PS207 Quantitative Causal Inference, Fall 2016

Regression Discontinuity Designs

Matto Mildenberger

UC Santa Barbara

`Special thanks to Chad Hazlett (UCLA) for select slides, used with permission`

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:
  - there is a rule influencing how people are assigned
  - this will give us boundaries where a trivial change in expected potential outcome corresponds to a big change in probability of treatment
- Widely applicable in a rule based world (administrative programs, elections, etc.)
- High internal validity (see e.g. Cook, Shadish, Wong 2008)

Sharp Regression Discontinuity Design

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

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

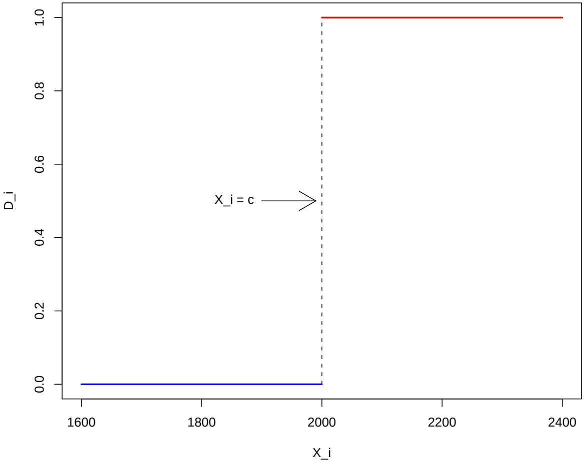
- X_i, called the forcing variable, may be correlated with potential outcomes, so comparing treated and untreated units does not provide causal estimates

- Situation arises often from administrative decisions, where policy decisions are made using sharp rules rather than administrator discretion

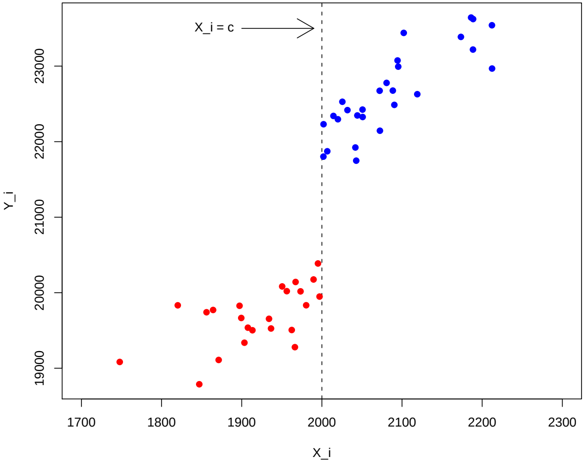
Hypothetical Example

- Thistlethwaite and Campbell (1960) study the effects of college scholarships on students' later achievements
- Scholarships are given on the basis of whether or not a student's test score exceeds some threshold c
  - Treatment D_i is scholarship
  - *Forcing* or *running* variable X_i is SAT score with cutoff c
  - Outcome Y_i is subsequent earnings
  - Y_{0i} denotes potential earnings without the scholarship
  - Y_{1i} denotes potential earnings with the scholarship
- Y_{1i} and Y_{0i} are correlated with X_i: on average, students with higher SAT scores obtain higher earnings

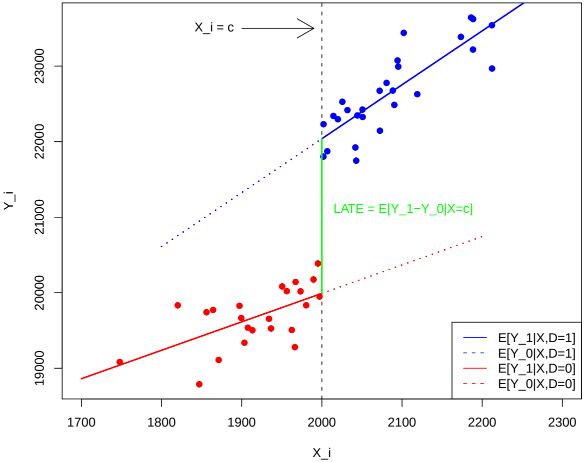
Sharp RDD: Graphical Illustration



Sharp RDD: Graphical Illustration



Sharp RDD: Graphical Illustration



Sharp RDD: Identification

Identification Assumption

$\mathbb{E}[Y_{1i}|X_i]$ and $\mathbb{E}[Y_{0i}|X_i]$ are continuous in X_i around the threshold $X_i = c$

Identification Result

The treatment effect is identified at the threshold as:

$$\begin{aligned} \alpha_{SRDD} &= \mathbb{E}[Y_{1i} - Y_{0i}|X_i = c] \\ &= \mathbb{E}[Y_{1i}|X_i = c] - \mathbb{E}[Y_{i0}|X_i = c] \\ &= \lim_{x \downarrow c} \mathbb{E}[Y_{1i}|X_i = c] - \lim_{x \uparrow c} \mathbb{E}[Y_{i0}|X_i = c] \end{aligned}$$

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

Comments on Identification

The continuity assumption says a lot:

- By saying $\mathbb{E}[Y_{1i}|X_i]$ and $\mathbb{E}[Y_{0i}|X_i]$ are continuous in X near c, we mean they are "smooth", or do not "jump" at $X = c$.
- This is why we can say:

$$\mathbb{E}[Y_{1i}|X_i = c] = \lim_{\epsilon \downarrow 0} \mathbb{E}[Y_{i1}|X_i = c + \epsilon, D_i = 1]$$
$$\mathbb{E}[Y_{0i}|X_i = c] = \lim_{\epsilon \uparrow 0} \mathbb{E}[Y_{io}|X_i = c - \epsilon, D_i = 0]$$

  where we estimate (rather than observe) the RHS of each identity above

This gives us ability to estimate:

$$\hat{\alpha} = \hat{\mathbb{E}}[Y_{1i}|X_i = c] - \hat{\mathbb{E}}[Y_{0i}|X_i = c]$$

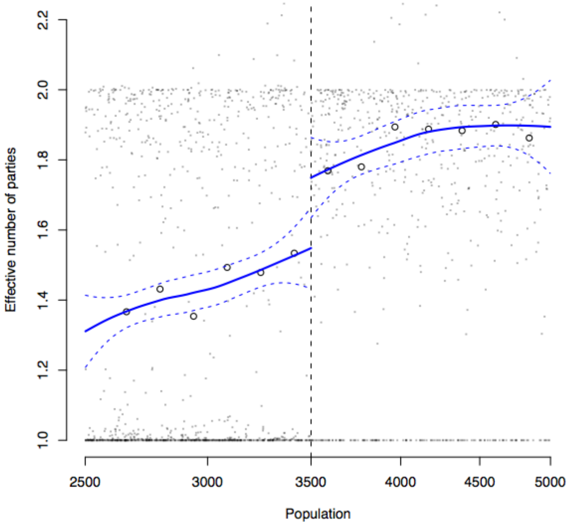
- Implies any *unobserved influences on the potential outcomes* are also continuous in X (at $X = c$)
- You can also see it as our good old:

$$\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp D_i | X_i = c$$

Duverger's Law (Eggers, 2010)

- Duverger (1972): "a majority vote on one ballot is conducive to a two-party system; proportional representation is conducive to a multiparty system"
- Therefore, we expect the number of parties to increase when going from a majority to a proportional electoral system
- In French municipalities, the electoral rule used to elect the municipal council depends on the city's population:
  - cities with fewer than 3,500 people use first-past-post
  - cities with a population of 3,500 or more use a form of PR rule

Sharp RDD: Duverger's Law



Sharp RDD: Electronic Voting (Hidalgo, 2012)

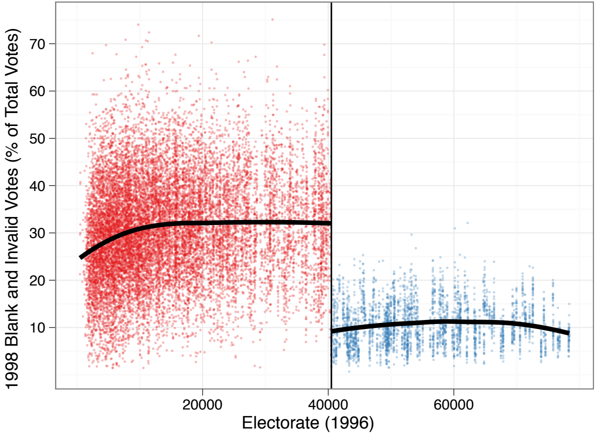
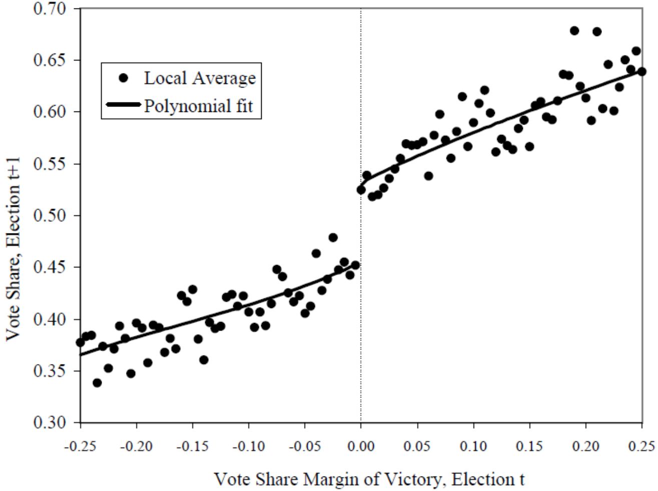


Figure 6: The effect of electronic voting on the percent of null and blank votes. Each dot is a polling station. Polling stations to the left of the vertical black line used paper ballots and polling stations to the right used electronic voting. The black horizontal line is the conditional mean of the outcome estimated with a loess regression.

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



Other Recent Examples

- Effect of class size on student achievement (Angrist and Lavy 1999)
- Effect of access to credit on development outcomes (loan offer is determined by credit score threshold)
- Effect of wage increase on performance of mayors (Ferraz and Finan 2011, Gagliarducci and Nannicini 2013)
- Effect of wining office on wealth of politicians (Eggers and Hainmueller 2009)
- Effect of school district boundaries on home values (Black 1999)
- RD that exploits “close" elections has become a workhorse model for causal inference in electoral research
  - ...Lee, Moretti and Butler 2004, DiNardo and Lee 2004, Hainmueller and Kern 2008, Leigh 2008, Pettersson-Lidbom 2008, Broockman 2009, Butler 2009, Dal Bó, Dal Bó and Snyder 2009, Eggers and Hainmueller 2009, Ferreira and Gyourko 2009, Uppal 2009, 2010, Cellini, Ferreira and Rothstein 2010, Gerber and Hopkins 2011, Trounstine 2011, Boas and Hidalgo 2011, Folke and Snyder Jr. 2012, and Gagliarducci and Paserman 2012...
- many more...

Estimating $\alpha_{SRDD} = \mathbb{E}[Y_{1i}|X_i = c] - \mathbb{E}[Y_{0i}|X_i = c]$

1. Trim sample to small window around the threshold c (discontinuity sample):
   - For window width $h > 0$, use $c - h \leq X_i \leq c + h$,
   - h may be determined by cross-validation
2. Recode running variable to deviations from threshold: $\tilde{X} = X - c$.
3. Decide on a model for $\mathbb{E}[Y|X]$ on each side
   - linear, same slope for $\mathbb{E}[Y_0|X]$ and $\mathbb{E}[Y_1|X]$
   - linear, but different slopes
   - non-linear
   - good to start with visual inspection (scatter plot with kernel/lowess) to check which model is appropriate

SRRD Estimation 1: Linear with Same Slope

- $\mathbb{E}[Y_{0i}|X]$ is linear and treatment effect, α, does not depend on X_i:

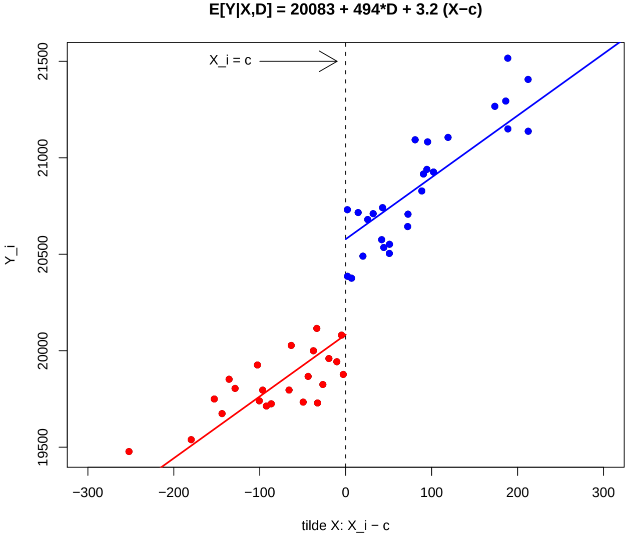
$$\mathbb{E}[Y_{0i}|X_i] = \mu + \beta X_i\,, \qquad \mathbb{E}[Y_{1i} - Y_{0i}|X_i] = \alpha$$

- Therefore $\mathbb{E}[Y_{1i}|X_i] = \alpha + \mathbb{E}[Y_{0i}|X_i] = \alpha + \mu + \beta X_i$
- Since D_i is determined given X_i, we have that:

$$\begin{aligned}
\mathbb{E}[Y_i|X_i, D_i] &= D \cdot \mathbb{E}[Y_{1i}|X_i] + (1 - D_i) \cdot \mathbb{E}[Y_{0i}|X_i] \\
&= \mu + \alpha D_i + \beta X_i \\
&= (\mu + \beta c) + \alpha D_i + \beta (X_i - c) \\
&= \gamma + \alpha D_i + \beta \tilde{X}_i
\end{aligned}$$

- So we just run a regression of Y on D and $\tilde{X}$

SRDD: Linear with Same Slope



Sharp RDD Estimation 2: Differential Slopes

- Or, let $\mathbb{E}[Y_0|X]$ and $E[Y_1|X]$ be different linear functions of X, so the average effect of the treatment $E[Y_1 - Y_0|X]$ varies with X:

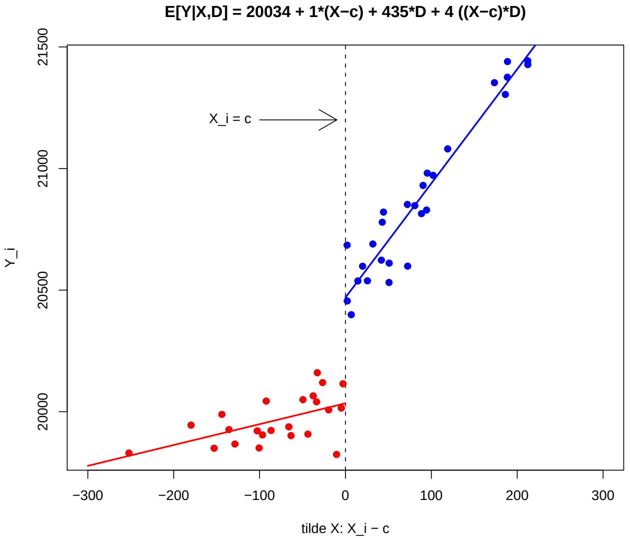
$$\mathbb{E}[Y_{0i}|X_i] = \mu_0 + \beta_0 X_i, \qquad E[Y_{1i}|X_i] = \mu_1 + \beta_1 X_i$$

- So $\alpha(X) = \mathbb{E}[Y_{1i} - Y_{0i}|X_i] = (\mu_1 - \mu_0) + (\beta_1 - \beta_0)X_i$ we have

$$\begin{aligned}\mathbb{E}[Y_i|X_i, D_i] =& D_i \cdot \mathbb{E}[Y_{1i}|X_i] + (1 - D_i) \cdot \mathbb{E}[Y_{0i}|X_i] \\ &= \mu_1 D_i + \beta_1 (X_i \cdot D_i) + \mu_0 (1 - D_i) + \beta_0 X_i \cdot (1 - D_i) \\ &= \gamma + \beta_0 \tilde{X}_i + \alpha D_i + \beta_1 \tilde{X}_i \cdot D_i\end{aligned}$$

- Regress Y_i on $\tilde{X}_i$, D_i, and the interaction $\tilde{X}_i \cdot D_i$.
- What does each coefficient in the above represent?

SRDD: Linear with Differential Slope



Sharp RDD Estimation 3: Non-Linear Case

Suppose $\mathbb{E}[Y_0|X]$ and $\mathbb{E}[Y_1|X]$ are separate *non-linear* functions of X and the average effect of the treatment $\mathbb{E}[Y_1 - Y_0|X]$ varies with X

Can use any non-linear model:

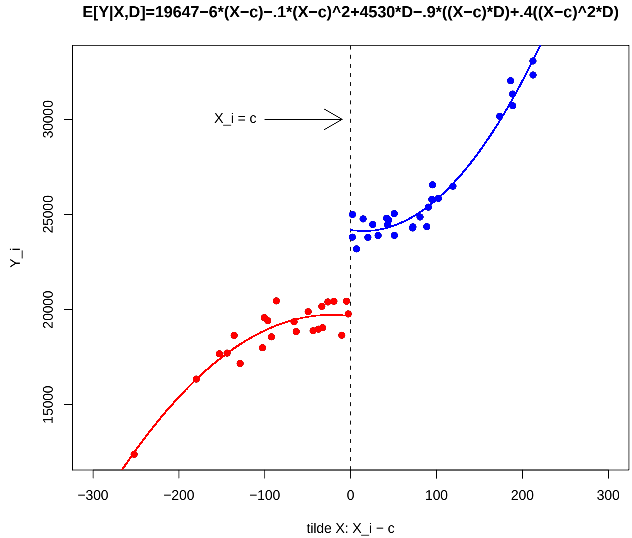
- At minimum, include higher-order terms in $(X - c)$ and their interactions with D, e.g.

$$\begin{aligned}\mathbb{E}[Y_i|X_i, D_i] =& \gamma_0 + \gamma_1 \tilde{X}_i + \gamma_2 \tilde{X}_i^2 \\ &+ \alpha_0 D_i + \alpha_1 (\tilde{X}_i \cdot D_i) + \alpha_2 (\tilde{X}_i^2 \cdot D_i)\end{aligned}$$

- In such cases $\alpha_0 = \mathbb{E}[Y_1 - Y_0|X = c]$ estimates effect at $X = c$
- Non-parametric smoothers and locally-weighted regressions (eg. LOESS) also good: just model on each side then take difference at $X = c$

These models differ only in how we borrow information from neighboring areas of X to guess $\mathbb{E}[Y|X = c]$ from each side.

SRDD: Non-Linear Case



Comments on Estimation

No ex-ante correct model for $\mathbb{E}[Y|X]$ on each side.

- bias variance tradeoff: more parameters (or smaller bandwidth) will give more flexible model but depends heavily on small changes in data
- standard practice: show results are not sensitive to changes in model or in the window around $X = c$.
- especially important to show you can cut window very narrowly and get similar result (though variance will explode)

Automated selection procedures:

- Imbens-Kalyanaraman (2012): local linear regression with a choice of bandwidth designed to minimize an approximation of

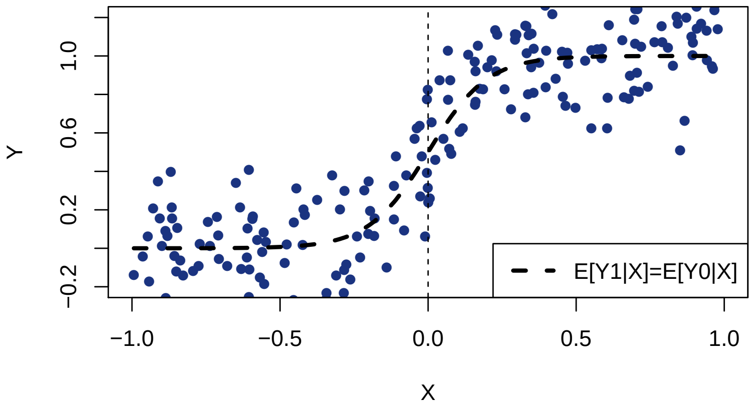
  $$MSE(h) = \mathbb{E}[(\hat{\alpha}_{SRDD} - \alpha_{SRDD})^2]$$

  See `IKbandwidth` in `RDD` package
- Also see Calonico, Cattaneo, and Titiunik for automatic bandwidth selection for local polynomial models, and automatic algorithms for a range of plots (see `rdrobust`)

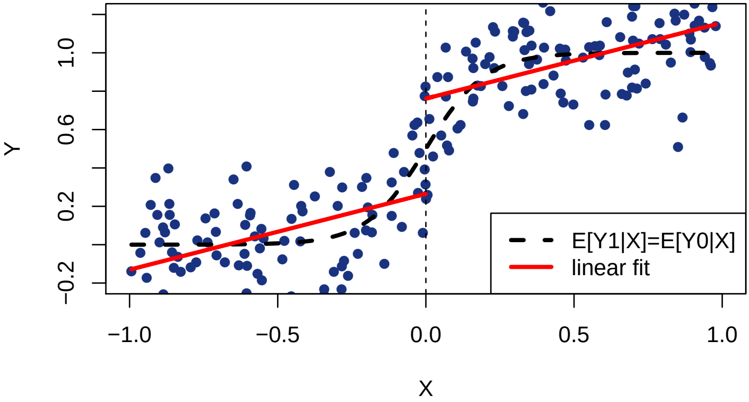
Sharp RDD: Falsification Checks

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

Sharp RDD: Risk of Mispecification

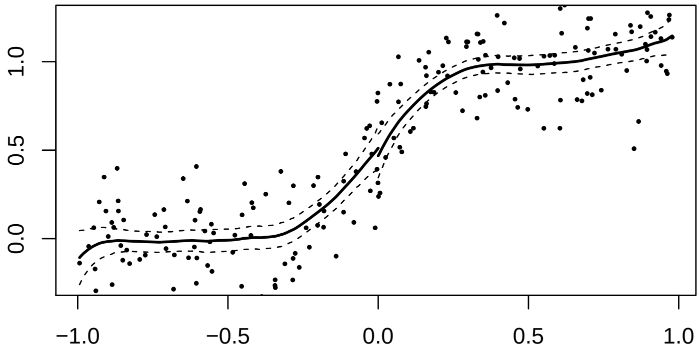


Sharp RDD: Risk of Mispecification



Sharp RDD: Risk of Mispecification

Using Imbens-Kalyanaraman bandwidth in `RDD`



SRDD: Balance Checks Test

Test for comparability of agents around the cut-off:

- For potential outcomes to be continuous, any influencer (observed and unobserved) must also be continuous
- Visual tests: for observed Z_i, plot $\mathbb{E}[Z_i|X_i]$ and look for jumps
- Run the RDD regression using Z_i as the outcome:

$$\mathbb{E}[Z_i|X_i, D_i] = \beta_0 + \beta_1(X_i - c) + \alpha_z D_i + \beta_3(X_i - c) \cdot D_i$$

  should yield $\alpha_z = 0$ if Z is balanced at the threshold.

Finding a discontinuity in Z poses a challenge to the identification strategy, but does not necessarily invalidate it:

- if it is just Z_i that is imbalanced, add it as a control in the main RDD regression; see if if changes point-estimate or just SE
- but an imbalanced observable often weakens claim that there are no imbalanced unobservables at the threshold
- if you see imbalance at threshold on a post-treatment variable, is that a problem?

As usual, balance checks address only observables, not unobservables.

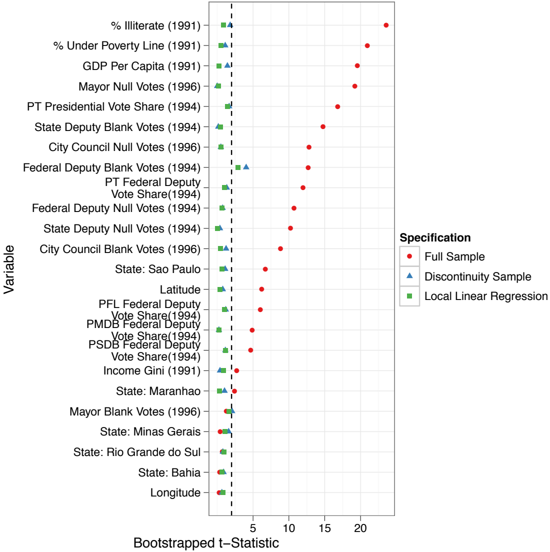
SRDD: You Can Add Covariates

TABLE 4. Regression Discontinuity Design Results: Effect of Serving in House of Commons on (Log) Wealth at Death

	Conservative Party		Labour Party	
Effect of serving	0.61	0.66	−0.20	−0.25
Standard error	(0.27)	(0.37)	(0.26)	(.26)
Covariates		x		x
Percent wealth increase	83	94	−18	−23
95% Lower bound	8	−7	−52	−65
95% Upper bound	212	306	31	71

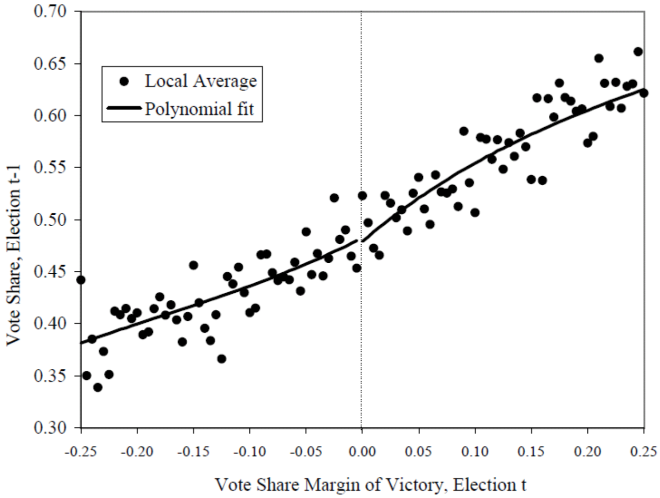
Note: Effect estimates at the threshold of winning $\tau_{RDD} = E[Y(1) - Y(0) \mid Z = 0]$. Estimates without covariates from local polynomial regression fit to both sides of the threshold with bootstrapped standard errors. Estimates with covariates from local linear regression with rectangular kernel (equation 2); bandwidth is 15 percentage point of vote share margin with

SRDD: Falsification Test



SRDD: Falsification Test

Especially convincing to do this with placebo outcome, i.e. covariates likely related to potential outcomes but not influence by treatment.



SRDD: Placebo Thresholds

Let $c^\star$ be a placebo threshold value. Run models

$$\mathbb{E}[Y_i|X_i, D_i] = \beta_0 + \beta_1(X_i - c^\star) + \alpha_{fake}D_i + \beta_3(X_i - c^\star) \cdot D_i$$

and check if α_{fake} is large and significant.

- Usually we run separate models below and above c, with bandwidth to exclude c, to avoid misspecification due to inclusion of the actual jump.
- Detecting a jump at placebo threshold requires an explanation.
  - e.g. if other treatments occurred near $c^\star$, it could be okay
  - but problematic if it implies unknown treatments or unobserved influences may account for effect at c
  - can also reveal specification problems, as your model could be guessing a discontinuous jump occurs where it was actually smooth

SRDD: Sorting and Bundling as Discontinuities

Can subjects self-sort around threshold?

- Can they exercise fine control over the running variable?
- Can administrators strategically choose what is used as running variable or as threshold?
- Either can allow sorting of agents around the cut-off, s.t. those below may differ on average form those just above
- Does not necessarily invalidate the design unless sorting is (a) precise, and (b) widespread
  - (a) imprecise sorting is okay as long as you can find narrow enough bandwidth around $X = c$ s.t. those to left and right of c are no different
  - (b) if some people can sort very precisely, RDD may be biased but less biased then naive comparison

Bundled Treatments. What else changes at c? Bundled treatments can invalidate continuity assumption.

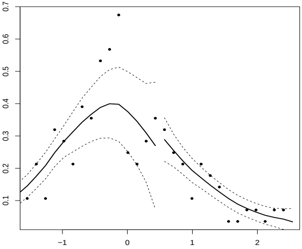
SRDD: Sorting Around the Threshold

Test for discontinuity in density of forcing variable:

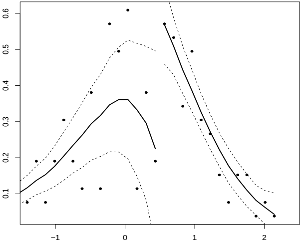
- Visual Histogram Inspection:
  - Construct equal-sized non-overlapping bins of the forcing variable s.t. no bin includes points both above and below c
  - For each bin, compute the number of observations and plot the bins to see if there is a discontinuity at the cut-off
- Formal test: McCrary (2008)

Self-sorting simulation

```
>N=200
>runvar=rnorm(N, mean = 0, sd=1)
>c=.5
>DCdensity(runvar, cutpoint=c)
[1] 0.5857803
```




```
> sortprob=rep(0,N)
> sortprob[runvar<c]=.25*exp(runvar[runvar<c]-c)
> selfsorts=rbinom(N, size=1, prob = sortprob)
> runvar[selfsorts==1]=.5+
runif(sum(selfsorts),.001,.5)
> DCdensity(runvar, cutpoint=c)
[1] 0.0001952333
```



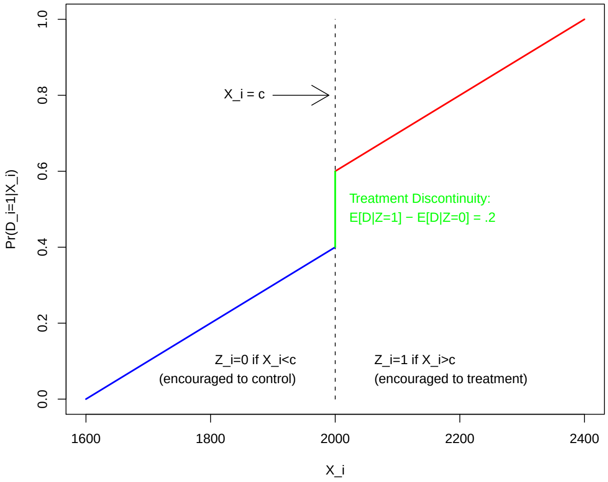
Fuzzy Regression Discontinuity Design

- Threshold may not perfectly determine treatment exposure, but it creates a discontinuity in the probability of treatment exposure
- Incentives to participate in a program may change discontinuously at a threshold, but the incentives are not powerful enough to move all units from non-participation to participation
- Should remind you of IV! We can use discontinuities to produce instrumental variable estimators of the effect of the treatment (close to the discontinuity)

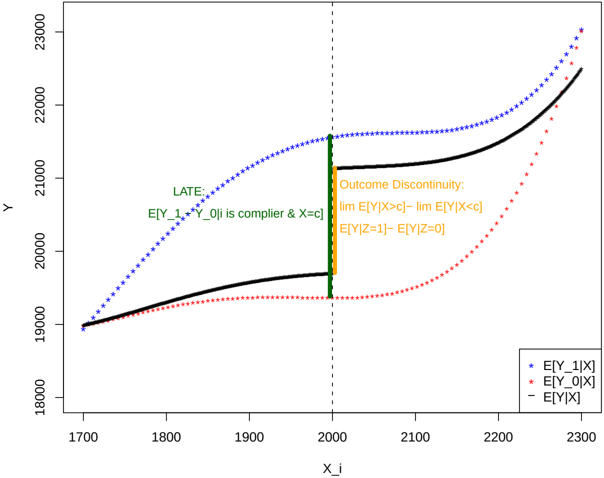
Fuzzy RDD Example:

- Probability of being offered a scholarship may jump at a certain SAT score threshold (when applicants are given “special consideration”)
- For applicants with scores very close to the threshold:
  - scholarship eligibility should be random (so $Z_i \perp\!\!\!\perp D_{1i}, D_{0i}$)
  - we may think tiny change in score, being nearly random, only effects outcome through scholarship chances ($Z_i \perp\!\!\!\perp Y_{0i}, Y_{1i}$)
  - If both, we have a good instrument.
- Who are compliers in this framework?

Fuzzy RDD: Discontinuity in $\mathbb{E}[D|X]$



Fuzzy RDD: Discontinuity in $\mathbb{E}[Y|X]$



Fuzzy RDD: Identification

Identification Assumption

For sufficiently small ϵ s.t. $X_i - \epsilon < c < X_i + \epsilon$, we have usual IV assumption:

1. *Ignorability (and Exclusion):* $\{Y_{1i}, Y_{0i}, D_{1i}, D_{0i}\} \perp\!\!\!\perp Z_i$
2. *Relevance:* $\mathbb{E}[D_{1i}] \neq \mathbb{E}[D_{0i}]$
3. *Monotoniticy:* $D_{1i} \geq D_{0i}\ \forall i$

Identification Result

$$
\begin{aligned}
\alpha_{FRDD} &= \mathbb{E}[Y_{1i} - Y_{0i} | X_i = c \text{ and } i \text{ is a complier}] \\
&= \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = c] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = c]}{\lim_{x \downarrow c} \mathbb{E}[D_i | X_i = c] - \lim_{x \uparrow c} \mathbb{E}[D_i | X_i = c]} \\
&= \frac{\text{outcome discontinuity}}{\text{treatment discontinuity}} \\
&\approx \frac{\hat{\mathbb{E}}[Y | Z = 1, X = c] - \hat{\mathbb{E}}[Y | Z = 0, X = c]}{\hat{\mathbb{E}}[D | Z = 1, X = c] - \hat{\mathbb{E}}[D | Z = 0, X = c]}
\end{aligned}
$$

Fuzzy RDD: Estimation

- Cut the sample to a small window above and below the threshold (discontinuity sample)
- Code instrument $Z_i = \mathbb{1}_{\{X>c\}}$
- Fit 2SLS/IV: $Y_i = \beta_0 + \beta_1(X_i - c) + \beta_2 Z_i(X_i - c) + \alpha D_i$ where D_i is instrumented with Z_i
- Better: use polynomials
- Also good: separately estimate the outcome discontinuity and treatment discontinuity, then divide. Can use any estimator this way.

Early Release Program (HDC)

- Prison system in many countries is faced with overcrowding and high recidivism rates after release.
- Early discharge of prisoners with electronic monitoring has become a popular policy
- Difficult to estimate impact of early release program on future criminal behaviour: best behaved inmates are usually the ones to be released early.
- Marie (2008) considers Home Detention Curfew (HDC) scheme in England and Wales:
    - Fuzzy RDD: Only offenders sentenced to more than 88 days in prison are eligible for HDC, but obviously, not all those with longer sentences are offered HDC
- How valid are the identification assumptions?

Table 1: Descriptive Statistics for Prisoners Released by Length of Sentence and HDC and Non HDC Discharges

Panel A - Released Before 3 Months:			
Discharge Type	**Non HDC**	**HDC**	**Total**
Percentage Female	12.2	-	12.2
Mean Age	29.5	-	29.5
Percentage Incarcerated for Violence	17.6	-	17.6
Mean Number Previous Offences	8.8	-	8.8
Recidivism within 12 Months	52.4	-	52.4
Sample Size	42,987	0	42,987
Panel B - Released Between 3 and 6 Months:			
Discharge Type	**Non HDC**	**HDC**	**Total**
Percentage Female	8.8	8.8	8.8
Mean Age at Release	27.6	30.8	28.4
Percentage Incarcerated for Violence	20.3	18.3	19.8
Mean Number Previous Offences	10	6.5	9.1
Recidivism within 12 Months	60	30.2	52.6
Sample Size	52,091	17,222	69,313

Table 2: Descriptive Statistics for Prisoners Released by Length of Sentence and HDC and Non HDC Discharges and +/-7 Days Around Discontinuity Threshold

Panel A - Released +/- 7 Days of 3 Months (88 Days) Cut-off:			
Discharge Type	**Non HDC**	**HDC**	**Total**
Percentage Female	10.5	9.7	10.3
Mean Age at Release	28.9	30.7	29.3
Percentage Incarcerated for Violence	19.8	18.2	19.4
Mean Number Previous Offences	9.5	5.7	8.7
Recidivism within 12 Months	54.6	28.1	48.8
Sample Size	18,928	5,351	24,279
Panel B - Released +/- 7 Days of 3 Months (88 Days) Cu-off:			
Day of Release around Cut-off	**- 7 Days**	**+ 7 Days**	**Total**
Percentage Female	11	10.2	10.3
Mean Age at Release	28.8	29.4	29.3
Percentage Incarcerated for Violence	17.1	19.7	19.4
Mean Number Previous Offences	9.1	8.6	8.7
Recidivism within 12 Months	56.8	47.9	48.8
Percentage Released on HDC	0	24.4	22
Sample Size	2,333	21,946	24,279

Figure 1: Proportion Discharged on HDC by Sentence Length

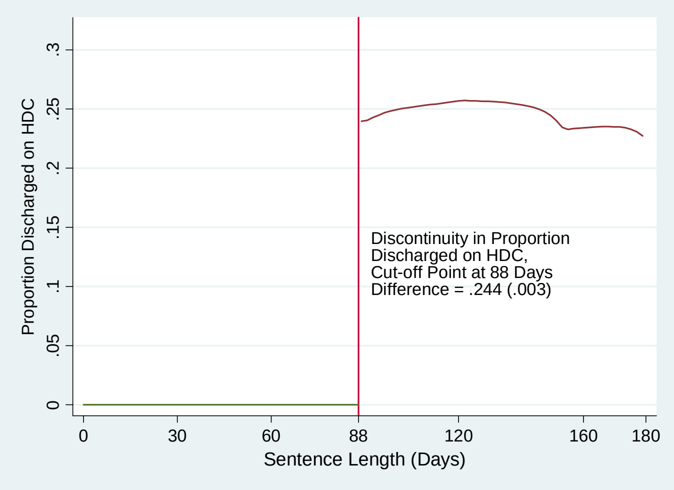


Figure 2: Mean Number of Previous Offence by Sentence Length

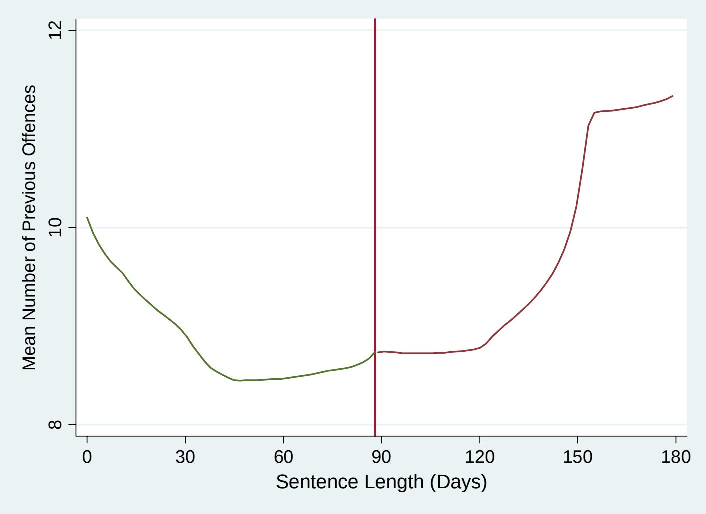


Table 4: RDD Estimates of HDC Impact on Recidivism – Around Threshold

	Dependent Variable = Recidivism Within 12 Months Estimation on Individuals Discharged +/- 7 Days of 88 Days Threshold		
	(1)	(2)	(3)
Estimated Discontinuity of HDC Participation at Threshold ($HDC^+ - HDC^-$)	.243 (.009)	.223 (.009)	.243 (.003)
Estimated Difference in Recidivism Around Threshold ($Rec^+ - Rec^-$)	-.089 (.011)	-.059 (.009)	-.044 (.014)
Estimated Effect of HDC on Recidivism Participation ($Rec^+ - Rec^-$)/ ($HDC^+ - HDC^-$)	-.366 (.044)	-.268 (.044)	-.181 (n.a.)
Controls	No	Yes	No
PSM	No	No	Yes
Sample Size	24,279	24,279	24,279

For what group is this the ATE?

Internal and External Validity

- At best, Sharp and Fuzzy RDD estimate the average effect of the sub-population with X_i close to c
- Fuzzy RDD further restricts this to the compliers with X_i close to c
- Only with strong assumptions (e.g., homogenous treatment effects) can we estimate the overall average treatment effect
- So, RDDs have strong internal validity but may have weak external validity