Regression Discontinuity Design

Prof. Tzu-Ting Yang
楊子霆

Institute of Economics, Academia Sinica
中央研究院經濟研究所

May 13, 2026

Main Idea

- A major problem of estimating causal effect of treatment is the threat of **selection bias**
- In many situations, individuals can **select into treatment** so those who get treatment could be very different from those who are untreated
- The best to deal with this problem is conducting a randomized controlled trial (RCT)

Main Idea of Regression Discontinuity Design

- In an RCT, researchers can eliminate selection bias by **controlling treatment assignment process**
  - An RCT randomizes who receives a treatment –the treatment group - and who does not –the control group
  - Since we randomly assign treatment, the probability of getting treatment is unrelated to other confounding factors
- But conducting an RCT is very expensive and may have ethical issue

Main Idea of Regression Discontinuity Design

- Instead of controlling treatment assignment process, if researchers have **detailed institutional knowledge of treatment assignment process**
- Then we could use this information to create an "experiment"

Main Idea of Regression Discontinuity Design

- Regression Discontinuity Design (RDD) exploits the facts that:
  - Some rules can generate a discontinuity in treatment assignment
    - The treatment assignment is determined based on whether a unit exceeds some threshold on a variable.
    - Such variable is called **assignment variable**, **running variable** or **forcing variable**
  - Assume other factors do NOT change abruptly at threshold
  - Then any change in outcome of interest can be attributed to the assigned treatment

Main Idea of Regression Discontinuity Design

A Motivating Example

- A large number of studies have shown that graduates from more selective programs or schools earn more than others
  - In Taiwan, many students want to enter elite schools
  - Students graduated from NTU earn more than those graduated from other schools

- But it is difficult to know whether the positive earnings premium is due to
  - true "causal" impact of human capital acquired in the academic program
  - a spurious correlation linked to the fact that good students selected in these programs would have earned more no matter what
- The latter point reflects **selection bias**
- We need to untangle the **causal effect** and **selection bias**

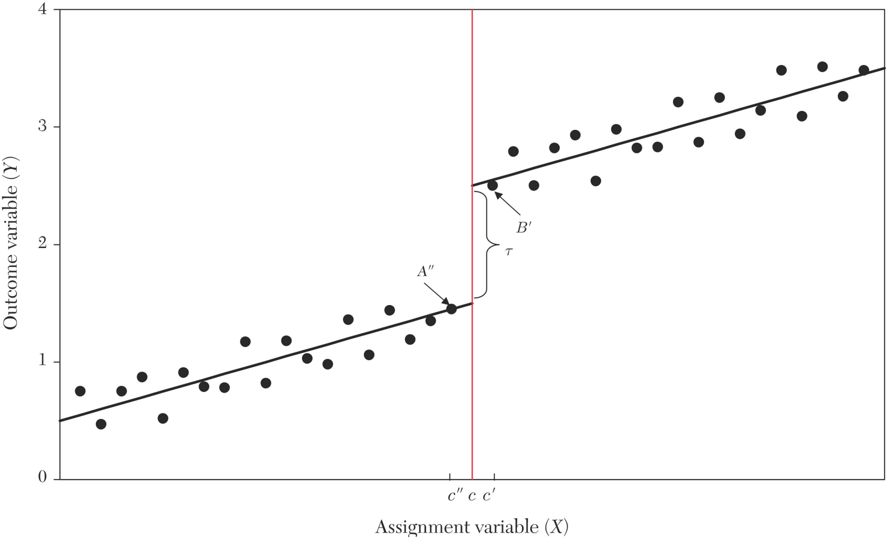
- A great way to answer that question would be to run an experiment:
  - Take students applying both to NTU and other schools
  - Instead of admitting them the regular way, just flip a coin to decide whether they get into NTU or other schools
  - Follow them up 10 years later to see whether those admitted to NTU earn more than those admitted to other schools
- Great idea, but nobody will let me run that experiment...

Main Idea of Regression Discontinuity Design

A Motivating Example

- But say that the entry cutoff for a score of entrance exam is 400 at NTU
- They would perhaps let me flip a coin for those with scores of 399 or 400
- Since the those get 399 and those get 400 are essentially identical
- They get different scores due to some random events
- **RD strategy**: I can do “as well” as in a randomized experiment by tracking down the long term outcomes for the 400 (admitted to NTU) and the 399 (admitted at other schools)

Test Score and Earnings



Source: Lee and Lemieux (2010)

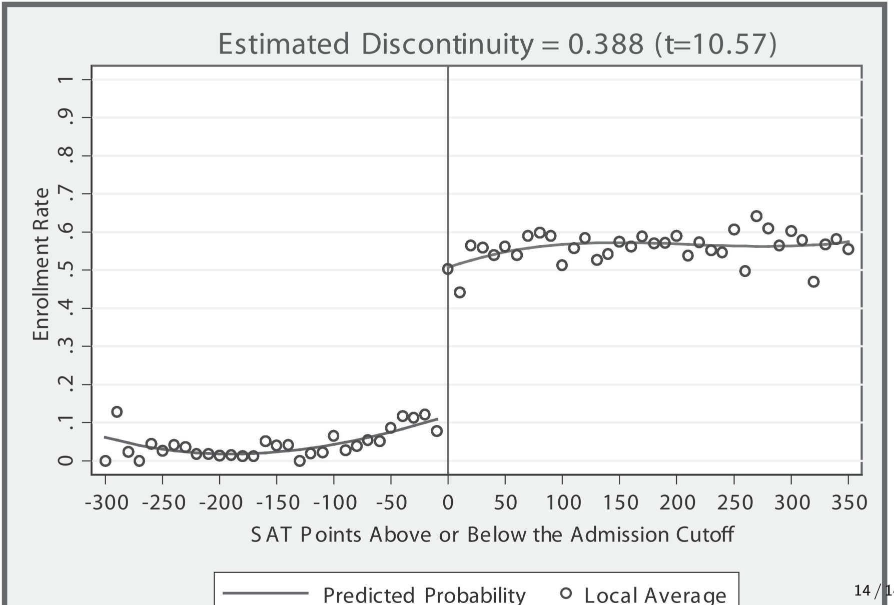
Mark Hoekstra (2009) "**The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach**" Review of Economics and Statistics

- This paper demonstrates the above RD idea by examining the economic return of attending the most selective public state university
- In the United States, most schools used SAT (or ACT) scores in their admission process
- For example, the flagship state university considered here uses a strict cutoff based on SAT score and high school GPA

- For the sake of simplicity, Hoekstra just focuses on the SAT score (adjusted depending on GPA)
- The author is then able to match (using social security numbers) students applying to the flagship university in 1986-89 to their administrative earnings data for 1998 to 2005
- As in any good RD study, pictures tell it all, so let's just focus on those

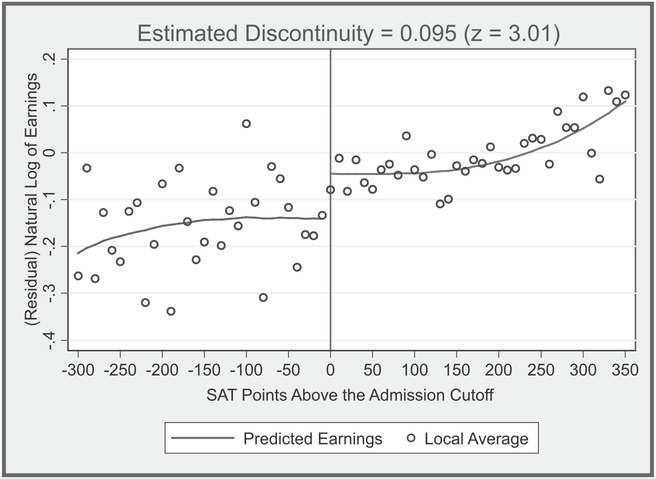
SAT Score and Enrollment

Figure 1.—Fraction Enrolled at the Flagship State University



SAT Score and Earnings

Figure 2.—Natural Log of Annual Earnings for White Men Ten to Fifteen Years after High School Graduation (Fit with a Cubic Polynomial of Adjusted SAT Score)



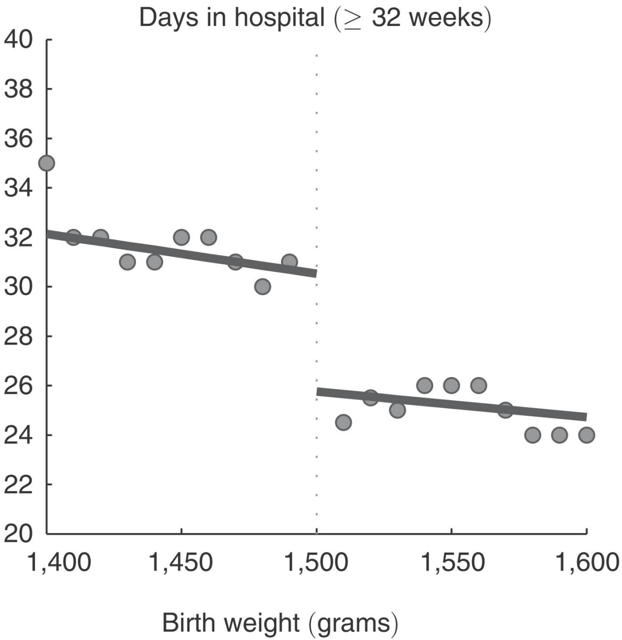
More Examples

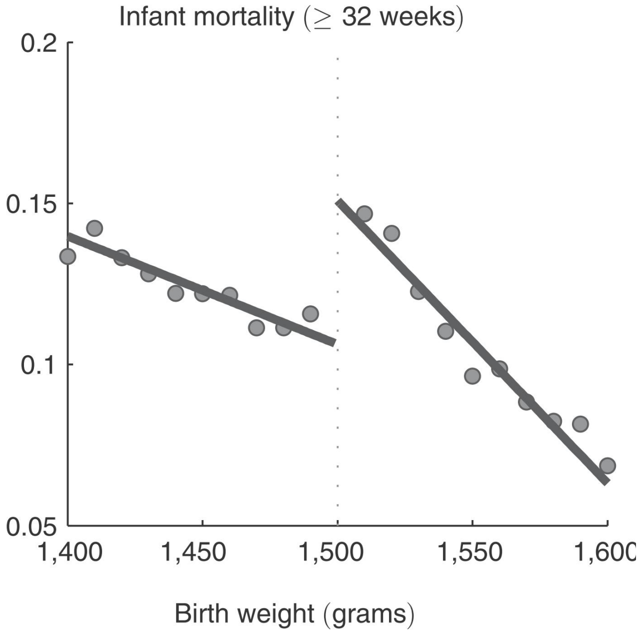
- **Where there is a cutoff there is a RD**

Prashant Bharadwaj, Katrine Vellesen Løken, and Christopher Neilson (2013) "**Early Life Health Interventions and Academic Achievement**" AER

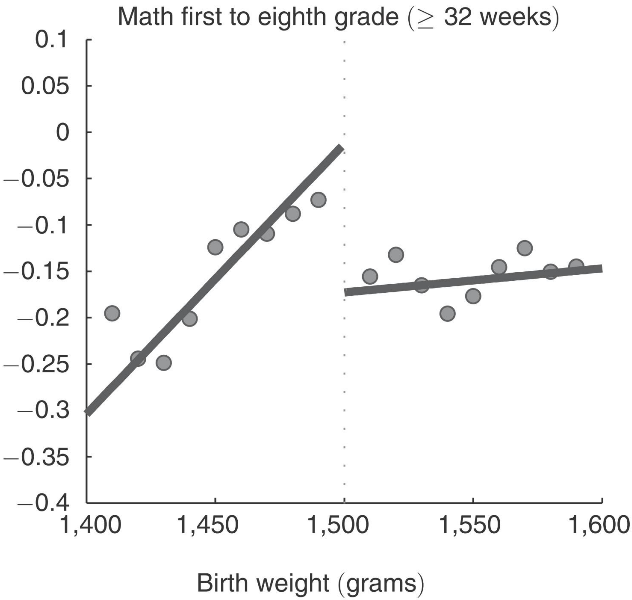
- The effect of **health intervention** in early childhood on **later life outcomes**
- **Selection bias**: children who need health intervention in early childhood could be very sick and might have bad later life outcome (e.g. low educational attachment)

- **RDD solution**:
    - Infants with a birth weight below **1500 grams** were eligible for additional healthcare while those with a birth weight just above the cutoff were not eligible
- Compares mortality rates and academic achievement between those infants **just below and above the cutoff of 1500 grams**


Days in hospital (≥ 32 weeks)
40
38
36
34
32
30
28
26
24
22
20
1,400
1,450
1,500
1,550
1,600
Birth weight (grams)



Infant mortality (≥ 32 weeks)
0.2
0.15
0.1
0.05
1,400
1,450
1,500
1,550
1,600
Birth weight (grams)



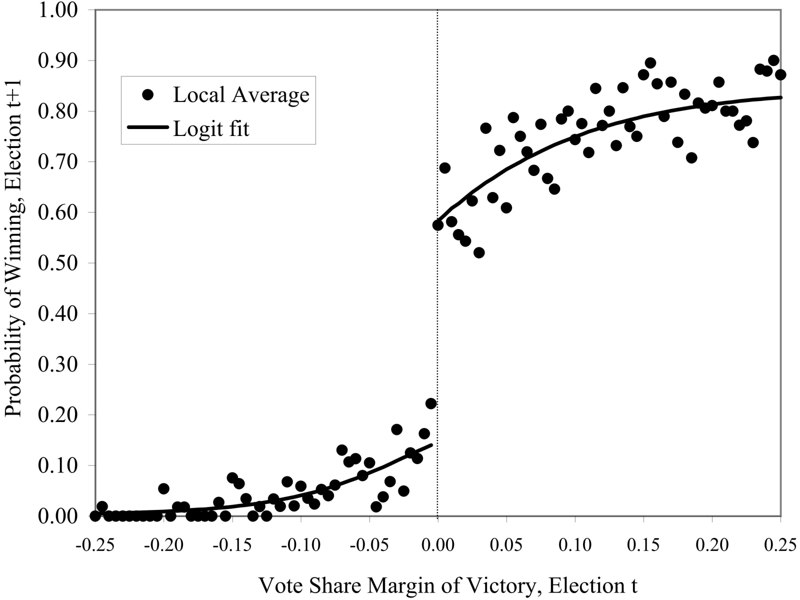
Math first to eighth grade (≥ 32 weeks)
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
−0.25
−0.3
−0.35
−0.4
1,400
1,450
1,500
1,550
1,600
Birth weight (grams)

David Lee (2007) "**Randomized Experiments from Non-random Selection in U.S. House Elections**" Journal of Econometrics

- Does **political incumbency** provide an **electoral advantage**
- **Selection bias**: people who win election (incumbency) should be more popular

- **RDD solution**:
  - Candidates who just **barely won** an election (barely became the incumbent) are likely to be ex ante comparable in all other ways to candidates who **barely lost**
  - So their differential electoral outcomes in the **next election** should represent a true **incumbency advantage**

Figure IIa: Candidate's Probability of Winning Election t+1, by Margin of Victory in Election t: local averages and parametric fit



Rafael Lalive (2008), "**How do extended benefits affect unemployment duration? A regression discontinuity approach**", Journal of Econometrics

- This paper studies a targeted program that extends the maximum duration of unemployment benefits from 30 weeks to 209 weeks in Austria

- Sharp discontinuities in treatment assignment at age 50 and at the border between eligible regions and control regions identify the effect of **extended benefits** on **unemployment duration**

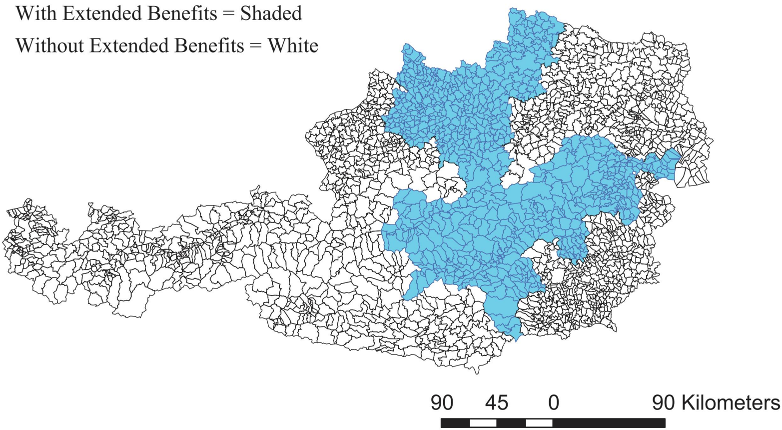
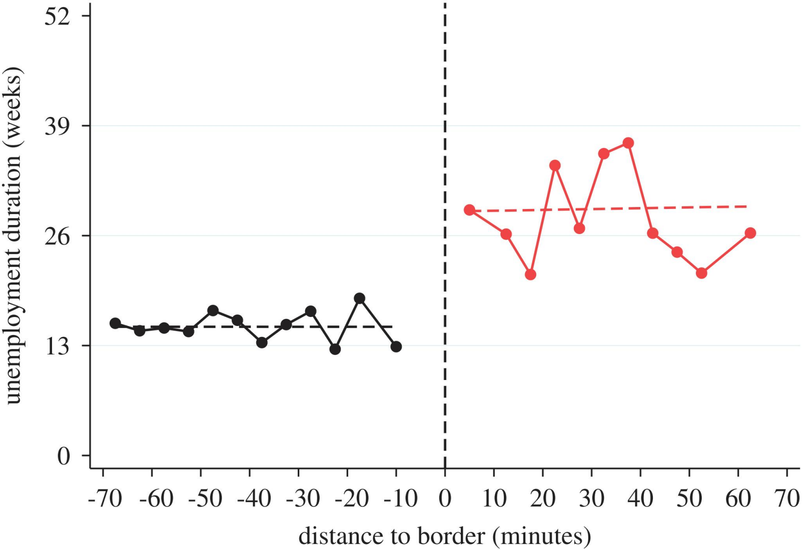


Fig. 1. Regional distribution of REBP.



Discontinuity at threshold = 13.622; with std. err. = 2.988.

RDD Became Popular since late 1990s

- The first RDD paper is Thistlethwaite and Campbell (1960), "RD Analysis: An Alternative to Ex Post Fact Experiments," Journal of Education Psychology
- RDD was not used much in economics until the late 1990s
- But hundreds of studies since then, starting with Van der Klaauw (2002)
- Two possible explanations:
  - Cutoff rules are very wide spread···
  - Much more data available now, especially administrative data sets

RDD Became Popular since late 1990s

- An important advantage of RD designs is that they are well suited to large administrative data sets with
  - Few covariates
  - Lots of observations and all the relevant information about cutoffs and assignment variables
  - Since those have to be used in the administration of programs

Sharp RDD and Fuzzy RDD

- In general, depending on enforcement of treatment assignment, RDD can be categorized into two types:
  1. Fuzzy RDD: the probability of getting the treatment jumps discontinuously at the cutoff
  2. Sharp RDD: the probability of getting the treatment jumps from 0 to 1 at the cutoff

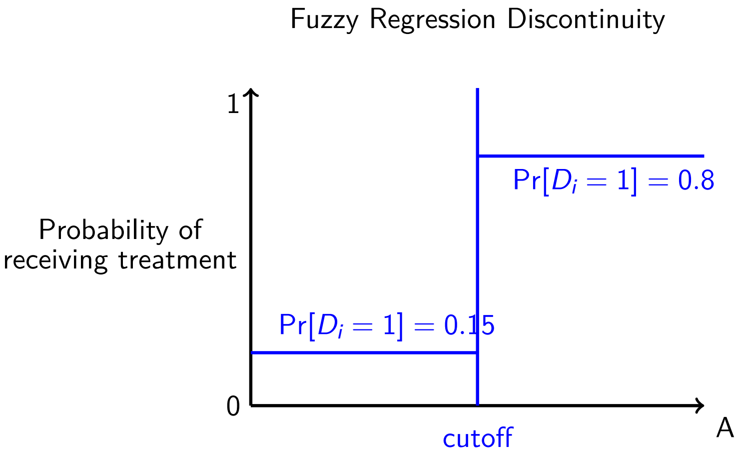
Fuzzy RDD

$$\lim_{\varepsilon \to 0} \Pr[D_i = 1 | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \Pr[D_i = 1 | A_i = c - \varepsilon] \neq 0$$

- The probability of getting the treatment jumps discontinuously at the cutoff
  - Some individuals above cutoff do NOT get treatment and some individuals below cutoff do receive treatment

Treatment Probability and assignment variable

Fuzzy RDD



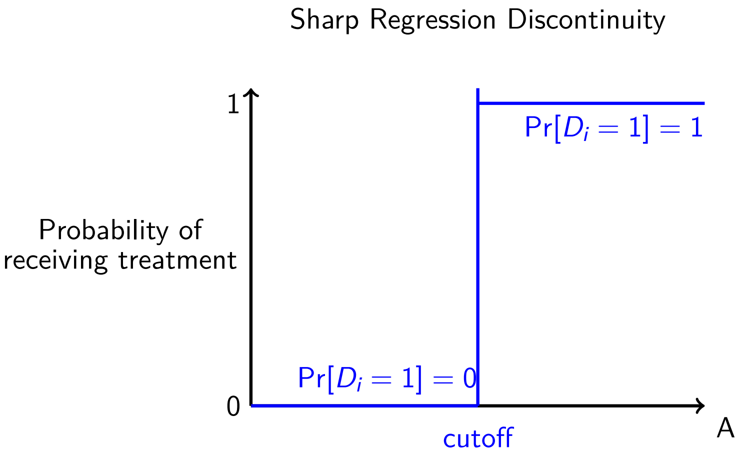
Sharp RDD

$$\lim_{\varepsilon \to 0} \Pr[D_i = 1 | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \Pr[D_i = 1 | A_i = c - \varepsilon] = 1$$

- Sharp RDD is a special case of fuzzy RDD
- It further required the jump in probability to be from 0 to 1
  - Nobody below the cutoff gets the “treatment”, everybody above the cutoff gets it

Treatment Probability and assignment variable

Sharp RDD



Identification

Sharp RDD

Christopher Carpenter and Carlos Dobkin (2009) "**The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age**" AEJ:Applied Economics

- We first discuss sharp RDD, which is less complicated and then move to more general design – fuzzy RDD
- This paper examine the effect of **legal access to alcohol** on **mortality** using a sharp RDD
  - I will use this paper as an example to go through the key concept of sharp RDD

Potential Outcomes Framework

- Assignment variable
  - Assignment variable (running variable): $A_i \in \mathbb{R}$
  - Threshold (cutoff) for treatment assignment: $c \in \mathbb{R}$

Potential Outcomes Framework

- Treatment Eligibility

$$Z_i = \begin{cases} 1 & \text{if } A_i \geq c, \text{ eligible for a treatment} \\ 0 & \text{if } A_i < c, \text{ not eligible for a treatment} \end{cases}$$

Potential Outcomes Framework

- In sharp RDD, treatment assignment is a deterministic function of the assignment variable A_i and the threshold c
- D_i: a dummy that indicate whether individual i receive treatment or not
- Treatment assignment

$$D_i = 1\{A_i \geq c\} \quad \forall i$$

$$D_i = \begin{cases} D_i = 1 & \text{if } A_i \geq c \\ D_i = 0 & \text{if } A_i < c \end{cases}$$

Potential Outcomes Framework

- In sharp RDD, the **eligible for a treatment** Z_i is the same as **getting a treatment** D_i
  - $Z_i = D_i$
  - We do not need to distinguish treatment eligibility Z_i and treatment assignment D_i
- In fuzzy RDD, the **eligible for a treatment** Z_i does NOT necessarily represent the **getting a treatment** D_i
  - $Z_i \neq D_i$
  - We need to estimate how treatment eligibility Z_i affects treatment assignment D_i (first-stage relationship)

Potential Outcomes Framework

- Potential Outcomes
  - Y_i^1: Potential outcome for an individual i if he would receive treatment
  - Y_i^0: Potential outcome for an individual i if he would not receive treatment
- Observed Outcomes
  - Observed outcomes Y_i are realized as:

$$\mathrm{Y}_i = \mathrm{Y}_i^1 D_i + \mathrm{Y}_i^0 (1 - D_i)$$

$$\mathrm{Y}_i = \begin{cases} \mathrm{Y}_i^1 & \text{if } D_i = 1 \ (A_i \geq c) \\ \mathrm{Y}_i^0 & \text{if } D_i = 0 \ (A_i < c) \end{cases}$$

Identification Results for Sharp RDD

- Ideally, for each individual *i*, if we could observe two potential outcomes at the same time, we can estimate **average treatment effect (ATE)**:

$$\alpha_{\mathsf{ATE}} = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

- But it is **impossible** to observe two potential outcomes at the same time

Identification Results for Sharp RDD

- Instead, we can use sharp RDD to investigate the behavior of the outcome around the threshold:

$$\alpha_{\mathsf{SRD}} = \lim_{\varepsilon \to 0} \mathrm{E}[\mathrm{Y}_i | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[\mathrm{Y}_i | A_i = c - \varepsilon]$$

- Under certain assumptions, this quantity identifies the **ATE at the threshold**:

$$\alpha_{\text{ATE at c}} = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | A_i = c]$$

Deterministic Assumption

$$D_i = 1\{A_i \geq c\} \quad \forall i$$

- Treatment assignment is a deterministic function of the assignment variable A_i and the threshold c
  - An individual's age is above 21st $\rightarrow$ get legal access to alcohol
  - An individual's age is below 21st $\rightarrow$ not get legal access to alcohol
- At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see non-treated units:

$$\Pr(D_i = 1 | A_i = c) = 1$$
$$\Pr(D_i = 1 | A_i = c - \varepsilon) = 0$$

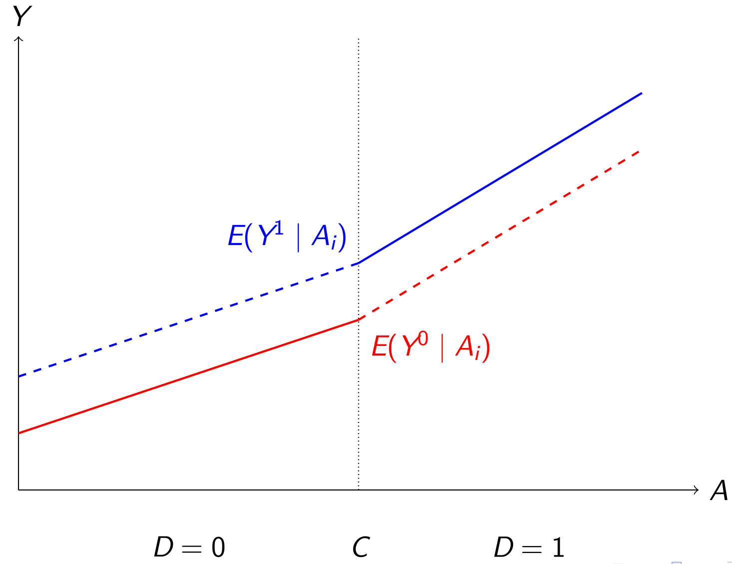
Continuity Assumption

$E[\mathrm{Y}_i^1|A_i]$ and $E[\mathrm{Y}_i^0|A_i]$ are continuous at $A_i = c$

- Assume potential outcomes do NOT change at cutoff c
  - This means that except **treatment assignment**, all other unobserved determinants of Y_i are continuous at cutoff c
  - This implies no other confounding factor affects outcomes at cutoff c
- **Any observed discontinuity in the outcome can be attributed to treatment assignment**

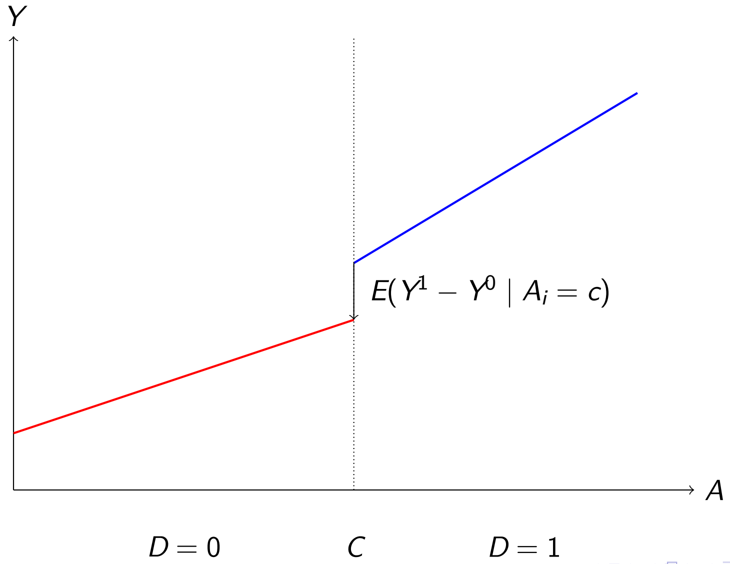
Identification Results for Sharp RDD

Graphical Interpretation



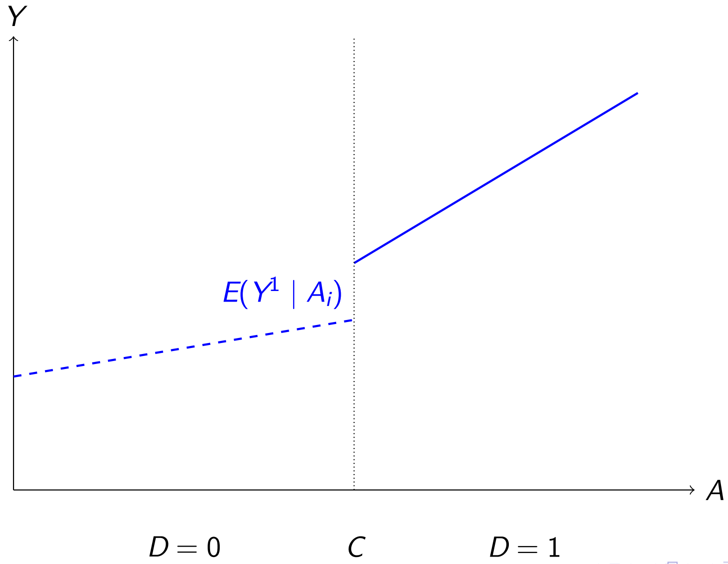
Identification Results for Sharp RDD

Graphical Interpretation



Identification Results for Sharp RDD

Graphical Interpretation



- Remember we want to identify the **ATE at the threshold**:

$$\begin{aligned}\alpha_{\text{ATE at c}} &= \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | A_i = c] \\ &= \mathrm{E}[\mathrm{Y}_i^1 | A_i = c] - \mathrm{E}[\mathrm{Y}_i^0 | A_i = c]\end{aligned}$$

- But we don't observe $\mathrm{E}[\mathrm{Y}_i^0 | A_i = c]$ ever due to the design, so we're going to extrapolate from $\mathrm{E}[\mathrm{Y}_i | A_i = c - \varepsilon]$.
- We want to construct counterfactural of $\mathrm{E}[\mathrm{Y}_i^0 | A_i = c]$ using observed data:

$$\mathrm{E}[\mathrm{Y}_i^0 | A_i = c] = \lim_{\varepsilon \to 0} \mathrm{E}[\mathrm{Y}_i | A_i = c - \varepsilon]$$

- The continuity assumption and deterministic assumption imply the following:

$$\begin{aligned} \mathrm{E}[\mathrm{Y}_i^0|A_i = c] &= \lim_{\varepsilon\to 0} \mathrm{E}[\mathrm{Y}_i^0|A_i = c - \varepsilon] && \text{(Conti)} \\ &= \lim_{\varepsilon\to 0} \mathrm{E}[\mathrm{Y}_i^0|D_i = 0, A_i = c - \varepsilon] && \text{(Deter)} \\ &= \lim_{\varepsilon\to 0} \mathrm{E}[\mathrm{Y}_i|A_i = c - \varepsilon] \end{aligned}$$

- Note that this is the same for the treated group:

$$\mathrm{E}[\mathrm{Y}_i^1|A_i = c] = \lim_{\varepsilon\to 0} \mathrm{E}[\mathrm{Y}_i|A_i = c + \varepsilon]$$

- This allows us to use average outcomes of units just below the cutoff as a valid counterfactual for units right above the cutoff

Identification Results for Sharp RDD

- The treatment effect is identified at the threshold as:

Identification Results for Sharp RDD

$$\begin{aligned}\alpha_{\text{ATE at c}} &= E[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | A_i = c] \\ &= E[\mathrm{Y}_i^1 | A_i = c] - E[\mathrm{Y}_i^0 | A_i = c] \\ &= \lim_{\varepsilon \to 0} \mathrm{E}[\mathrm{Y}_i | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[\mathrm{Y}_i | A_i = c - \varepsilon] = \alpha_{\text{SRD}}\end{aligned}$$

- Under the above assumptions, we can identify the **ATE at the threshold (unobserved)** using **sharp RD estimates (observed)**

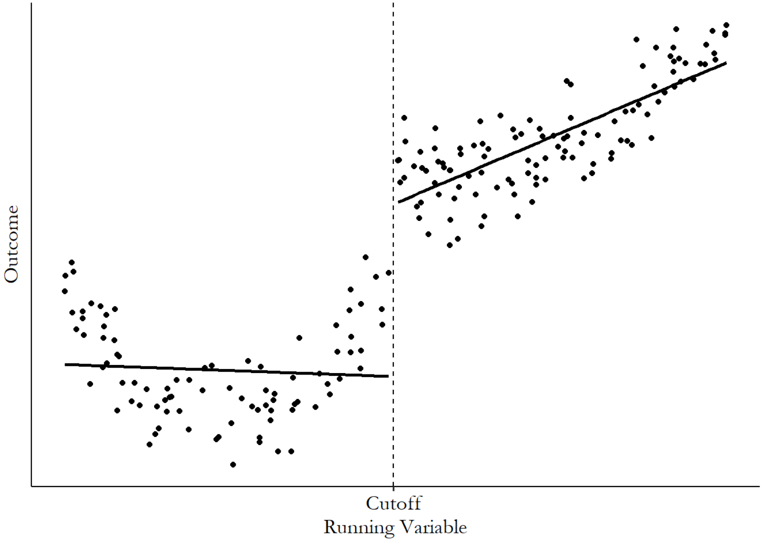
Estimation

Sharp RDD Estimation

Overview

- There are two strategies for getting RD estimates:
  1. Parametric/global method:
     - ★ Use all available observations
     - ★ Estimate treatment effects based on a **specific functional form** for the outcome and assignment variable relationship
  2. Nonparametric/local method:
     - ★ Use the observations around cutoff
     - ★ Compare the outcome of treated and untreated observations that lie **within specific bandwidth**

Estimate Discontinuity in Outcome



Source: Nick Huntington-Klein, The Effect: An Introduction to Research Design and Causality, Chapter 20

- To **estimate the discontinuity at cutoff**, we need to model the relationship between assignment variable A and outcome Y
- Suppose that potential outcomes can be described by some reasonably smooth function $\mathrm{f}(A_i)$:

$$E[\mathrm{Y}_i^0|A_i] = \beta + \mathrm{f}(A_i)$$
$$\mathrm{Y}_i^1 = \mathrm{Y}_i^0 + \alpha$$

- We can get RD estimates by fitting:

$$Y_i = \beta + \alpha D_i + \mathrm{f}(A_i) + \eta_i$$

Sharp RDD Estimation

Parametric/Global Approach

- Assume $\mathrm{f}(A_i)$ is a linear function of A_i
- We usually do the following two settings:
  1. Allow the A_i terms to differ on both sides of the threshold
     - ★ Include A_i both individually and interacting them with D_i
     - ★ By doing this, we can estimate $\mathrm{f}(A_i)$ on each side
  2. Re-center A_i at c:
     - ★ This step ensures that the treatment effect at $A_i = c$ is the coefficient on D_i in a regression model

Sharp RDD Estimation

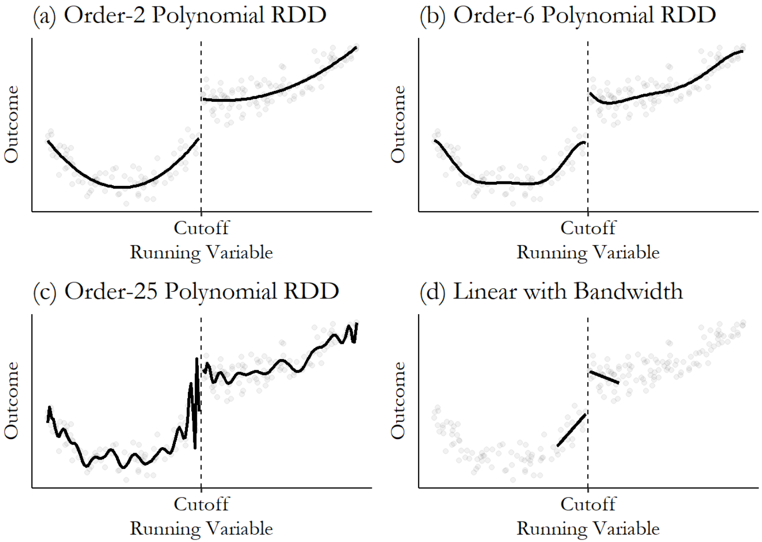
Parametric/Global Approach

- Therefore, we estimate the following regression model:

$$Y_i = \beta + \alpha D_i + \gamma_1(A_i - c) + \gamma_2(A_i - c) \times D_i + \eta_i$$

  - The intercept at untreated side around the cutoff: β
  - The intercept at treated side around the cutoff: $\beta + \alpha$
  - The discontinuity in outcome at cutoff: $(\beta + \alpha)$ - $\beta = \alpha$
    - ★ $\alpha = \lim_{\varepsilon \to 0} \mathrm{E}[\mathrm{Y}_i | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[\mathrm{Y}_i | A_i = c - \varepsilon]$

More Flexible Function



Source: Nick Huntington-Klein, The Effect: An Introduction to Research Design and Causality, Chapter 20

Sharp RDD Estimation

Parametric/Global Approach

- We can use more flexible function $\mathrm{f}(A_i)$ to capture the relationship between assignment variable A and outcome Y
  - For example, use second-order polynomial of A
  - However, more flexible function $\mathrm{f}(A_i)$ might not give better estimate of discontinuity
- Gelman and Imbens (2019):
  - It's not a great idea to go above the second-order term when performing RDD
  - If there's a complex shape that needs fitting, instead try limiting the range of the data with a bandwidth and use a simpler function

- The core idea of RDD is to compare outcomes just above and just below the cutoff c.
- Nonparametric/local method:
  - We do **not** assume a specific global functional form for the relationship between the outcome Y and the assignment variable A over their entire range.
    - ★ This is what "nonparametric" means in this context.
  - Instead, we focus the analysis **locally**, using only observations close to the cutoff c.
  - We compare treated and untreated observations that fall within a specific **bandwidth (h)** around the cutoff.

Sharp RDD Estimation

Nonparametric/Local Approach

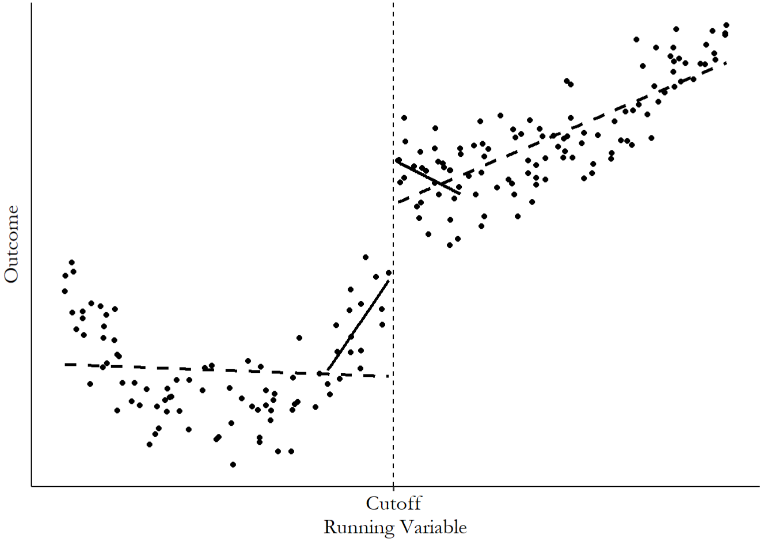
- Key insight: within a **sufficiently small local window (bandwidth h)** around the cutoff c, even if the true relationship is complex globally.
  - It can often be well **approximated by a simple linear function**.
- Therefore, we use **Local linear regression**:
  - This involves fitting separate linear regressions on each side of the cutoff, but *only* using data within the bandwidth h.
  - Optionally, kernel functions can be used to give more weight to observations closer to the cutoff.

- Specifically, we estimate the following regression model **using only observations within the chosen bandwidth** h:

$$Y_i = \beta + \alpha D_i + \gamma_1 (A_i - c) + \gamma_2 (A_i - c) \times D_i + \eta_i$$

- This linear model is just a **local approximation** of the relationship near the cutoff c, within the bandwidth h.
  - It is *not* a global assumption about the functional form.
- The estimated coefficient $\hat{\alpha}$ provides the RDD estimate of the treatment effect.

Bandwidth



Source: Nick Huntington-Klein, The Effect: An Introduction to Research Design and Causality, Chapter 20

- The main challenge of nonparametric approach is to **choose a bandwidth**
- There is essentially a trade-off between **bias** and **precision** (efficiency)
- Use a larger bandwidth:
  - Since more data points are used in the regression
  - Get more **precise** treatment effect estimates
  - But use data points far from cutoff
  - The estimated treatment effect could be **biased**

Sharp RDD Estimation

How to Choose Bandwidth

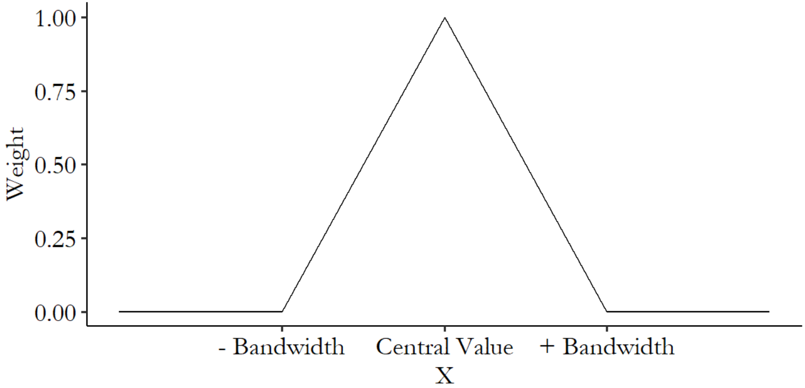
1. **Cross-Validation (CV) Procedure**:
   - Aims to select h that minimizes prediction errors of the RDD model specifically in the neighborhood of the cutoff c.

2. **Plug-In Procedure**:
   - Derives a formula for the optimal bandwidth (h_{opt}) by analytically minimizing an (Asymptotic) Mean Squared Error for the RDD treatment effect estimator (e.g., $\hat{\alpha}$)
     - ⋆ The h_{opt} formula depends on some unknown data characteristics at the cutoff c (like local trends and data variability).
     - ⋆ These must first be estimated from your data, often using a "pilot" bandwidth.
   - See, e.g., Imbens and Kalyanaraman (2012), Calonico, Cattaneo, and Titiunik (2014).

- In practice, we also use a specific kernel function to weight observations more heavily around cutoff
- A very commonly-used kernel in regression discontinuity is the **triangular kernel**

$$K(A) = \begin{cases} 1 - \frac{|A_i - c|}{h} & \text{for } c - h < A_i < c + h, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Triangular Kernel Function



Source: Nick Huntington-Klein, The Effect: An Introduction to Research Design and Causality, Chapter 20

- In alcohol example, we can estimate the following regression **within 180 days before and after age 21**:

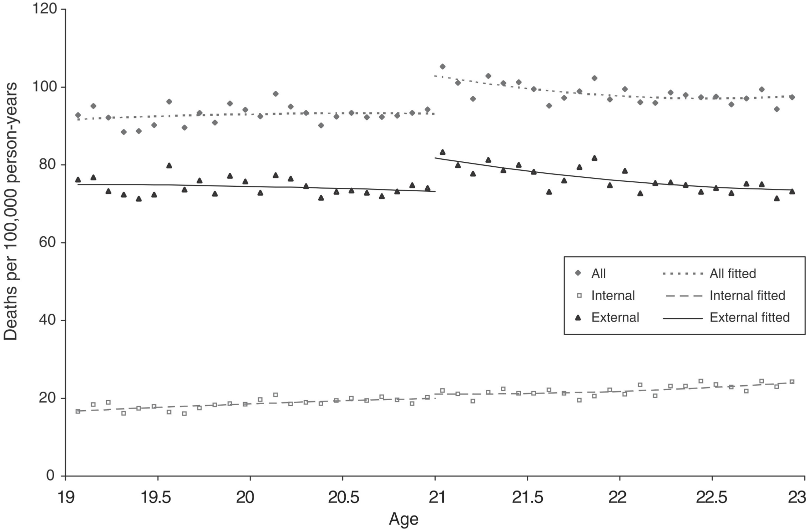
$$Y_i = \beta + \alpha D_i + \gamma_1 (A_i - 21) + \gamma_2 D_i (A_i - 21) + \eta_i$$

  - $h = 180$
  - Use a triangular kernel as a weight and implement a weighted OLS
  - The effect of legal access to alcohol on mortality rate at age 21 is α
  - Usually, we would present the RD estimates by different choices of bandwidth

TABLE 4—DISCONTINUITY IN LOG DEATHS AT AGE 21

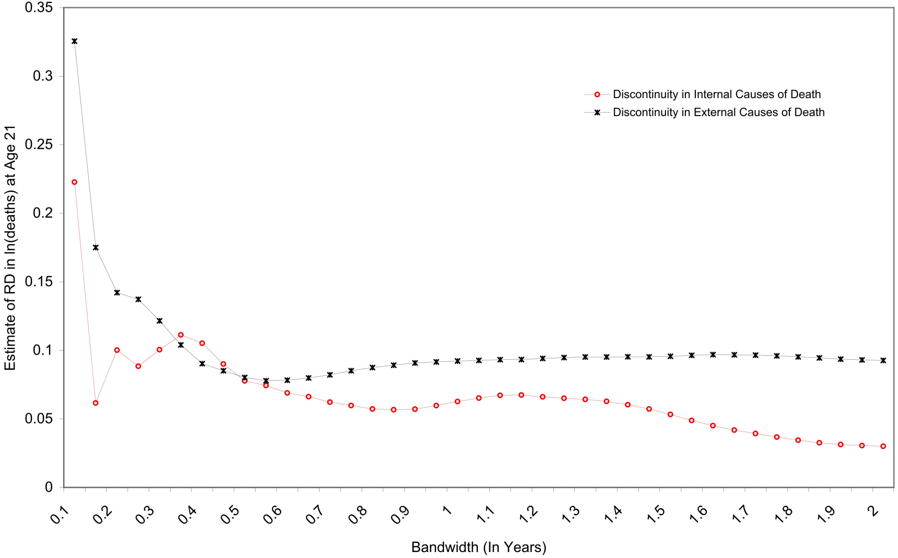
	(1)	(2)	(3)	(4)
Deaths due to all causes				
Over 21	0.096	0.087	0.091	0.074
	(0.018)	(0.017)	(0.023)	(0.016)
Observations	1,460	1,460	1,460	1,458
R^2	0.04	0.05	0.05	
Prob > Chi-Squared		0.000	0.735	
Deaths due to external causes				
Over 21	0.110	0.100	0.096	0.082
	(0.022)	(0.021)	(0.028)	(0.021)
Observations	1,460	1,460	1,460	1,458
R^2	0.06	0.08	0.08	
Prob > Chi-Squared		0.000	0.788	
Deaths due to internal causes				
Over 21	0.063	0.054	0.094	0.066
	(0.040)	(0.040)	(0.053)	(0.031)
Observations	1,460	1,460	1,460	1,458
R^2	0.10	0.10	0.10	
Prob > Chi-Squared		0.000	0.525	
Covariates	N	Y	Y	N
Quadratic terms	Y	Y	Y	N
Cubic terms	N	N	Y	N
LLR	N	N	N	Y

Notes: See Notes from Table 1. The dependent variable is the log of the number of deaths that occurred *x* days from the person's twenty-first birthday. External deaths include all deaths with mention of an injury, alcohol use, or drug use. The Internal Death category includes all deaths Nt coded as external. Please see Web Appendix C for the ICD codes for each of the categories above. The first three columns give the estimates from polyNmial regressions on age interacted with a dummy for being over 21.



Deaths per 100,000 person-years
120
100
80
60
40
20
0
19
19.5
20
20.5
21
21.5
22
22.5
23
Age
All
Internal
External
All fitted
Internal fitted
External fitted

Appendix J: Nonparametric Estimates of Discontinuity in Internal and External Deaths with Different Bandwidths



Examine Validity of Identification Assumptions

1. Examine “Discontinuity” in Nonoutcome Variables

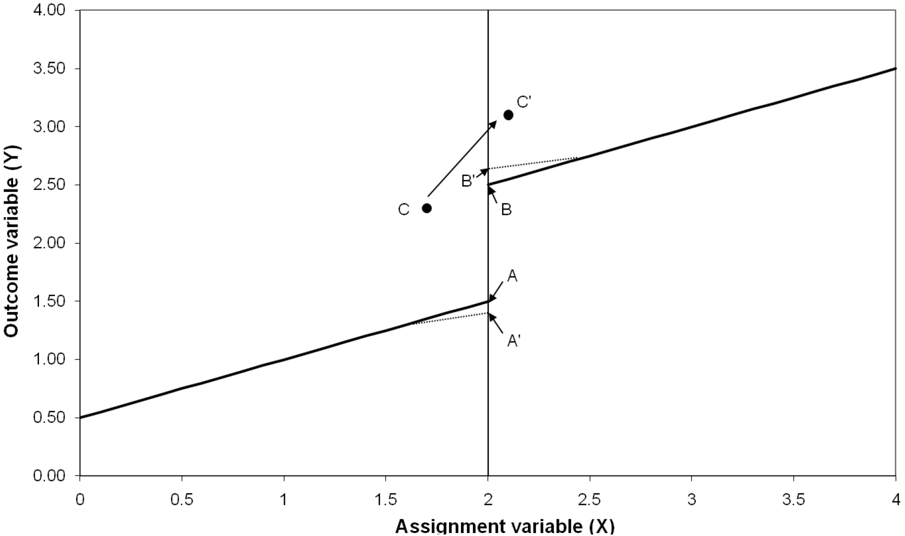
- Construct a similar graph to the one before but using a covariate as the “outcome”
- There should be **NO jump in other covariates**
- If the covariates would jump at the cutoff one would doubt the identifying assumption

2. Examine "Discontinuity" in Density of the assignment variable

- Individuals may invalidate the **continuity assumption** if they strategically manipulate assignment variable A to be just above or below the cutoff
- That is, people just above and just below the cutoff are no longer comparable

Consequence of Sorting Behavior

Example



Sorting Behavior

Example

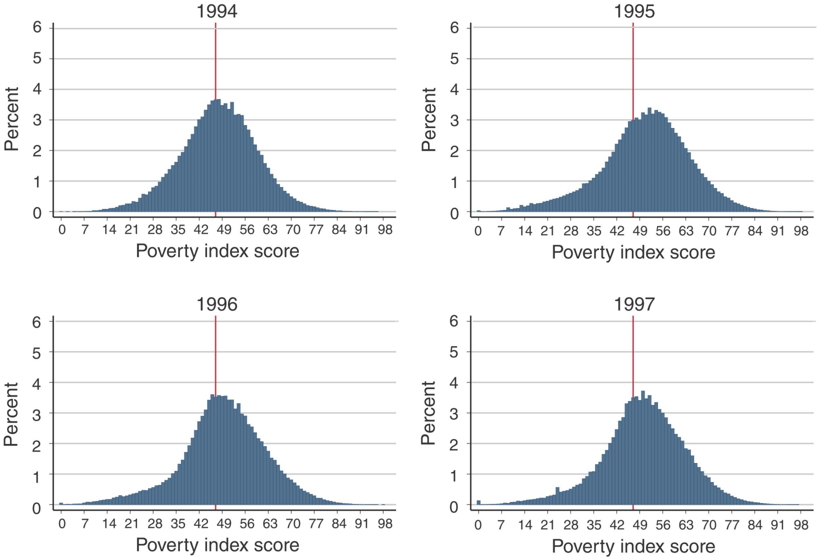
- This is a concern especially if the exact value of the cutoff is known to the individuals in advance
  - Such sorting behavior may create a **discontinuity in the distribution of A at the cutoff**
  - That is, "bunching" to the right or to the left of the cutoff

Adriana Camacho and Emily Conover (2011) "**Manipulation of Social Program Eligibility**" AEJ: Economic Policy

- Manipulation of a poverty index in Colombia
  - A poverty index is used to decide eligibility for social programs
- The algorithm to create the poverty index **becomes public** during the second half of 1997

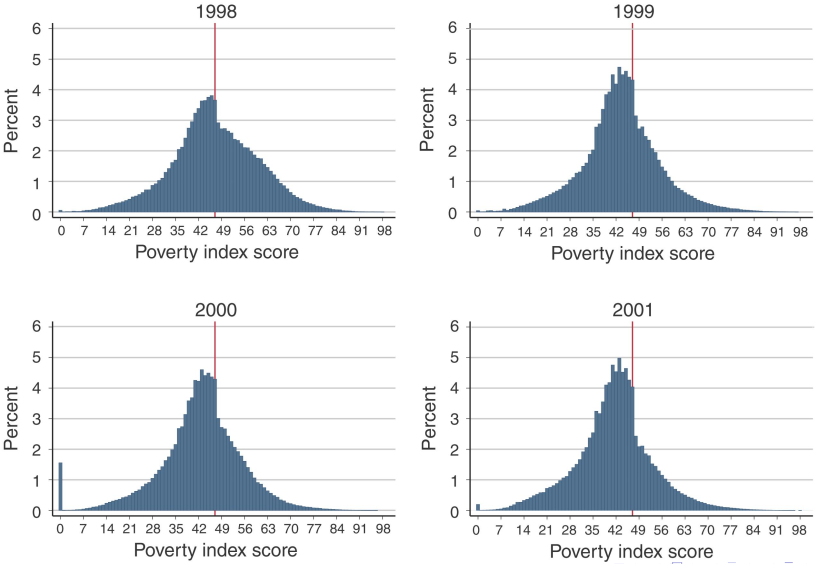
Sorting Behavior

Example



Sorting Behavior

Example



How to Examine "Discontinuity" in Density of the assignment variable

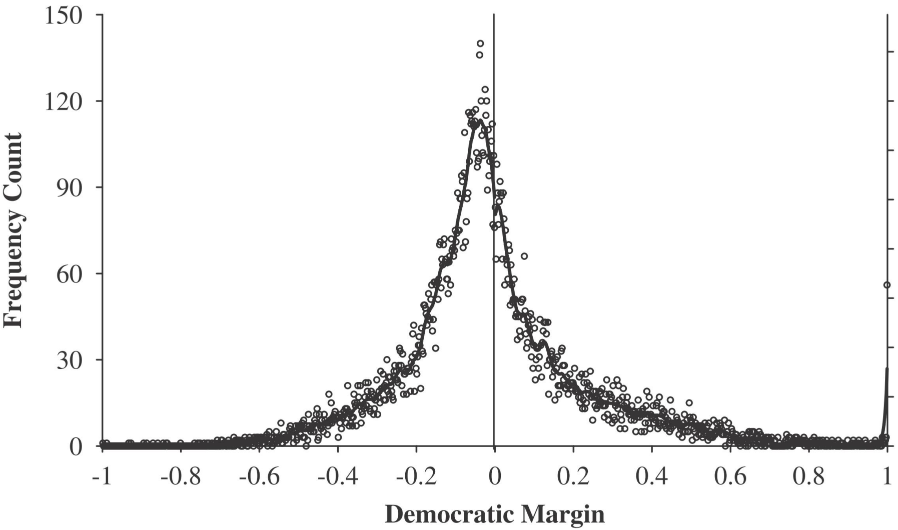
- Plot the number of observations in each bin of assignment variable
- Investigate whether there is a **discontinuity in the distribution of the assignment variable** at the threshold
  - A discontinuity in the density suggests that people might manipulate the assignment variable around the threshold

Test Internal Validity of RDD

Examine Density of the assignment variable

- A formal test is provided by McCrary (2008)
    1. Partition the assignment variable into bins and calculate frequencies (number of observations) in each bins
    2. Calculate frequencies (number of observations) in each bin
    3. Ensure that no bin overlaps the cutoff
    4. **Run two local linear regressions**, one to the right and one to the left of the cutoff
    5. In these regressions, the bin midpoints are the regressors and **number of observations is outcome**
    6. Test whether log difference of the intercepts of the two regressions is statistically different from zero

McCrary Test



STATA Example

David S. Lee, Enrico Moretti, and Matthew J. Butler (2004) "**Do Voters Affect or Elect Policies? Evidence from the U.S. House**" QJE

- They test two different theories of the role of elections in policy formation
  1. Voters can affect candidates' policy choices: competition for votes induces politicians to move toward the center
  2. Voters merely elect policies: politicians cannot make credible promises to moderate their policies
- We will use a part of their paper to implement sharp RDD using STATA
- Examine the the effect of **Democratic membership** on **congressman's voting behavior**

- This study uses a Sharp RDD to test whether voters primarily "affect policies" or "elect policies."
- The strategy compares outcomes in districts where the Democrat candidate barely won (e.g., 50.5% of the vote) with districts where the Democrat candidate barely lost (e.g., 49.5% of the vote).
- **Key Idea**:
  - In such extremely close elections, the winner is determined by factors that are *as if random* right around the 50% vote share cutoff

- **RDD Core Logic**:
  - "As-if random" assignment near the cutoff means districts where Democrats **barely won** vs. **barely lost** should, on average, have similar underlying characteristics (e.g., voter preferences).
- Any systematic differences in the voting behavior of representatives elected from these comparable districts can be primarily attributed to:
  - The **effect of the winning candidate's party affiliation**.
- This helps to clarify the "affect policies" vs. "elect policies" debate.

$$Y_i = \alpha + \rho D_i + \mathrm{f}(A_i) + \eta_i$$

$$D_i = \begin{cases} D_i = 1 & \text{if } A_i \geq c \\ D_i = 0 & \text{if } A_i < c \end{cases}$$

- Y_i (outcome): a liberal voting score from the Americans for Democratic Action (ADA)
- A_i (assignment variable): Democratic vote share
- D_i (treatment variable): If $A_i > 0.5$ then Democratic candidate is elected

- See RDD-LMB-data.do
- Use RDD-LMB-data.dta
- Install the following ado files:
  - binscatter.ado
  - rdrobust.ado
  - cmogram.ado
  - DCdensity.ado

- Plot **outcome** (ADA score) by **assignment variable** (Democrat vote share)
  - This is the standard graph showing the discontinuity in the outcome variable
  - Construct bins and average the outcome within bins on both sides of the cutoff
  - You may also want to plot a relatively flexible regression line on top of the bin means
  - Inspect whether there is a discontinuity at cutoff (0.5)

STATA Command: binscatter

- Syntax:

```
binscatter varlist [if] [in] [weight] [, options]
```

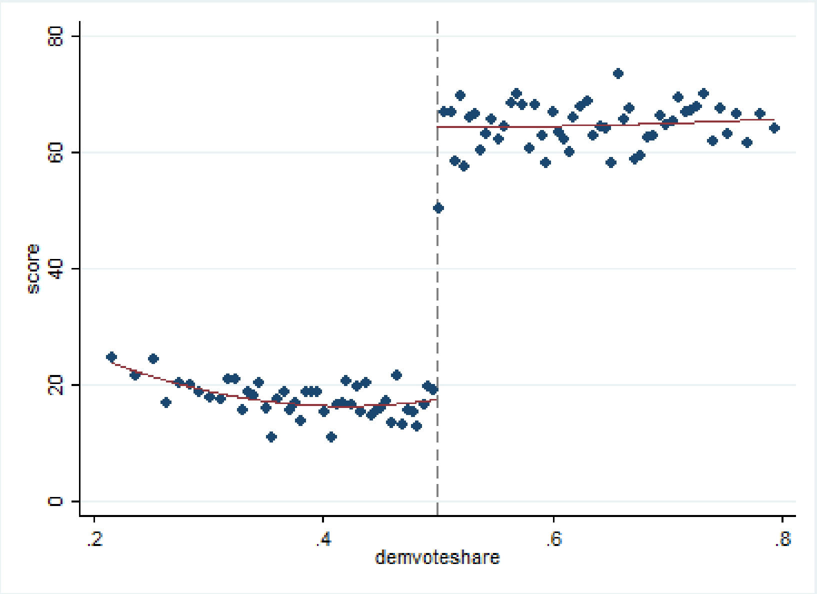
- Example:

```
binscatter score demvoteshare if demvoteshare>=0.2 &
    demvoteshare<=0.8, n(100) rd(0.5) linetype(qfit)
graph export f1.png, replace
```

- **varlist**: the list of outcome and assignment variable
- **n()**: number of equal-sized bins to be created
- **rd()**: the cutoff in the assignment variable
- **linetype(linetype)**: type of fit line
- **graph export**: save your graph

Outcome by assignment variable

ADA score by Democrat vote share



Step 1: Graphical Analysis

- Construct a similar graph but using a **covariate** as the "outcome"
- There should be no jump in other covariates
- If the covariates would jump at the discontinuity one would doubt the continuity assumption

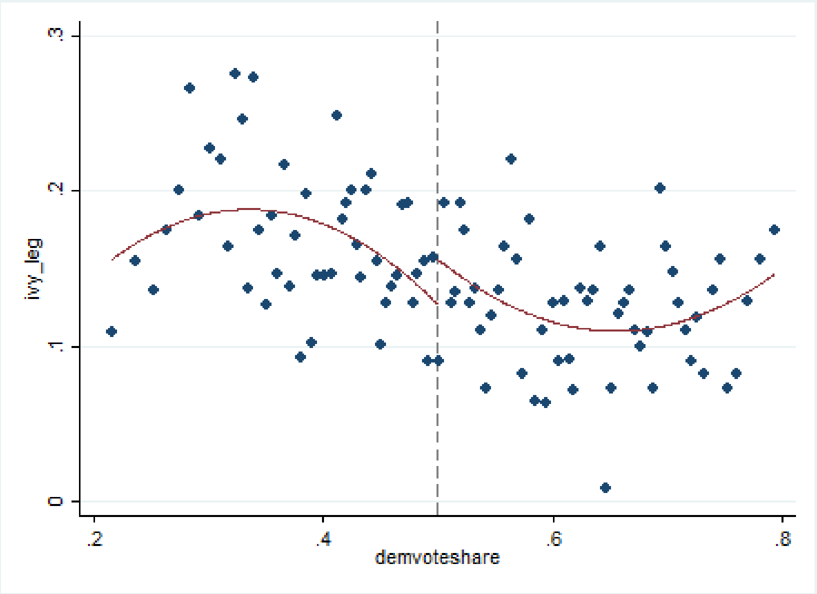
STATA Command: binscatter

- Example:

```
binscatter ivy_leg demvoteshare if demvoteshare>=0.2 &
    demvoteshare<=0.8, n(100) rd(0.5) linetype(qfit)
graph export f3.png, replace
```

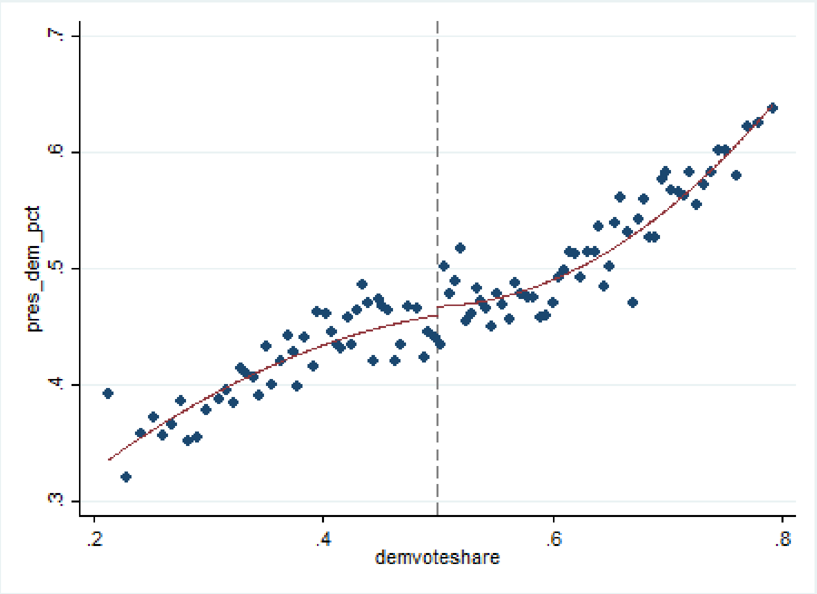
Covariates by assignment variable

Ivy league graduate by Democrat vote share



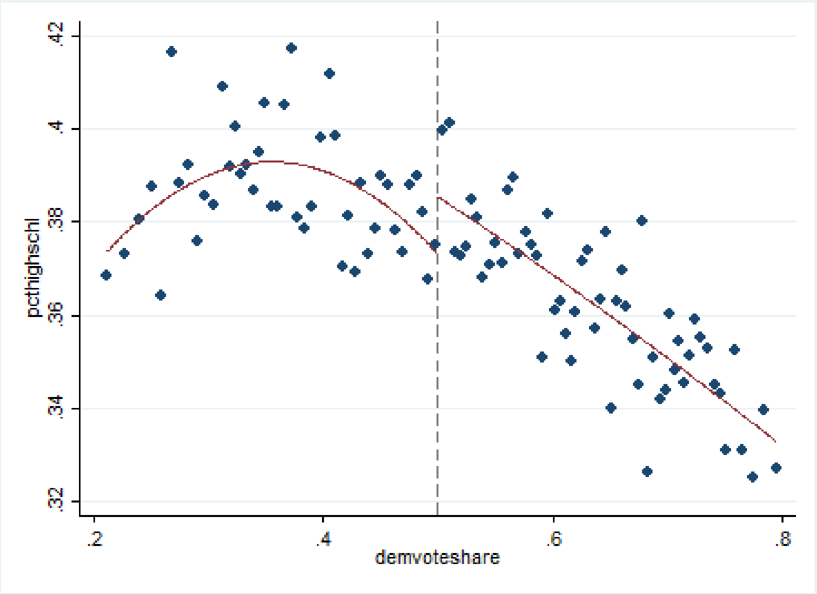
Covariates by assignment variable

Presidential vote share by Democrat vote share



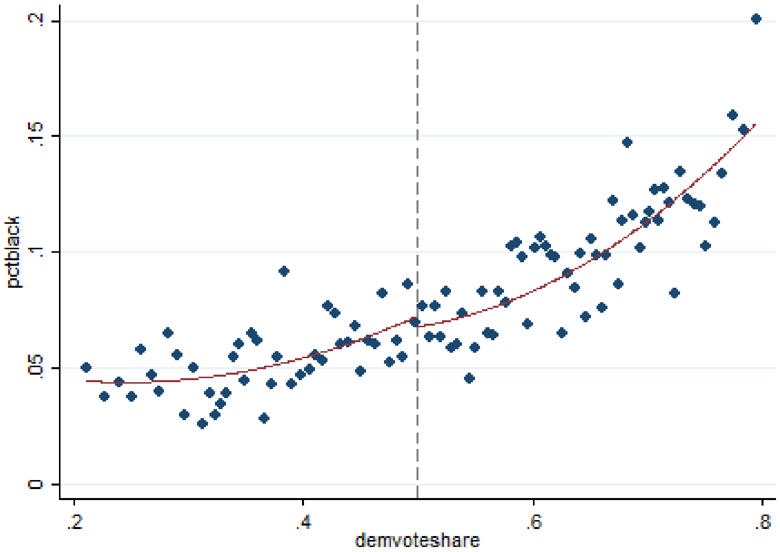
Covariates by assignment variable

High school ratio by Democrat vote share



Covariates by assignment variable

African American ratio by Democrat vote share



Step 2: Test Sorting Behavior

- Plot the **number of observations** in each bin of assignment variable
- This graph can investigate whether there is a discontinuity in the **distribution of the assignment variable** at the threshold
- If it is the case, this would imply people can manipulate the assignment variable around the threshold

STATA Command: rddensity

- Syntax:

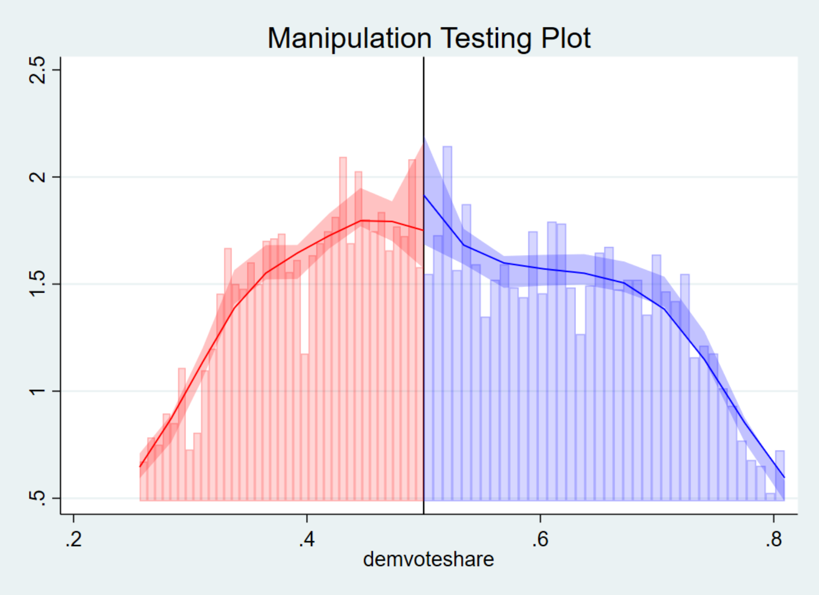
```
rddensity runvar [if] [in] [, options]
```

- Example:

```
rddensity demvoteshare, c(0.5) all plot
graph export f7.png, replace
```

- **runvar**: assignment variable
- **c()**: the cutoff in the assignment variable

Test Sorting Behavior



Step 3: Preparation for Estimation

- Generate treatment variable D_i

```
gen democrat = demvoteshare>=0.5
```

- Generate some Polynomials (recenter to cutoff 0.5)

```
gen x_c = demvoteshare - 0.5
gen x2_c = x_c^2
```

- Generate some Polynomials interacted with D_i

```
gen d_x_c = democrat*x_c
gen d_x2_c = democrat*x2_c
```

Step 4: Estimation

- Parametric Approach using all available data

```
reg score democrat x_c-x2_c d_x_c-d_x2_c,cl(id2)
```

- Nonparametric Approach using sample around cutoff (local linear regression)

- use “rdrobust”

```
rdrobust score demvoteshare, c(0.5) h(0.05) kernel(tri)
rdrobust score demvoteshare, c(0.5) bwselect(mserd)
```

Step 5: Robustness Checks

- Try different bandwidth values
- Compare results across different polynomial orders
- Check sensitivity to kernel choices

R Example

- Install and load required R packages:
  - rdrobust
  - haven (for reading Stata data)
  - rdd
- Use RDD-LMB-data.dta

Package Installation and Data Loading

- Install and load packages:

```
install.packages('rdrobust')
install.packages('haven')
install.packages('rdd')
library(rdrobust)
library(haven)
library(rdd)
```

- Load data:

```
# Read Stata data file
RDD_LMB_data <- read_dta("RDD-LMB-data.dta")
summary(RDD_LMB_data)
```

Step 1: Graphical Analysis

- Plot outcome variable (ADA score):

```
rdplot(y = RDD_LMB_data$score,
x = RDD_LMB_data$demvoteshare,
binselect = "es",
ci = 95,
c = 0.5,
p = 1,
h = 0.5,
title = "RD Plot: U.S. Senate Election Data")
```

- Key parameters:
  - **binselect**: bin selection method
  - **ci**: confidence interval level
  - **c**: cutoff point
  - **p**: order of polynomial
  - **h**: bandwidth

Step 2: Testing for Sorting Behavior

- Test for manipulation of running variable:

```
DC_test <- DCdensity(RDD_LMB_data$demvoteshare, 0.5)
summary(DC_test)
```

- **DCdensity**: Implements McCrary (2008) density test
- **0.5**: Specifies the cutoff point

Step 3: Data Preparation for Estimation

- Generate treatment and polynomial variables:

```
# Create treatment indicator
RDD_LMB_data$democrat[RDD_LMB_data$demvoteshare >= 0.5]
    <- 1
RDD_LMB_data$democrat[RDD_LMB_data$demvoteshare < 0.5]
    <- 0

# Center running variable
RDD_LMB_data$x_c <- RDD_LMB_data$demvoteshare - 0.5

# Generate polynomials
RDD_LMB_data$x2_c <- RDD_LMB_data$x_c^2
RDD_LMB_data$d_x_c <- RDD_LMB_data$democrat *
    RDD_LMB_data$x_c
RDD_LMB_data$d_x2_c <- RDD_LMB_data$democrat *
    RDD_LMB_data$x2_c
```

Step 4: RDD Estimation

- Parametric approach using all data:

```
# Linear model with quadratic terms
rdd_lm <- lm(score ~ democrat + x_c + x2_c +
d_x_c + d_x2_c, data = RDD_LMB_data)
summary(rdd_lm)
```

- Nonparametric approach using rdrobust:

```
# Default bandwidth selection
rdd_robust <- rdrobust(y = RDD_LMB_data$score,
x = RDD_LMB_data$x_c)
summary(rdd_robust)

# Specific bandwidth
rdd_robust_h <- rdrobust(y = RDD_LMB_data$score,
x = RDD_LMB_data$x_c,
h = 0.05)
summary(rdd_robust_h)
```

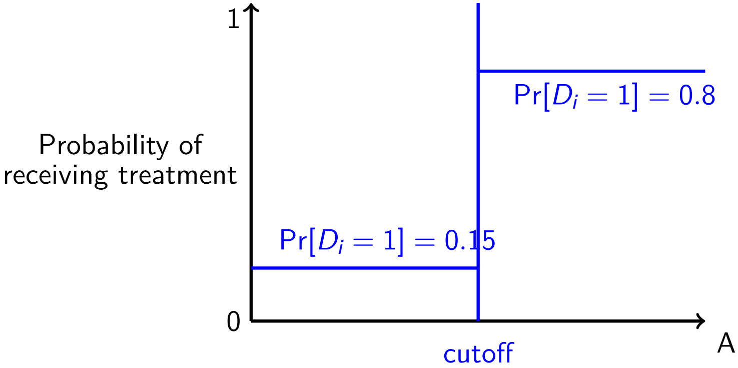
Extension: Fuzzy RDD

Fuzzy RDD

$$\lim_{\varepsilon \to 0} \Pr[D_i = 1 | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \Pr[D_i = 1 | A_i = c - \varepsilon] \neq 0$$

- The probability of getting the treatment jumps discontinuously at the cutoff
  - Some individuals above cutoff do NOT get treatment and some individuals below cutoff do receive treatment

Fuzzy Regression Discontinuity



Fuzzy RDD: Identification

Potential Outcomes Framework

- Treatment Eligibility

$$Z_i = \begin{cases} 1 & \text{if } A_i \geq c, \text{ eligible for a treatment} \\ 0 & \text{if } A_i < c, \text{ not eligible for a treatment} \end{cases}$$

Potential Outcomes Framework

- Potential Treatments
  - D_i^z: treatment status given the value of Z
  - D_i^1: treatment status if eligible for a treatment (above cutoff c)

$$\mathrm{D}_i^1 = \begin{cases} 1 & \text{if getting a treatment} \\ 0 & \text{if not getting a treatment} \end{cases}$$

  - D_i^0: treatment status if not eligible for a treatment (below cutoff c)

$$\mathrm{D}_i^0 = \begin{cases} 1 & \text{if getting a treatment} \\ 0 & \text{if not getting a treatment} \end{cases}$$

Potential Outcomes Framework

- Observed Treatment

$$D_i = \begin{cases} \mathrm{D}_i^1 & \text{if } Z_i = 1, A_i \geq c \\ \mathrm{D}_i^0 & \text{if } Z_i = 0, A_i < c \end{cases}$$

- or, in a more compact notation: $\mathrm{D}_i = Z_i \mathrm{D}_i^1 + (1 - Z_i)\mathrm{D}_i^0$

Potential Outcomes Framework

- In sharp RDD, the **eligible for a treatment** Z_i is the same as **getting a treatment** D_i
  - $Z_i = D_i$
- In fuzzy RDD, the **eligible for a treatment** Z_i does NOT represent the **getting a treatment** D_i
  - $Z_i \neq D_i$

Use Passing Cutoff as an IV

- The discontinuity in outcome is actually the **average causal effect** of **treatment eligibility** $Z_i = 1(A_i \geq c)$ at cutoff c

$$\lim_{\varepsilon \to 0} \mathrm{E}[Y_i | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[Y_i | A_i = c - \varepsilon] \tag{2}$$

- To recover the causal effect of **receiving treatment** D_i
- Divide (2) by the jump in the treatment probability at cutoff

$$\lim_{\varepsilon \to 0} \mathrm{E}[D_i | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[D_i | A_i = c - \varepsilon]$$

Fuzzy RDD is IV

- The the average causal effect of **receiving treatment** defined in fuzzy RDD :

$$\alpha_{FRD} = \frac{\lim_{\varepsilon \to 0} \mathrm{E}[Y_i | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[Y_i | A_i = c - \varepsilon]}{\lim_{\varepsilon \to 0} \mathrm{E}[D_i | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[D_i | A_i = c - \varepsilon]}$$

- This is a **Wald estimate** at cutoff c
- So we can consider fuzzy RDD as an IV estimate
- Use treatment eligibility $Z_i = 1(A_i \geq c)$ as an instrument for treatment received D_i
    - Use an indicator for the test score above threshold as an instrument for attending NTU

Fuzzy RDD is IV

Assumptions

- **First-Stage Relationship**: $Z_i = 1(A_i \geq c)$ affects treatment probability
- **Local Independent Assumption**: In a neighborhood of cutoff c

$$(\mathrm{Y}_i^1, \mathrm{Y}_i^0, \mathrm{D}_i^1, \mathrm{D}_i^0) \perp\!\!\!\perp Z_i$$

  - In a neighborhood of cutoff c, the assignment to treatment is random

Fuzzy RDD is IV

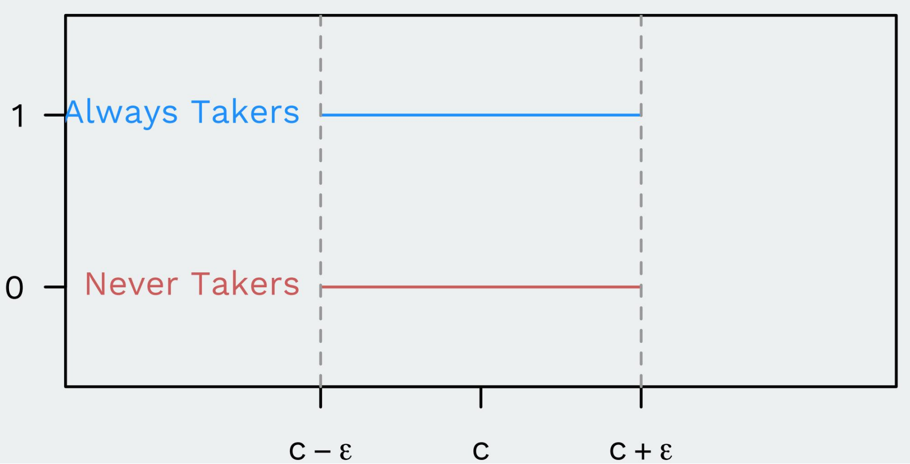
Assumptions

- **Exclusion Restriction**: $Z_i = 1(A_i \geq c)$ affects outcome Y_i only through changing treatment status D_i
- **Monotonicity Assumption**: $\mathrm{D}_i^1 \geq \mathrm{D}_i^0$
  - No one is discouraged from taking the treatment by crossing the threshold

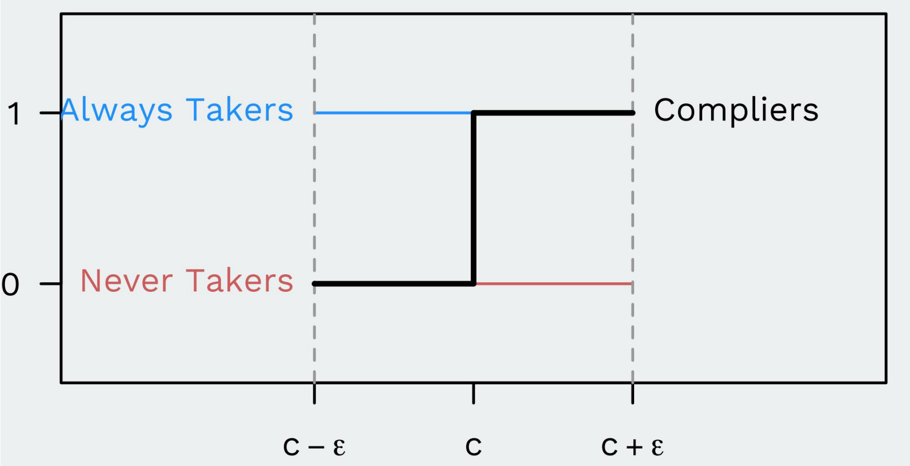
Fuzzy RDD and Compliers

- We can define four types of individuals based on whether they fellow the treatment assignment:
  - **Compliers**: $\mathrm{D}_i^1 > \mathrm{D}_i^0$ ($\mathrm{D}_i^0 = 0$ and $\mathrm{D}_i^1 = 1$)
    - ★ David's test score is above NTU cutoff and enrolled in NTU
    - ★ David's test score is below NTU cutoff and did not enroll in NTU
  - **Always-takers**: $\mathrm{D}_i^1 = \mathrm{D}_i^0 = 1$
    - ★ John always can enroll in NTU (whether or not his test score is above NTU cutoff)
  - **Never-takers**: $\mathrm{D}_i^1 = \mathrm{D}_i^0 = 0$
    - ★ Hank never enroll in NTU (whether or not his test score is above NTU cutoff)
  - **Defiers**: $\mathrm{D}_i^1 < \mathrm{D}_i^0$ ($\mathrm{D}_i^0 = 1$ and $\mathrm{D}_i^1 = 0$)
    - ★ Jimmy's test score is above NTU cutoff and did NOT enroll in NTU
    - ★ Jimmy's test score is below NTU cutoff but enrolled in NTU

Fuzzy RDD and Compliers



Fuzzy RDD and Compliers



Identification Results for Fuzzy RDD

Fuzzy RDD Identify LATE at cutoff c

$$\alpha_{FRD} = \frac{\lim_{\varepsilon \to 0} \mathrm{E}[Y_i | A_i = x + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[Y_i | A_i = x - \varepsilon]}{\lim_{\varepsilon \to 0} \mathrm{E}[D_i | A_i = x + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[D_i | A_i = x - \varepsilon]}$$
$$= \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | \mathrm{D}_i^1 > \mathrm{D}_i^0, A_i = c]$$

- The estimate in fuzzy RDD represents the **causal effect for compliers** (local average treatment effect, LATE) at cutoff c
- **Compliers** are those who receive the treatment when they follow treatment eligibility rule ($A_i \geq c$), but would not otherwise receive it ($A_i < c$)

Fuzzy RDD: Estimation

Fuzzy RDD

Estimation

- Fuzzy RDD implies the discontinuity in treatment assignment Z_i can act as an instrument for actual receipt of treatment D_i
- Therefore, we can estimate treatment effect in fuzzy RDD by using two-stage least squares (TSLS)

$$D_i = \alpha_1 + \rho_1 Z_i + \mathrm{f}_1(A_i) + \upsilon_i$$
$$Y_i = \alpha_2 + \rho_2 D_i + \mathrm{f}_2(A_i) + \eta_i$$

- As in the sharp RDD setting, we can use either parametric or nonparametric approaches to estimate treatment effect

Fuzzy RDD: Empirical Example

Zachary Bleemer and Aashish Mehta (2022) "**Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major**", American Economic Journal: Applied Economics

- Examine the causal effect of majoring in economics on early-career earnings
- Key Research Questions:
  - What is the wage return to studying economics?
  - How does access to economics major affect students' career paths?
  - Do observational wage differences reflect causal effects?

Empirical Example: Bleemer & Mehta (2022)

Motivation

- College graduates with economics degrees earn substantially higher wages
  - Median wage of $90,000 for economics majors vs. $65,000 for other social sciences
- However, estimating causal effects is challenging due to:
  - Students' nonrandom selection into majors
  - Universities' admissions and grade requirements
- Study exploits a GPA threshold policy at UC Santa Cruz that restricted access to the economics major
  - Students needed 2.8 GPA in Economics 1 & 2 to declare major

- **Fuzzy RDD**: Exploits GPA threshold (2.8) in Economics 1 and 2 for major access

$$\alpha_{FRD} = \frac{\lim_{\varepsilon \to 0} \mathrm{E}[Y_i | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[Y_i | A_i = c - \varepsilon]}{\lim_{\varepsilon \to 0} \mathrm{E}[D_i | A_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[D_i | A_i = c - \varepsilon]}$$

- Where:
  - Y_i: Early-career wages (2017-2018)
  - A_i: Average GPA in Economics 1 & 2
  - D_i: Economics major indicator
  - c: GPA threshold (2.8)

- **First Stage**:

$$D_i = \alpha_1 + \rho_1 Z_i + \mathrm{f}_1(A_i) + \upsilon_i$$

- **Reduced Form**:

$$Y_i = \alpha_2 + \rho_2 Z_i + \mathrm{f}_2(A_i) + \eta_i$$

- Where:
  - D_i: Economics major indicator
  - $Z_i = 1(A_i \geq 2.8)$: Treatment eligibility
  - A_i: Average GPA in Economics 1 & 2
  - $\mathrm{f}_1(.), \mathrm{f}_2(.)$: Linear functions of GPA

- **Fuzzy RD Estimate**:

$$\alpha_{FRD} = \frac{\rho_2}{\rho_1} = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | \mathrm{D}_i^1 > \mathrm{D}_i^0, A_i = 2.8]$$

Empirical Example: Bleemer & Mehta (2022)

First Stage

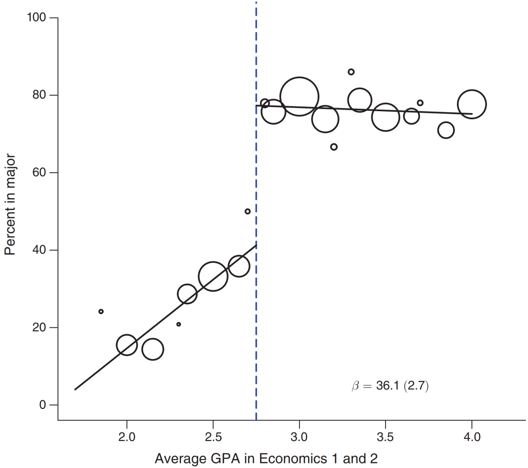


Figure 1. The Effect of the UCSC Economics GPA Threshold on Majoring in Economics

Empirical Example: Bleemer & Mehta (2022)

First Stage

- Students just above 2.8 GPA threshold were:
    - 36 percentage points more likely to major in economics
    - Most would have otherwise majored in other social sciences

Empirical Example: Bleemer & Mehta (2022)

Second Stage

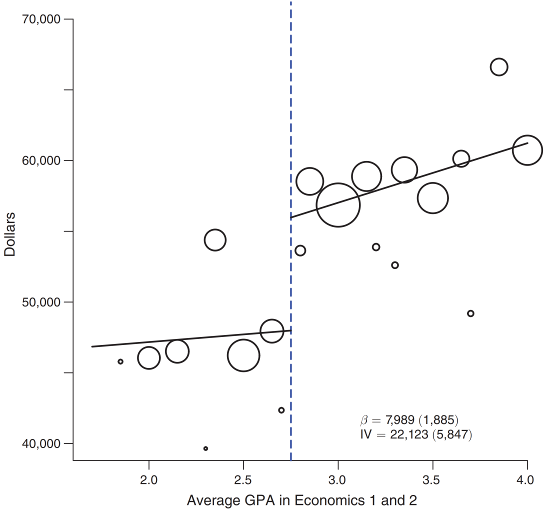
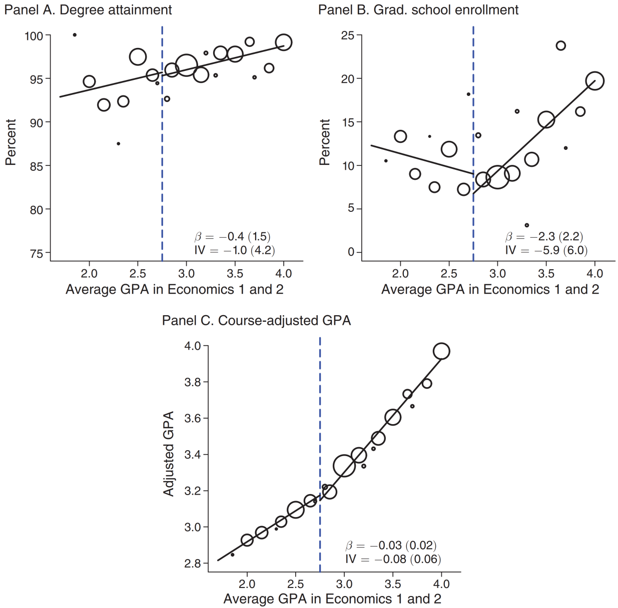


FIGURE 2. THE EFFECT OF THE UCSC ECONOMICS GPA THRESHOLD ON ANNUAL WAGES

- **Large Returns**: Majoring in economics caused:
  - 46% increase in early-career wages ($22,000)
  - Similar effects for male and female students
  - Possibly larger effects for URM students
    - URM: Underrepresented Minority includes Black, Hispanic, and Native American students

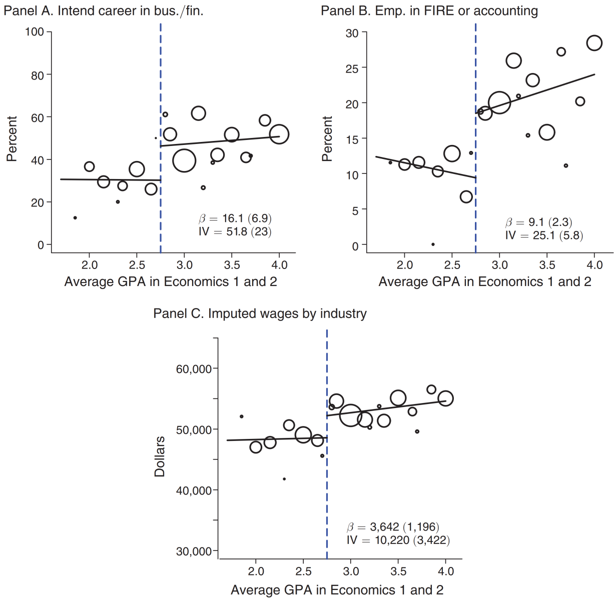
Empirical Example: Bleemer & Mehta (2022)

Mechanisms: Educational Investment



Empirical Example: Bleemer & Mehta (2022)

Mechanisms: Industry Effects



- **Educational Investment**:
  - No effect on graduation rates or graduate school enrollment
  - No change in study time or course-adjusted grades
  - 13 more economics courses, 9 fewer other social science courses
- **Industry Effects**:
  - 52 pp more likely to prefer business/finance careers
  - 25 pp more likely to work in FIRE or accounting
  - About half ($10,220) of wage effect explained by industry sorting

Empirical Example: Bleemer & Mehta (2022)

Alternative Specifications

- **Robustness Checks**:
  1. Quadratic running variable terms
  2. Adding controls:
     - ★ Demographic variables
     - ★ High school fixed effects
  3. Narrower bandwidth ($\pm$0.5 GPA points)
- Results robust across all specifications

Suggested Readings

- Chapter 4, *Mastering 'Metrics: The Path from Cause to Effect*
- Chapter 6, *Mostly Harmless Econometrics: An Empiricist's Companion*
- Chapter 6, *Causal Inference: The Mixtape*