4. Regression Discontinuity

PhD Applied Methods

Duncan Webb
NovaSBE

Spring 2026

What is regression discontinuity?

- Regression discontinuity design (RDD) is one of the most powerful and credible research designs for causal inference

- The core idea: take advantage of **institutional features** that generate a **discontinuous change** in treatment at some threshold

- Examples:
  - Scholarship eligibility based on income threshold
  - Election outcomes determined by vote share
  - Class size rules based on enrollment thresholds
  - Program eligibility based on age cutoffs

Example 1: Financial aid and college enrollment

Research question: Does receiving financial aid increase college enrollment?

Setting: Many scholarship programs have sharp income cutoffs

- Example: Students from families earning below $30,000 receive a $1,500 scholarship
- Students from families earning $30,001 receive nothing

Key insight: Students just below vs. just above the income threshold are likely very similar in all respects *except* scholarship receipt

If enrollment rates differ discontinuously at the threshold, this difference can be attributed to the scholarship

Example 2: Electoral advantage and incumbency

Research question: Does barely winning an election give a party an advantage in future elections?

Setting: Close elections (Lee, 2008)

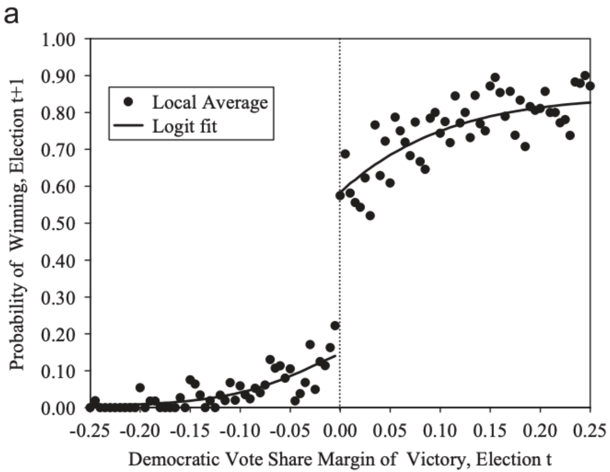
- A Democratic candidate wins if they get $> 50\%$ of the vote
- They lose if they get $< 50\%$ of the vote

Key insight: Elections decided by a very small margin (e.g., 50.1% vs 49.9%) are essentially random

- The parties competing in such close races should be very similar
- Any discontinuous change in future electoral success at the 50% threshold reveals the causal effect of incumbency

Graphical intuition

D.S. Lee / Journal of Econometrics 142 (2008) 675–697



Why is RDD so popular?

RDD has **exploded in popularity** in empirical economics. Key advantages:

- **Credibility**: Exploits institutional rules, not researcher assumptions
  - Often considered nearly as credible as randomized experiments
- **Transparency**: Visual evidence makes results compelling and hard to manipulate
  - The ”eyeball test” is very informative
- **Policy relevance**: Treatment itself is often the policy of direct interest
  - Can also be used as an instrument (fuzzy RDD = IV setup)

Outline

Setup: Notation

Using the potential outcomes framework from our first lecture:

- $Y_i(0)$, $Y_i(1)$: potential outcomes for individual i
- $D_i \in \{0, 1\}$: treatment status
- Observed outcome: $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$

New element: The **running variable** (or **forcing variable**) Z_i

- Also called the "assignment variable" or "score"
- Examples: test score, age, income, vote share
- Treatment assignment determined by cutoff value z^*

The forcing variable

Key assumption: Treatment status D_i is determined (at least partly) by whether Z_i crosses a threshold z^*

Two cases:

- **Sharp RDD**: Treatment changes *deterministically* at z^*
  - Everyone with $Z_i \geq z^*$ is treated; everyone with $Z_i < z^*$ is untreated
- **Fuzzy RDD**: Treatment *probability* changes at z^*
  - Crossing z^* increases the chance of treatment, but doesn't guarantee it

We'll first derive the **general identification result** that applies to both cases, then distinguish between them

The identifying assumption

Core identification assumption: The conditional expectation functions $\mathbb{E}[Y_i(0)|Z_i = z]$ and $\mathbb{E}[Y_i(1)|Z_i = z]$ are **continuous** in z at the cutoff $z = z^*$

What does this mean?

- The average potential outcomes change **smoothly** as Z_i changes
- There are no **other factors** that jump discontinuously at $Z_i = z^*$
- The only thing that changes discontinuously at the threshold is treatment D_i

Intuition: People just below vs. just above the threshold are essentially identical *except* for their treatment status

Identification: The Wald estimand

Define the limits of the conditional expectation of Y on either side of the cutoff:

$$Y^+ := \lim_{z \to (z^*)^+} \mathbb{E}[Y_i | Z_i = z] \tag{1}$$

$$Y^- := \lim_{z \to (z^*)^-} \mathbb{E}[Y_i | Z_i = z] \tag{2}$$

Similarly for treatment:

$$D^+ := \lim_{z \to (z^*)^+} \mathbb{E}[D_i | Z_i = z] \tag{3}$$

$$D^- := \lim_{z \to (z^*)^-} \mathbb{E}[D_i | Z_i = z] \tag{4}$$

Identification: The Wald estimand

Under our model $Y_i = Y_i(0) + D_i(Y_i(1) - Y_i(0))$ and the continuity assumption:

$$
\begin{aligned}
Y^+ &= \lim_{z \to (z^*)^+} \mathbb{E}[Y_i(0) + D_i(Y_i(1) - Y_i(0)) | Z_i = z] && (5)\\
&= \lim_{z \to (z^*)^+} \mathbb{E}[Y_i(0) | Z_i = z] + D^+ \cdot \mathbb{E}[Y_i(1) - Y_i(0) | Z_i = z^*] && (6)\\
&= \mathbb{E}[Y_i(0) | Z_i = z^*] + D^+ \cdot \tau && (7)
\end{aligned}
$$

where $\tau = \mathbb{E}[Y_i(1) - Y_i(0) | Z_i = z^*]$ is the treatment effect at the cutoff

Similarly:

$$
Y^- = \mathbb{E}[Y_i(0) | Z_i = z^*] + D^- \cdot \tau \qquad (8)
$$

Identification: The Wald estimand

Taking the difference:

$$\begin{aligned} Y^+ - Y^- &= \mathbb{E}[Y_i(0)|Z_i = z^*] + D^+ \cdot \tau - \mathbb{E}[Y_i(0)|Z_i = z^*] - D^- \cdot \tau \qquad (9) \\ &= (D^+ - D^-) \cdot \tau \qquad (10) \end{aligned}$$

Therefore, the treatment effect at the cutoff is:

$$\boxed{\tau = \frac{Y^+ - Y^-}{D^+ - D^-}}$$

This is the **Wald estimand** - the ratio of the jump in outcomes to the jump in treatment

What are we identifying?

The Wald formula identifies the **local average treatment effect** at the cutoff:

$$\tau_{RDD} = \mathbb{E}[Y_i(1) - Y_i(0) | Z_i = z^*] \tag{11}$$

Interpretation depends on sharp vs. fuzzy:

- **Sharp RDD**: This is the ATE for everyone at the cutoff (since everyone's treatment changes)
- **Fuzzy RDD**: This is the LATE for compliers at the cutoff (those induced to take treatment by crossing z^*)

Key limitation: Very local parameter (measure-zero set of population)

RDD for reduced form vs. as an instrument

Important distinction

1. RDD for direct/reduced form effects:

- Interested in effect of treatment D itself on outcome Y
- Example: Effect of scholarship receipt on college enrollment
- Example: Effect of winning election on policy outcomes

2. RDD as an instrument:

- Treatment D is not ultimate interest; it affects some other variable S
- Use RDD assignment as instrument for S to estimate effect of S on Y
- Example: Scholarship $\rightarrow$ years of education $\rightarrow$ labor outcomes
- RDD identifies effect of years of education (not just scholarship)

Both uses can occur in sharp or fuzzy RDD!

RDD as a local randomized experiment

Alternative interpretation: Think of RDD as a randomized experiment in a neighborhood of the cutoff

- For individuals very close to $Z_i = z^*$, whether they end up just above or just below the threshold is essentially random
- Example: In a close election with $Z_i =$ vote margin, whether you get 50.01% or 49.99% is basically random
- This makes treated and untreated individuals near the cutoff comparable
- This ”local randomization” interpretation is increasingly popular
- Helps with thinking about inference and design

Outline

Sharp regression discontinuity

Definition: In a **sharp RDD**, treatment status changes *deterministically* at the cutoff

Treatment assignment is a **deterministic function** of the running variable:

$$D_i = \begin{cases} 1 & \text{if } Z_i \geq z^* \\ 0 & \text{if } Z_i < z^* \end{cases} \tag{12}$$

This means: $D^+ = 1$ and $D^- = 0$

The Wald formula simplifies:

$$\tau_{Sharp} = \frac{Y^+ - Y^-}{D^+ - D^-} = \frac{Y^+ - Y^-}{1 - 0} = Y^+ - Y^- \tag{13}$$

The treatment effect is simply the jump in outcomes!

Sharp RDD: Examples

- **Age-based eligibility**: Pension eligibility at age 65
  - Z_i = age, cutoff at 65
  - Everyone 65+ receives pension, nobody under 65 does
- **Test score cutoff**: Admission to selective program
  - Z_i = test score, cutoff at 70
  - Everyone scoring $\geq$ 70 admitted, everyone below rejected
- **Electoral threshold**: Winning an election
  - Z_i = vote share - 50%
  - Above 50%: you win; below 50%: you lose

Fuzzy regression discontinuity

Definition: In a **fuzzy RDD**, the *probability* of treatment changes discontinuously at the cutoff, but not from 0 to 1

Treatment probability jumps, but not perfectly:

$$\mathbb{P}(D_i = 1 | Z_i) = \begin{cases} P_1(Z_i) & \text{if } Z_i \geq z^* \\ P_0(Z_i) & \text{if } Z_i < z^* \end{cases} \tag{14}$$

where $P_1(z^*) > P_0(z^*)$ but the jump is not from 0 to 1

Equivalently: $0 < D^+ - D^- < 1$

This has the structure of instrumental variables:

- Eligibility: Crossing the cutoff $\mathbb{1}\{Z_i \geq z^*\}$
- Treatment: Actual treatment receipt D_i

Why does fuzzy RDD arise?

Reasons we see fuzzy rather than sharp:

Non-compliance (most common): Eligibility $\neq$ actual treatment

- Not everyone eligible takes up the program
- Some ineligible gain access through exceptions
- Example: Scholarship eligibility vs. actual receipt

When you have fuzzy RDD: Think of it as IV, where crossing z^* is an instrument for treatment D_i

Fuzzy RDD: Examples

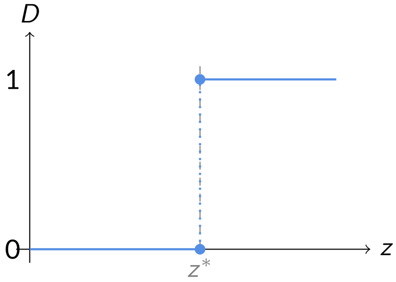
- **Scholarship eligibility**:
  - Z_i = family income, cutoff at $30,000
  - Not everyone eligible applies for or receives the scholarship
  - Some ineligible students receive aid from other sources

- **Class size rules**:
  - Z_i = enrollment, cutoff triggers new class
  - Rules may not be perfectly enforced
  - Some schools may combine classes despite the rule

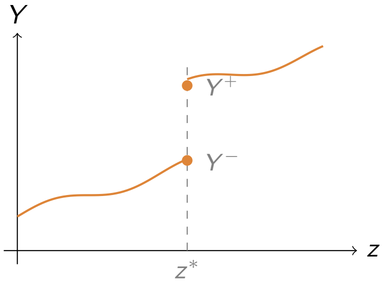
- **Medicaid eligibility**:
  - Z_i = income relative to poverty line
  - Not all eligible enroll
  - Some ineligible may be covered through other programs

Sharp RDD: Visual representation

Treatment assignment

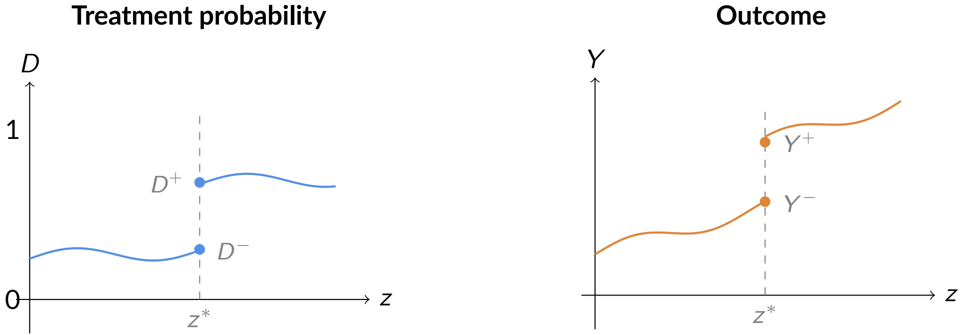


Outcome



Left: Treatment status jumps from 0 to 1 at cutoff. Right: Outcome shows discontinuous jump.

Fuzzy RDD: Visual representation



Left: Treatment probability jumps but not from 0 to 1. Right: Outcome shows discontinuous jump.

Fuzzy RDD and LATE

Since fuzzy RDD has IV structure (eligibility vs. take-up), all the IV results apply!

The Wald formula $\tau = \frac{Y^+ - Y^-}{D^+ - D^-}$ identifies a **Local Average Treatment Effect (LATE)**:

- Effect for **compliers**: those induced to take treatment by crossing z^*
- **Not** the effect for always-takers or never-takers

Compliers in fuzzy RDD:

- Would take treatment if $Z_i \geq z^*$ (eligible)
- Would not take treatment if $Z_i < z^*$ (ineligible)
- Their treatment status is *changed* by crossing the threshold

Standard IV assumptions needed: Relevance, exclusion, monotonicity

Outline

No exact empirical counterpart to the limit

Important practical issue: We want to estimate limits as $z \to z^*$, but:

- We only have a **finite sample** of observations
- We may not observe anyone *exactly* at $Z_i = z^*$
- We need to use data **away from the cutoff**

This creates a fundamental estimation challenge:

- How far from the cutoff should we use data?
- How do we approximate the conditional expectations near the cutoff?
- This is the focus of the next section!

From theory to estimation

Recall the identification formula:

$$\tau = \frac{Y^+ - Y^-}{D^+ - D^-} \tag{15}$$

where $Y^+ = \lim_{z \to (z^*)^+} \mathbb{E}[Y_i | Z_i = z]$, etc.

But: We need to estimate these limits from data

Approach: Assume $\mathbb{E}[Y_i | Z_i = z]$ can be approximated by a smooth function $g(z)$ near the cutoff

Basic regression approach

The simplest specification uses a linear function on each side:

$$Y_i = \alpha + \beta D_i + \gamma_1 Z_i + \gamma_2 D_i \cdot Z_i + \varepsilon_i \tag{16}$$

where $D_i = \mathbb{1}\{Z_i \geq z^*\}$

Interpretation:

- α: Intercept for untreated (extrapolated to $Z_i = z^*$)
- γ_1: Slope for untreated
- β: **Treatment effect** at the cutoff
- γ_2: Difference in slopes between treated and untreated

Understanding the specification

The regression can be written separately on each side:

Below the cutoff ($Z_i < z^*$, so $D_i = 0$):

$$\mathbb{E}[Y_i|Z_i = z] = \alpha + \gamma_1(z - z^*) \tag{17}$$

Above the cutoff ($Z_i \geq z^*$, so $D_i = 1$):

$$\mathbb{E}[Y_i|Z_i = z] = (\alpha + \beta) + (\gamma_1 + \gamma_2)(z - z^*) \tag{18}$$

At the cutoff ($Z_i = z^*$):

- From below: $Y^- = \alpha$
- From above: $Y^+ = \alpha + \beta$
- Difference: $\beta = Y^+ - Y^-$

Allowing for flexible functional forms

Linear may be too restrictive! We can use **polynomial specifications**:

$$Y_i = \alpha + \beta D_i + \sum_{p=1}^{P} \gamma_p Z_i^p + \sum_{p=1}^{P} \delta_p D_i \cdot Z_i^p + \varepsilon_i \tag{19}$$

Common choices:

- $P = 1$: Linear (most common)
- $P = 2$: Quadratic
- $P = 3$ or $P = 4$: Higher-order polynomials

Trade-off:

- Higher $P \rightarrow$ more flexible, fits data better
- But: risk of overfitting, especially near endpoints
- **Warning**: Gelman and Imbens (2019) show high-order polynomials can be problematic

Local linear regression

Better approach: Use data only *close to the cutoff* with linear specification

Estimate:

$$Y_i = \alpha + \beta D_i + \gamma_1 Z_i + \gamma_2 D_i \cdot Z_i + \varepsilon_i \tag{20}$$

but only using observations where $|Z_i| \leq h$ for some bandwidth h

Why is this better?

- Linear approximation is better when we're close to the cutoff
- Avoids relying on functional form assumptions far from cutoff
- More robust to misspecification

But: How do we choose the bandwidth h? (More on this soon!)

Weighted local linear regression

Possible other specification: weight observations based on distance from cutoff

Use a **kernel function** $K(u)$ that gives more weight to observations closer to the cutoff

Weight for observation i: $w_i = K\left(\frac{Z_i - z^*}{h}\right)$

Common kernels:

- **Uniform**: $K(u) = 0.5$ for $|u| \leq 1$, zero otherwise
- **Triangular**: $K(u) = 1 - |u|$ for $|u| \leq 1$
- **Epanechnikov**: $K(u) = 0.75(1 - u^2)$ for $|u| \leq 1$

In practice: Choice of kernel matters less than choice of bandwidth

Estimation with fuzzy RDD

For **fuzzy RDD**, we use instrumental variables (2SLS):

First stage (treatment on eligibility):

$$D_i = \pi_0 + \pi_1 \mathbb{1}\{Z_i \geq z^*\} + \pi_2(Z_i - z^*) + \pi_3 \mathbb{1}\{Z_i \geq z^*\} \cdot (Z_i - z^*) + \nu_i \tag{21}$$

Second stage (outcome on predicted treatment):

$$Y_i = \alpha + \beta \hat{D}_i + \gamma_1(Z_i - z^*) + \gamma_2 \mathbb{1}\{Z_i \geq z^*\} \cdot (Z_i - z^*) + \varepsilon_i \tag{22}$$

Or equivalently: Reduced form divided by first stage

$$\beta = \frac{Y^+ - Y^-}{D^+ - D^-} = \frac{\text{Reduced form effect}}{\text{First stage effect}} \tag{23}$$

Outline

The fundamental trade-off

Central challenge in RDD: choosing the bandwidth h

Narrow bandwidth (small h):

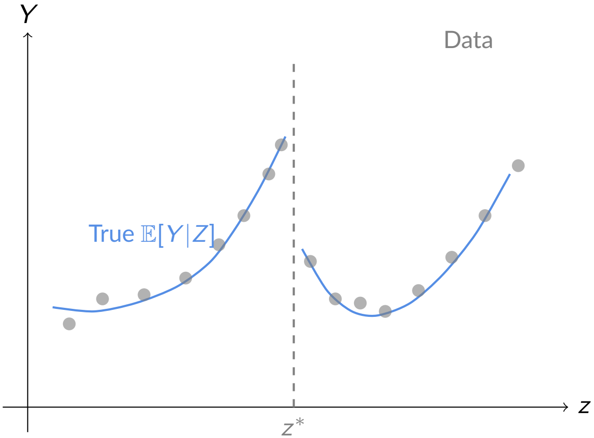
- Good: Better approximation (less bias from functional form)
- Bad: Fewer observations (more variance/noise)

Wide bandwidth (large h):

- Good: More observations (less variance)
- Bad: Worse approximation (more bias from functional form)

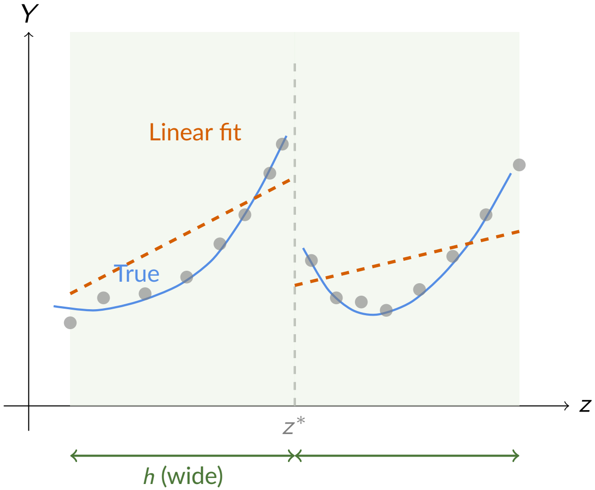
This is the classic **bias-variance trade-off**!

Visualizing the trade-off: True CEF



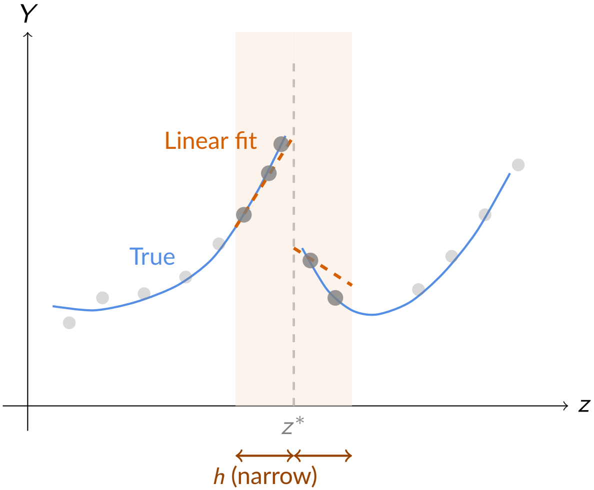
True conditional expectation function with curvature and observed data points

Wide bandwidth: Low variance, high bias



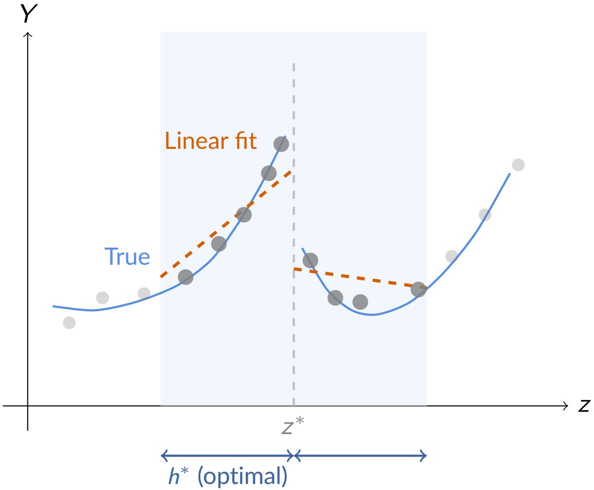
Many observations used, but poor approximation due to curvature

Narrow bandwidth: High variance, low bias



Few observations used, but good local approximation

Optimal bandwidth: Balance bias and variance



Optimal h^* balances moderate sample size with reasonable approximation

Optimal bandwidth selection

Question: Can we choose bandwidth optimally?

Yes! Modern approaches minimize mean squared error (MSE):

$$MSE(\hat{\tau}) = \text{Bias}^2(\hat{\tau}) + \text{Var}(\hat{\tau}) \tag{24}$$

Key insight:

- Bias typically grows with h^{p+1} (where p is polynomial order)
- Variance typically shrinks with $1/(nh)$ (where n is sample size)
- Optimal h balances these two

Practical bandwidth selection

In practice: Use data-driven bandwidth selection

Popular methods:

- **rdrobust** (Calonico, Cattaneo, Titiunik): Most common, includes bias correction
- **RDHonest** (Kolesar, Rothe): For discrete running variables
- Cross-validation approaches

Important: Always show **robustness** to bandwidth choice!

- Plot estimates for a range of bandwidths
- Show that results are not driven by one specific choice

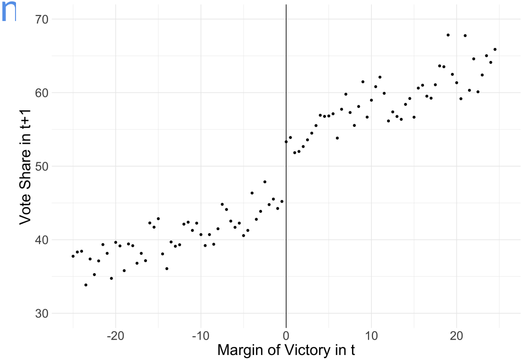
Graphical presentation: Choosing bin

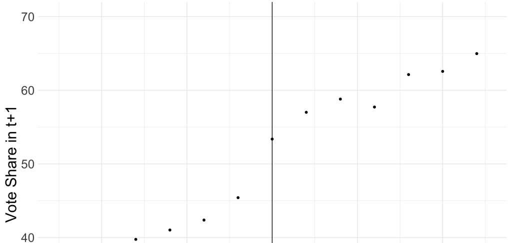
Similar trade-off for graphical presentation:

- Too many bins $\rightarrow$ noisy (hard to interpret)
- Too few bins $\rightarrow$ masks true pattern
- Need to find the right balance

Ideally: Plot binned means

- Similar intuition to binscatter
- But how do we choose bins?





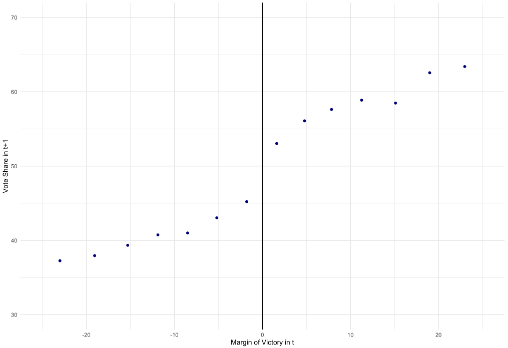
Graphical presentation: Optimal bin s

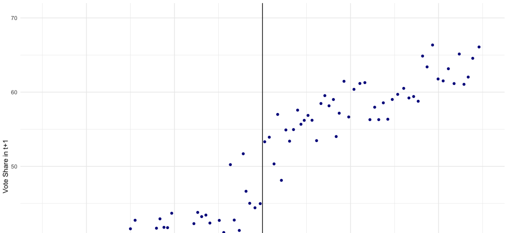
Modern approach: Choose bins optimally

Cattaneo et al. (2020) propose two approaches:

1. **IMSE-minimizing** ($\propto n^{1/3}$ bins)
   - Trades off bias vs. variance over whole range
2. **Mimicking-variance** ($\propto n/\log(n)^2$ bins)
   - Matches variance of raw data in binned plots
   - Typically more bins

Software: `rdplot` in `rdrobust` (Stata/R)





Graphical presentation: Bin placement

Other decision to make: How to place bins: equal-spaced vs. quantile

- Quantile binning more transparent
- Equal-spaced can mask underlying density

Outline

Validating the RDD

Recall: RDD relies on the assumption that potential outcomes are continuous at the cutoff

This assumption is not directly testable!

- We never observe both $Y_i(0)$ and $Y_i(1)$ for same individual
- Cannot verify continuity of counterfactuals

But: We can test **implications** of the assumption

Main threats to validity:

1. **Manipulation/sorting** of the running variable
2. **Imbalance** in covariates at cutoff
3. **Other discontinuities** at the same threshold

Threat 1: Manipulation and bunching

Problem: If individuals can manipulate their value of Z_i to cross the threshold, then treated and untreated are no longer comparable

Example (McCrary, 2008):

- Government announces income support for those earning $<$\$14,000
- Some people might underreport income to qualify
- Would see "bunching" just below the cutoff

Why is this a problem?

- Those who manipulate may be systematically different
- Example: More sophisticated, more desperate, etc.
- Violates continuity assumption

Testing for manipulation: McCrary density test

McCrary (2008) test: Test whether the *density* of the running variable is continuous at the cutoff

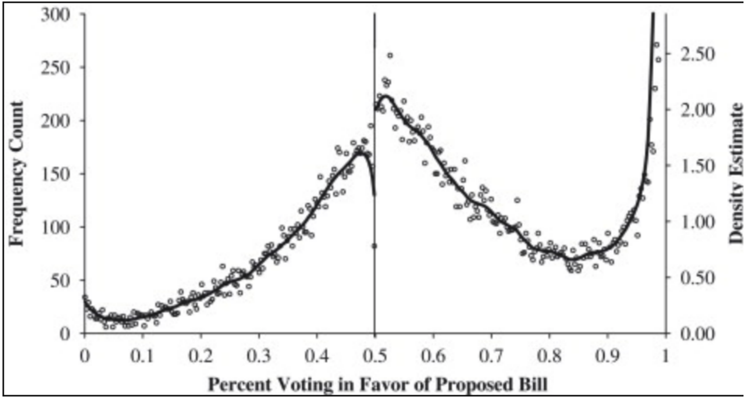
Intuition:

- If no manipulation, density should be smooth through cutoff
- Manipulation creates a ”jump” or ”hole” in the density

Implementation:

- Estimate density on each side of cutoff (using kernel methods)
- Test for discontinuity at $Z_i = z^*$
- Software: `rddensity` (Cattaneo, Jansson, Ma, 2020)

McCrary test: Example



Lee (2008): No evidence of manipulation in close elections

When manipulation happens anyway...

What if there IS bunching?

Option 1: Bounds on treatment effects

- Gerard, Rokkanen, and Rothe (2020): `rdbounds`
- Even with sorting, can bound the magnitude of effects
- Asks: how much could sorting explain the results?

Option 2: Donut-hole RDD

- Exclude observations very close to cutoff
- If manipulation only affects those right at threshold
- But: loses precision, changes estimand

Threat 2: Covariate imbalance

Covariate balance test: Check whether predetermined covariates change discontinuously at cutoff

Idea: If RDD is valid, things that were determined *before* treatment assignment shouldn't jump at the cutoff

What to test:

- Demographic characteristics (age, gender, race, etc.)
- Baseline/lagged outcomes (outcomes before treatment)
- Other predetermined variables

Implementation: Run the same RDD regression, but use covariate as outcome

$$X_i = \alpha + \beta \mathbb{1}\{Z_i \geq z^*\} + \gamma_1(Z_i - z^*) + \gamma_2 \mathbb{1}\{Z_i \geq z^*\} \cdot (Z_i - z^*) + \varepsilon_i \tag{25}$$

Test $H_0: \beta = 0$ for each covariate

Threat 3: Other discontinuities at cutoff

Problem: If *other things* besides treatment change at the cutoff, we can't isolate the treatment effect

Example: Retirement age

- At age 65: eligibility for pension, Medicare, mandatory retirement, etc.
- Hard to say which one drives any outcome changes

Solutions:

- Find a setting where only one thing changes
- Try to measure all the things that change
- Focus on outcomes that should only be affected by one channel

Additional robustness tests

Placebo cutoffs:

- Test for discontinuities at points *other than* the true cutoff
- If you find effects at fake cutoffs, something's wrong!
- Permutation tests: randomly assign cutoff location

Donut-hole specification:

- Exclude observations very close to cutoff
- Tests whether results driven by a few observations at boundary
- Also useful if worried about measurement error in Z_i

Alternative specifications:

- Different polynomial orders
- Different bandwidths
- With/without covariates

External validity: The local nature of RDD

Key limitation: RDD identifies $\tau_{RDD} = \mathbb{E}[Y_i(1) - Y_i(0)|Z_i = z^*]$ - a **very local** parameter

When is this a problem?

- Treatment effects are heterogeneous (marginal vs. inframarginal individuals)
- Policy question is about broader population, not just at cutoff
- Example: Class size effect at 41 vs. 40 may not apply to 30 vs. 20

When is it less of a concern?

- Marginal policy changes (moving the cutoff slightly)
- Effect at cutoff IS the policy-relevant parameter
- Using RDD to understand mechanisms, not for extrapolation

RDD estimation checklist

A credible RDD paper should present:

1. **Graphical evidence**: Plot outcome vs. running variable with discontinuity clearly visible (`rdplot`)
2. **Density test**: McCrary/`rddensity` test to rule out manipulation at cutoff
3. **Covariate balance**: Test for discontinuities in predetermined variables
4. **Main estimates**: Use local linear with MSE-optimal bandwidth (`rdrobust`)
   - For fuzzy RDD: report first stage (check F-stat > 10)
5. **Robustness**: Show sensitivity to bandwidth, polynomial order, covariates
6. **Placebo tests**: Check for effects at fake cutoffs, lagged outcomes, donut-hole

Outline

Summary

Key takeaways:

- RDD exploits discontinuous changes in treatment at a threshold
- **Identification**: Relies on continuity of potential outcomes at cutoff
- **Sharp vs. Fuzzy**: Deterministic vs. probabilistic treatment assignment
- **Estimation**: Local linear regression with optimal bandwidth
- **Validity**: Test for manipulation, covariate balance, other discontinuities
- **External validity**: Effects are local to the cutoff