Part E: Regression Discontinuity

E2: RDD Extensions

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2025

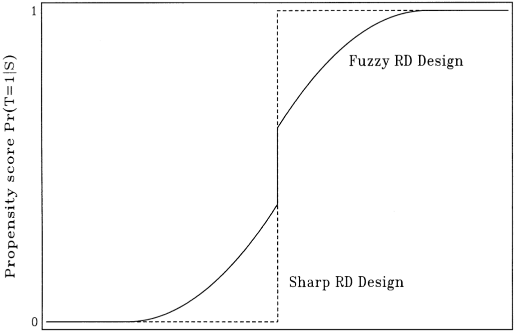
E2 Outline

1. Fuzzy RDD
2. Discrete running variables and RD in time
3. Adding covariates
4. Multiple cutoffs or running variables and Spatial RD
5. Local randomization approach
6. Extrapolating RD estimates

Reading: Cattaneo, Idrobo, and Titiunik (2023)

Fuzzy RDD

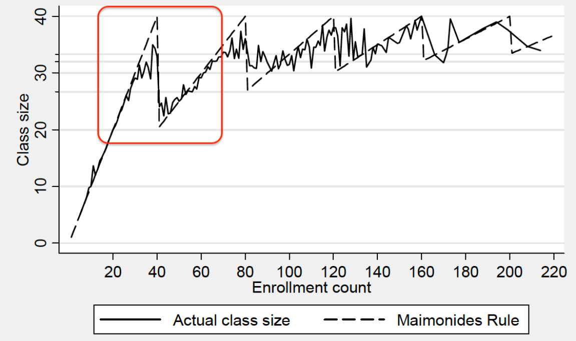
- Think of fuzzy RDD as using $Z_i = \mathbf{1}[X_i \geq c]$ as an instrument:
  - $D_i \neq Z_i$: treatment is not fully determined by X_i
  - But $\mathbb{E}[D_i \mid X_i]$ jumps at the cutoff $\Longrightarrow Z_i$ is a relevant IV around $X_i = c$



(Illustration from Van Der Klaauw (2002), Fig. 2)

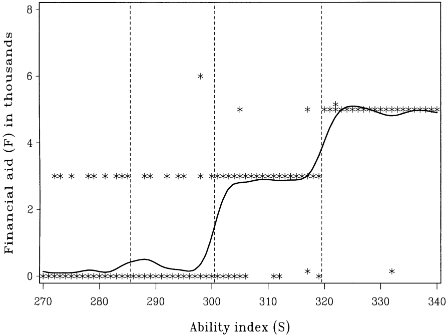
Fuzzy RDD examples

Class size "first stage"



(MHE, Fig. 6.2.1a, based on Angrist and Lavy (1999))

Financial aid "first stage"



(Van Der Klaauw (2002), Fig. 4)

Identification and estimation

- With binary treatment, need standard IV assumptions:
  - *Exclusion*: $Y_i(d, z) \equiv Y_i(d)$
  - *Independence*: continuity of $\mathbb{E}[D_i(z) \mid X_i = x]$ and $\mathbb{E}[Y_i(d) \mid X_i = x]$ at $x = c$
  - *Monotonicity*: $D_i(1) \geq D_i(0)$
- With binary D, reduced form/first stage $\equiv \tau_Y/\tau_D$ identifies LATE:

$$\frac{\tau_Y}{\tau_D} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0),\ X_i = c]$$

- Report $\hat{\tau}_D$, $\hat{\tau}_Y$ and fuzzy RD estimate $\hat{\tau}_Y/\hat{\tau}_D$ from local polynomial estimation with the same bandwidth for Y and D
  - `rdrobust` chooses bandwidth to minimize MSE for the IV
  - Show RD plots for the first stage and reduced form (ITT)

Outline

Discrete running variables

- In many RDD applications X_i is discrete
  - E.g. number of kids in a cohort *(Angrist and Lavy (1999))*
- Does this matter conceptually?
  - $\lim_{x \downarrow, \uparrow c} \mathbb{E}\left[Y_i \mid X_i = x\right]$ is not well-defined $\implies$ RD identification fails
  - As $N \to \infty$, we can't shrink bandwidth $h \to 0$
- Does this matter in practice?
  - If there are many mass points of X_i around c, can probably ignore the issue
  - If X_i is sparse around c, it's more salient

An “honest” (“bias-aware”) approach

- Armstrong and Kolesár (2018) and Kolesár and Rothe (2018) develop an “honest” approach to RDDs
  - Acknowledges that bias in local linear estimation is inevitable
  - With discrete X_i we cannot consistently estimate bias
- Instead, it bounds worst-case bias by assuming that $\mathbb{E}[Y_i \mid X_i]$ is sufficiently smooth on either side of c
  - Choose bound M on the curvature of $\mathbb{E}[Y_i \mid X_i] \Rightarrow$ get a partially identified set of τ
  - Reminds you of anything?
- Choosing M is annoying but ignoring discreteness does the same implicitly
  - A rule of thumb is available

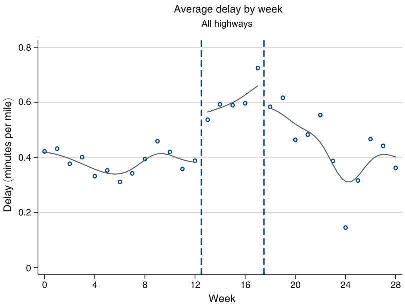
An “honest” (“bias-aware”) approach (2)

- `rdhonest` produces a “bias-aware” confidence interval
  - Centered around the local linear estimator
    - With a different bandwidth (optimized for CI length); in practice, similar to the Calonico et al. (2014) bandwidth
  - $\pm$ worst-case bias (rather than subtracting estimated bias), in addition to SE
- Approach applies even with continuous running variable
  - And can have good finite-sample properties because doesn't rely on h being small
- See Imbens and Wager (2019) for an honest approach not based on local linear estimation
  - More complex (via numerical optimization) but more generalizable

RD in time

Related problems arise with "**RD in time**"

- X_i = period; often no cross-sectional variation at all, just a time series
- E.g. Anderson (2014): the effect of a public transit strike in LA on highway congestion



(Anderson (2014), Fig. 2)

- Similar situation: X_i = age

How to think about RD in time?

- Theoretically, time is a continuous variable
  - Could measure the outcome a second before and after the policy change — like event studies in finance
  - Asymptotic with data frequency growing
- In practice, outcomes are measured at discrete intervals, and collecting more data involves going further in time from c
  - As $T \to \infty$, the bandwidth can't (and doesn't) shrink
  - Understanding check: is the McCrary test helpful here?

Outline

Adding covariates

As usual, covariates W_i can be added to increase efficiency or to avoid OVB

- If W_i are predetermined, i.e. $\mathbb{E}[W_i \mid X_i]$ is continuous at the cutoff:
  - Include W_i in the regression implementation of local linear estimator *without* interactions:

$$Y_i = \tau D_i + \gamma_0 + \gamma_1(X_i - c) + \gamma_2(X_i - c)D_i + \delta' W_i + \text{error}$$

  - This increases efficiency without changing the estimand *(Calonico et al. (2019))*
  - See Noack, Olma, and Rothe (2023) on flexible covariate adjustment
- If W_i jumps at the cutoff, the effects of D_i and W_i cannot be separated without further assumptions
  - Frölich and Huber (2019) make a selection-on-observables assumption; see also Peng and Ning (2021)

Outline

Multiple cutoffs or running variables

Four scenarios:

1. Scalar running variable, single heterogeneous cutoff
2. Scalar running variable, multiple cutoffs
3. Multiple running variables, single discontinuity — including spatial RD (see Cattaneo, Titiunik, and Yu (2025d) for a review)
4. Aggregating multiple discontinuities (Borusyak and Kolerman-Shemer (2024))

Note: package `rdmulti` provides commands for estimation and plotting in #1–3

#1: Scalar running variable, single heterogeneous cutoff

- E.g. states have different income cutoffs for a means-tested program: $D_i = \mathbf{1}[X_i \geq C_i]$ for $C_i \in \{c_1, \ldots, c_K\}$

- Obviously, we can RD by subgroup $C_i = c$:

$$\tau_c = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = c, C_i = c\right]$$

- Can also pool them by using "normalized" $\tilde{X}_i = X_i - C_i$ with a cutoff of zero

  - Pooled RDD identifies a weighted average of group-specific ones:

$$\tau_{\text{pooled}} = \frac{\sum_c \omega_c \tau_c}{\sum_c \omega_c}, \qquad \omega_c = f_{X|C}(c, c)$$

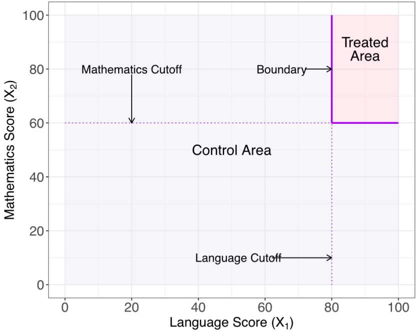
#2: Scalar running variable, multiple cutoffs

- E.g. rounding of Yelp stars *(Anderson and Magruder (2012))* generates cutoffs at $X_i = 3.25, 3.75, 4.25, 4.75$
- Or federal subsidies determined by local population, with several discontinuities
- Consider a sharp design with multi-valued $D_i = \begin{cases} d_0, & \text{if } X_i < c_1 \\ d_1, & \text{if } c_1 \leq X_i < c_2 \\ \dots & \\ d_K, & \text{if } c_K \leq X_i \end{cases}$
- RDD on subsample with $D_i \in \{d_{k-1}, d_k\}$ identifies $\mathbb{E}\left[Y_i(d_k) - Y_i(d_{k-1}) \mid X_i = c_k\right]$
- Mostly similar to the previous case
    - Except same observation can be used twice (unless bandwidth is small enough)

#3: Multiple running variables, single discontinuity

E.g. scholarship awarded to students scoring above a cutoff in *both* math and English:

$$\boldsymbol{X}_i = (Math_i, English_i), \qquad D_i = \mathbf{1}\left[Math_i \geq c_{Math}\right] \times \mathbf{1}\left[English_i \geq c_{English}\right]$$

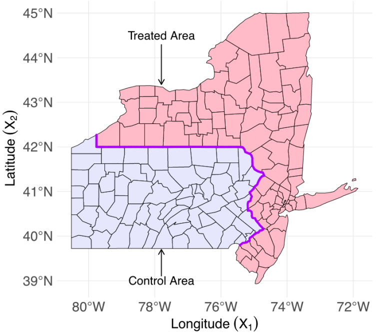


(Cattaneo et al. (2023), Fig. 5.5a)

- Note that a student needn't be close to the border on *both* running variables to be near the boundary

Special case: Spatial discontinuity designs

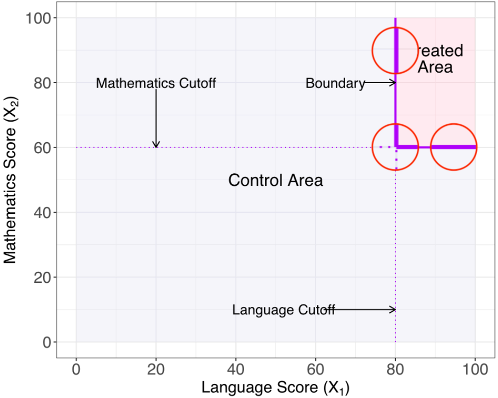
$\boldsymbol{X}_i = (Longitude_i, Latitude_i)$; border is often complex



(Cattaneo et al. (2023), Fig. 5.5b)

- E.g. Dell (2010) compares today's outcomes across the boundary of a forced mining labor system in Peru and Bolivia *"mita"*

Effects at a single boundary point



Effects at a single boundary point: Identification

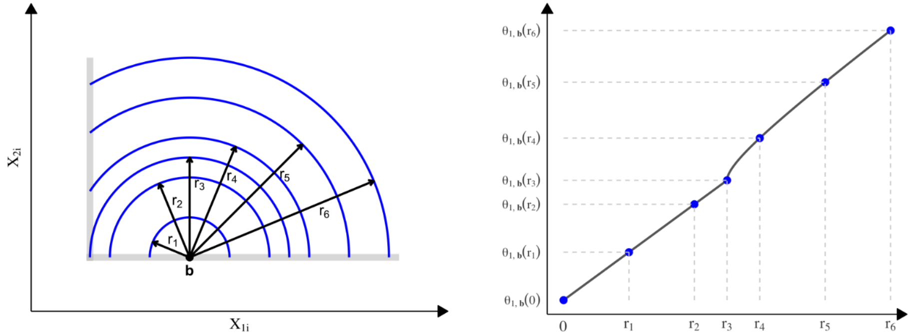
- Let $D_i = a(\boldsymbol{X}_i)$, $A_0 = \{x\colon a(x) = 0\}$, $A_1 = \{x\colon a(x) = 1\}$, and B = boundary between A_0 and A_1
- Assume $\mathbb{E}\left[Y_i(d) \mid \boldsymbol{X}_i = x\right]$ is continuous at $x \in B$
  - Violated when multiple outcome-relevant treatments jump at the same boundary
  - Or when location $\boldsymbol{X}_i$ or the boundary can be manipulated
- Average causal effect at point $b \in B$, $\tau(b)$, is identified by

$$\tau(b) = \lim_{x \to b,\ x \in A_1} \mathbb{E}\left[Y_i \mid \boldsymbol{X}_i = x\right] - \lim_{x \to b,\ x \in A_0} \mathbb{E}\left[Y_i \mid \boldsymbol{X}_i = x\right]$$

Effects at a single boundary point: Estimation

1. Let $d(\boldsymbol{X}_i, b)$ denote some distance metric (e.g. Euclidean)
   - Use "signed distance" as the running variable: $\tilde{X}_i = \begin{cases} d(\boldsymbol{X}_i, b), & D_i = 1 \\ -d(\boldsymbol{X}_i, b), & D_i = 0 \end{cases}$ (with a cutoff of zero)
   - Use local linear estimation, controlling for $\tilde{X}_i$ and $D_i\tilde{X}_i$
   - But Cattaneo et al. (2025b) show poor behavior near irregular points of the boundary (e.g., kinks)
2. Location-based approach (Cattaneo et al. (2025c)): within $d(\boldsymbol{X}_i, b) \leq h$, regress Y_i on D_i controlling for a polynomial of $\boldsymbol{X}_i - b$ and its interaction with D_i
   - Does not suffer from problems of the distance-based approach

Problems of distance-based estimation



(a) Distance to $\mathbf{b} \in \mathscr{B}$.

(b) Distance-based Conditional Expectation.

Cattaneo et al. (2025b), Fig.1

Pooled effect

- How can we estimate the average effect pooling across boundary points, $\mathbb{E}\left[Y_i(1) - Y_i(0) \mid \mathbf{X}_i \in B\right]$?
- Naive "pre-aggregation" approach: Compute distance $d_{min}(\mathbf{X}_i)$ to the *closest* boundary point
    - Use signed distance $\tilde{X}_i = \begin{cases} d_{min}(\mathbf{X}_i), & D_i = 1 \\ -d_{min}(\mathbf{X}_i), & D_i = 0 \end{cases}$ and a cutoff of zero, controlling for $\tilde{X}_i$ and $\tilde{X}_i D_i$
- But in finite samples this may not be enough
    - Observations on the two sides of the border may be geographically imbalanced

Pooled effect: better estimators

- Better pre-aggregation:
  - Black (1999) further controls for FEs of boundary segments to improve geographic balance
  - Dell (2010) further controls for polynomials in $\mathbf{X}_i$ (better: interacted with D_i!)
- Imbens and Zajonc (2009) and Cattaneo et al. (2025a) "post-aggregation": manually average estimated $\hat{\tau}(b)$ over the boundary $b \in B$
- The honest approach of Imbens and Wager (2019)
  - Directly chooses the optimal estimator (via numerical optimization), accounting for worst-case bias under a bound on two-dimensional curvature of $\mathbb{E}\left[Y_i(d) \mid \mathbf{X}_i\right]$

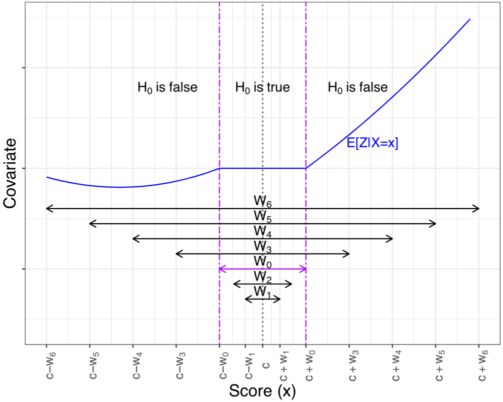
Outline

Local randomization approach to RDDs

- Lee (2008) title: *"Randomized experiments from non-random selection in U.S. House elections"*
- In the continuity approach this idea is a heuristic
  - D_i is only approximately independent from $Y_i(d) \Longrightarrow$ local polynomial adjustments
- Local randomization approach (Cattaneo et al. (2015)) takes this idea seriously
  - Assume $X_i \perp\!\!\!\perp Y_i(d) \mid X_i \in \mathcal{X}$ in a (potentially unknown) window $\mathcal{X} = [c-h, c+h]$
  - And that $F(X_i \mid X_i \in \mathcal{X})$ is known, e.g. uniform in $\mathcal{X}$ or across permutations
  - For no good reason!
- Under these assumptions, can use all RCT machinery
  - E.g. randomization inference that is valid in finite-samples

Choosing the window

Cattaneo et al. (2015) propose to start from smallest h and increase it until you reject balance of some predetermined W_i



(Cattaneo et al. (2023), Figure 2.4)

Outline

1. Fuzzy RDD
2. Discrete running variables and RD in time
3. Adding covariates
4. Multiple cutoffs or running variables and Spatial RD
5. Local randomization approach
6. Extrapolating RD estimates

Extrapolating RD estimates

A key limitation of RDDs is the local nature of $\tau = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = c\right]$

- Kids who barely receive financial aid and politicians who barely win may be unusual
- When can we identify effects away from the cutoff to improve external validity?

Idea #1 (Dong and Lewbel (2015)): local linear estimation automatically yields

$$\phi = \frac{\partial \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = x\right]}{\partial x} \Big|_{x=c}$$

$=$ Difference in regression slopes on the right and left

- Thus, $\mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = x\right] \approx \tau + \phi(x - c)$ for $x \approx c$

Extrapolating RD estimates (2)

Idea #2 (Angrist and Rokkanen (2015)): suppose X_i is a noisy version of observable W_i

- E.g. $D_i =$ offer for a selective school in Boston, $X_i =$ admission test score
  - Assume X_i is random noise conditionally on pre-application test score
- Conditional independence assumption:

$$\mathbb{E}\left[Y_i(d) \mid X_i, W_i\right] = \mathbb{E}\left[Y_i(d) \mid W_i\right] \quad \Longrightarrow \quad \mathbb{E}\left[Y_i(d) \mid D_i, W_i\right] = \mathbb{E}\left[Y_i(d) \mid W_i\right]$$

  - Given W_i, we can compare treated and untreated, as long as there is overlap
  - Can use standard CIA methods to get the ATE (on everyone — nothing local)
  - *Note*: we used that D_i is a deterministic function of X_i, but not its discontinuity

Extrapolating RD estimates (3)

Angrist and Rokkanen's CIA assumption is falsifiable: it implies

$$\mathbb{E}\left[Y_i \mid X_i, D_i, W_i\right] = \mathbb{E}\left[Y_i \mid D_i, W_i\right]$$

- Among the treated, X_i should not predict Y_i given W_i; same for the untreated
- *Exercise*: prove this!

References I

Anderson, M. and J. Magruder (2012): "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *The Economic Journal*, 122, 957–989.

Anderson, M. L. (2014): "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion," *American Economic Review*, 104, 2763–2796.

Angrist, J. D. and V. Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics*, 114, 533–575.

Angrist, J. D. and M. Rokkanen (2015): "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away From the Cutoff," *Journal of the American Statistical Association*, 110, 1331–1344.

Armstrong, T. B. and M. Kolesár (2018): "Optimal Inference in a Class of Regression Models," *Econometrica*, 86, 655–683.

Black, S. E. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *The Quarterly Journal of Economics*, 114, 577–599.

Borusyak, K. and M. Kolerman-Shemer (2024): "Regression Discontinuity Aggregation, with an Application to the Union Effects on Inequality," *Mimeo*.

References II

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019): "Regression Discontinuity Designs Using Covariates," *The Review of Economics and Statistics*, 101, 442–451.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs: Robust Nonparametric Confidence Intervals," *Econometrica*, 82, 2295–2326.

Cattaneo, M. D., B. R. Frandsen, and R. Titiunik (2015): "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate," *Journal of Causal Inference*, 3, 1–24.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2023): "A Practical Introduction to Regression Discontinuity Designs: Extensions," *arXiv preprint*, arXiv:2301.08958 [econ, stat].

Cattaneo, M. D., R. Titiunik, and R. Yu (2025a): "Causal Treatment Effect Aggregation in Boundary Discontinuity Designs," *Working Paper*.

——— (2025b): "Estimation and Inference in Boundary Discontinuity Designs: Distance-Based Methods," *Working Paper*.

References III

——— (2025c): "Estimation and Inference in Boundary Discontinuity Designs: Location-Based Methods," *Working Paper*.

Cattaneo, M. D., R. Titiunik, and R. R. Yu (2025d): "Boundary Discontinuity Designs: Theory and Practice," *Working Paper*, arXiv:2511.06474 [econ].

Dell, M. (2010): "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78, 1863–1903.

Dong, Y. and A. Lewbel (2015): "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models," *Review of Economics and Statistics*, 97, 1081–1092.

Frölich, M. and M. Huber (2019): "Including Covariates in the Regression Discontinuity Design," *Journal of Business & Economic Statistics*, 37, 736–748.

Imbens, G. and S. Wager (2019): "Optimized Regression Discontinuity Designs," *The Review of Economics and Statistics*, 101, 264–278.

Imbens, G. and T. Zajonc (2009): "Regression Discontinuity Design with Vector-Argument Assignment Rules," .

Kolesár, M. and C. Rothe (2018): "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 108, 2277–2304.

References IV

Lee, D. S. (2008): "Randomized experiments from non-random selection in U.S. House elections," *Journal of Econometrics*, 142, 675–697, iSBN: 03044076.

Noack, C., T. Olma, and C. Rothe (2023): "Flexible Covariate Adjustments in Regression Discontinuity Designs," ArXiv:2107.07942 [econ, stat].

Peng, S. and Y. Ning (2021): "Regression Discontinuity Design under Self-selection," *International Conference on Artificial Intelligence and Statistics*, 118–126.

Van Der Klaauw, W. (2002): "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression–Discontinuity Approach*," *International Economic Review*, 43, 1249–1287.