Causal Inference

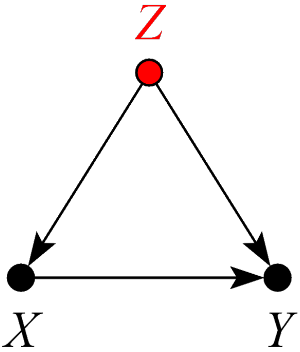
Lecture #4

Juraj Medzihorsky

SGIA & RMC
Durham University

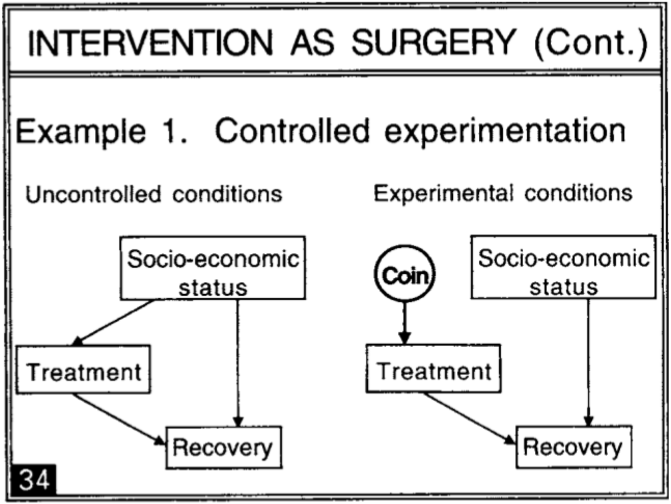
22 February 2024

Recap again



Z is a *confounder* w.r.t. X's effect on Y

Recap again



Source: Pearl, *Causality*, 2nd edition

RDD often takes advantage of the fact that rules allocating units to different meaningful groups are often arbitrary and based on thresholds

The argument is that units just above and just below the threshold are comparable w.r.t. $\{Y(1), Y(0)\}$

There is some X based on which the units are allocated to $\mathcal{D} = \{0, 1\}$. X may be related to Y even asides of this.

Two types of RDD

1. Sharp: allocation to D is a *deterministic* function of X
2. Fuzzy: allocation to D is a *probabilistic* function of X

Regression discontinuity captures a causal effect by distinguishing between a smooth function that specifies the relationship between X and Y and a discontinuous part of that function

Sharp RDD: Examples

Effect of ...

- school district boundary changes on home prices
- class size on student achievement (Maimonides' Rule)
- an additional night in the hospital on health outcomes (babies born just before and just after midnight)
- offer of financial aid that is offered based on exceeding an SAT cut-point
- electoral systems that jump at population thresholds for municipalities
- party: center-left versus center-right mayor and municipal budgets
- wage increases for mayors that jump at population cutoffs on policy performance

Sharp RDD: The General Framework

Suppose a binary treatment D set by the value of a predictor, X_i, being compared to a fixed cutoff point, c:

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

X, the"forcing variable," may be associated with Y

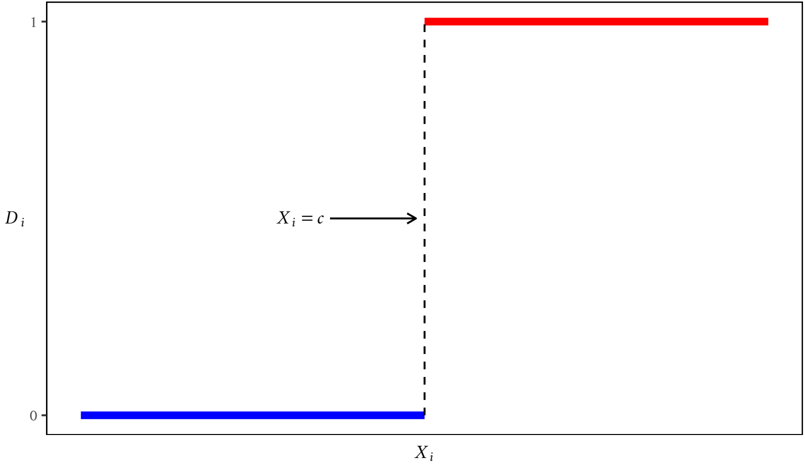
- If so, a simple comparison of outcomes between treatment and control does not provide a *causal* estimate

Design often arises from administrative decisions, where

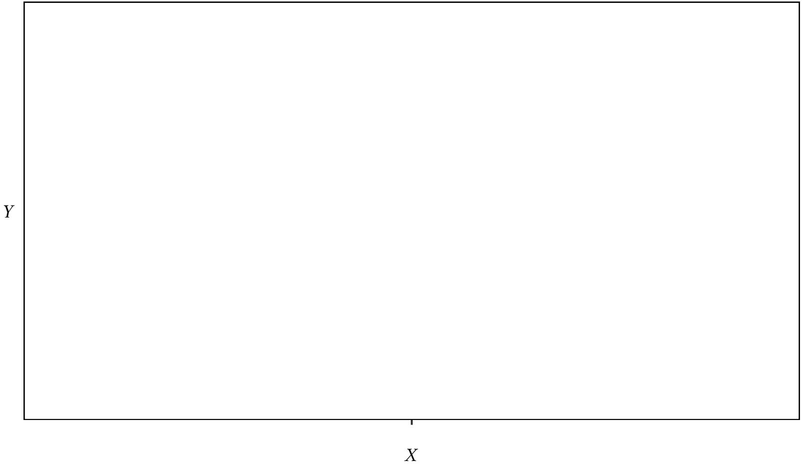
- Allocation of units to a program is partly limited because of resource constraints or need
- Sharp rules, and not discretion by administrators, are used for allocation decisions

If the relationship between X and Y is "smooth" around the cutoff, c we can use the discontinuity created by the treatment at the cutoff point to estimate the causal effect of D on Y

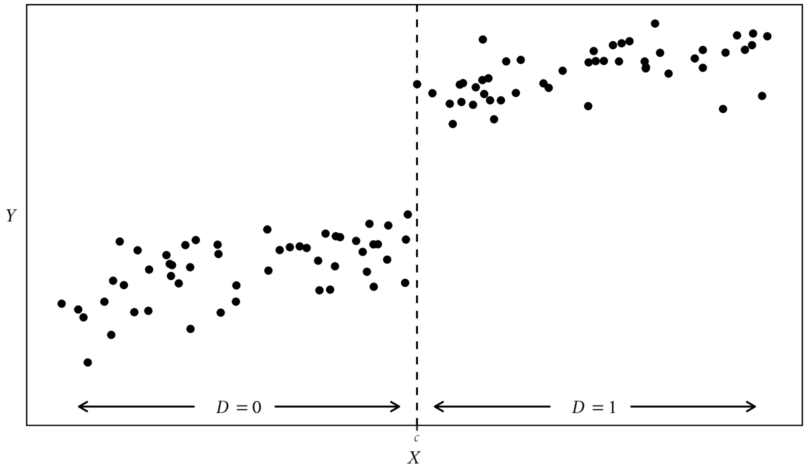
Regression Discontinuity Design



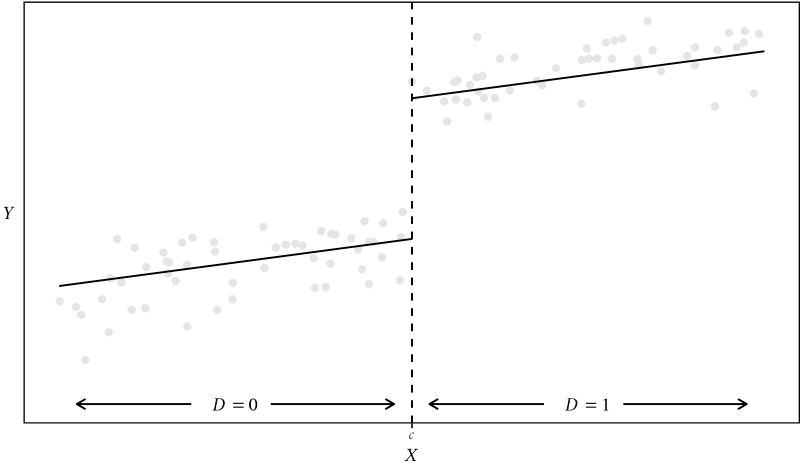
Regression Discontinuity Design



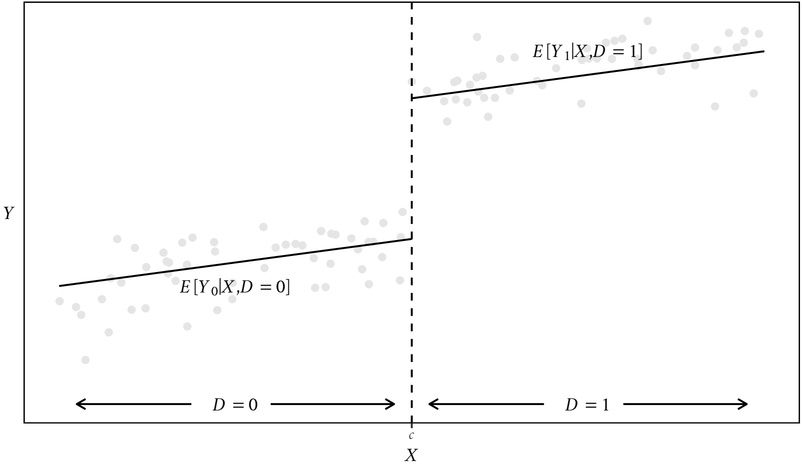
Regression Discontinuity Design



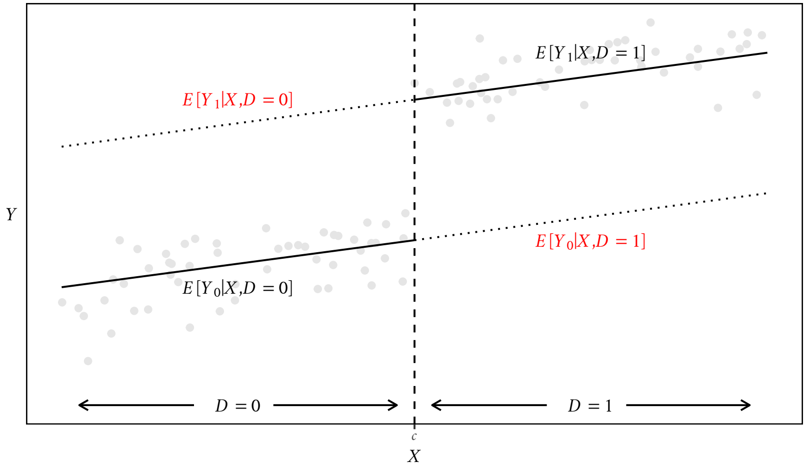
Regression Discontinuity Design



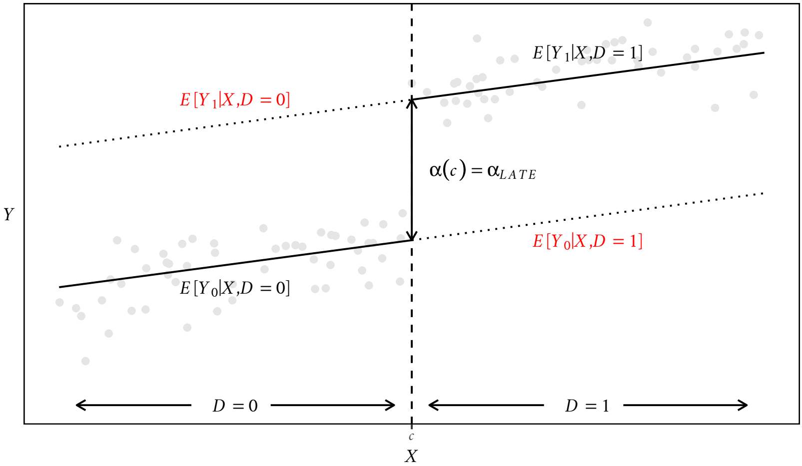
Regression Discontinuity Design



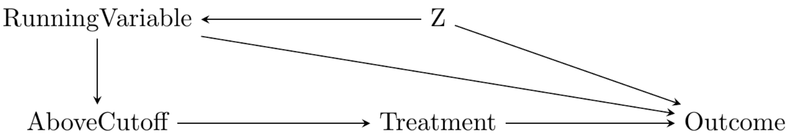
Regression Discontinuity Design



Regression Discontinuity Design



- Allocation driven by an observed covariate know as the forcing (or running) variable
- *Sharp RDD*: deterministic allocation based on a threshold
- *Fuzzy RDD*: conditional on the covariate, some residuals w.r.t. allocation



Source: `https://theeffectbook.net/ch-RegressionDiscontinuity.html`

Sharp RDD: Identification

1. $\{Y(0), Y(1)\} \perp\!\!\!\perp D|X$ is trivially met
2. $0 < P(D = 1|X = x) < 1$ is always violated in Sharp RDD
3. $\mathbb{E}[Y(1)|X, D]$ and $\mathbb{E}[Y(0)|X, D]$ are continuous in X around the threshold, $X = c$ (to compensate for failure of common support)

The treatment effect is identified at the threshold as:

$$\begin{aligned}
\alpha_{SRDD} &= \mathbb{E}[Y(1) - Y(0)|X = c] \\
&= \mathbb{E}[Y(1)|X = c] - \mathbb{E}[Y(0)|X = c] \\
&= \lim_{x \downarrow c} \mathbb{E}[Y(1)|X = c] - \lim_{x \uparrow c} \mathbb{E}[Y(0)|X = c]
\end{aligned}$$

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

Sharp RDD: Identification

Main additional assumption required:
Exogeneity of forcing variable values and outcomes

- No manipulation by units about their own values X
- No manipulations by creators and applicators of c
- If such control is precise, exogeneity is violated
- If control is imprecise—i.e., with random error—this is enough to produce random assignment at the cutpoint
  - Lee (2008), Lee and Lemieux (2010)

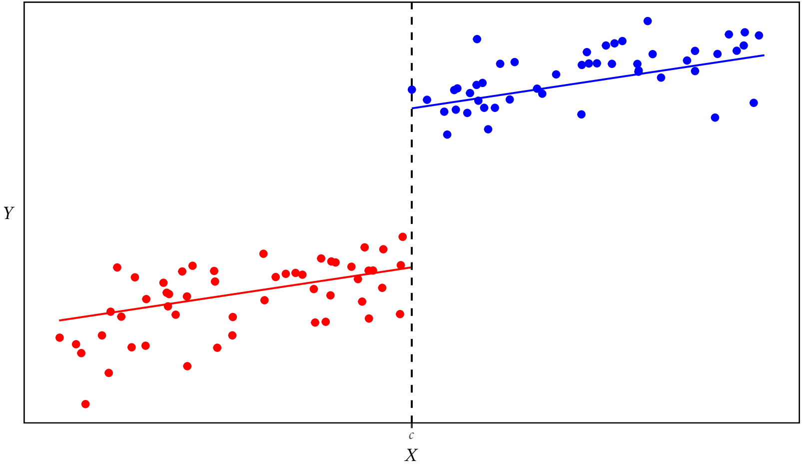
Sharp RDD: Estimation Steps

1. Trim the sample to a reasonable window around the threshold c
   - The "discontinuity sample"
   - $c - h \leq X_i \leq c + h$, where $h \geq 0$ sets the size of the window
   - h may be chosen by a cross-validation procedure

2. Code an alternative forcing variable as deviations from the threshold: $\widetilde{X} = X - c$
   - $\widetilde{X} = 0$ if $X = c$
   - $\widetilde{X} > 0$ if $X > c \Longrightarrow D = 1$
   - $\widetilde{X} < 0$ if $X < c \Longrightarrow D = 0$

3. Decide on a model for $\mathbb{E}[Y|X]$
   - Linear, same slope for $\mathbb{E}[Y(0)|X]$ and $\mathbb{E}[Y(1)|X]$
   - Linear, different slopes for $\mathbb{E}[Y(0)|X]$ and $\mathbb{E}[Y(1)|X]$
   - Non-linear
   - ALWAYS begin with visual inspection (scatterplots, kernel densities, lowess smoothers, piecewise linear fits) to check which model is appropriate

Sharp RDD: Linear, Same Slope



Sharp RDD: Linear, Same Slope

- $\mathbb{E}[Y(0)|X]$ is linear and the ATE does not depend on X:

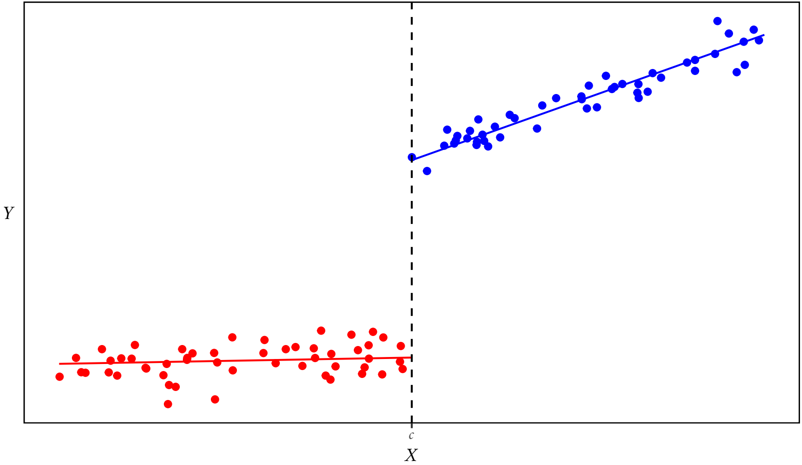
$$\mathbb{E}[Y(0)|X] = \mu + \beta X, \qquad \mathbb{E}[Y(1) - Y(0)|X] = \alpha$$

- Therefore $\mathbb{E}[Y(1)|X] = \alpha + \mathbb{E}[Y(0)|X] = \alpha + \mu + \beta X$
- Since D is completely determined given X, we have that:

$$\begin{aligned} \mathbb{E}[Y|X, D] &= D \cdot \mathbb{E}[Y(1)|X] + (1 - D) \cdot \mathbb{E}[Y(0)|X] \\ &= \mu + \alpha D + \beta X \\ &= (\mu + \beta c) + \alpha D + \beta(X - c) \\ &= \gamma + \alpha D + \beta \widetilde{X} \end{aligned}$$

- So we just run a regression of Y on D and $\widetilde{X}$
  - Regression of Y on D and $X - c$ yields the same result

Sharp RDD: Linear, Different Slopes



Sharp RDD: Linear, Different Slopes

- $\mathbb{E}[Y(0)|X]$ and $\mathbb{E}[Y(1)|X]$ are distinct linear functions of X, so the average effect of the treatment, $\mathbb{E}[Y(1) - Y(0)|X]$, varies with X:

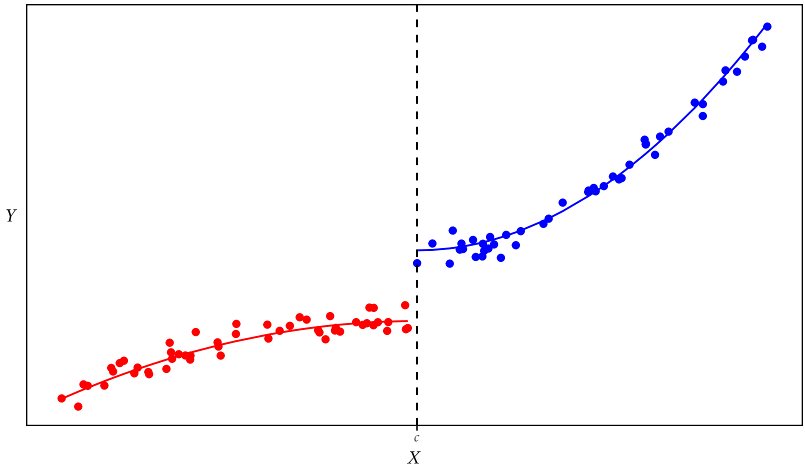
$$\mathbb{E}[Y(0)|X] = \mu_0 + \beta_0 X, \qquad \mathbb{E}[Y(1)|X] = \mu_1 + \beta_1 X$$

- So $\alpha(X) = \mathbb{E}[Y(1) - Y(0)|X] = (\mu_1 - \mu_0) + (\beta_1 - \beta_0)X$
- Since D is completely determined given X, we have:

$$\begin{aligned}
\mathbb{E}[Y|X, D] &= D \cdot \mathbb{E}[Y(1)|X] + (1 - D) \cdot \mathbb{E}[Y(0)|X] \\
&= \mu_1 D + \beta_1 (X \cdot D) + \mu_0 (1 - D) + \beta_0 (X \cdot (1 - D)) \\
&= \gamma + \beta_0 (X - c) + \alpha D + \beta_1 ((X - c) \cdot D)
\end{aligned}$$

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

Sharp RDD: Non-Linear, Different Slopes



Sharp RDD: Non-Linear

- $\mathbb{E}[Y(0)|X]$ and $\mathbb{E}[Y(1)|X]$ are distinct non-linear functions of X and the average effect of the treatment $\mathbb{E}[Y(1) - Y(0)|X]$ 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

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

  The specification with cubic terms is

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

- In both cases $\alpha_0 = \mathbb{E}[Y(1) - Y(0)|X = c]$

Sharp RDD: Cross-Validation

How do we choose the bandwidth h in an automated way?

- h is a "tuning" parameter
- Standard approach is to choose a tuning parameter that minimizes prediction error based on a training data set and then check its predictive accuracy on a separate validation dataset
- One approach is to use an algorithm that picks a bandwidth suited for the conditional expectation of the outcome near the cutpoint (Imbens and Kalyanaraman 2011)
- Since training datasets are not available in the typical RDD case, one technique for predictive accuracy is *cross-validation*

Sharp RDD: Cross-Validation

The typical procedure:

1. Select a bandwidth, h_1
2. Start with an observation A to the left of the threshold with measures on the forcing variable X_A and outcome variable Y_A
3. Run a regression of Y on X using all of the observations that are located to the left of observation A and have a value of X that ranges from $X_A - h_1$ to X_A (not including X_A)
4. Get the predicted value of the outcome variable for observation A based on this regression, $\widehat{Y}_A$
5. Shift the "band" slightly over to the left and repeat the process for observation B, C, and so on for all observations to the left of the threshold; stop when there are fewer than two observations
6. Repeat the process for observations to the right of the threshold
7. Calculate the cross-validation (CV) criterion for bandwidth h_1:
    - Most commonly, the mean-squared error: $CV(h_1) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \widehat{Y}_i)^2$
8. Repeat the process for other bandwidth choices, $h_2, h_3, \ldots$
9. Choose the bandwidth that minimizes the cross-validation criterion

Sharp RDD: Falsification Checks

1. *Sensitivity*: Are the results sensitive to alternative specifications?

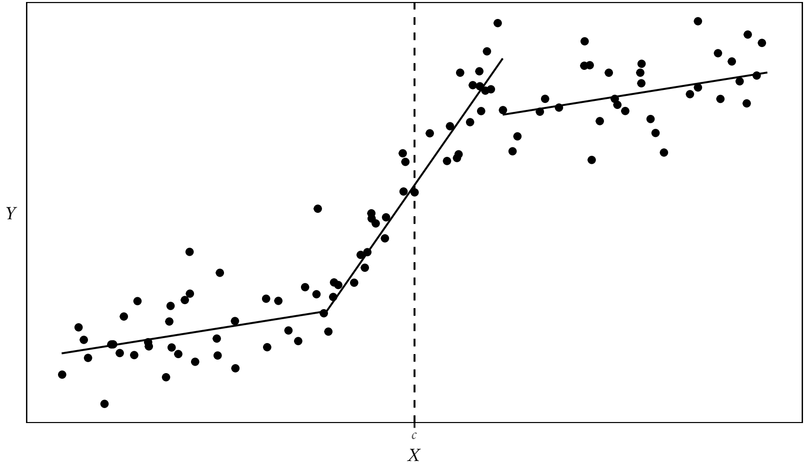
2. *Balance Checks*: Do covariates Z jump at the threshold?

3. *Placebo Tests*: Do jumps occur at placebo thresholds c^*?

4. *Sorting*: Do units sort around the threshold?

Sharp RDD: Sensitivity to Specification

Nonlinearity Mistaken for Discontinuity



Sharp RDD: Sensitivity to Specification

Nonlinearity Mistaken for Discontinuity

- $Y = f(X) + \alpha D + \epsilon$
  - A misspecified control function $f(X)$ can lead to a spurious jump
  - Need to take care not to confuse a non-linear relationship with a discontinuity
- More flexibility in $f(.)$ reduces bias, but decreases efficiency
- Check sensitivity to bandwidth

Sharp RDD: Balance Checks

Test for comparability of units around the threshold

- Visual tests: Plot $\mathbb{E}[Z|X, D]$ and look for jumps
  - Ideally the relation between covariates and treatment should be smooth around the threshold
- Run the RDD regression using Z as the outcome:
  $\mathbb{E}[Z|X, D] = \beta_0 + \beta_1(X - c) + \alpha_Z D + \beta_3((X - c) \cdot D)$
  - Should yield $\alpha_Z = 0$ if Z is balanced at the threshold

Finding a discontinuity at Z does not necessarily invalidate the RDD

- Can incorporate Z as additional controls in the main RDD regression
  - Should only impact SE/CI, not magnitude of treatment effect
- Alternatively, regress Y on controls and use the residuals in the RDD instead of Y itself

As usual, balance checks only cover observables

Sharp RDD: Placebo Threshold

- Test whether the treatment effect is zero when it should be
- Let c^* be a placebo threshold value and estimate:

$$\mathbb{E}[Y|X,D] = \beta_0 + \beta_1(X - c^*) + \alpha D + \beta_3((X - c^*) \cdot D)$$

  - Check α
  - Usually we split the sample to the left and right of the theoretical threshold, c, in order to avoid misspecification by imposing no 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 threshold is contaminated by other factors
- Maybe data exists in a period where there was no program

Sharp RDD: Sorting Around the Threshold

Can units' behavior invalidate the local continuity assumption?

- Can they exercise control over their values on the assignment variable?
- Can administrators strategically choose the assignment variable to use or the threshold in order to include and/or exclude certain units?
- Either can invalidate the comparability of subjects near the threshold because of sorting of agents around the threshold, where those below may differ on average from those just above
- Does not necessarily invalidate the design unless sorting is very precise

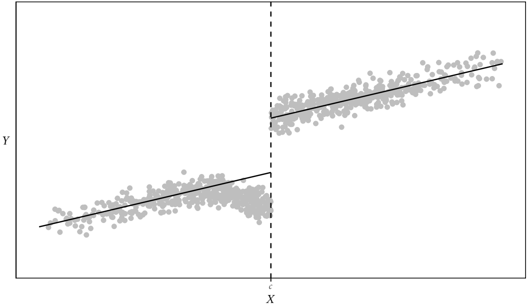
What else changes at c?

- Continuity violated in the presence of other programs that use a discontinuous assignment rule with the exact same assignment variable and threshold

Sharp RDD: Sorting Around the Threshold

Example: Beneficial job training program offered to people with income less than c

- Concern: people will withhold labor to lower their incomes below c to gain access to the program



Sharp RDD: Sorting Around the Threshold

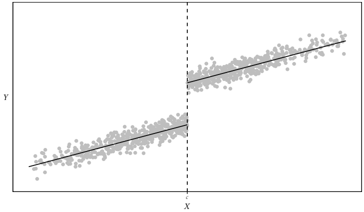
Check for discontinuity in the density of the forcing variable

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 threshold
- For each bin, compute the number of observations and plot the bins to determine whether there is a discontinuity at the threshold

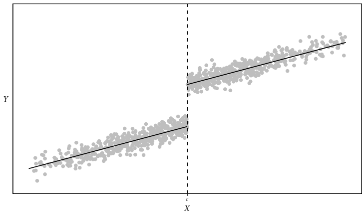
Sharp RDD: Sorting Around the Threshold

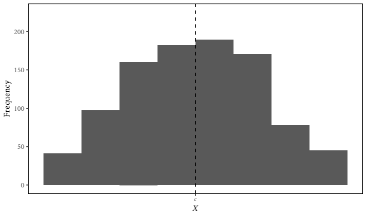
Visual histogram inspection:



Sharp RDD: Sorting Around the Threshold

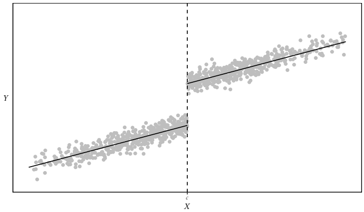
Visual histogram inspection:

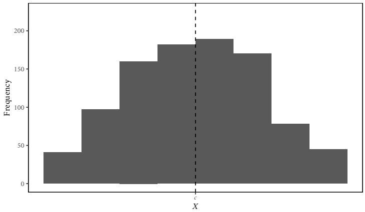


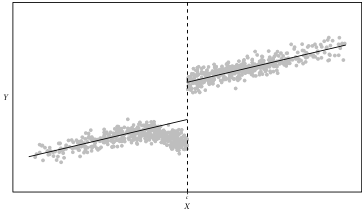


Sharp RDD: Sorting Around the Threshold

Visual histogram inspection:

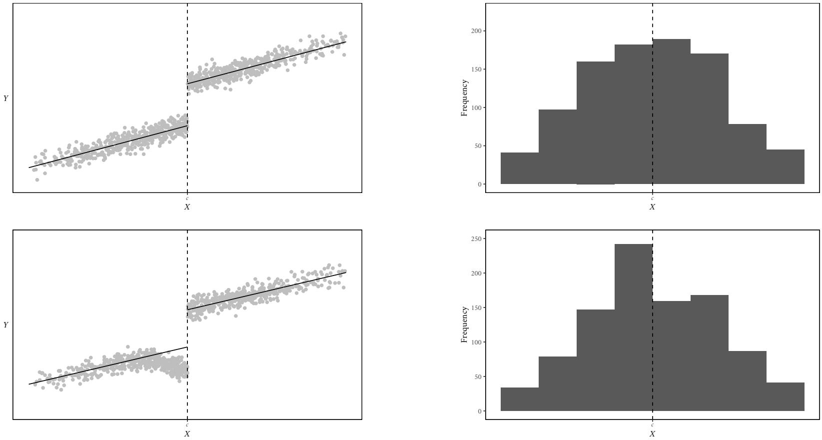






Sharp RDD: Sorting Around the Threshold

Visual histogram inspection:



Sharp RDD: Sorting Around the Threshold

Check for discontinuity in the density of the 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 threshold
- For each bin, compute the number of observations and plot the bins to determine whether there is a discontinuity at the threshold

Formal tests (e.g., McCrary 2008)

Fuzzy Regression Discontinuity: Basic Framework

- Threshold does not perfectly determine treatment exposure, but 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
- Discontinuities can be used to produce instrumental variables estimators of the effect of treatment (close to the discontinuity)

Fuzzy RDD: Example

- Probability of being offered a scholarship may jump at a certain test score threshold
  - Applicants given “special consideration”
- We should not compare recipients with non-recipients (even close to the threshold)
  - Applicants are likely to differ along unobservables related to the outcome
- For applicants with scores close to the threshold, we can exploit the discontinuity as an instrument to estimate the LATE for the subgroup of applicants for whom scholarship receipt depends on the difference between their scores and the threshold
  - A complier is a student who switches from non-recipient to recipient if her scores crosses the threshold

Fuzzy RDD: Identification

- Binary instrument Z with $Z = \mathbb{I}\{X > c\}$
- Restrict sample to observations close to discontinuity where $\mathbb{E}[Y|X, D]$ jumps so that $X \approx c$ and thus $\mathbb{E}[X|Z = 1] - \mathbb{E}[X|Z = 0] \approx 0$
- Usual instrumental variables assumptions (ignorability, first stage, monotonocity)

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

Fuzzy RDD: Estimation Steps

1. Cut the sample into a small window above and below the threshold (the discontinuity sample)
2. Code instrument: $Z = \mathbb{I}\{X > c\}$
3. Fit 2SLS: $Y = \beta_0 + \beta_1(X - c) + \beta_2(Z \cdot (X - c)) + \alpha D$, where D is instrumented with Z
4. Specification can be made more flexible by adding polynomials
5. Also helpful to separately plot and estimate the outcome discontinuity and treatment discontinuity

Internal and External Validity

- At best, sharp and fuzzy RDD estimate the average effect of the treatment on the subpopulation with X_i close to c
- Fuzzy RDD restricts this subpopulation even further to that of compliers with X_i close to c
- Only with strong assumptions (e.g., homogeneous treatment effects) can we estimate the overall average treatment effect
- For these reasons, RDD may have strong internal validity but weak external validity

Regression Discontinuity: Key Points

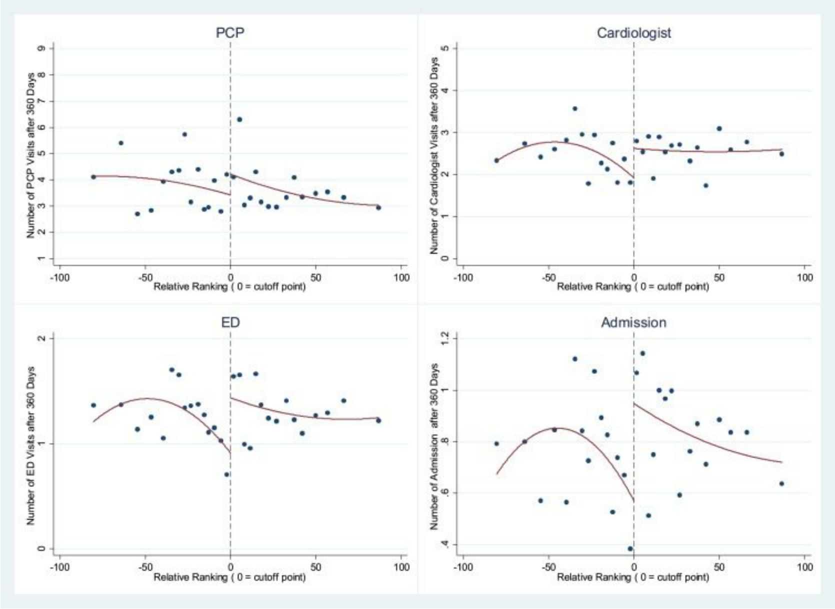
RDD can be used when our research setting satisfies two conditions

1. A continuous measure exists that defines eligibility for a treatment (often a program or policy)
2. A clearly defined cutoff point on this continuous measure that serves as a rule for assigning units in the population to be part of the program or policy

Under some additional assumptions, we can estimate a $LATE$

- We try to meet some of these assumptions by focusing on data within a small window around the threshold
- We can perform some basic checks to see whether we have evidence supporting the validity of the regression discontinuity design

RDD: *but, but, watch it strut*



Source: `https://statmodeling.stat.columbia.edu/2021/11/21/sausage-notice/`

?

Thank you!

Further Reading

- Holland, Paul W. 1986. “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81(396): 945–960.
- Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton: Princeton University Press.
- Cook, Thomas D., and Vivian C. Wong. 2008. “Empirical Tests of the Validity of the Regression Discontinuity Design.” *Annales d’Économie et de Statistique* 91/92: 127–150.
- Lee, David S. 2008. “Randomized Experiments from Non-random Selection in U.S. House Elections.” *Journal of Econometrics* 142(2): 675–697.
- Lee, David S., and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48: 281–355.
- Ludwig, Jens, and Douglas L. Miller. 2007. “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design.” *Quarterly Journal of Economics* 122(1): 159–208.
- Imbens, Guido, and Karthik Kalyanaraman. 2012. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *Review of Economic Studies* 79(3): 933–959.
- McCrary, Justin. 2008. “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test.” *Journal of Econometrics* 142(2): 698–714.