) + \beta_0(X \cdot (1 - D)) \\ =& \gamma + \beta_0(X - c) + \alpha D + \beta_1((X - c) \cdot D) \end{aligned}$$

- Regress Y on $(X - c)$, D and the interaction $(X - c) \cdot D$, the coefficient of D reflects the average effect of the treatment at $X = c$

Linear with Same Slope



SRDD: Linear with Differential Slope



Sharp RDD Estimation: Non-Linear Case

- $E[Y_0|X]$ and $E[Y_1|X]$ are distinct non-linear functions of X and the average effect of the treatment $E[Y_1 - Y_0|X]$ varies with X
- Include quadratic and cubic terms in $(X - c)$ and their interactions with D in the equation
- The specification with quadratic terms is

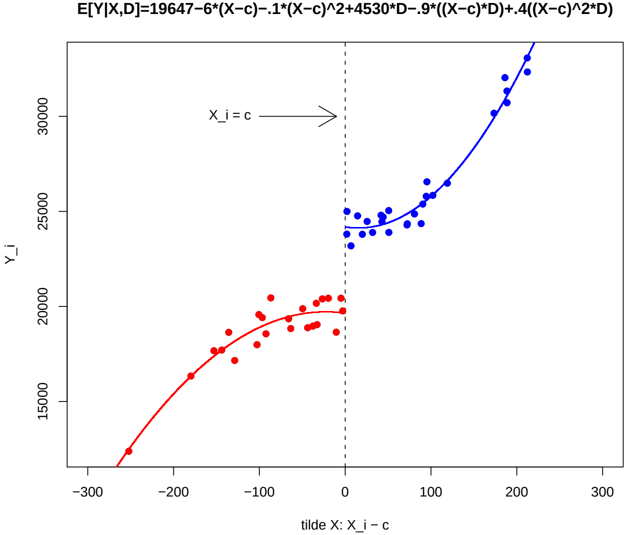
$$\begin{aligned} E[Y|X, D] =& \gamma_0 + \gamma_1(X - c) + \gamma_2(X - c)^2 \\ &+ \alpha_0 D + \alpha_1((X - c) \cdot D) + \alpha_2((X - c)^2 \cdot D) \end{aligned}$$

The specification with cubic terms is

$$\begin{aligned} E[Y|X, D] =& \gamma_0 + \gamma_1(X - c) + \gamma_2(X - c)^2 + \gamma_3(X - c)^3 + \alpha_0 D \\ &+ \alpha_1((X - c) \cdot D) + \alpha_2((X - c)^2 \cdot D) + \alpha_3((X - c)^3 \cdot D) \end{aligned}$$

- In both cases $\alpha_0 = E[Y_1 - Y_0|X = c]$

SRDD: Non-Linear Case



Local Polynomial Methods

- **Idea**: approximate regression functions for control and treatment units *locally*.
- "Local-linear" ($p = 1$) estimator (w/ weights $K(\cdot)$):

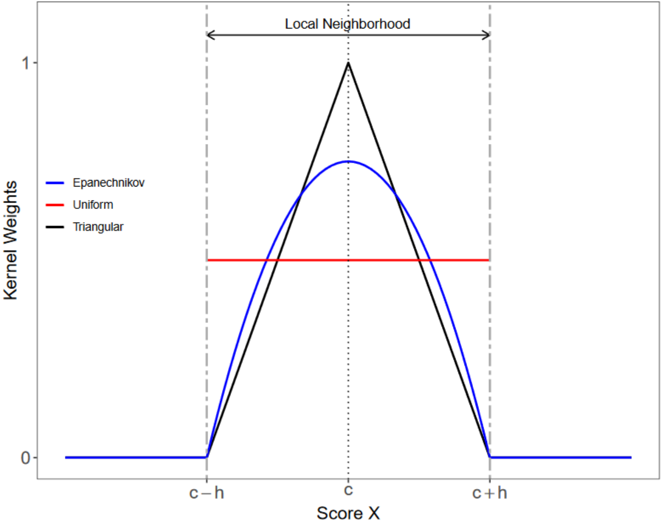
$-h \leq X_i < c:$	$c \leq X_i \leq h:$
$Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i}$	$Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i}$

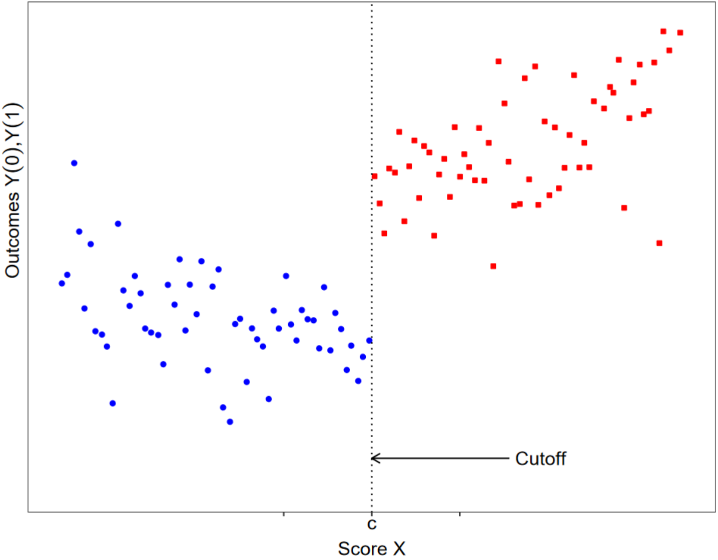
  - Treatment effect (at the cutoff): $\hat{\tau}_{\mathsf{SRD}}(h) = \hat{\alpha}_+ - \hat{\alpha}_-$

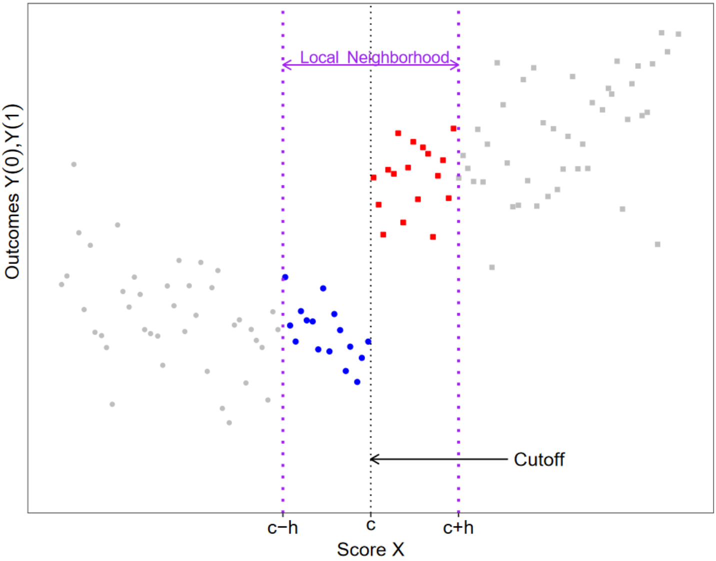
- Can be estimated using linear models (w/ weights $K(\cdot)$):

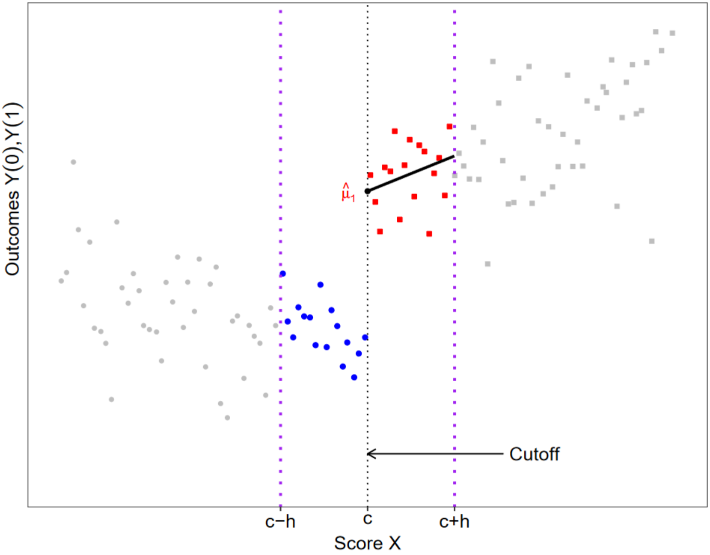
$$Y_i = \alpha + \tau_{\mathsf{SRD}} \cdot T_i + (X_i - c) \cdot \beta_1 + T_i \cdot (X_i - c) \cdot \gamma_1 + \varepsilon_i, \qquad |X_i - c| \leq h$$

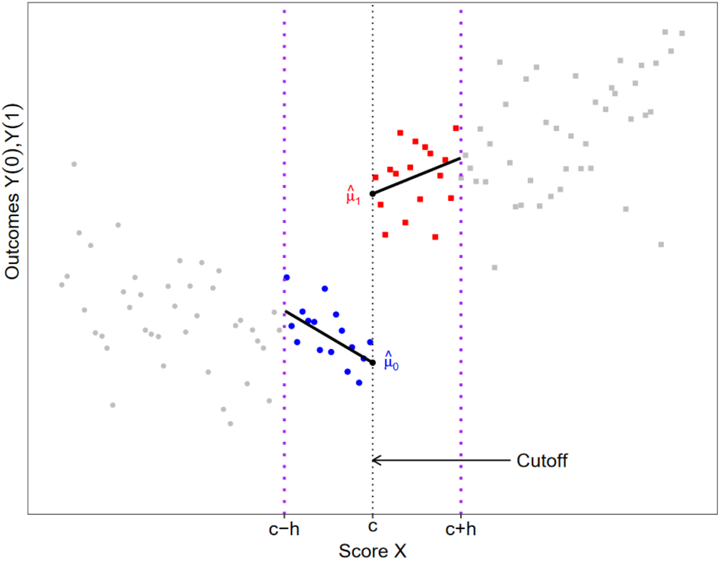
- Given p, K, h chosen $\implies$ weighted least squares estimation.

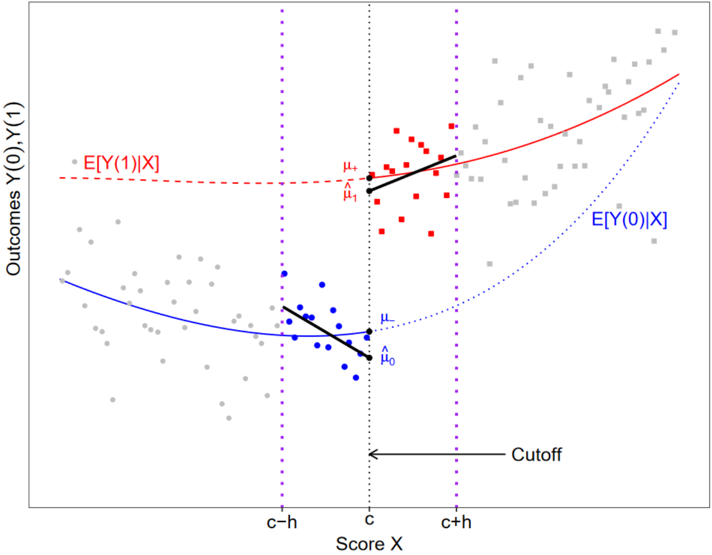

Local Neighborhood
1
0
Kernel Weights
Epanechnikov
Uniform
Triangular
c−h
c
c+h
Score X

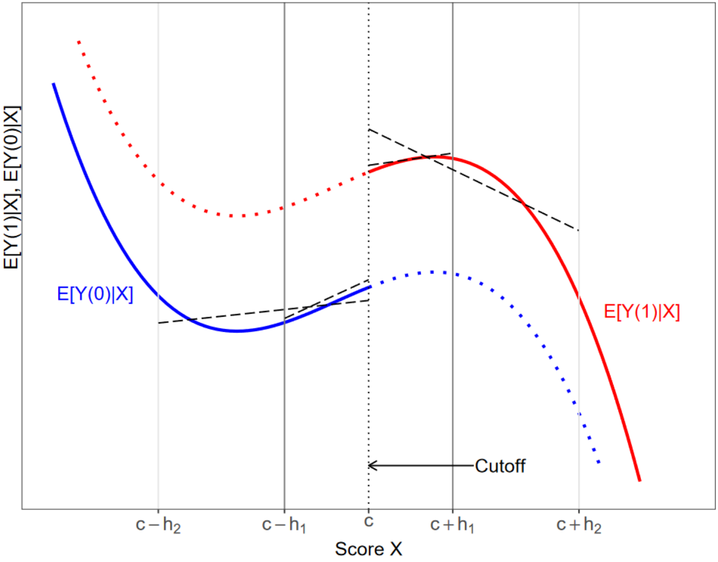

Outcomes Y(0),Y(1)
Cutoff
c
Score X


Local Neighborhood
Cutoff
Outcomes Y(0),Y(1)
c−h
c
c+h
Score X


Outcomes Y(0),Y(1)
$\hat{\mu}_1$
Cutoff
c−h
c
c+h
Score X


Outcomes Y(0),Y(1)
$\hat{\mu}_1$
$\hat{\mu}_0$
Cutoff
c−h
c
c+h
Score X


E[Y(1)|X]
μ_+
$\hat{\mu}_1$
E[Y(0)|X]
μ_-
$\hat{\mu}_0$
Outcomes Y(0),Y(1)
Cutoff
c−h
c
c+h
Score X


E[Y(1)|X], E[Y(0)|X]
E[Y(0)|X]
E[Y(1)|X]
Cutoff
c − h2
c − h1
c
c + h1
c + h2
Score X

Which window size h should we use?

- How do we pick the bandwidth size?
- Regression discontinuity design is a prediction problem in that we want to predict the value of the outcome at a specific point
- However there is a trade-off between bias and efficiency:
  - Closer to the cut-off there is less bias in the estimator, but the sample size is smaller (= larger variance)
  - However, farther from the cut-off there is a larger sample size but there is a larger bias in the estimator
- Recent proposals for automatic bandwith selection include Imbens and Kalyanaraman (2011) for MSE-optimal rule
- And robust confidence interval-corrected approach by Calonico, Cattaneo and Titiunik (2014) for the MSE-optimal bandwidth
- The later approach seems to work well in validation studies
- Report conventional estimates with robust confidence intervals and p-values

Which window size h should we use?

- How do we pick the bandwidth size?
- Regression discontinuity design is a prediction problem in that we want to predict the value of the outcome at a specific point
- However there is a trade-off between bias and efficiency:
  - Closer to the cut-off there is less bias in the estimator, but the sample size is smaller (= larger variance)
  - However, farther from the cut-off there is a larger sample size but there is a larger bias in the estimator
- Recent proposals for automatic bandwith selection include Imbens and Kalyanaraman (2011) for MSE-optimal rule
- And robust confidence interval-corrected approach by Calonico, Cattaneo and Titiunik (2014) for the MSE-optimal bandwidth
- The later approach seems to work well in validation studies
- Report conventional estimates with robust confidence intervals and p-values

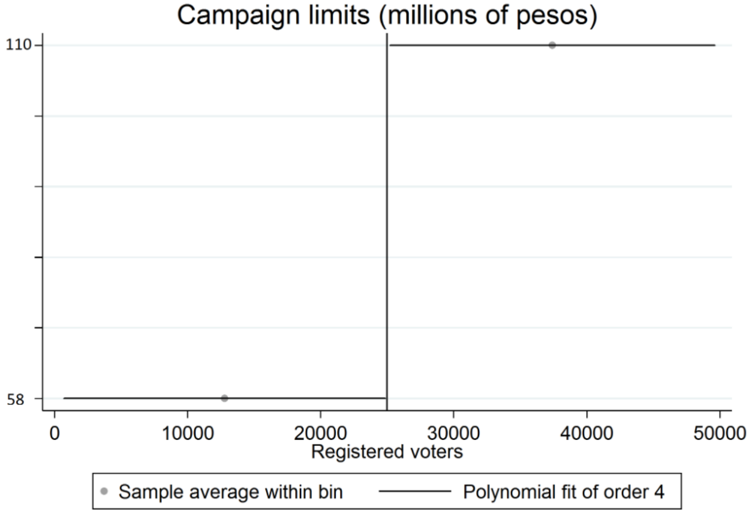
Example: Campaign Limits. Gulzar, Rueda and Ruiz (2020)

- Policymakers have enacted limits on private funding of campaigns to reduce the influence of donors
  - Over 40% of countries have some form of limit on how much donors are allowed to contribute to electoral campaigns IDEA (2014)
- But are they effective? What are the effects of campaign finance limits?
  - This paper: Effects of *campaign limits* on public *contracts*
  - They use a novel dataset from Colombia that links individual donations to public contracts and exploit institutional rules where campaign limits jump *discontinuously* at specific number of registered voters

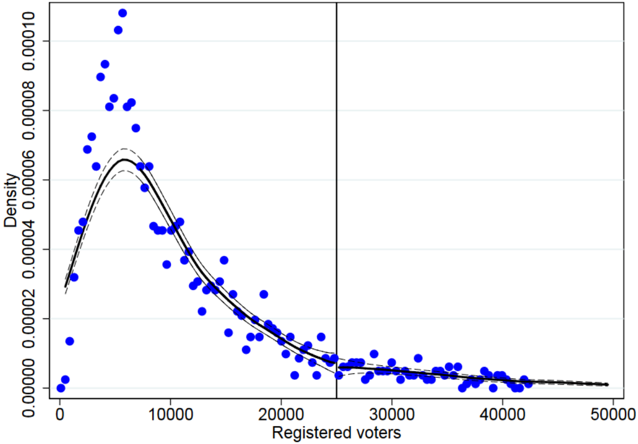
Context

- 75% of donors give contributions that are *larger* than the municipality average monthly wage
- Out of 6,658 donors in the dataset, 6,542 contribute only in one municipality
- There are campaign finance limits on total campaign contributions and on *individual contributions* (10% of total)
- Limits jump *discontinuously* at arbitrary registered voter cut-offs
- For example at 25,000 registered voters the campaign limit increases from 58M (17,000 USD) to 110M COP (36,400 USD)

Sharp Discontinuity



RDD-Distribution of registered voters



Notes: Figure shows the density of the running variable on both sides of the threshold, binned averages and 95% confidence intervals. The discontinuity

Empirical approach: RDD campaign limits

- Focus on <u>similar</u> municipalities where there are tighter **or** looser campaign limits by a narrow registered voter margin
- V_i = number of registered voters centered around 0 (25,000 cut-off)
- When $CampaignLimit_i = 1$ then ($V_i > 0$)

$$CampaignLimit_i = \begin{cases} CampaignLimit_i = 1 & \text{if } V_i > 0 \\ CampaignLimit_i = 0 & \text{if } V_i < 0 \end{cases} \tag{1}$$

- Note that the treatment implies **looser campaign limits**

Empirical approach: RDD campaign limits

- Focus on <u>similar</u> municipalities where there are tighter **or** looser campaign limits by a narrow registered voter margin

- V_i = number of registered voters centered around 0 (25,000 cut-off)
- When $CampaignLimit_i = 1$ then $(V_i > 0)$

$$CampaignLimit_i = \begin{cases} CampaignLimit_i = 1 & \text{if } V_i > 0 \\ CampaignLimit_i = 0 & \text{if } V_i < 0 \end{cases} \tag{1}$$

- Note that the treatment implies **looser campaign limits**

Empirical approach: RDD campaign limits

$$Y_i = \beta_1 CampaignLimit_i + \beta_2 f(V_i) + \beta_3 CampaignLimit_i \times f(V_i) + \varepsilon_i \tag{2}$$

- Treatment effect:

$$\beta_1 = lim_{v \downarrow c} E\left[Y_1 | V = c\right] - lim_{v \uparrow c} E\left[Y_0 | V = c\right] \tag{3}$$

- $f(V_i)$: local polynomial that may vary for $CampaignLimit_i = 0$ or $CampaignLimit_i = 1$
- Y_i are outcomes like the number contracts for donors and average donation in election i
- We estimate for a narrow bandwidth h Imbens and Kalyanaraman (2011) with robust p-values by Calonico, Cattaneo and Titiunik (2014)

Looser limits and rewards for donors

Table: Effect of looser campaign contribution limits on contracts assigned to donors to the mayor

Outcome:	# Contracts (1)	ln(Value All) (2)	# Min. Value Contracts (3)	ln(Value Min. Value) (4)
Looser contribution limit	3.091	0.819	2.030	0.638
Robust p-value	0.012	0.234	0.023	0.061
CI 95%	[0.735,5.994]	[-0.513,2.098]	[0.306,4.200]	[-0.030,1.364]
Observations	2,049	2,049	2,049	2,049
Bandwidth obs.	457	366	366	341
Mean	0.280	0.205	0.210	0.101
Effect mean(%)	1,103.93	399.51	966.67	631.68
Bandwidth	6,980	5,312	5,292	5,190

Local linear estimates of average treatment effects at cutoff estimated with triangular kernel weights and optimal MSE bandwidth. 95% robust confidence intervals and robust p-values with clustering at the municipality level are computed following Calonico, Cattaneo and Titiunik (2014). Bandwidth obs. denotes number of observations in the optimal MSE bandwidth for each dependent variable. Each observation is a donor.

Outline

Sharp RDD: Falsification Checks

1. Sensitivity: Are results sensitive to alternative specifications?
2. Continuity Checks: Do predetermined covariates X jump at the threshold?
   - Do placebo outcomes jump at the threshold?
3. Sorting: Do units sort around the threshold?
4. Check if jumps occur at placebo thresholds $c^{\star}$ (?)

Sharp RDD: Falsification Checks

1. Sensitivity: Are results sensitive to alternative specifications?
2. Continuity Checks: Do predetermined covariates X jump at the threshold?
    - Do placebo outcomes jump at the threshold?
3. Sorting: Do units sort around the threshold?
4. Check if jumps occur at placebo thresholds $c^{\star}$ (?)

Sharp RDD: Falsification Checks

1. Sensitivity: Are results sensitive to alternative specifications?

2. Continuity Checks: Do predetermined covariates X jump at the threshold?
   - Do placebo outcomes jump at the threshold?

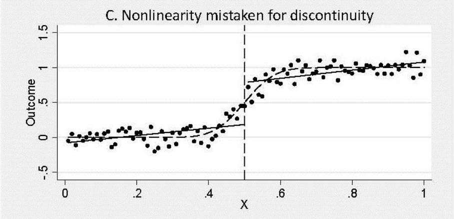
3. Sorting: Do units sort around the threshold?

4. Check if jumps occur at placebo thresholds $c^{\star}$ (?)

Sharp RDD: Falsification Checks

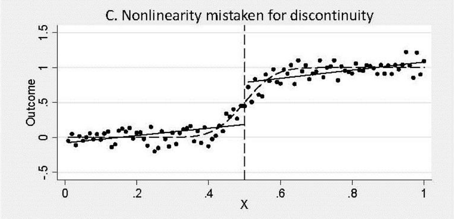
1. Sensitivity: Are results sensitive to alternative specifications?
2. Continuity Checks: Do predetermined covariates X jump at the threshold?
   - Do placebo outcomes jump at the threshold?
3. Sorting: Do units sort around the threshold?
4. Check if jumps occur at placebo thresholds $c^{\star}$ (?)

Sharp RDD: Sensitivity to Specification



- $Y = f(X) + \alpha D + \varepsilon$: A miss-specified control function $f(X)$ can lead to a spurious jump: Take care not to confuse a nonlinear relation with a discontinuity
- More flexibility reduces bias, decreases efficiency. But higher flexibility leads to disappearance of actual discontinuities
- Local linear and local polynomials are standard. Gelmans and Imbens (2015)

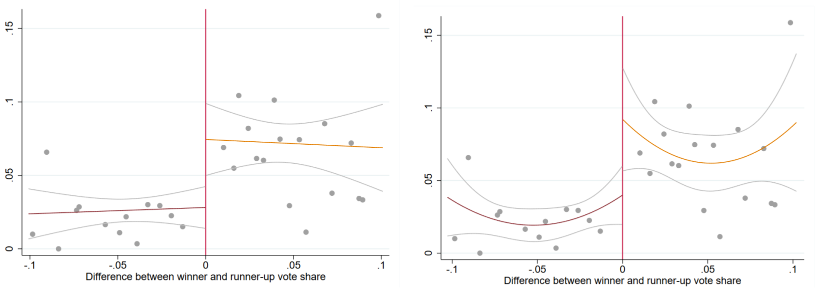
Sharp RDD: Sensitivity to Specification



- $Y = f(X) + \alpha D + \varepsilon$: A miss-specified control function $f(X)$ can lead to a spurious jump: Take care not to confuse a nonlinear relation with a discontinuity
- More flexibility reduces bias, decreases efficiency. But higher flexibility leads to disappearance of actual discontinuities
- Local linear and local polynomials are standard. Gelmans and Imbens (2015)

Example of graphical representation of result

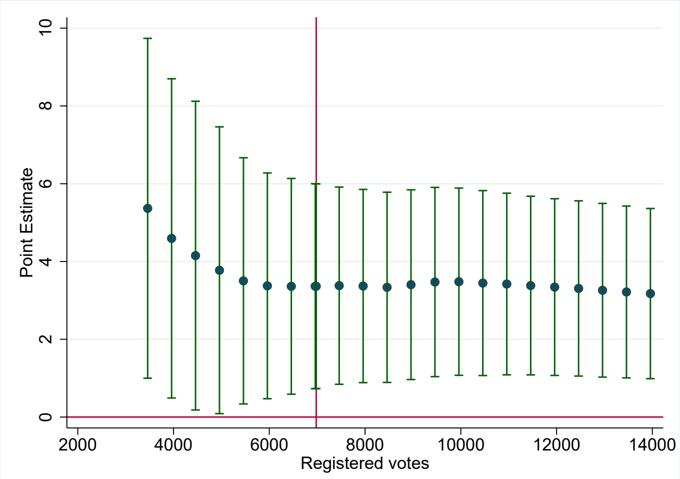
Effect of barely electing a politician on probability of donor getting a contract



Note: Observations within Calonico et. al (2014) bandwidth displayed. Left: linear fit. Right: quadratic fit.

Estimate for different bandwidth sizes

Effect of loser campaign limits on number of contracts for donors



SRDD: Continuity Checks Test

- Test for comparability of agents around the cut-off:
  - What covariates (Z) could vary at the threshold and correlated to the outcome?
  - Run the RDD regression using Z as the outcome
- Finding a discontinuity in Z does not necessarily invalidate the RDD
  - Can be a multiple hypothesis testing problem
  - Can incorporate Z as additional controls into our main RDD regression. Ideally, this should only impact statistical significance, not the treatment effect
- Smoothness checks address only observables, not unobservables
  - On expectations unobservables should be similar across the cut-off, if many observables do not jump discontinuously

SRDD: Continuity Checks Test

- Test for comparability of agents around the cut-off:
  - What covariates (Z) could vary at the threshold and correlated to the outcome?
  - Run the RDD regression using Z as the outcome
- Finding a discontinuity in Z does not necessarily invalidate the RDD
  - Can be a multiple hypothesis testing problem
  - Can incorporate Z as additional controls into our main RDD regression. Ideally, this should only impact statistical significance, not the treatment effect
- Smoothness checks address only observables, not unobservables
  - On expectations unobservables should be similar across the cut-off, if many observables do not jump discontinuously

Testing of discontinuities - Smooth covariates

Table: Municipality characteristics around campaign contribution limits cutoff

	Looser limits (1)	CI 95% (2)	Obs. (3)	Band. Obs. (4)	Bandwidth (5)	p-value (6)
Discretionary revenue	592.716	[-9000,8867.212]	970	76	4518.17	0.986
Municipal category	0.111	[-0.196,0.479]	999	61	3528.11	0.412
Mayor wages	-0.222	[-0.955,0.396]	999	61	3524.19	0.417
Council size	-0.354	[-1.134,0.286]	999	62	3563.62	0.241
Population	-448.213	[-4.5000,4011.520]	999	171	8786.33	0.907
Schools	60.152	[-34.789,176.287]	999	103	5767.05	0.189
Contracts	-87.087	[-880.423,686.785]	992	106	5989.31	0.809

Column 1 reports local linear estimates of average treatment effects at cutoff estimated with triangular kernel weights and optimal MSE bandwidth (reported in column 7). Columns 2 and 6 report 95% robust confidence intervals and robust p-values computed following Calonico, Cattaneo and Titiunik (2014). Columns 4 and 5 report total observations and observations in optimal MSE bandwidth. Discretionary income scaled in # of minimum monthly wages. Schools denotes all educational establishments.

SRDD: Sorting Around the Threshold

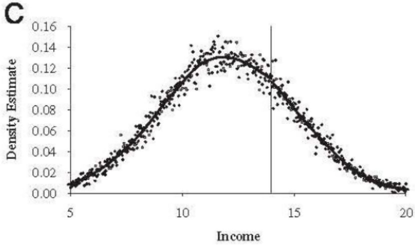
- Can subjects behavior invalidate the local continuity assumption?
  - Can they exercise control over their values of the assignment variable?
  - Can administrators strategically choose what assignment variable to use or which cut-off point to pick?
  - Either can invalidate the comparability of subjects near the threshold because of sorting of agents around the cut-off, where those below may differ form those just above
  - Does not necessarily invalidate the design unless sorting is very widespread and very precise
  - We look at the continuity of the distribution of the running variable at the cut-off
- What else changes at c? Continuity violated in the presence of other programs that use a discontinuous assignment rule with the exact same assignment variable and cut-off
  - This can lead to a compound treatment

SRDD: Sorting Around the Threshold

- Can subjects behavior invalidate the local continuity assumption?
  - Can they exercise control over their values of the assignment variable?
  - Can administrators strategically choose what assignment variable to use or which cut-off point to pick?
  - Either can invalidate the comparability of subjects near the threshold because of sorting of agents around the cut-off, where those below may differ form those just above
  - Does not necessarily invalidate the design unless sorting is very widespread and very precise
  - We look at the continuity of the distribution of the running variable at the cut-off
- What else changes at c? Continuity violated in the presence of other programs that use a discontinuous assignment rule with the exact same assignment variable and cut-off
  - This can lead to a compound treatment

SRDD: Sorting Around the Threshold

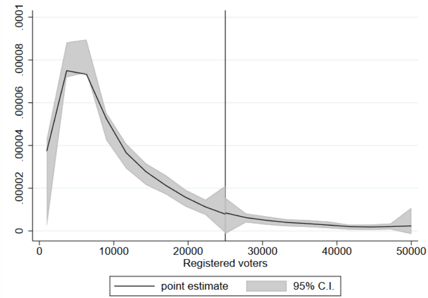
Example: Beneficial job training program offered to agents with income $< c$. Concern, people will withhold labor to lower their income below the cut-off to gain access to the program.

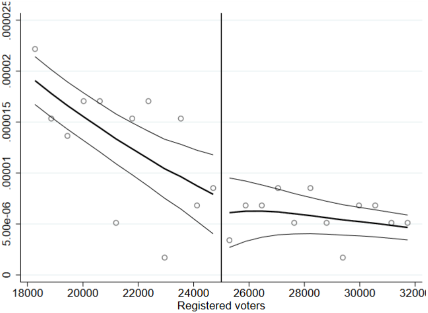


SRDD: Sorting Around the Threshold

- Are individuals able to influence the forcing variable - if so, how?
- Test for discontinuity in density of forcing variable:
  - Visual Histogram Inspection:
    - Construct equal-sized non-overlapping bins of the forcing variable such that no bin includes points to both the left and right of the cut-off
    - For each bin, compute the number of observations and plot the bins to see if there is a discontinuity at the cut-off
  - Formal tests (e.g. McCrary, 2008)
  - Cattaneo, Jansson and Ma (2018). [Not dependent on the binning]

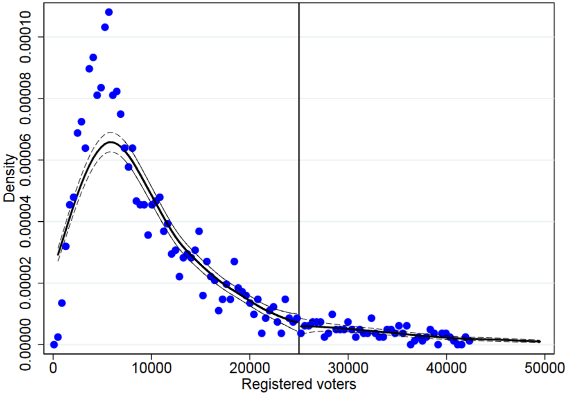
Sorting Tests





The left figure shows the density of the running variable. The test of no discontinuity at the cutoff (Cattaneo,Janson and Ma 2019) gives a statistic of -0.128 and a p-value of 0.98). The right figure presents the density graph in a narrower band around the cutoff. Dots represent averages of multiple observations.

Sorting Checks: The case of campaign finance limits



Notes: Figure shows the density of the running variable on both sides of the threshold, binned averages and 95% confidence intervals. The discontinuity estimate (log difference in height) is −0.188 with standard error of 0.321

SRDD: Placebo Threshold

- Test whether the treatment effect is zero when it should be (at policy cut-off c), instead of other $c^\star$
- Let $c^\star$ be a placebo thershold value. Run the regression of: $E[Y|X,D] = \beta_0 + \beta_1(X - c^\star) + \alpha D + \beta_3((X - c^\star)\cdot D)$ and check if α large and significant?
    - We split the sample to the left and the right of the actual threshold c in order to avoid miss-specification by imposing a zero jump at c
- The existence of large placebo jumps does not invalidate the RDD, but does require an explanation
- Concern is that the relation is fundamentally discontinuous and jump at cut-off is contaminated by other factors

RD check list

1. Understand the context: Is there an arbitrary cut-off?
2. Why was the cut-off designed? Is there compliance?
3. Can you obtain the data?
4. Are there enough observations close to the cut-off?
5. Are there possibilities for manipulation?
   - Check the manipulation tests
6. Plot the regression discontinuity - patterns should striking
7. Run the first regression discontinuity, are the results robust?
   - To different bandwidth sizes around the optimal bandwidth?
   - To different local polynomials? Usually order 1 and 2
   - Are other covariates jumping discontinously at the cut-off? Are they potentially related to treatment and outcome?

RD Packages

`https://rdpackages.github.io/`

- **rdrobust**: estimation, inference and graphical presentation using local polynomials, partitioning, and spacings estimators.
  - `rdrobust`, `rdbwselect`, `rdplot`.
- **rddensity**: discontinuity in density tests (manipulation testing) using both local polynomials and binomial tests.
  - `rddensity`, `rdbwdensity`.