ECON 672

Week 7: Regression Discontinuity Design

Samuel Rowe, PhD 1/12/2023

Overview

- Takeaway
- DAG
- Visualization of RDD
- Sharp RDD
- Fuzzy RDD
- Challenges to Identification
- Example: Close Elections

Overview of RDD

- Intuition of RDD
    - Units around an arbitrarily defined cutoff are similar to one another along a continuous variable
    - We expect to see a discontinuous jump in the outcome at the cutoff along the running variable
    - We can calculate an average treatment effect for this subpopulation around the cutoff (LATE)
- Growth in popularity in recent years
    - Ever since Black (1999) utilizes a RDD with geographic discontinuities for school zoning districts to estimate willingness to pay for “better” schools, RDD has grown in popularity
    - Lee and Lemieux (2010) practical guide has been cited thousands of times

The Takeaway

- Pros
  - The most credible quasi-experimental design strategy
  - Most assumptions are directly or indirectly testable
  - Graphical approach to an identification strategy
- Cons
  - You need deep institutional knowledge about the treatment, cutoff/threshold, running variable, and outcomes
  - LATE cannot be extrapolated to the general population, only those around the cutoff
  - You need lots of data around the cutoff/threshold

The Takeaway

Assumptions of RDD

- Continuity Assumption (Sharp and Fuzzy)
    - This is indirectly testable by looking for bunching around the cutoff
- Fuzzy RDD Only
    - Exclusion Restriction Assumption - not testable
    - Monotonicity Assumption - testable
    - Non-Zero 1st Stage Assumption - testable
    - Stable Unit Treatment Value Assumption (SUTVA) - not testable

The Takeaway

Assumptions of RDD

- Testing Assumptions - Continuity Assumption
- McCrary Test
  - Test for bunching (density) on one side of the cutoff compared to the other side
- Covariate Balance Test
  - Covariates on observables should be balanced on both sides of the cutoff
  - There should be no discontinuous jumps in covariates, only the outcome
- Placebo Test
  - Testing for discontinuous jumps in the running variable when there shouldn’t be any

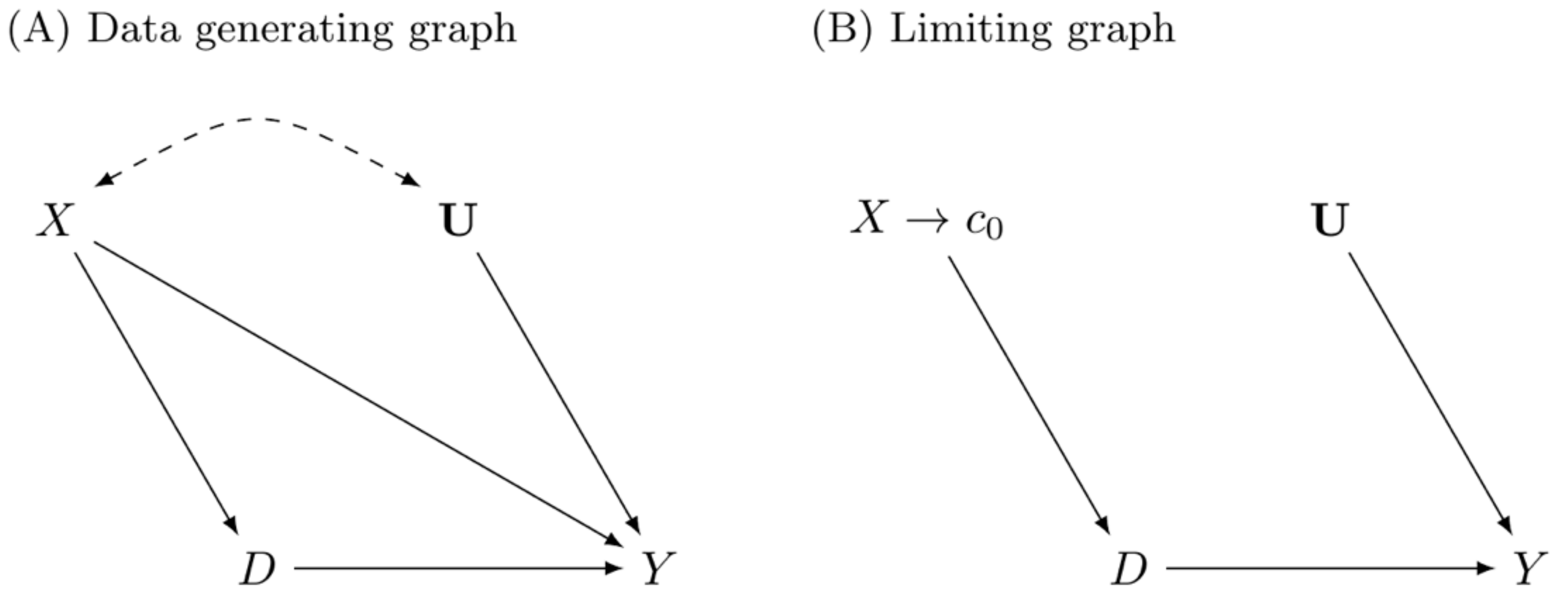
Definitions

- Running variable or forcing variable
  - The continuous variable that determines treatment and control
- Cutoff/threshold
  - The assignment rule into treatment or control by an arbitrary rule along the running variable
- Bandwidth/Windows
  - The area or window around the cutoff that you are comparing
- Bins
  - The regular intervals to estimate means within the specified intervals along the running variable
- Sharp RDD
  - The running variable is deterministic of assignment into treatment and control
- Fuzzy RDD
  - The running variable is probabilistic of assignment into treatment and control

Regression Discontinuity DAG

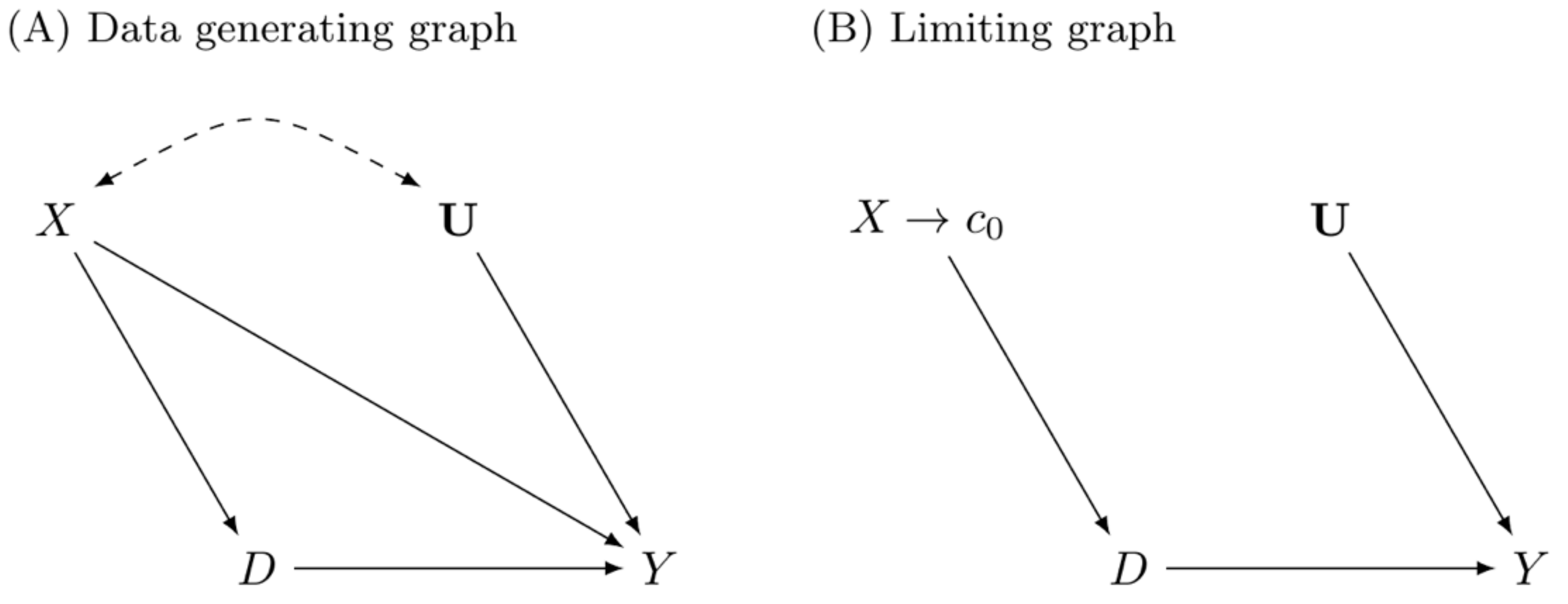
Regression Discontinuity DAG

- RDD is a popular design because it is very convincing for the identification of the causal effect by eliminating selection bias on observed and unobserved confounders
- If RDD is successful in eliminating selection bias, we are able to retrieve the LATE for a subpopulation



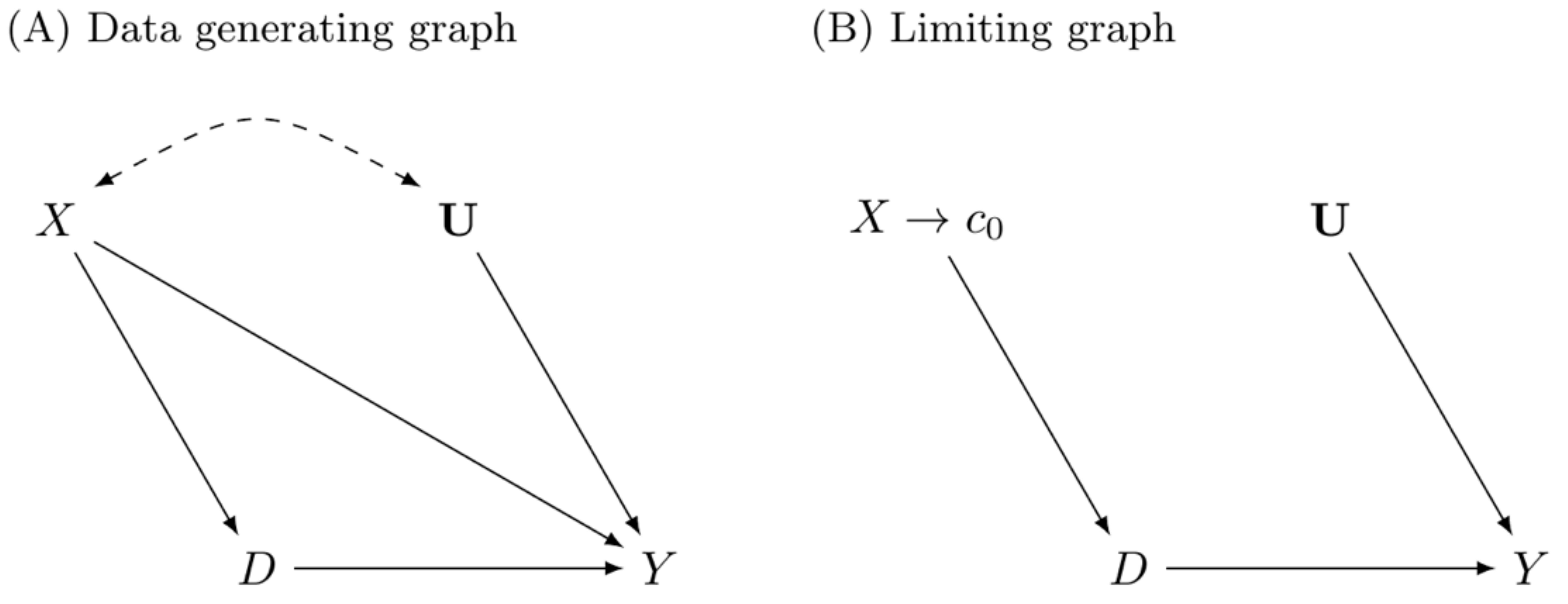
Regression Discontinuity DAG

- Data Generating Graph
- Without a cutoff (A) we have several backdoor that prevent us from identifying the impact of D
- Key Point
    - The running variable X assigns treatment so we do have have overlap
    - We cannot just control for X even if U was not present - no overlap



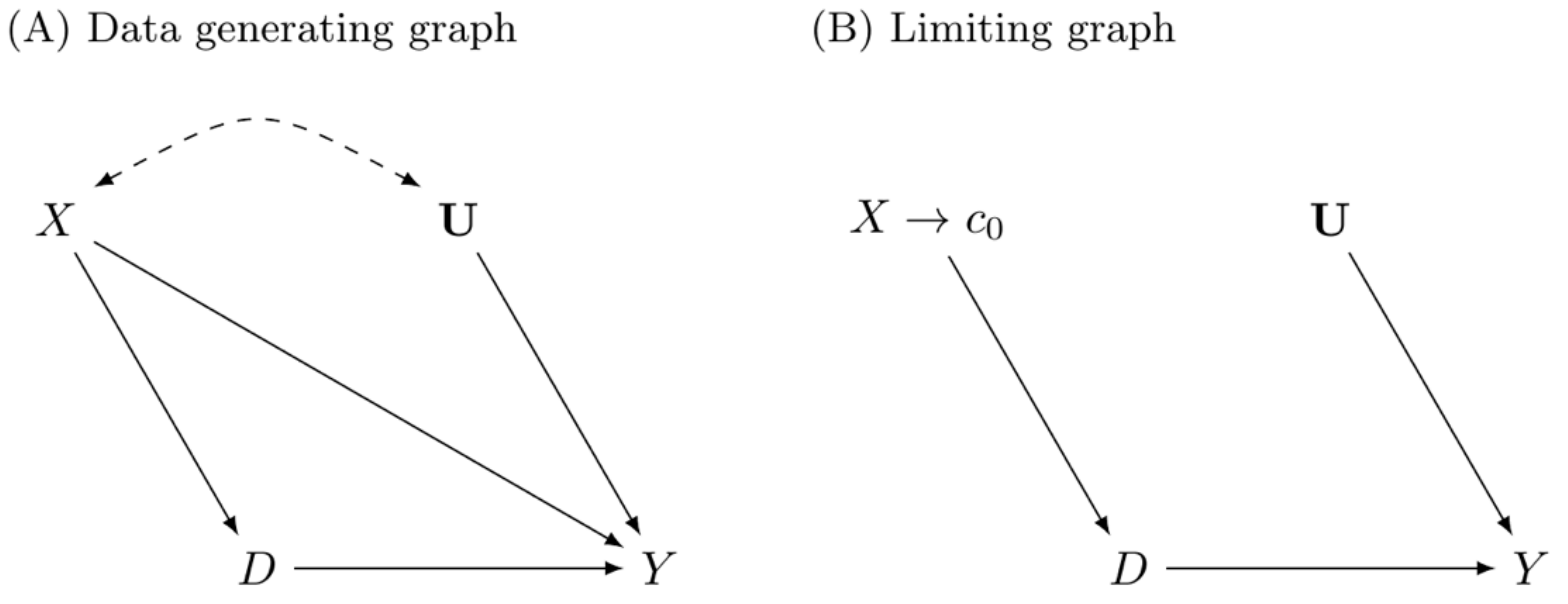
Regression Discontinuity DAG

- (A) Direct Pathway
  - $D \rightarrow Y$
- (A) Indirect Pathways
  - X is a continuous variable that assigns treatment and control
  - $D \leftarrow X \rightarrow Y$
  - We also have U, which may be observed or unobserved
  - $D \leftarrow - U - \rightarrow Y$



Regression Discontinuity DAG

- (B) Limiting Graph - Add Cutoff
  - We can identify the causal effect $(D \rightarrow Y)$ for those around the cutoff c_0
  - The cutoff blocks the backdoor pathway from $X \rightarrow Y$
  - We cutoff off the backdoor pathway by utilizing the limit near c_0
- We still have the running variable determine treatment and control
- Cutoff X onto Y
  - $D \leftarrow X \rightarrow c_0$
  - Only works if continuity assumption is satisfied

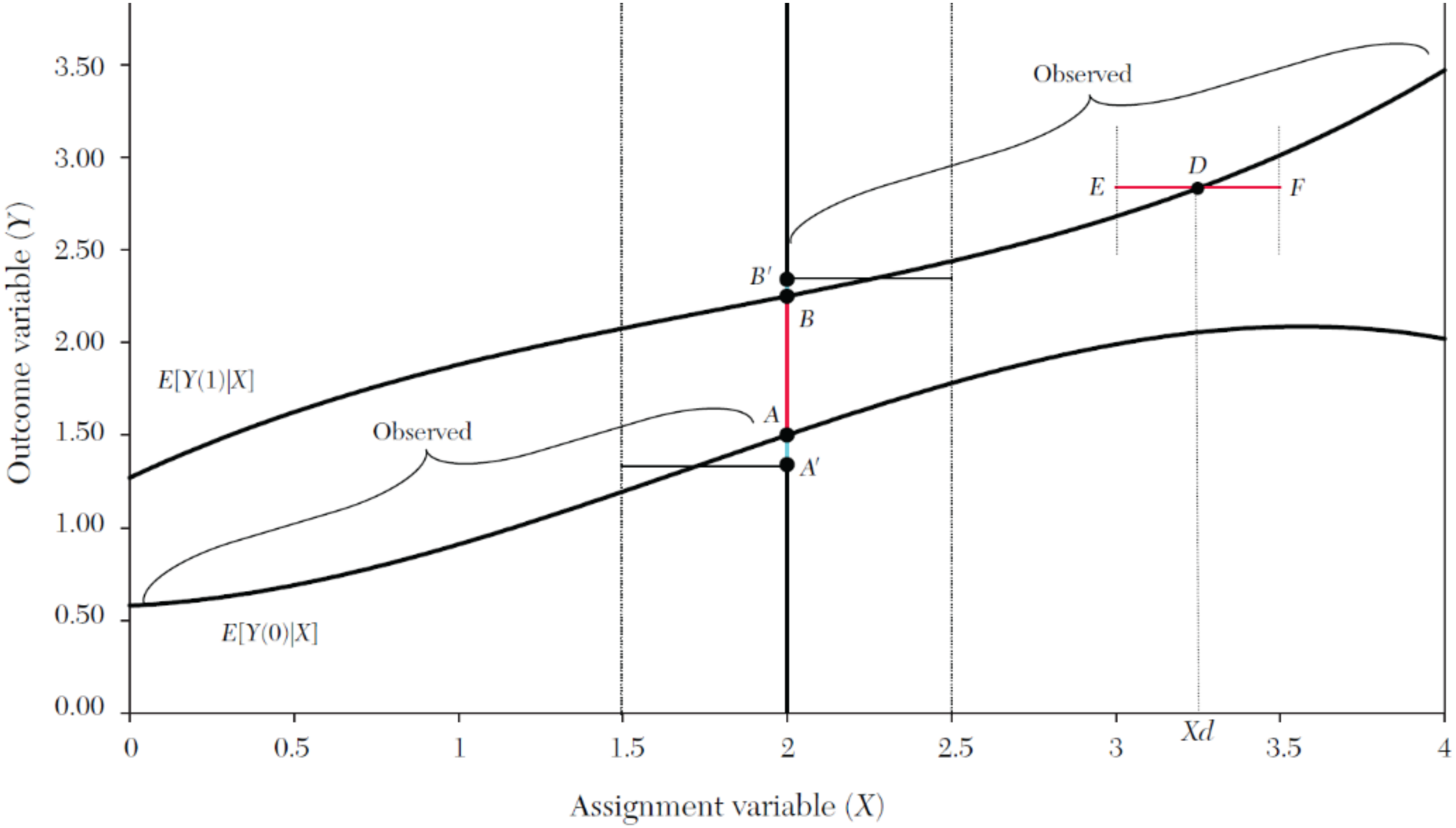


Cutoff

- The cutoff c_0 rule
  - An assignment along the running variable X determines assignment into treatment and control depending upon which side of the cutoff c_0 the observation is on
- For example
  - If someone scores an 0.08 blood alcohol content then assigned drunk driving, but if someone is 0.078 then not
  - 250K in receipts will be assigned one treatment
  - Test score will have cutoffs for admissions
- Key Point
  - The continuity assumption must be satisfied at the cutoff

Continuity Assumption

- The continuity assumption states that potential outcomes need to be continuous at the cutoff c_0
  - The cutoff rule cannot be endogenous to other treatments near the cutoff
  - $E[Y^0]$ and $E[Y^1]$ need to be smooth at c_0
- “Nature does not make jumps” or a turtle on a fencepost is not natural
  - A discontinuous jump will be from $E[Y^0 \mid D = 0]$ to $E[Y^1 \mid D = 1]$ due to treatment
- We never observe $E[Y^0 \mid D = 1]$ or $E[Y^1 \mid D = 0]$



Continuity Assumption

- We have to extrapolate $E[Y^0 | D = 1]$ and $E[Y^1 | D = 0]$
- We assume a smooth continuous function of $E[Y^1 | X = c_0]$ and $E[Y^0 | X = c_0]$
- The treatment causes the jump from the $E[Y^0 | X]$ line to the $E[Y^1 | X]$ line at c_0

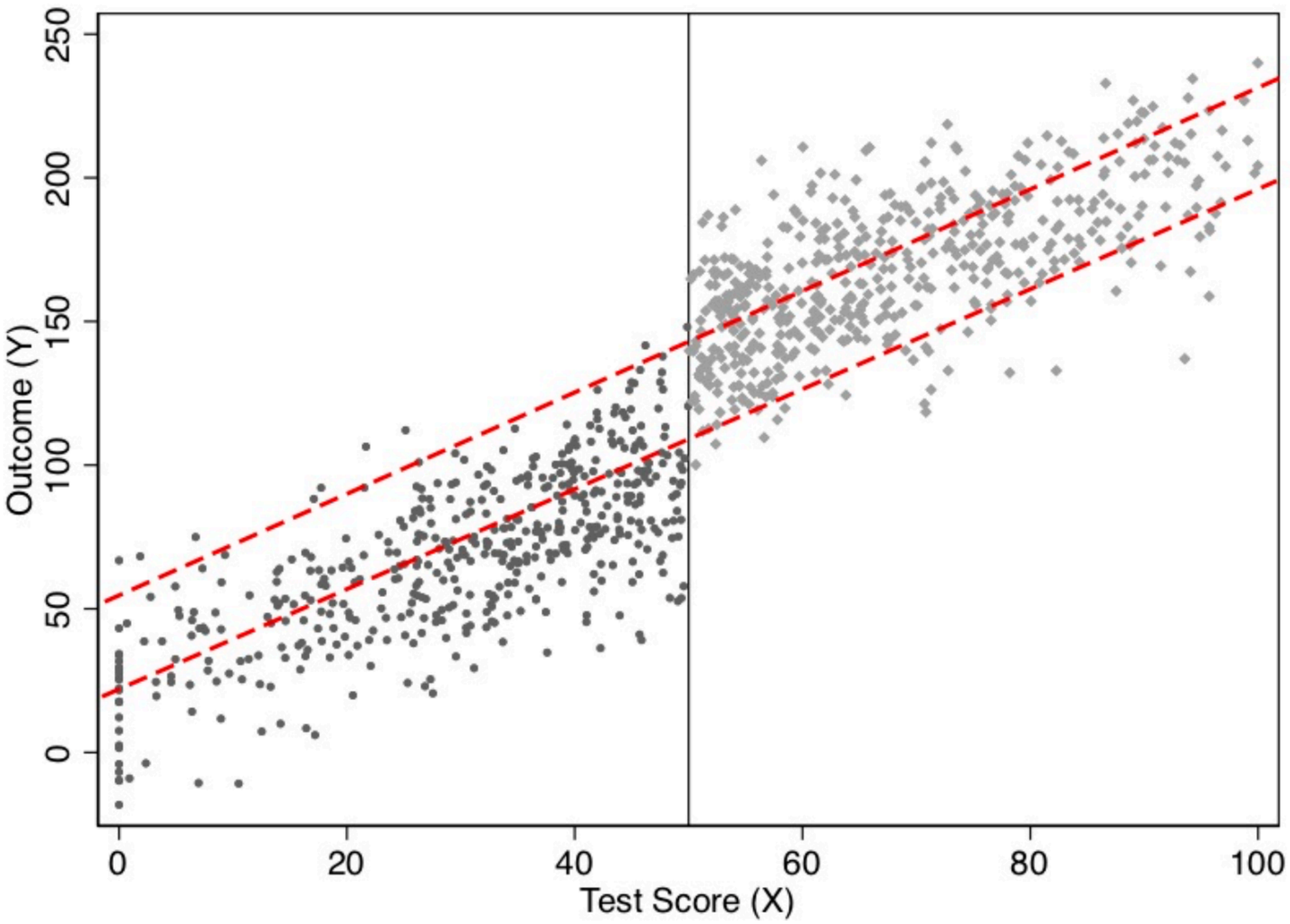


Figure 6.5: Simulated data representing observed data points along a running variable below and above some binding cutoff.
Notes: Dashed lines are extrapolations

Visualization of RDD

- Example from Hoekstra (2009)

Visualization of RDD Example

- RDD is a very visual research design
- Hoekstra (2009) provides an example of using a Fuzzy RDD
  - He wants to assess heterogenous returns across public colleges
  - We might expect self-selection into different universities with students with higher abilities self-selecting into state flagship universities
  - Higher returns to education from state flagships vs state regional might just be a reflection of student’s unobserved ability

Visualization of RDD Example

- There are 4 things we need to consider
  - 1) the running variable
  - 2) bins
  - 3) functional form
  - 4) the discontinuous jump at the cutoff

Visualization of RDD Example

- 1st Stage of Fuzzy RDD is the probability of assignment
- Enrollment increases by 0.388 percentage points at c_0

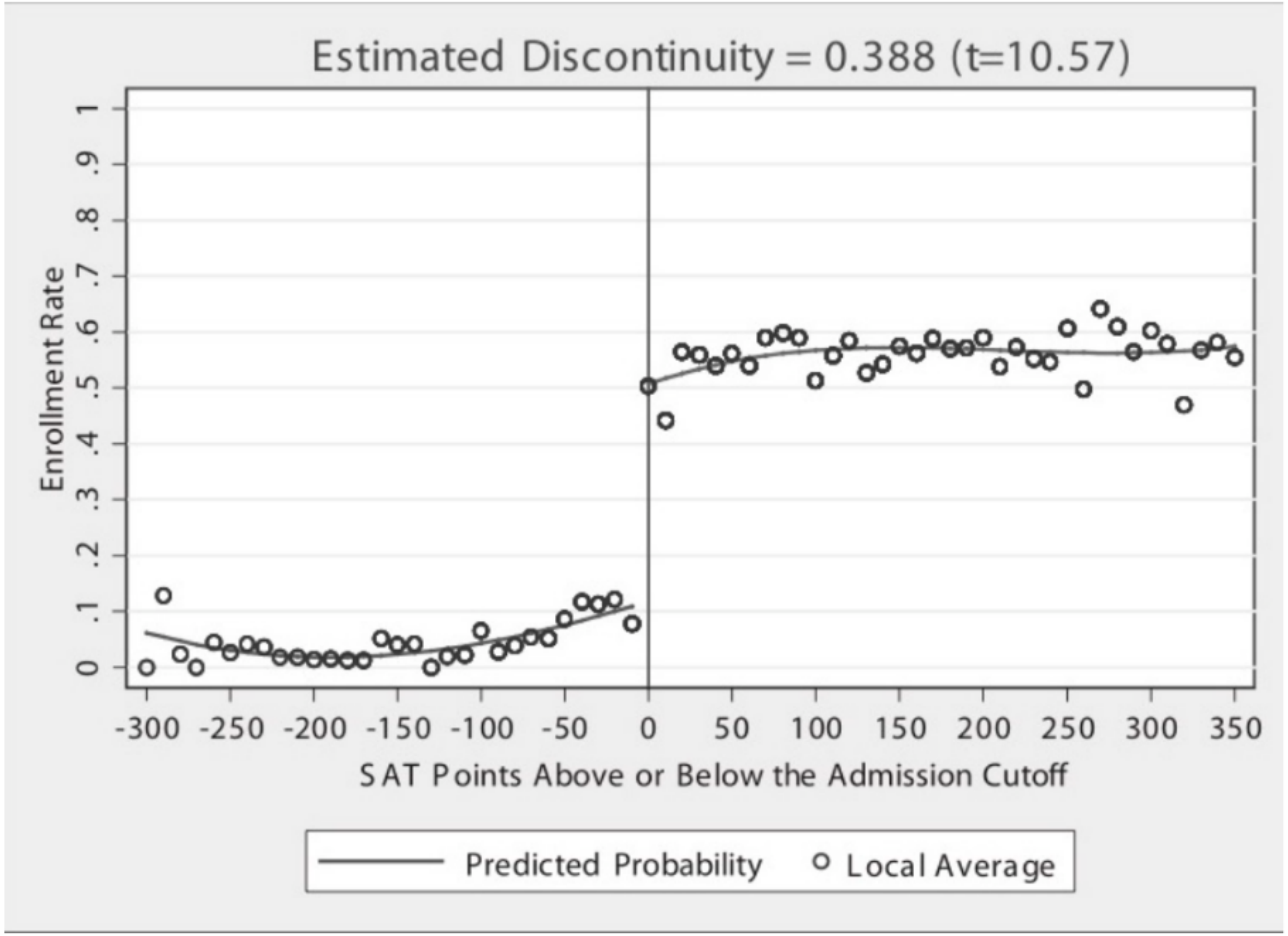


Figure 6.2: Attending the state flagship university as a function of re-centered standardized test scores.

Visualization of RDD Example

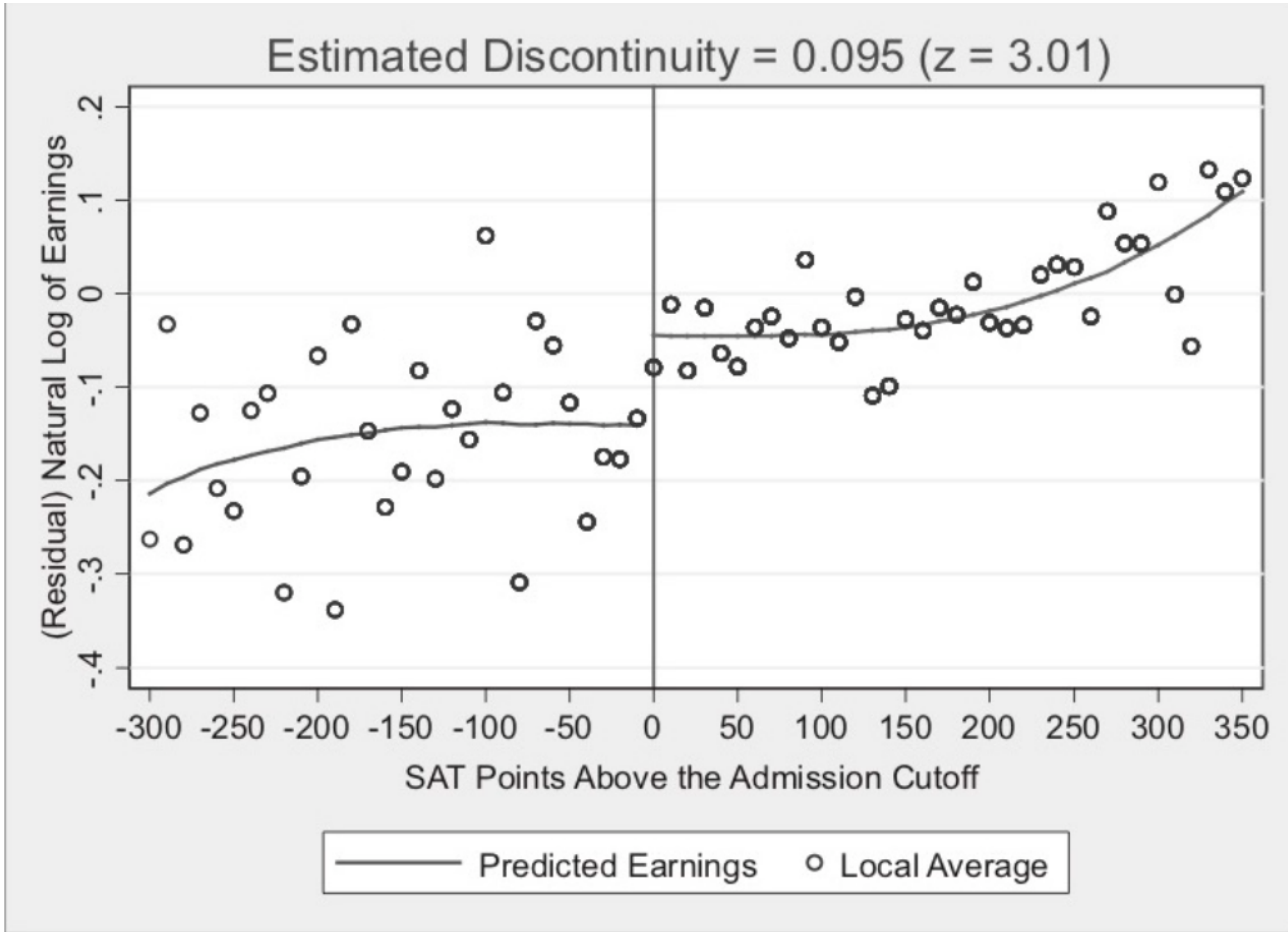
- Hoekstra (2009) utilizes running variable
  - The SAT score as the running variable and the admission cutoff as the cutoff rule c_0, where c_0 is determined by the admissions department and re-centered to zero
- The bins size
  - Regular intervals to assess the probability of admission along the running variable
  - Each dot represents the probability of admission within the bin/interval of SAT scores
- The functional form
  - There are two curved lines (not linear), with one line fitted on each side of the cutoff score
  - The functional form here is takes on a higher order terms (such as $x^2, x^3, \ldots, x^p$)

Visualization of RDD Example

- Imagine two students with an SAT score of 1240 and 1250
  - It is likely that these two particular student differ quite a bit
  - However, if we have hundreds of students with SAT scores of 1240 and 1250, these students are less likely to have differences among observed and unobserved characteristics
- Reflection Questions
  - Should students with SAT scores of 1240 and 1250 differ that much in large samples?
  - Do administrators know of a large jump in ability at the cutoff or is the natural ability continuous around the cutoff?

Visualization of RDD Example

- 2nd Stage with Fuzzy RDD
- Admission into the state flagship university increased earnings by $(e^{0.095} - 1) \approx 0.1$
- Depending upon the window size, earnings increased by 7.4% to 11.1% ten to fifteen years after admission



Data Requirement for RDD

- Acquiring Data
  - Acquiring can be very challenges, especially for administrative data
  - Hoekstra (2009) had gain trust and build relationships with program administrators
  - This is not too different from federal evaluations when working with program staff
- The secret sauce of RDD
  - Gaining access to subject-matter experts on the data and the data itself requires a lot of soft skills (e.g.: building trust and relationships)
  - It take a lot time compared to downloading publicly available data and reading documentation

Data Requirement of RDD

- Where do we find jumps?
  - We can find jumps embedded in arbitrary rules set by people
  - It is possible that firms and government agencies are sitting unknowingly on top of lots of data that can be utilized for an RDD
- Key Points
  - The assignment rule or cutoff rule (c_0) needs to be free from manipulation
  - We need lots of data around the cutoff and its a main feature of RDD

Sharp RDD

Sharp RDD

- The first type of RDD we will discuss is the Sharp RDD
- Why is it called Sharp RDD?
  - Because the assignment rule is ***deterministic*** of assignment
  - $E[D\,|\,X >= c_0] = 1$ and $E[D\,|\,X < c_0] = 0$

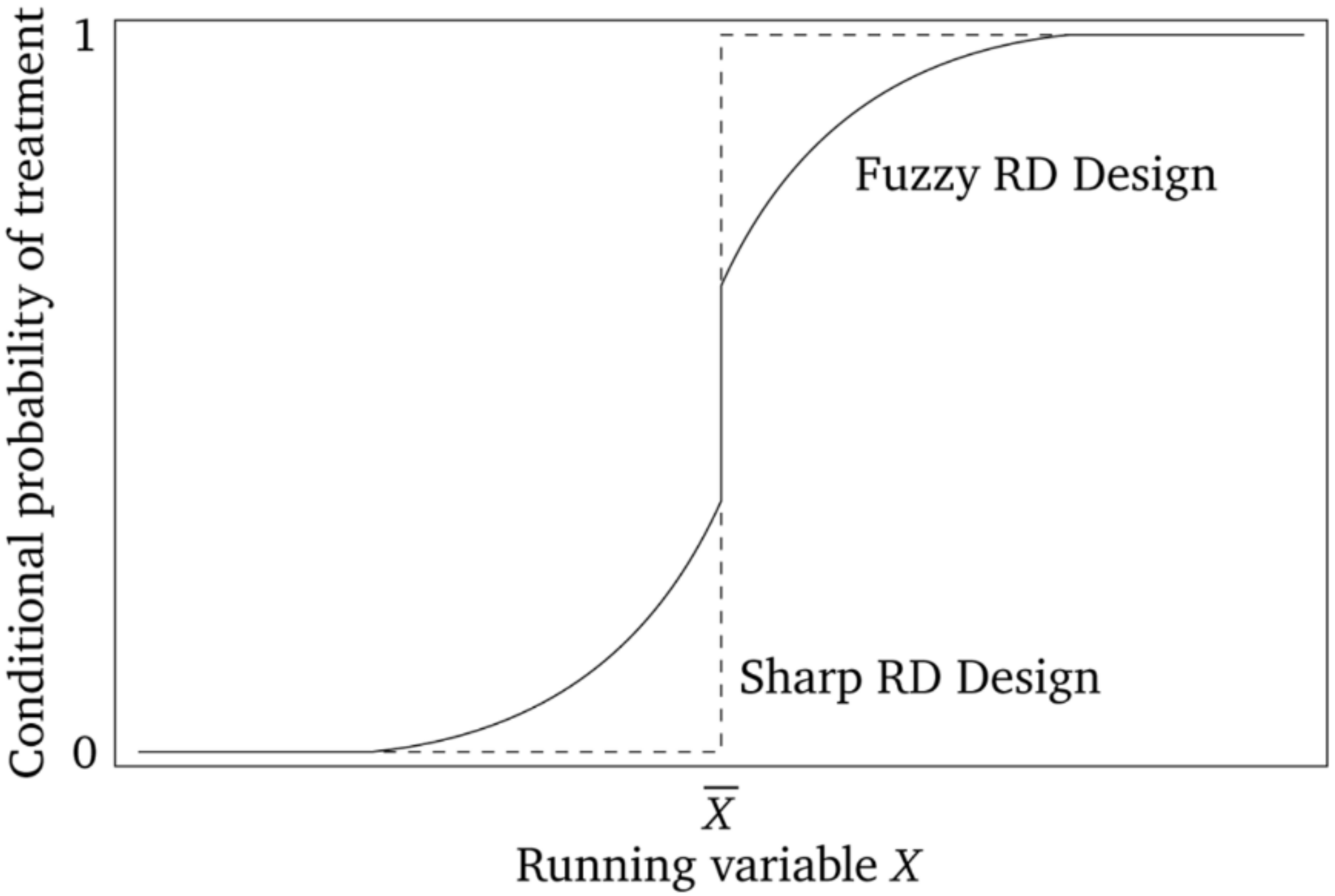


Figure 6.4: Sharp vs. Fuzzy RDD

Sharp RDD

- Sharp RDD status is deterministic and a discontinuous function of X
  - $D_i = \begin{cases} 1 \ \mathit{if} \ X_i \geq c_o \\ 0 \ \mathit{if} \ X_i < c_0 \end{cases}$
  - Where c_0 is our known threshold or cutoff rule
  - X perfectly predicts assignment into treatment or control
- Key Point
  - There is ***no overlap*** between treatment and control in Sharp RDD

Sharp RDD

Potential Outcome Notation

- Assume constant treatment effects or homogenous treatment effects
    - $Y_i^0 = \alpha + \beta X_i$
    - $Y_i^1 = Y_i^0 + \delta$
- This gives us a switching equation
    - $Y_i = Y_i^0 + (Y_i^1 - Y_i^0)D_i$
    - $Y_i = \alpha + \beta X_i + \delta D_i + \varepsilon_i$

Sharp RDD

Potential Outcome Notation

- The treatment effect parameter, δ, is the discontinuity in the conditional expectation function

- $$\begin{aligned}\delta &= \lim_{c_0 \leftarrow x_i} E[Y_i^1 \,|\, X_i = c_0] - \lim_{x_i \rightarrow c_0} E[Y_i^0 \,|\, X_i = c_0] \\ &= \lim_{c_0 \leftarrow x_i} E[Y_i \,|\, X_i = c_0] - \lim_{x_i \rightarrow c_0} E[Y_i \,|\, X_i = c_0]\end{aligned}$$

- We can identify the causal effect by approaching the limit from both sides of the running variable at the cutoff c_0

Sharp RDD

Local Average Treatment Effects (LATE)

- δ_{SRD} is interpreted as a local estimate of the treatment effects
  - Why? Because we ***only have overlap at the limit*** between treatment and control
- Only identifying causal effects on a subpopulation is very similar to IV compliers
  - In Sharp RDD, our subpopulation is not compliers,
  - Our subpopulation is at the limit as $X \rightarrow c_0$ from both sides
- We do not have overlap beyond the limit
  - $\delta_{SRD} = E[Y_i^1 - Y_i^0 | X_i = c_0]$

Sharp RDD

Continuity Assumption plus Stata Exercise

- The key assumption in Sharp RDD is the continuity assumption
    - That is saying that $E[Y^1 | X = c_0]$ and $E[Y^0 | X = c_0]$ are smooth continuous functions of X
        - $E[Y^0 | D = 1] = E[Y^0 | D = 0]$ and $E[Y^1 | D = 1] = E[Y^1 | D = 0]$
    - The continuity assumption rules out observed and unobserved omitted variable bias at the cutoff, since confounders are continuous at the cutoff
- Key Point
    - ***In the absence of treatment*** the functions would remain ***smooth and continuous*** and no jumps would have occurred
- If a discontinuous jumps occurs in any confounder (observed or unobserved), then the continuity assumption is violated

Sharp RDD

Testing the Continuity Assumption

- We cannot directly test the continuity assumption, but there are a few method we can employ
  - We don't have potential outcomes to test directly, but we can test indirectly
  - An SME with deep-institutional knowledge should know if something else besides the treatment is causing a jump at the cutoff c_0
- Placebo tests
  - Testing a fake cutoff, such as $c_0 + n$ or $c_0 - n$
- Testing continuity in covariates at the cutoff
  - Covariates should be smooth and continuous at the cutoff c_0
- McCrary Test
  - There should be do difference in the density of observations at the cutoff c_0

Sharp RDD

Estimating the LATE using Sharp RDD

- There are a few ways to estimate the LATE using RDD
    - This will require specification tests
- Transforming the cutoff
- Global vs Local
    - What is your window and what are your bins
- Functional Forms
    - Linear
    - Pth-order polynomial
    - Non-parametric kernel

Sharp RDD

Estimating Sharp RDD

- 3 Steps
  - Step 1: Recenter your running variable
  - Step 2: Local vs Global (Variance-Bias Trade Off)
  - Step 3: Functional Form

Sharp RDD

1st Step: Transform running variable

- The first step is to transform the running variable
  - Recenter the running variable around the cutoff c_0
  - $Y_i = \alpha + \beta(X_i - c_0) + \delta D_i + \varepsilon_i$
  - This only changes the intercept, not δ_{SRD}
- For example
  - $Y = \beta_0 + \beta_1(Age - 65) + \beta_2 Edu + \varepsilon$
  - $Y = \beta_0 + \beta_1 Age_i + \beta_1 65 + \beta_2 Edu + \varepsilon = (\beta_0 + \beta_1 65) + \beta_1 Age + \beta_2 Edu + \varepsilon$
  - $Y = \alpha + \beta_1 Age + \beta_2 Edu + \varepsilon$, where $\alpha = \beta_0 + \beta_1 65$

Sharp RDD

Step 2: Global vs Local

- Also known as Variance-Bias Trade-off
- Global typically means all of the data you have
- Local typically means restricting the data to a specified window around the cutoff
    - We lose observations
- You typically want to focus on the local
    - However, it is good to include for specification tests
    - Global regression may influence your averages on both sides of the cutoff

Sharp RDD

Step 3: Functional Form

- What is the functional form, specification, or data-generation process?
  - If we do not properly specify the functional form, we may estimate spurious discontinuous jumps
  - The jumps may not exist since the data-generation process was in a different function form
- The data-generation process could have been:
  - Linear specification
  - p^{th} order specification
  - Non-parametric specification

Sharp RDD

Step 3: Functional Form

- Figure 6.9 shows what happens when you incorrectly specify the functional form
- Here we use a linear functional form when the data-generation process was non-linear
- Using a linear functional form on a p^{th} order polynomial will result in a spurious discontinuous jump at c_0

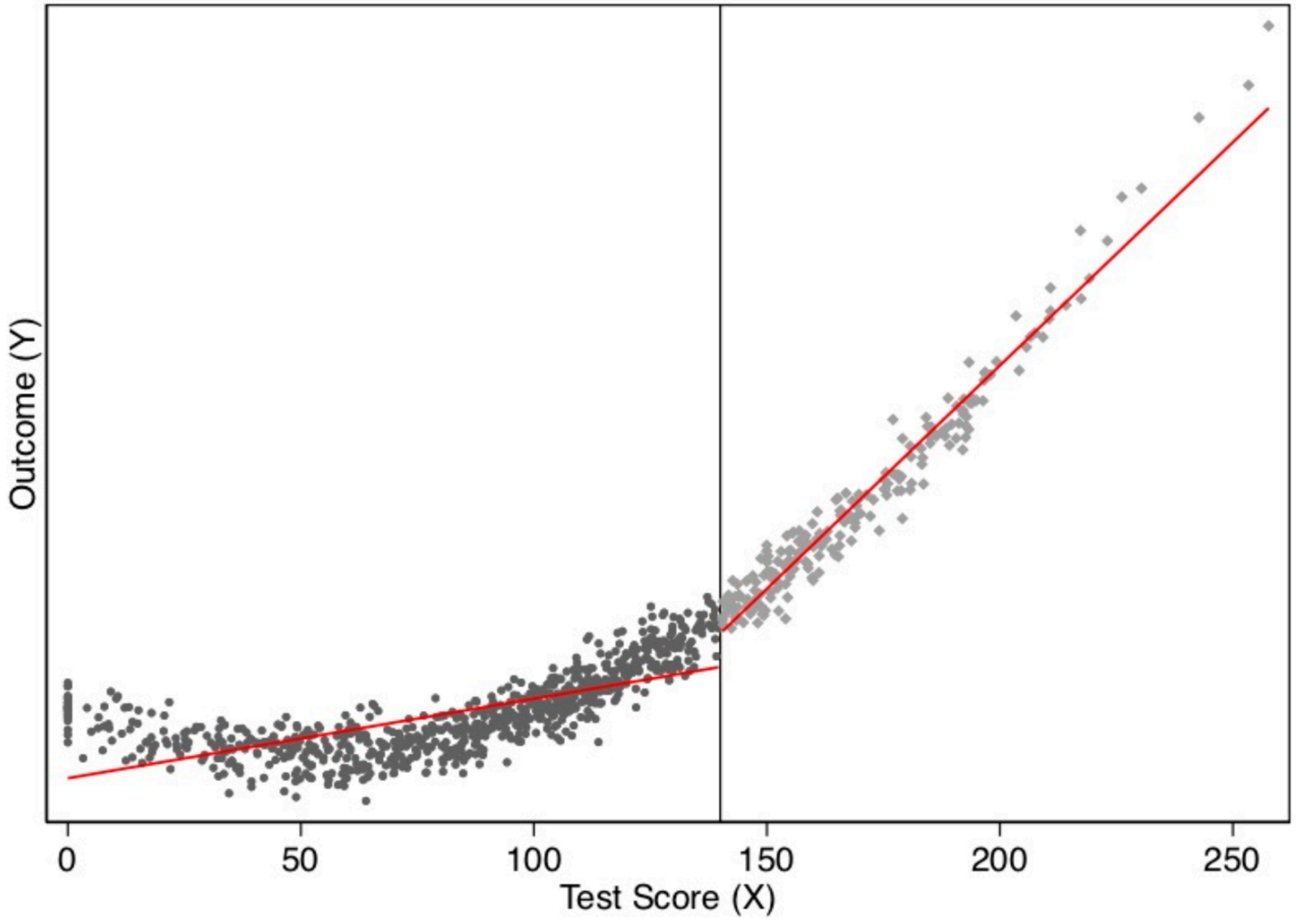


Figure 6.9: Simulated nonlinear data from Stata. Figure attributed to Marcelo Perraillon.

Sharp RDD
Step 3: Functional Form

- What happens when the data-generation process is non-linear?
  - Use p^{th} order polynomial
  - Use non-parametric kernel
- If we have a non-linear specification (or relationship)
  - $E[Y_i^0 \mid X_i] = f(X_i)$
- If we have $f(X_i)$ as a smooth, continuous function, we fit the model
  - $Y_i = f(X_i) + \delta D_i + \eta_i$
  - Use p^{th} order polynomial or non-parametric kernel

Sharp RDD

Step 3: Functional Form - p^{th} order

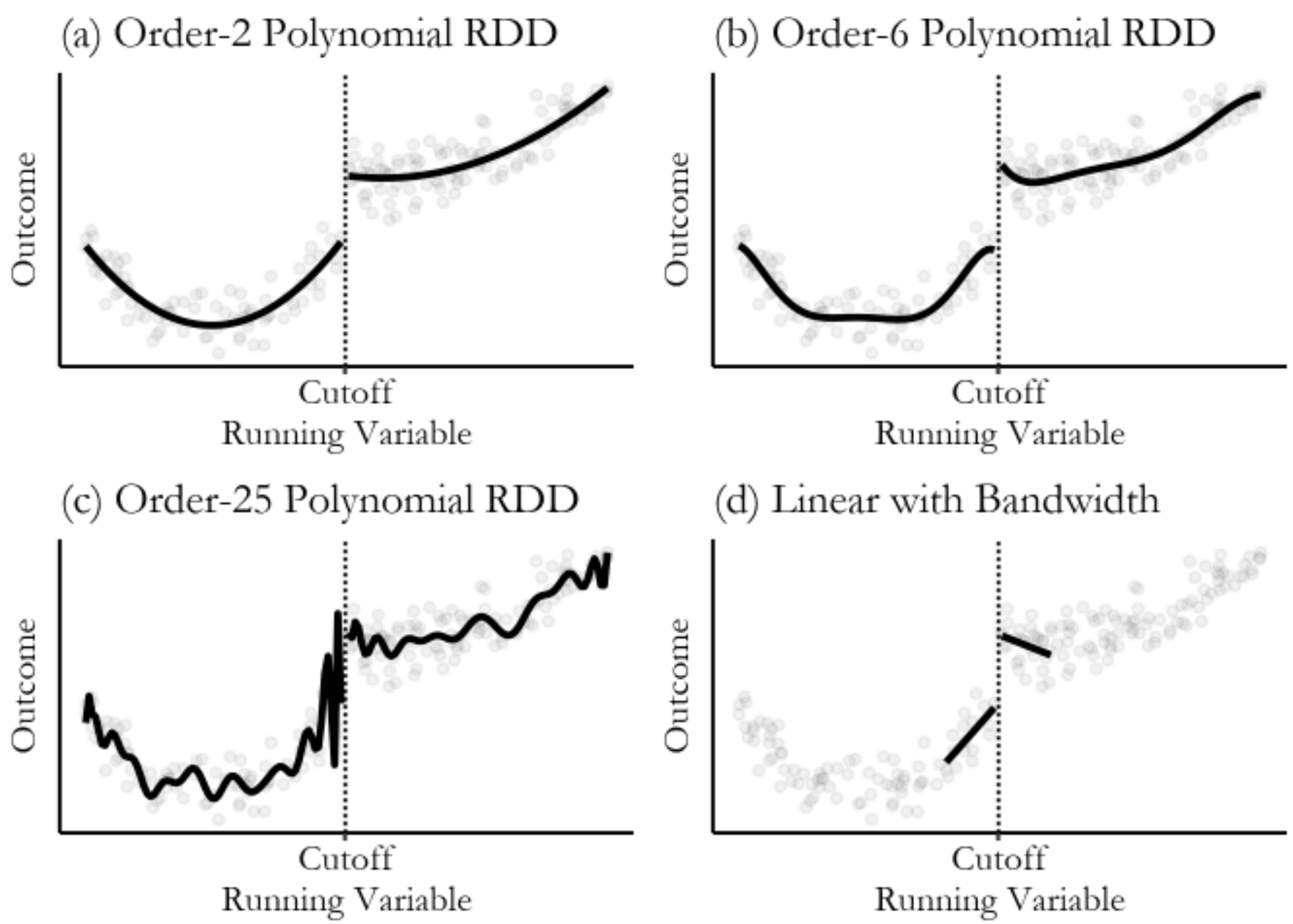
- Using a p^{th} order model for $f(X_i)$
  - $Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \delta D_i + \eta_i$
- Key Note:
  - Higher-order polynomials may result in overfitting which introduces bias
  - German and Imbens (2019) suggest using linear or quadratic $(p = 2)$ for local linear regressions
-

Sharp RDD

Step 3: Functional Form - p^{th} order

- Huntington-Klein (2021) provides a nice example of overspecification with higher polynomial orders

- Try to stick with linear or quadratic



Sharp RDD

Step 3: Functional Form - p^{th} order

- Allow for $f(X_i)$ to differ on both sides of the cutoff
  - $E[Y_i^0 | X_i] = \alpha + \beta_{01}\tilde{X}_i + \ldots + \beta_{0p}\tilde{X}_i^p$
  - $E[Y_i^1 | X_i] = \alpha + \delta + \beta_{11}\tilde{X}_i + \ldots + \beta_{1p}\tilde{X}_i^p$
  - Where $\tilde{X}_i = (X_i - c_o)$
- Using potential outcomes switching equation
  - $E[Y | X] = E[Y^0 | X] + \left(E[Y^1 | X] - E[Y^0 | X]\right)D$

Sharp RDD

Step 3: Functional Form - p^{th} order

- Derive the regression model
  - $Y_i = \alpha + \beta_{01}\tilde{X}_i + \ldots + \beta_{0p}\tilde{X}_i^p + \delta D_i + \beta_{11}D_i\tilde{X}_i + \ldots + \beta_{1p}D_i\tilde{X}_i^p + \varepsilon_i$
  - We interact the continuous variable, $\tilde{X}$ with our treatment variable D
  - Allows for a discontinuous jump and p^{th} order polynomial
  - Our treatment effect of interest is δ at c_0

Sharp RDD

Step 3: Functional Form - Non-parametric Kernel

- The other non-linear option is a non-parametric kernel
  - German and Imbens (2019) suggest avoiding p^{th} order polynomials beyond quadratic when estimating local regressions
  - Kernel regressions are an alternative to p^{th} order polynomials, but there are draw backs
- Kernel Regression
  - A kernel is a function describing how values are weighted
  - A kernel regression is a weighted regression within a specified window

Sharp RDD

Step 3: Functional Form - Non-parametric Kernel

- A rectangle kernel gives the same weight within the bandwidth (window) along the running variable X
- You can see that the true discontinuous jump is AB, but the rectangle kernel estimates $A'B'$, which is upwards biased

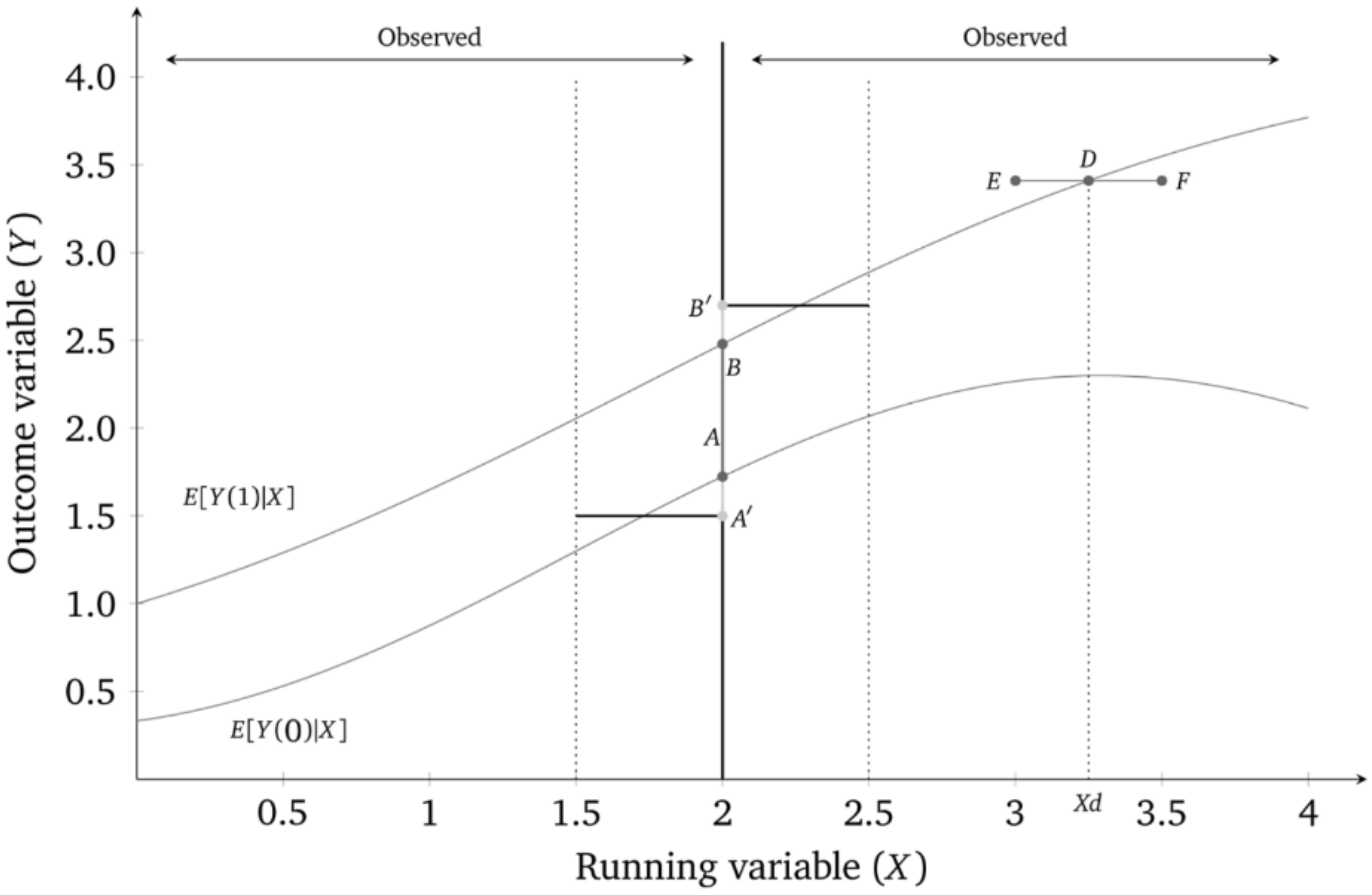


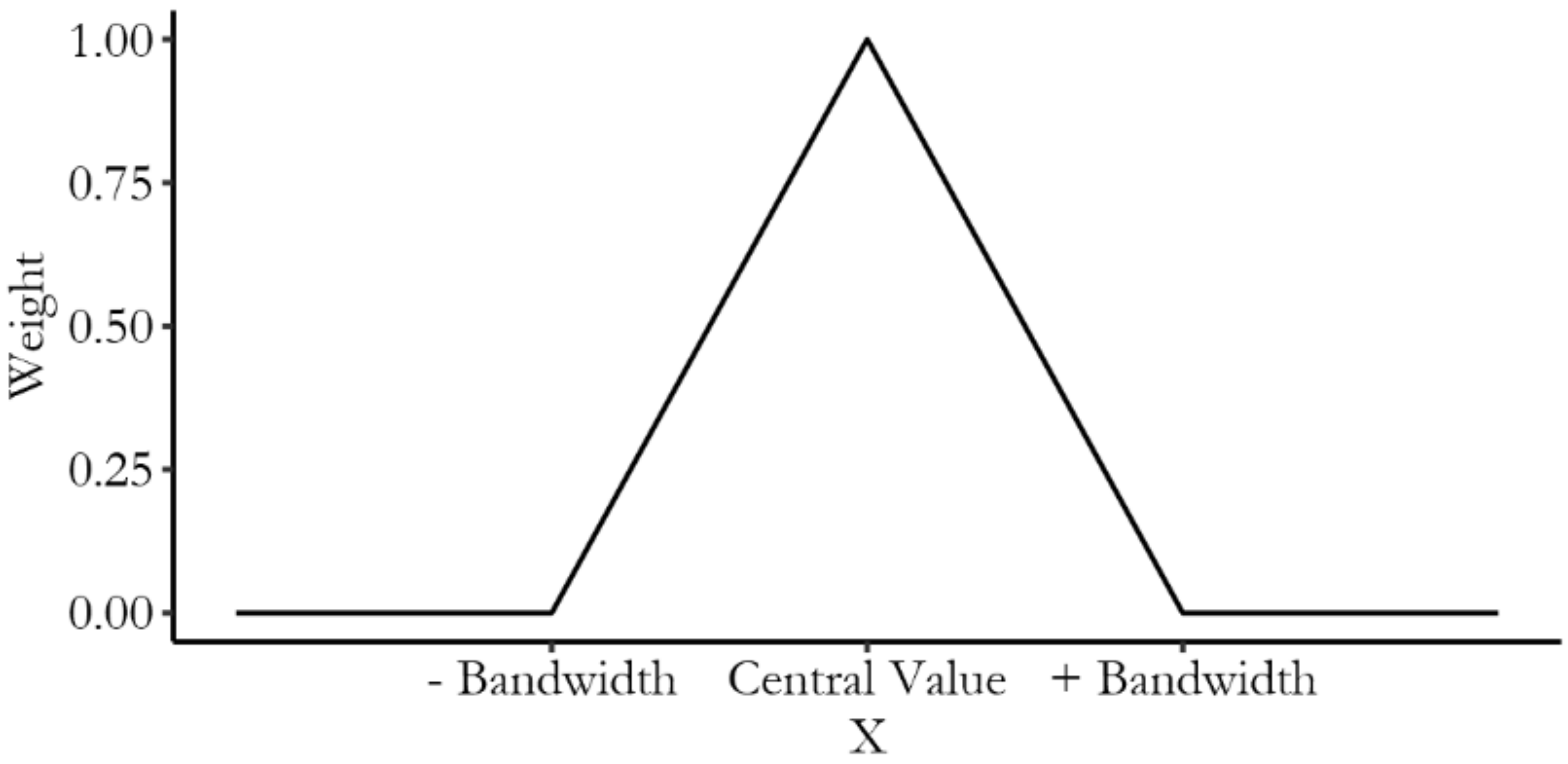
Figure 6.11: Simulated nonlinear data from Stata. Figure attributed to Marcelo Perraillon.

Sharp RDD

Step 3: Functional Form - Non-parametric Kernel

- A triangular kernel gives more weight within the bandwidth (window) as the running variable X gets closer to the cutoff

- One way to deal with the boundary issue is to run a local linear nonparametric regression (Hahn, Todd, and Klaauw, 2001)



Sharp RDD

Step 3: Functional Form - Non-parametric Kernel

- A typical kernel regression is

$$(\hat{a}, \hat{b}) = argmin_{a,b} \sum \left(y_i - a - b(x_i - c_0)\right)^2 K\left(\frac{x_i - c_0}{h}\right) 1(x_i > c_0)$$

  - Where $argmin_{a,b} \sum \left(y_i - a - b(x_i - c_0)\right)^2$ is just minimizing the sum of squares
  - Where $K\left(\frac{x_i - c_0}{h}\right)$ is our kernel function and h is our width of the window
  - Where $1(x_i > c_0)$ is a binary function (on-off switch)

Sharp RDD

Inference

- The standard practice in RDD is to estimate the causal effects using local polynomial regressions around the cutoff (this include p=1)
- Variance-Bias Trade-off
  - If you narrow a window enough, you will minimize misspecification bias, but you will have larger variance due to smaller sample size
  - Expand the window will reduce variance, but increase misspecification bias
- What if the window cannot be narrowed
  - Heteroskedasticity-robust confidence intervals (RDHonest in R)
  - May use random inference

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- What is the impact of universal health insurance for the elderly?
  - Does better insurance cause better health outcomes?
    - There is a lot self-selection into health insurance or higher quality insurance
  - Card, et al. (2008) find that Medicare increases access to health insurance for underserved groups
- Running variable and cutoff
  - Age is the running variable
  - Sharp cutoff at age of 65 for Medicare eligibility
  - If the continuity assumption holds, then there is a discontinuous jump in health insurance at age 65

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- The authors specify the model for causal effect of health-care insurance on health-care utilization
  - $y_{ija} = X_{ija}\alpha + f_j(\alpha, \beta) + \sum_K C^k_{ija}\delta^k + u_{ija}$
  - Where i is individual, j is socioeconomic status, and a is age
  - y_{ija} is health care usage
  - X_{ija} is a set of covariates
  - $f_j(\alpha, \beta)$ is a smooth function representing the age profiles of outcome y for group j
  - C^k_{ija} are k number of characteristics of insurance coverage, such as copayments, etc.

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- Let's assume that insurance can be characterized by two characteristics (k=2)
    - C^1 is any coverage (0,1) and C^2 is insurance generosity (0,1)
- Card, Dobkin, and Maestas (2008) estimate the following two equations
    - $C^1_{ija} = X_{ija}\beta^1_j + g^1_j(a) + D_a\pi^1_j + v^1_{ija}$
    - $C^2_{ija} = X_{ija}\beta^2_j + g^2_j(a) + D_a\pi^2_j + v^2_{ija}$
    - Where β^k_j are group-specific coefficients of the set of covariates
    - Where $g^k_j(a)$ is a smooth age profiles of group j
    - Where D_a is a binary equal to 1 if respondent is greater than or equal to 65

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- Let's combine our 3 equations
  - $y_{ija} = X_{ija}\left(\alpha_j + \beta_j^1\delta_j^1 + \beta_j^2\delta_j^2\right) + h_j(a) + D_a\pi_j^y + v_{ija}^y$
  - Where $h_j(a) = f_j(a) + \delta^1 g_j^1(a) + \delta^2 g_j^2(a)$ is the reduced form age profile
  - Where $\pi_j^y = \pi_j^1\delta^1 + \pi_j^2\delta^2$
  - The magnitude depends upon the size of insurance changes at 65 $(\pi_j^1$ and $\pi_j^2)$ and on the associated causal effects of $(\delta^1$ and $\delta^2)$
  - π_j^y is a linear combination of discontinuities in coverage and generosity
    - $\pi_j^y = \delta^0 + \delta^1\pi_j^1 + \delta^2\pi_j^2 + e_j$

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- 12.3% of the sample of 63-64 year olds had Medicare which is likely due to disability (1) and Medicare jumps almost 60 percentage points at 65 (2)
- Any insurance jumps 9.5 percentage points (3) and -2.9 percentage point drop in private coverage
- Non-white high school dropouts experience a 21.5 percentage points increase in any insurance

	On Medicare		Any insurance		Private coverage		2+ Forms coverage		Managed care	
	Age 63–4 (1)	RD at 65 (2)	Age 63–4 (3)	RD at 65 (4)	Age 63–4 (5)	RD at 65 (6)	Age 63–4 (7)	RD at 65 (8)	Age 63–4 (9)	RD at 65 (10)
Overall sample	12.3	59.7 (4.1)	87.9	9.5 (0.6)	71.8	−2.9 (1.1)	10.8	44.1 (2.8)	59.4	−28.4 (2.1)
Classified by ethnicity and education:										
White non-Hispanic:										
High school dropout	21.1	58.5 (4.6)	84.1	13.0 (2.7)	63.5	−6.2 (3.3)	15.0	44.5 (4.0)	48.1	−25.0 (4.5)
High school graduate	11.4	64.7 (5.0)	92.0	7.6 (0.7)	80.5	−1.9 (1.6)	10.1	51.8 (3.8)	58.9	−30.3 (2.6)
At least some college	6.1	68.4 (4.7)	94.6	4.4 (0.5)	85.6	−2.3 (1.8)	8.8	55.1 (4.0)	69.1	−40.1 (2.6)
Minority:										
High school dropout	19.5	44.5 (3.1)	66.8	21.5 (2.1)	33.2	−1.2 (2.5)	11.4	19.4 (1.9)	39.1	−8.3 (3.1)
High school graduate	16.7	44.6 (4.7)	85.2	8.9 (2.8)	60.9	−5.8 (5.1)	13.6	23.4 (4.8)	54.2	−15.4 (3.5)
At least some college	10.3	52.1 (4.9)	89.1	5.8 (2.0)	73.3	−5.4 (4.3)	11.1	38.4 (3.8)	66.2	−22.3 (7.2)
Classified by ethnicity only:										
White non-Hispanic (all)	10.8	65.2 (4.6)	91.8	7.3 (0.5)	79.7	−2.8 (1.4)	10.4	51.9 (3.5)	61.9	−33.6 (2.3)
Black non-Hispanic (all)	17.9	48.5 (3.6)	84.6	11.9 (2.0)	57.1	−4.2 (2.8)	13.4	27.8 (3.7)	48.2	−13.5 (3.7)
Hispanic (all)	16.0	44.4 (3.7)	70.0	17.3 (3.0)	42.5	−2.0 (1.7)	10.8	21.7 (2.1)	52.9	−12.1 (3.7)

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- Testing continuity assumption indirectly
- We observe $E[Y^0 | age]$ before 65 and $E[Y^1 | age]$ for 65+
- $\lim_{65 \leftarrow a} E[Y^1 | a] - \lim_{a \rightarrow 65} E[Y^0 | a]$
- Test employment, since it could be confounding and discontinuous at 65
- Jumps in employment measures are small and insignificant

TABLE 2—ESTIMATED DISCONTINUITIES IN EMPLOYMENT MEASURES AT AGE 65

	Data from NHIS		Data from March CPS			
	Employed (1)	Full time (2)	Employed (3)	Hours/Wk (4)	Retired (5)	Earnings (1000s) (6)
Overall sample	0.3	0.8	1.8	1.3	−1.6	−0.2
	(0.8)	(1.2)	(1.1)	(0.7)	(1.7)	(0.4)
Classified by ethnicity and education:						
White non-Hispanic:						
High school dropout	1.0	1.1	1.9	1.8	−3.7	0.3
	(1.4)	(1.9)	(1.1)	(0.5)	(2.4)	(0.8)
High school graduate	1.7	2.4	2.3	1.4	−2.5	−0.2
	(1.0)	(1.8)	(1.5)	(0.7)	(2.3)	(0.5)
At least some college	−1.6	−0.4	0.9	0.9	−0.6	−1.5
	(1.5)	(2.0)	(1.4)	(1.0)	(1.7)	(0.8)
Minority:						
High school dropout	2.6	1.5	1.2	0.7	−0.3	0.1
	(1.6)	(2.0)	(1.3)	(0.5)	(2.4)	(0.4)
High school graduate	0.0	0.2	6.1	2.5	−5.3	0.7
	(3.2)	(2.6)	(2.3)	(1.1)	(1.5)	(0.6)
At least some college	−4.6	−2.6	−1.2	0.1	2.4	0.7
	(2.6)	(3.0)	(1.5)	(0.6)	(1.8)	(1.4)
Classified by ethnicity only:						
White non-Hispanic	0.2	0.9	1.6	1.3	−1.7	−0.5
	(0.9)	(1.3)	(1.2)	(0.7)	(2.0)	(0.5)
Black non-Hispanic (all)	1.3	0.1	5.7	2.2	−4.6	1.3
	(2.6)	(2.0)	(1.4)	(0.5)	(0.9)	(0.4)
Hispanic (all)	−0.5	0.7	−0.4	0.8	−0.7	−1.1
	(2.0)	(2.0)	(1.7)	(0.9)	(1.7)	(1.2)
Classified by gender:						
Men	1.7	3.0	2.8	1.7	−1.6	−0.5
	(1.3)	(1.6)	(1.5)	(0.9)	(1.6)	(0.6)
Women	−1.0	−1.2	0.9	0.9	−1.4	0.1
	(1.1)	(1.2)	(1.0)	(0.6)	(2.0)	(0.3)

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- Testing continuity assumption indirectly with employment

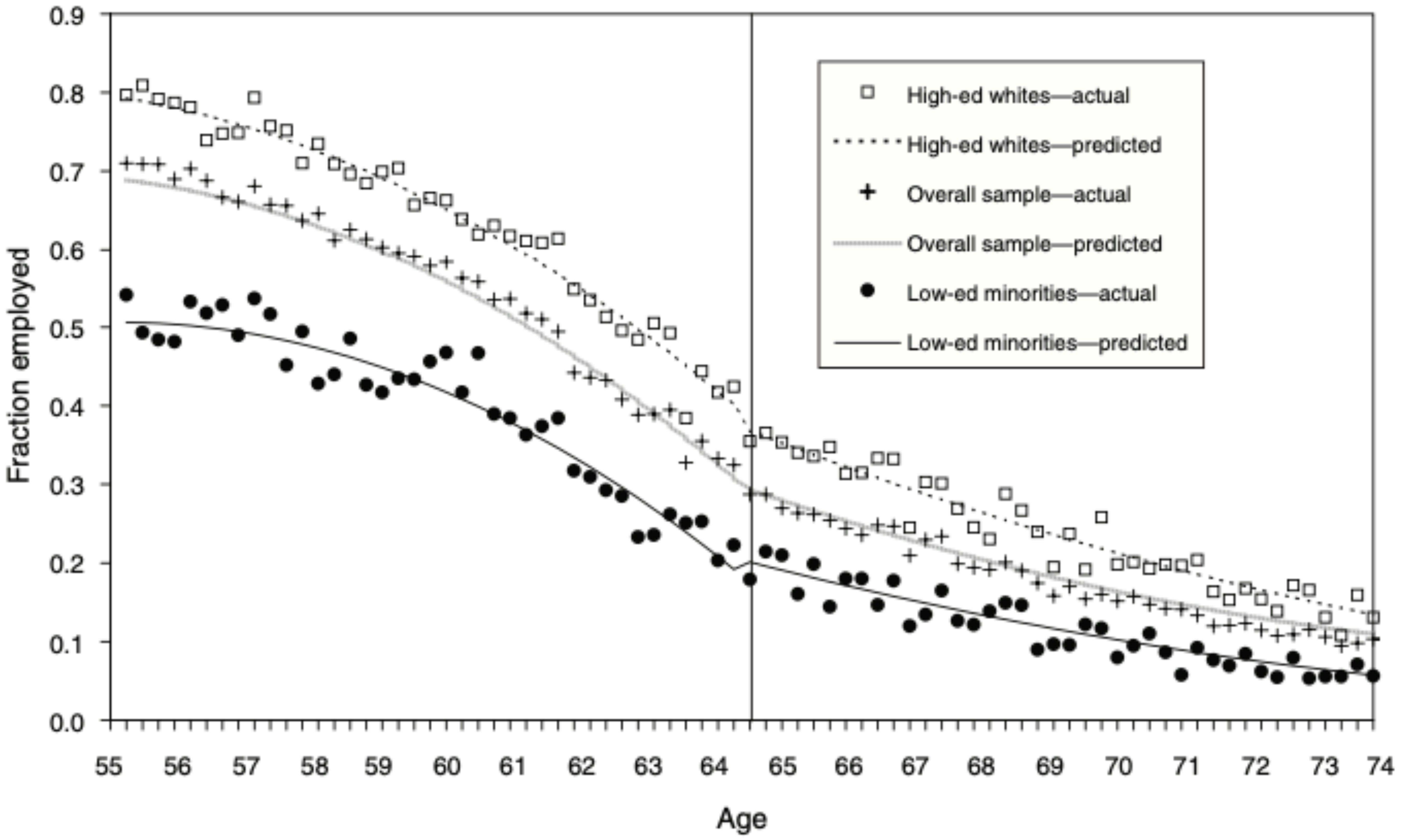


FIGURE 2. EMPLOYMENT RATES BY AGE AND DEMOGRAPHIC GROUP (1992–2003 NHIS)

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- Jumps in access to healthcare at the cutoff but much smaller magnitudes
- Delaying care last year drops 1.8 percentage points and not getting care drops 1.3 percentage points

TABLE 3—MEASURES OF ACCESS TO CARE JUST BEFORE 65 AND ESTIMATED DISCONTINUITIES AT 65

	1997–2003 NHIS				1992–2003 NHIS			
	Delayed care last year		Did not get care last year		Saw doctor last year		Hospital stay last year	
	Age 63–64 (1)	RD at 65 (2)	Age 63–64 (3)	RD at 65 (4)	Age 63–64 (5)	RD at 65 (6)	Age 63–64 (7)	RD at 65 (8)
Overall sample	7.2	−1.8 (0.4)	4.9	−1.3 (0.3)	84.8	1.3 (0.7)	11.8	1.2 (0.4)
Classified by ethnicity and education:								
White non-Hispanic:								
High school dropout	11.6	−1.5 (1.1)	7.9	−0.2 (1.0)	81.7	3.1 (1.3)	14.4	1.6 (1.3)
High school graduate	7.1	0.3 (2.8)	5.5	−1.3 (2.8)	85.1	−0.4 (1.5)	12.0	0.3 (0.7)
At least some college	6.0	−1.5 (0.4)	3.7	−1.4 (0.3)	87.6	0.0 (1.3)	9.8	2.1 (0.7)
Minority:								
High school dropout	13.6	−5.3 (1.0)	11.7	−4.2 (0.9)	80.2	5.0 (2.2)	14.5	0.0 (1.4)
High school graduate	4.3	−3.8 (3.2)	1.2	1.5 (3.7)	84.8	1.9 (2.7)	11.4	1.8 (1.4)
At least some college	5.4	−0.6 (1.1)	4.8	−0.2 (0.8)	85.0	3.7 (3.9)	9.5	0.7 (2.0)
Classified by ethnicity only:								
White non-Hispanic	6.9	−1.6 (0.4)	4.4	−1.2 (0.3)	85.3	0.6 (0.8)	11.6	1.3 (0.5)
Black non-Hispanic (all)	7.3	−1.9 (1.1)	6.4	−0.3 (1.1)	84.2	3.6 (1.9)	14.4	0.5 (1.1)
Hispanic (all)	11.1	−4.9 (0.8)	9.3	−3.8 (0.7)	79.4	8.2 (0.8)	11.8	1.0 (1.6)

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- There is a jump in hospital admissions at the cutoff of 65
- This is statistically significant across racial/ethnic groups
- There are bigger jumps in non-ER admissions across racial/ethnic groups

TABLE 4—HOSPITAL ADMISSIONS AND INSURANCE COVERAGE AT AGE 65: CALIFORNIA, FLORIDA, AND NEW YORK

	All		Whites		Hispanics		Blacks	
	Rate age 60–64 (1)	RD at 65 (2)	Rate age 60–64 (3)	RD at 65 (4)	Rate age 60–64 (5)	RD at 65 (6)	Rate age 60–64 (7)	RD at 65 (8)
Hospital admissions								
All admissions	1,443	7.57	1,407	7.74	1,262	9.47	2,008	4.39
		(0.29)		(0.33)		(0.55)		(0.71)
By route into hospital								
ER admission	761	3.30	688	3.70	774	2.63	1,313	1.93
		(0.39)		(0.40)		(0.92)		(0.95)
Non–ER admission	682	12.16	718	11.51	488	19.89	695	8.92
		(0.46)		(0.49)		(1.05)		(1.04)
By admission diagnosis								
Chronic ischemic heart disease	83	11.58	90	11.05	59	18.45	66	8.29
		(0.96)		(1.16)		(2.45)		(2.78)
AMI	48	4.41	50	5.31	38	3.90	45	−3.43
		(1.43)		(1.65)		(3.33)		(4.78)
Heart failure	56	0.44	45	2.33	62	−4.85	130	−1.47
		(1.11)		(1.24)		(2.63)		(2.43)
Chronic bronchitis	34	7.50	36	6.50	19	9.76	38	13.05
		(1.51)		(1.52)		(5.58)		(4.43)
Osteoarthrosis	34	26.97	38	27.16	18	29.27	27	22.08
		(1.39)		(1.64)		(5.05)		(4.01)
Pneumonia	34	2.44	32	2.05	30	3.39	51	3.81
		(1.42)		(1.74)		(4.34)		(3.21)

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- We can visualize the discontinuous jumps in admission and admission for hip and knee replacements

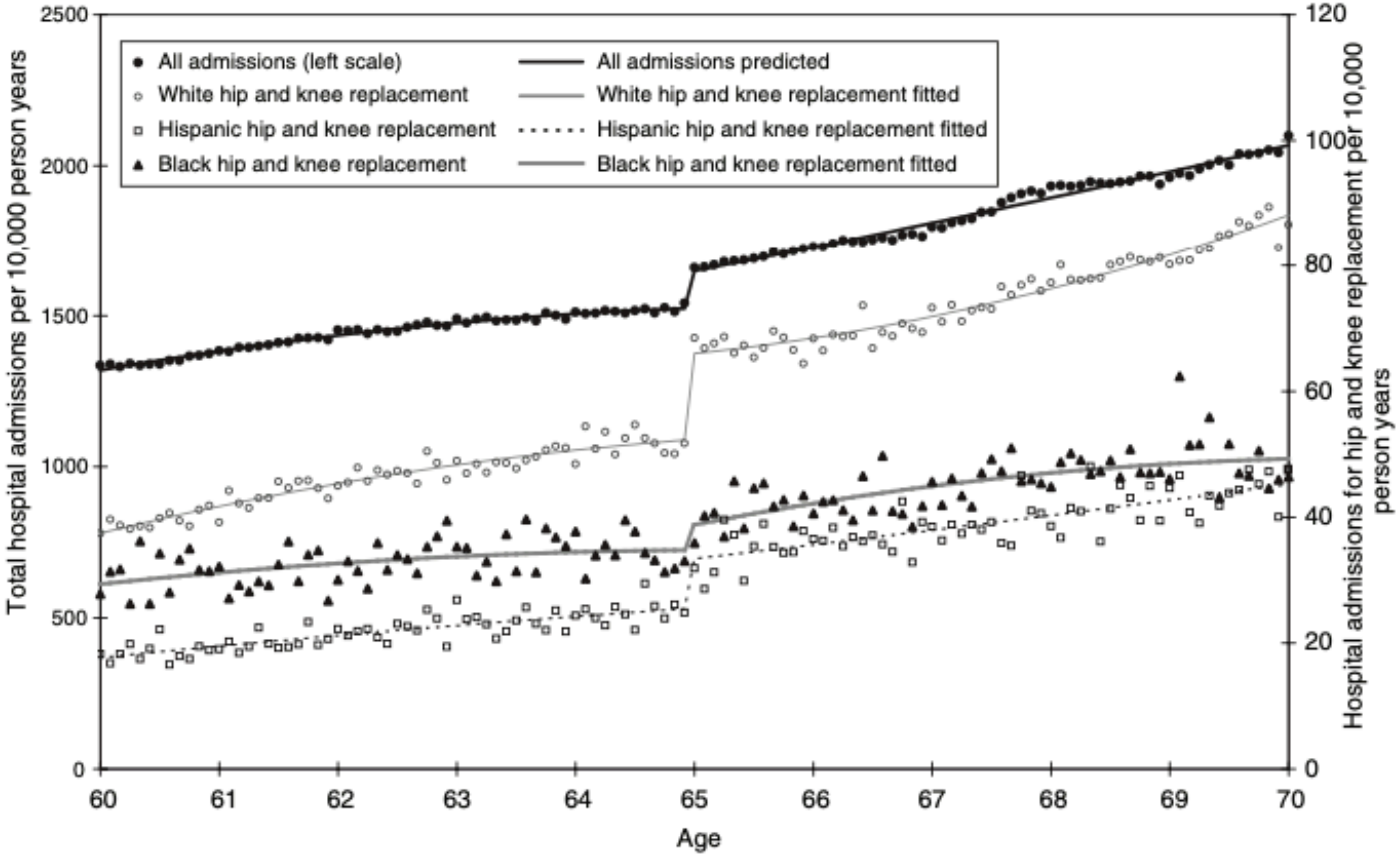


FIGURE 3. HOSPITAL ADMISSION RATES BY RACE/ETHNICITY

Example: Medicare and Universal Health Insurance

Card, Dobkin, and Maestas (2008)

- We see jumps in admission for private hospitals in CA
- We see two other patterns
    - A big declines in county hospital admissions, a redistribution of patients
    - No change in Kaiser HMO admissions and remain in managed care before and after 65 - no incentive to get high-cost Medicare procedures

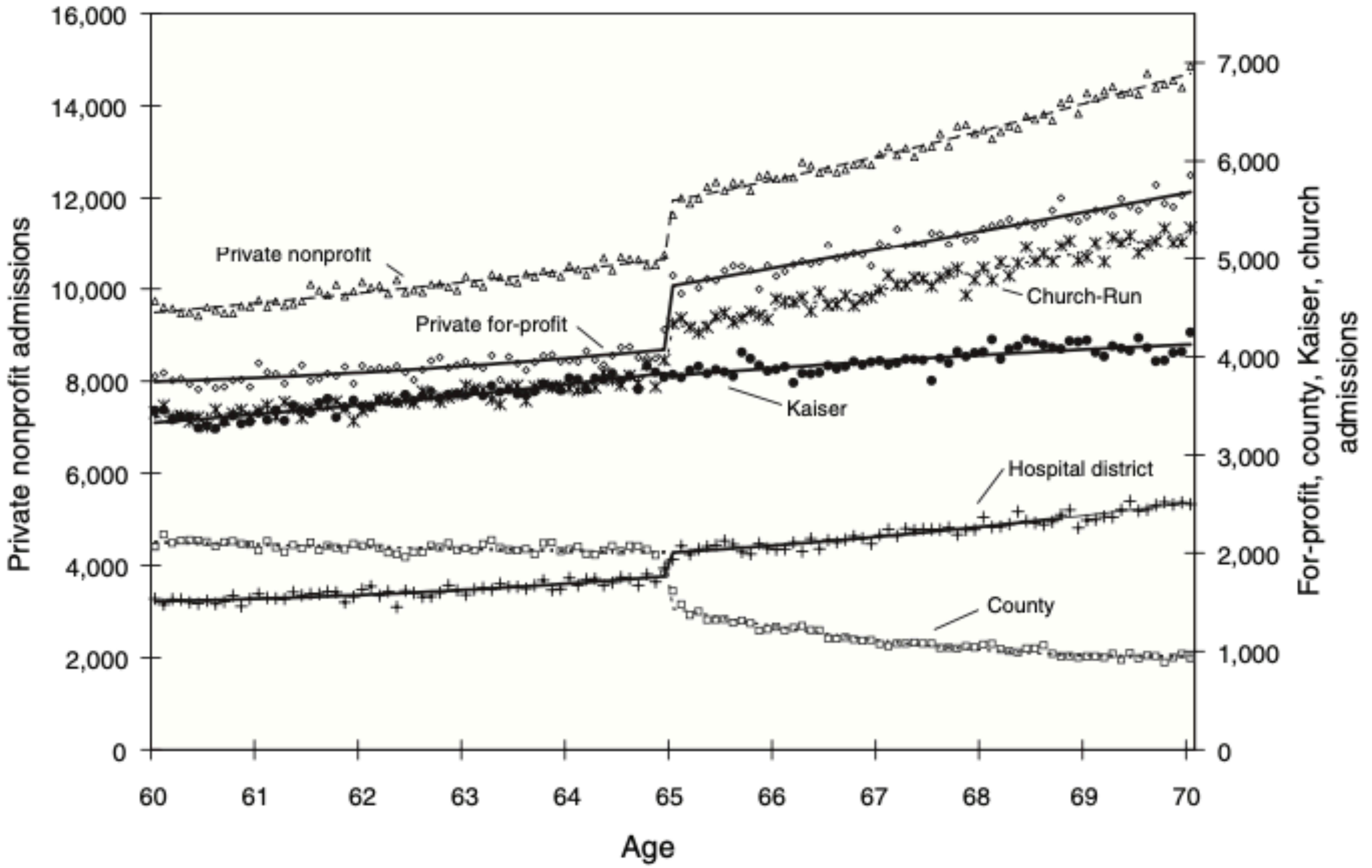


FIGURE 4. HOSPITAL ADMISSION IN CALIFORNIA BY OWNERSHIP TYPE (1992–2002)

Fuzzy RDD

Fuzzy RDD

Overview

- Fuzzy RDD differs from Sharp RDD
  - The running variable is no longer deterministic of assignment to treatment
  - The running variable is probabilistic in assignment to treatment
  - The conditional probability of treatment is discontinuous as the running variable approaches at the limit c_0
- Fuzzy RDD has a lot of similarities to Instrumental Variables
  - Very similar assumptions

Fuzzy RDD
Overview

- The running variable is associated with the probability of assignment to treatment
  - There is a discontinuous jump in probability to assignment to treatment at c_0

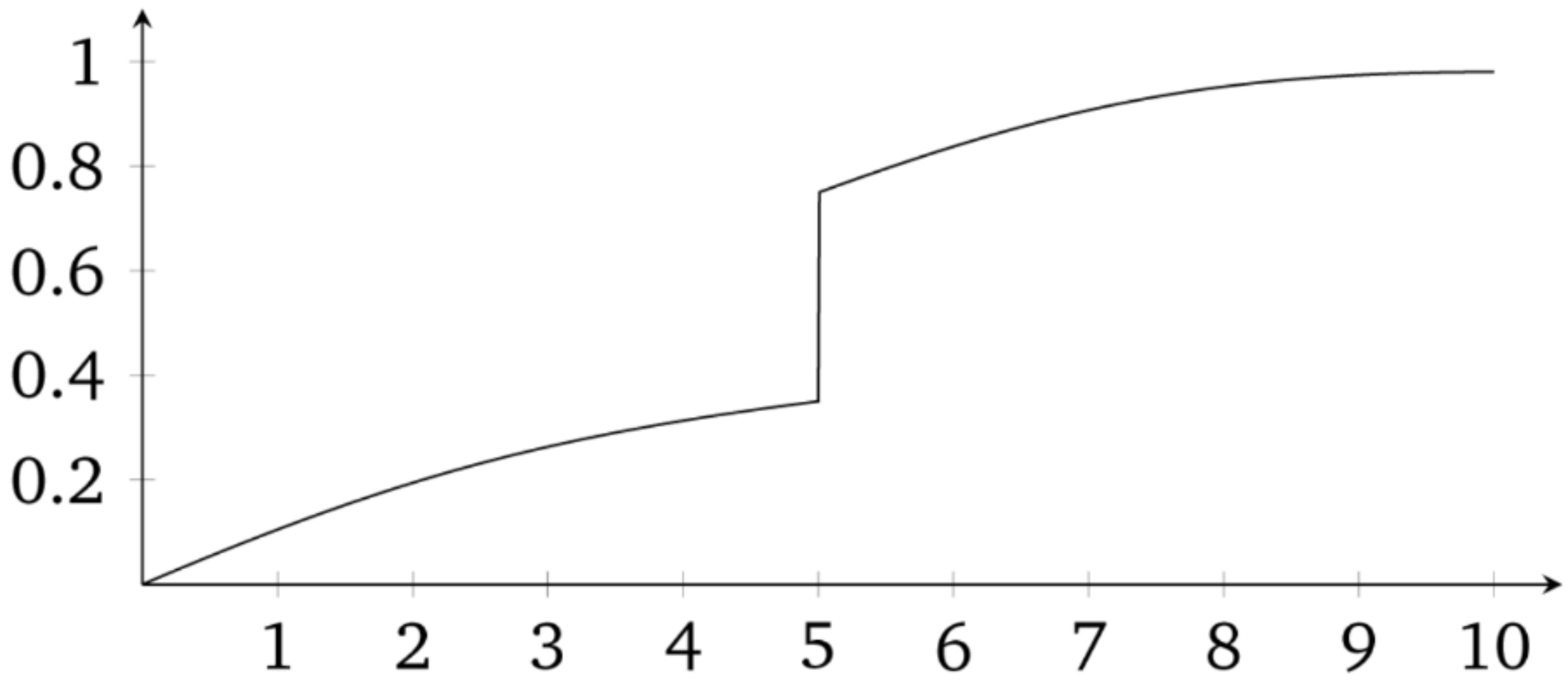
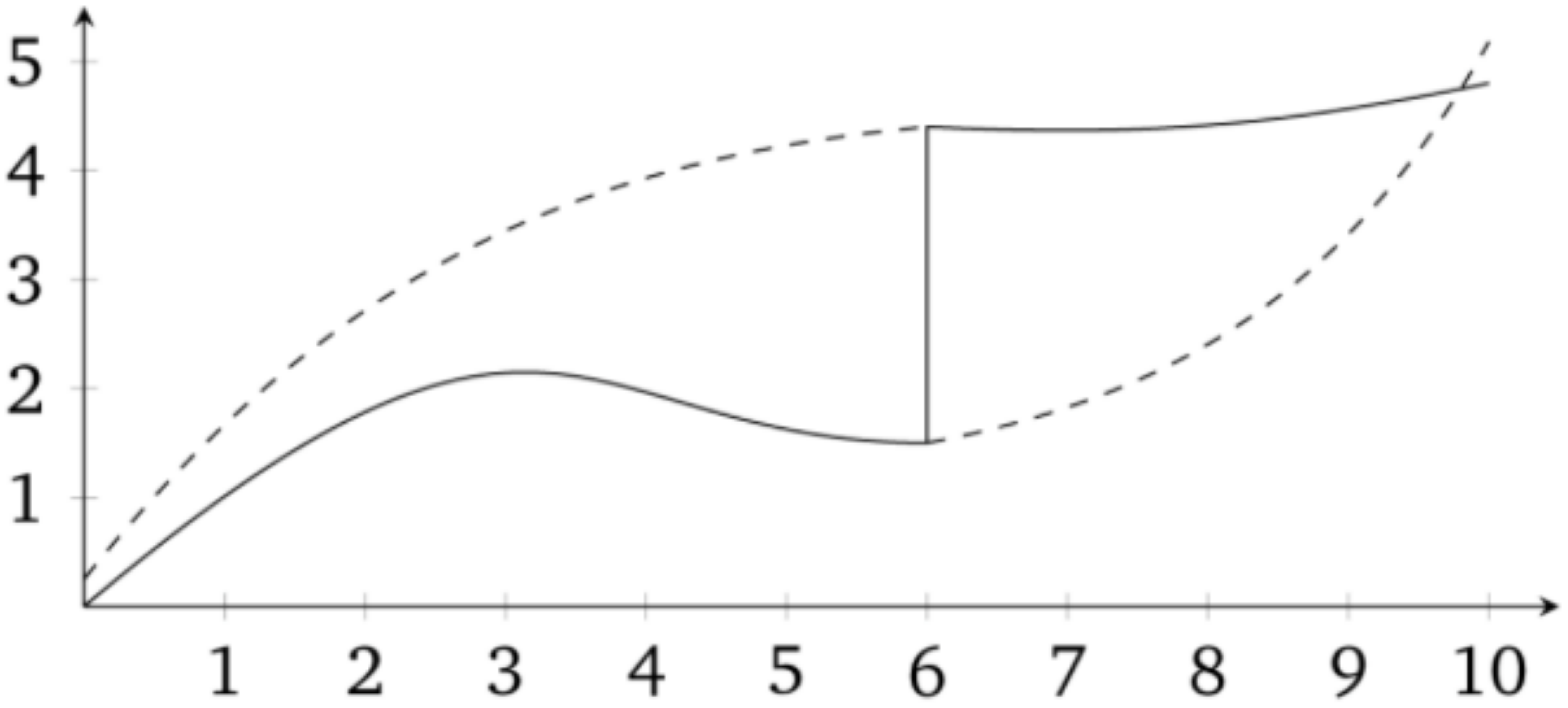


Figure 6.14: Vertical axis is the probability of treatment for each value of the running variable.

Fuzzy RDD
Overview

- When we have an increase in probability of treatment, we have a Fuzzy RDD
  - $\lim_{c_0 \leftarrow x_i} \Pr(D_i = 1 \,|\, X = c_0) \neq \lim_{x_i \rightarrow c_0} \Pr(D_i = 1 \,|\, X = c_0)$
  - This means that the conditional probability of treatment is discontinuous at the cutoff in the limit
- We still have our continuity assumption, where $E[Y^0 \,|\, X < c_0]$ and $E[Y^1 \,|\, X \geq c_0]$ are smooth and continuous



Fuzzy RDD

The Estimator

- Estimating the Fuzzy RDD is similar to IV
  - $$\delta_{Fuzzy\ RDD} = \frac{\lim_{c_0 \leftarrow X} E[Y \mid X = c_0] - \lim_{X \rightarrow c_0} E[Y \mid X = c_0]}{\lim_{c_0 \leftarrow X} E[D \mid X = c_0] - \lim_{X \rightarrow c_0} E[D \mid X = c_0]}$$
  - Where our first stage is seen in the denominator and reduced form in the numerator
- Estimate
  - We can use a 2SLS for estimating the first stage and then the 2nd stage
  - Think of Z as the instrument that is 1 when $X \geq c_0$ and 0 when $X < c_0$

Fuzzy RDD

Assumptions

- Continuity Assumption (Indirectly Testable)
  - $E[Y^0]$ and $E[Y^1]$ are smooth and continuous functions at $X = c_0$
- Exclusion Restriction Assumption
  - Z works through D and not other sources; the running variable is not associated with other covariates
  - Not directly testable, but we can look for discontinuous jumps in observed covariates like Card, et al. (2008)
- Monotonicity Assumption (Testable)
  - Conditional probability of treatment with the running variable The running variable is increasing or decreasing but not both
- Stable Unit Treatment Value Assumption (Not Testable)
  - No spillovers
- Non-Zero 1st Stage Assumption (Testable)
  - There is an associated between the running variable and the conditional probability of treatment

Fuzzy RDD
1st Stage

- We can estimate the Fuzzy RDD in two stages or with a reduced form
- 1st Stage with no interaction
    - $D_i = \gamma_0 + \gamma_1 X_i + \gamma_2 X_i^2 + \ldots + \gamma_p X_i^p + \pi Z_i + \zeta_i$
    - Where Z_i is 1 if $X \geq c_0$ and 0 if $X < c_0$
    - π is the causal effect of Z_i on D_i
- 1st Stage with interaction
    - We can interact Z_i with D_i to change the slope and the intercept at the cutoff
    - $D_i = \gamma_{00} + \gamma_{01}\tilde{X}_i + \gamma_{02}\tilde{X}_i^2 + \ldots + \gamma_{0p}\tilde{X}_i^p + \pi Z_i + \gamma_{11}\tilde{X}_i D_i + \gamma_{12}\tilde{X}_i^2 D_i + \ldots + \gamma_{1p}\tilde{X}_i^p D_i + \zeta_i$

Fuzzy RDD

2nd Stage

- Once we have fitted our treatment $\hat{D}_i$ in our 1st stage we can estimate the LATE at the cutoff
- 2nd Stage with no interaction
  - Intercept only changes at the cutoff
  - $Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \ldots + \beta_p X_i^p + \delta \hat{D}_i + \eta_i$
- 2nd Stage with interaction
  - Slope and intercept change at the cutoff
  - $Y_i = \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \ldots + \beta_{0p}\tilde{X}_i^p + \delta \hat{D}_i + \beta_{11}\tilde{X}_i\hat{D}_i + \beta_{12}\tilde{X}_i^2\hat{D}_i + \ldots + \beta_{1p}\tilde{X}_i^p\hat{D}_i + \eta_i$

Fuzzy RDD
Reduced Form

- If we want to skip over the full IV model, we can use a reduced form
  - This identifies the impact of the instrument on the outcome of interest
- No interaction reduced form
  - Only the intercept changes at the cutoff
  - $Y_i = \mu + \kappa_1 X_i + \kappa_2 X_i^2 + \ldots + \kappa_p X_i^p + \delta Z_i + \zeta_i$
- Interaction reduced form
  - Slope and the intercept changes at the cutoff
  - $Y_i = \mu + \kappa_{01} \tilde{X}_i + \kappa_{02} \tilde{X}_i^2 + \ldots + \kappa_{0p} \tilde{X}_i^p + \delta Z_i + \kappa_{11} \tilde{X}_i Z_i + \ldots + \kappa_{1p} \tilde{X}_i^p Z_i + \zeta_i$

Challenges to Identification

Challenges to Identification

Overview

- We have many potential avenues where agents can violate the continuity assumption
    - We cannot directly test the continuity assumption because ***we never observe*** $E[Y^0 | D = 1]$ and $E[Y^1 | D = 0]$
- We have three main tests at our disposal to indirectly test the continuity assumption
    - McCrary Test
    - Covariate Balance
    - Placebo Tests

Challenges to Identification

Violations of Continuity Assumption

- Assignment rule is known in advance
- Agents are interested in adjusting around the assignment rule
- Agents have time to adjust to the assignment rule
- Assignment rule is endogenous with other factors that can potentially impact the outcome
- There is nonrandom heaping along the running variable
- Examples
    - Retaking exams
    - Self-reported income
    - Turning 18 as a cutoff rule can be endogenous with graduating from high school, voting, etc.

Challenges to Identification

McCrary Test

- The McCrary Test is an indirect test of the continuity assumption
  - Is there bunching around the cutoff? Are people sorting around the cutoff?
- If there is bunching (or uneven distributions) then people are likely sorting along the running variable around the cutoff
  - There is manipulation
- Null Hypothesis
  - There is continuous density at the cutoff along the running variable and there is no bunching
  - Similar to continuous expected outcomes

Challenges to Identification

McCrary Test

- Test the Null Hypothesis
  - The McCrary Test tests this null hypothesis
  - If we reject the null hypothesis, then there is possible sorting and bunching, which implies selection bias
- How to Estimate McCrary Test
  - Partition the running variable into bins
  - Calculate the frequencies of observations in each bins
  - Treat the frequencies as a dependent variable in a local linear regression
  - You needs lots of observations to distinguish a discontinuity in the density from the noise
- Use RDDensity in Stata

Challenges to Identification

McCrary Test

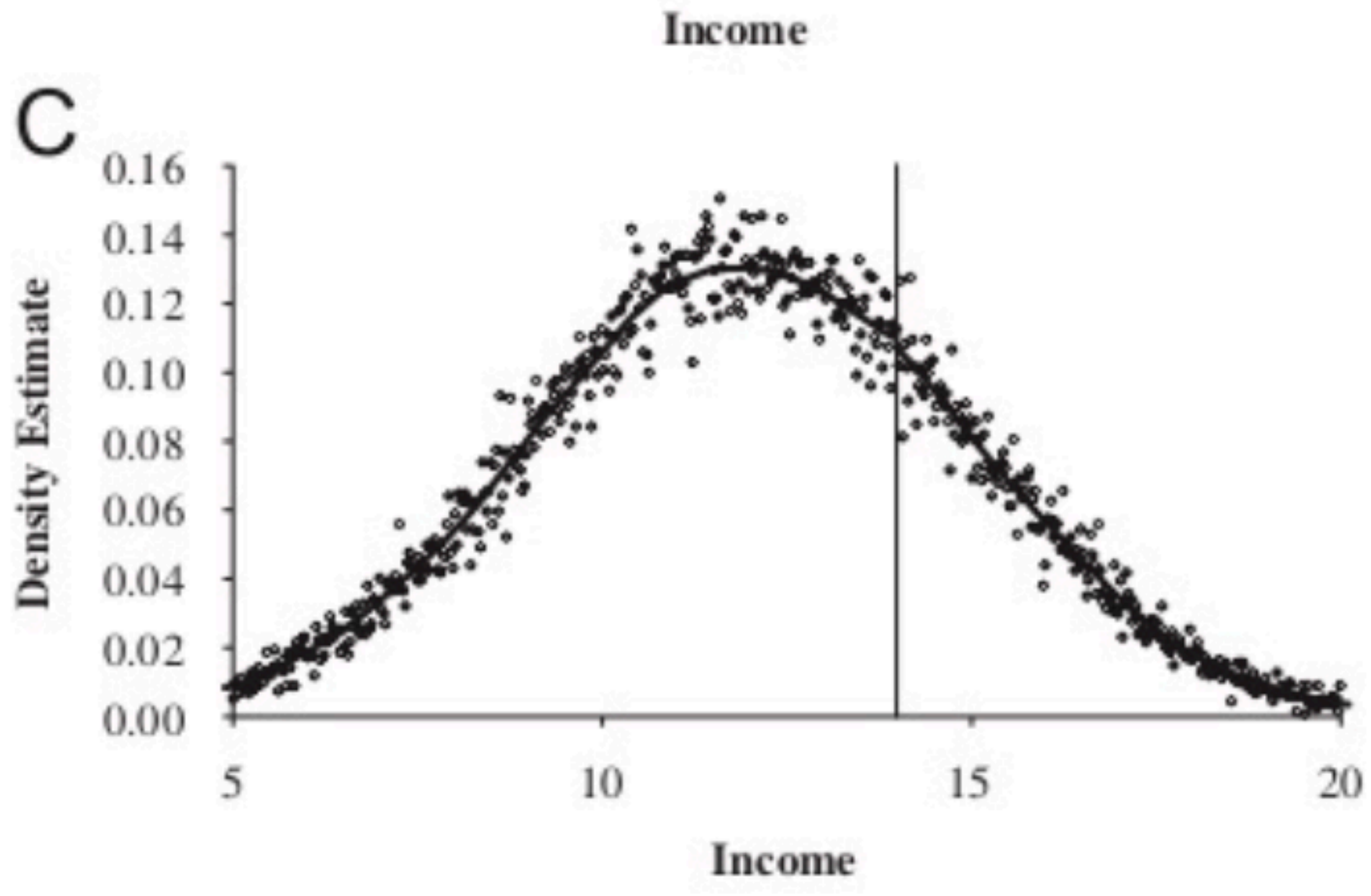


Figure 6.16: A picture with and without a discontinuity in the density from McCrary (2008).

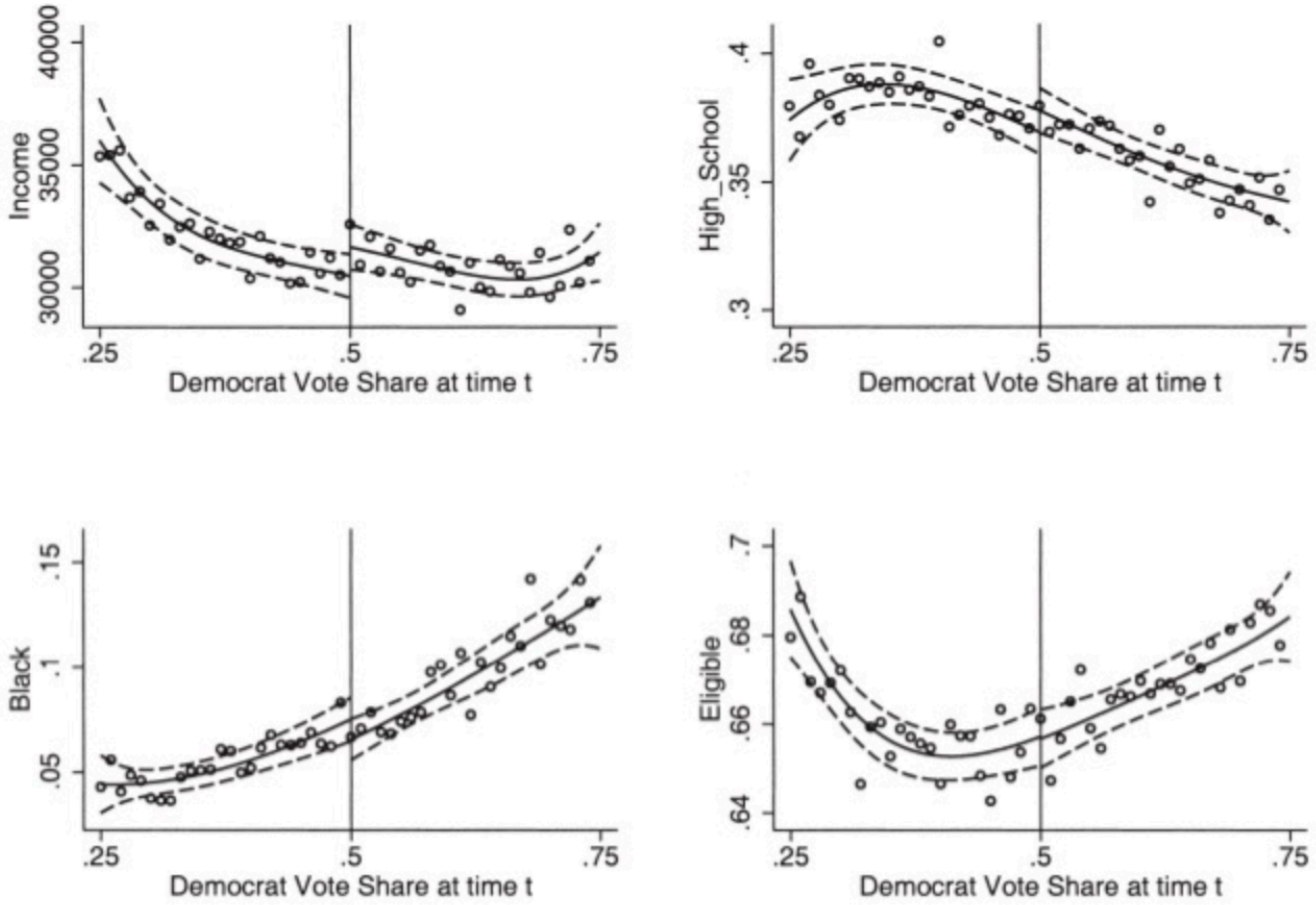
Challenges to Identification

Covariate Balance and Placebo Tests

- Covariate Balance Tests
  - We can indirectly test the continuity assumption with covariate balance
  - There should be no discontinuous jumps at the cutoff, only the treatment and outcome
- Placebo Tests
  - Fake Cutoffs are another indirect test of the continuity assumption
  - We should expect no discontinuous jumps at the fake cutoff in the outcome
  - Imbens and Lemieux (2008) suggest using the median running value as a fake cutoff or c_{fake}

Challenges to Identification

Covariate Balance and Placebo Tests



Challenges to Identification

Nonrandom Heaping

- We have seen how nonrandom sorting around the cutoff on the running variable contaminates the estimator with bias
- Almond, Doyle, Kowalski, and Williams (2010) are interested in medical expenditures on health outcomes
    - If a physician perceives that an intervention will be beneficial then it is likely that the physician will assign treatment to a patient
    - Babies with low-birth weight receive heighten medical attention
- The authors find that 1-year mortality decreases 1 percentage point and considering the mean of 5.5 percent chance, this is a large decrease

Challenges to Identification

Nonrandom Heaping

- Berreca, Guildi, Lindo, and Waddell (2011) note that heaping occurs with baby birth weights
- Heaping
  - When excessive number of units occur at certain points along the running variable
- Birth weights appear to suffer from heaping and round to the nearest 5
- Doctors may be rounding down to get low-birth weight intervention
- This highlights a shortcoming of the McCrary Test, since Almond, et al. (2010) failed to reject the null hypothesis
- A “donut hole” RDD is a potential solution to heaping

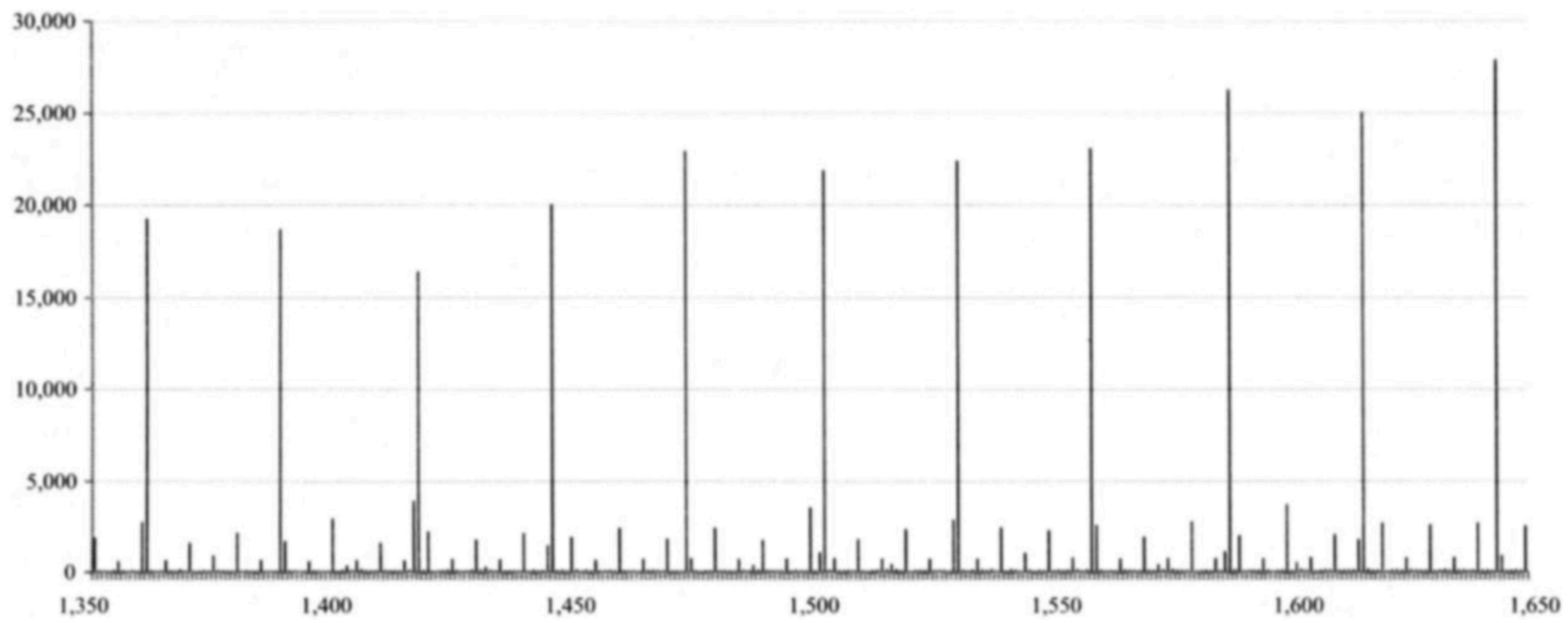


FIGURE I
Frequency of Births by Gram: Population of U.S. Births between 1,350 and 1,650 g
NCHS birth cohort linked birth/infant death files, 1983–1991 and 1995–2003, as described in the text.

Challenges to Identification

Nonrandom Heaping

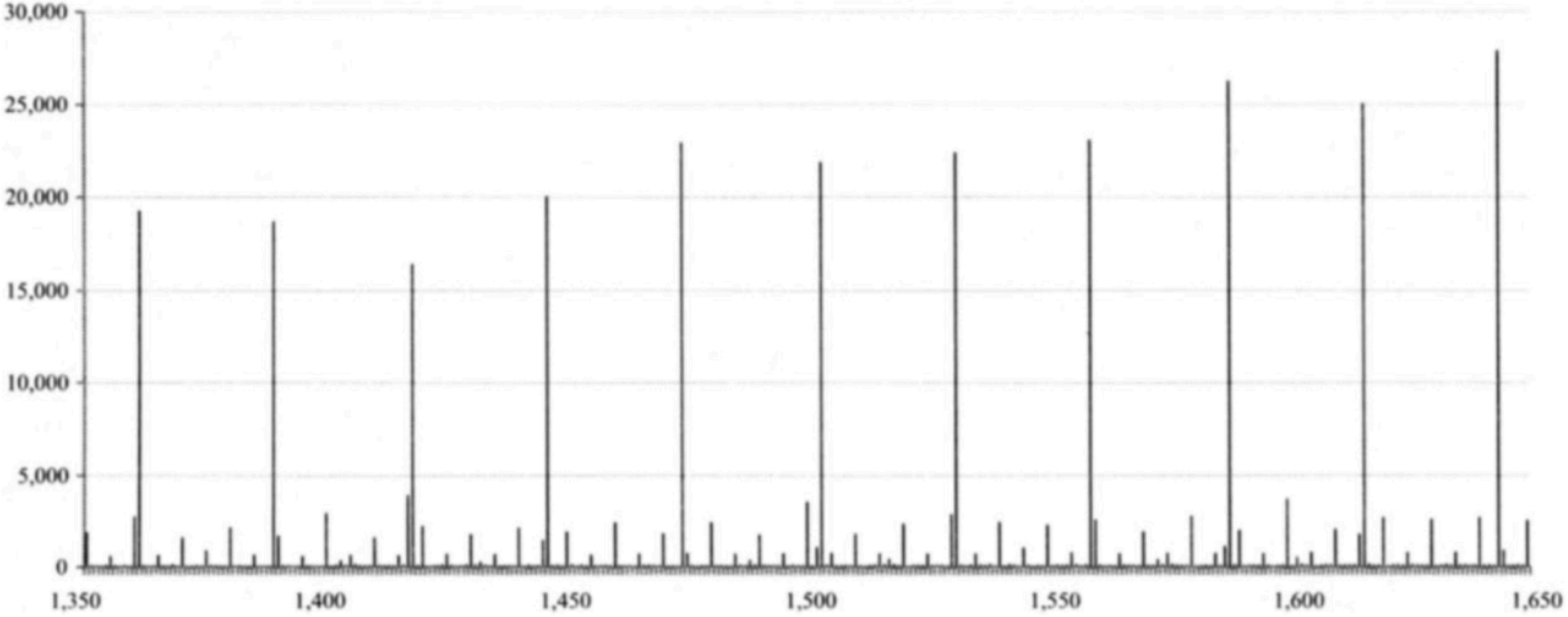


FIGURE I
Frequency of Births by Gram: Population of U.S. Births between 1,350 and 1,650 g

NCHS birth cohort linked birth/infant death files, 1983–1991 and 1995–2003, as described in the text.

RDD Example: Close Elections

Lee, Moretti, and Butler (2004)

- Lee, Moretti, and Butler (2004) want to answer the question
  - Do voters pick politicians or policies?
- Convergence Theory
  - Heterogenous voter ideology forces each candidate to moderate this or her position (similar to median voter theorem)
- Divergence Theory
  - The winning candidate pursues his or her most-preferred policies
- The authors focus on very narrow elections by margins less than 2% in 1992

RDD Example: Close Elections

Lee, Moretti, and Butler (2004)

- Theoretical Framework
  - For policy preferences under a quadratic loss function of $u(l) = -(1/2)(l-c)^2$ for Democratic party and $v(l) = -(1/2)l^2$ for Republican party
    - Democratic “bliss point is when $c > 0$ and Republican “bliss point” is when $c = 0$
  - Before election t, voters develop policy expectations denoted x^e and y^e for Democratic and Republican nominee, respectively
  - The probability of party D winning is P and $P(x^e, y^e)$ is a function of x^e and y^e
- t is the 1992 November election and RC_t means the 1993-1994 Congressional session
- $t+1$ is 1994 November election and RC_{t+1} is the 1994-1995 Congressional sessions

RDD Example: Close Elections

Lee, Moretti, and Butler (2004)

- Alesina (1988) shows that there is a efficient frontier by $x* = y* = \lambda c$ where $\lambda \in (0,1)$, where $x*$ and $y*$ are policy equilibrium points
  - There are 3 Nash equilibria
- Complete convergence
  - $x* = y* = \lambda * c$ where $\lambda *$ is Nash bargaining equilibrium
  - When $\frac{dx*}{dP*} > 0, \frac{dy*}{dP*} > 0$, where an exogenous marginal increase in party D popularity $P*$ pulls both parties, so voter's affect candidates' policy choices
  - It is irrelevant for policy which party is actually elected

RDD Example: Close Elections

Lee, Moretti, and Butler (2004)

- Partial convergence
  - $0 \leq y* \leq x* \leq c$
  - Is $\dfrac{dx*}{dP*} > 0, \dfrac{dy*}{dP*} > 0$ robust to minor deviation?
  - Both parties may be better off moving closer to their bliss points, but with an exogenously higher $P*$, party D has better "bargaining position" and can converge towards c, which is the party D bliss point
  - The authors note that partial convergence is a more likely equilibrium than complete convergence, so they focus on partial convergence moving forward

RDD Example: Close Elections

Lee, Moretti, and Butler (2004)

- Complete divergence
  - $x* = c$ and $y = 0$
  - In this equilibrium, voters expect parties to carry out their bliss points
  - $\dfrac{dx*}{dP*} = \dfrac{dy*}{dP*} = 0$
  - An exogenous marginal shock in party D popularity, $P*$, has no impact on candidates positions
  - Voters only elect politicians not policies

RDD Example: Close Elections

Lee, Moretti, and Butler (2004) Empirical Framework

- Roll-call voting record is denoted
  - $RC_t = (1 - D_t)y_t + D_t x_t$
  - Where D_t is an indicator of the Democratic nominee won election t, and only the winning candidate's policy is observable
- We can transform roll-call voting record from the prior expression to:
  - $RC_t = \alpha + \pi_0 P_t^* + \pi_1 D_t + \varepsilon_t$ and $RC_{t+1} = \alpha + \pi_0 P_{t+1}^* + \pi_1 D_{t+1} + \varepsilon_{t+1}$
    - Where $P*$ is a measure of the electoral strength "popularity" of party D, which is the probability of a party D victory at fixed platforms c and 0; P is the probability that party D will win
    - ε is the heterogeneity in bliss points across districts

RDD Example: Close Elections

Lee, Moretti, and Butler (2004) Empirical Framework

- Our interest is π_0
  - When $\pi_0 > 0$ we have partial convergence
  - When $\pi_0 = 0$, we have complete divergence
- The prior equations allow use to parameterize the derivatives $\dfrac{dx*}{dP*}$ and $\dfrac{dy*}{dP*}$ as π_0
  - We cannot directly estimate this equation because we never observe $P*$
  - We do observe D and can estimate π_1, which is an important determinant of roll-call voting records

RDD Example: Close Elections

Lee, Moretti, and Butler (2004) Empirical Framework

- Suppose we randomize D_t, then $D_t \perp P^*$ and $D_t \perp \varepsilon_t$
- If bliss points are exogenous and not influenced by who won the prior election, then D_t will have no impact on ε_{t+1}
- We have three equations to back out π_0
- $E[RC_{t+1} | D_t = 1] - E[RC_{t+1} | D_t = 0] = \pi_0[P^{*D}_{t+1} - P^{*R}_{t+1}] + \pi_1[P^D_{t+1} - P^R_{t+1}] = \gamma$
  - Where γ is the total effect on initial win on future roll-call votes and we don't observe P^* and π_0
  - P^D_{t+1} is the equilibrium probability of a Democratic victor in $t+1$; P^{*D}_{t+1} is electoral strength
- $E[RC_t | D_t = 1] - E[RC_t | D_t = 0] = \pi_1$
- $E[D_{t+1} | D_t = 1] - E[D_{t+1} | D_t = 0] = P^D_{t+1} - P^R_{t+1}$

RDD Example: Close Elections

Lee, Moretti, and Butler (2004) Empirical Framework

- 1) $E[RC_{t+1} | D_t = 1] - E[RC_{t+1} | D_t = 0] = \pi_0[P^{*D}_{t+1} - P^{*R}_{t+1}] + \pi_1[P^{D}_{t+1} - P^{R}_{t+1}] = \gamma$
  - This shows the total effect of γ of a Democratic victory in election t on roll-call voting in congressional session $t + 1$
  - The first component is the unobserved “affect” component
  - The second component is observed “elect” component
- 2) $E[RC_t | D_t = 1] - E[RC_t | D_t = 0] = \pi_1$
- 3) $E[D_{t+1} | D_t = 1] - E[D_{t+1} | D_t = 0] = P^{D}_{t+1} - P^{R}_{t+1}$

RDD Example: Close Elections

Lee, Moretti, and Butler (2004) Empirical Framework

- We can estimates of γ, π_1, and $P^D_{t+1} - P^R_{t+1}$ from our three equations
- We can estimate "affect" by backing out $\pi_0[P^{*D}_{t+1} - P^{*R}_{t+1}]$ by $\gamma - \pi_1[P^D_{t+1} - P^R_{t+1}]$
- If voters merely "elect" policies (complete divergence), then $\pi_0[P^{*D}_{t+1} - P^{*R}_{t+1}]$ should be small
- If voters affect policies (partial convergence), then $\pi_0[P^{*D}_{t+1} - P^{*R}_{t+1}]$ will be large
- How much of γ is explained by "elect" vs "affect"?

RDD Example: Close Elections

Lee, Moretti, and Butler (2004) Empirical Framework

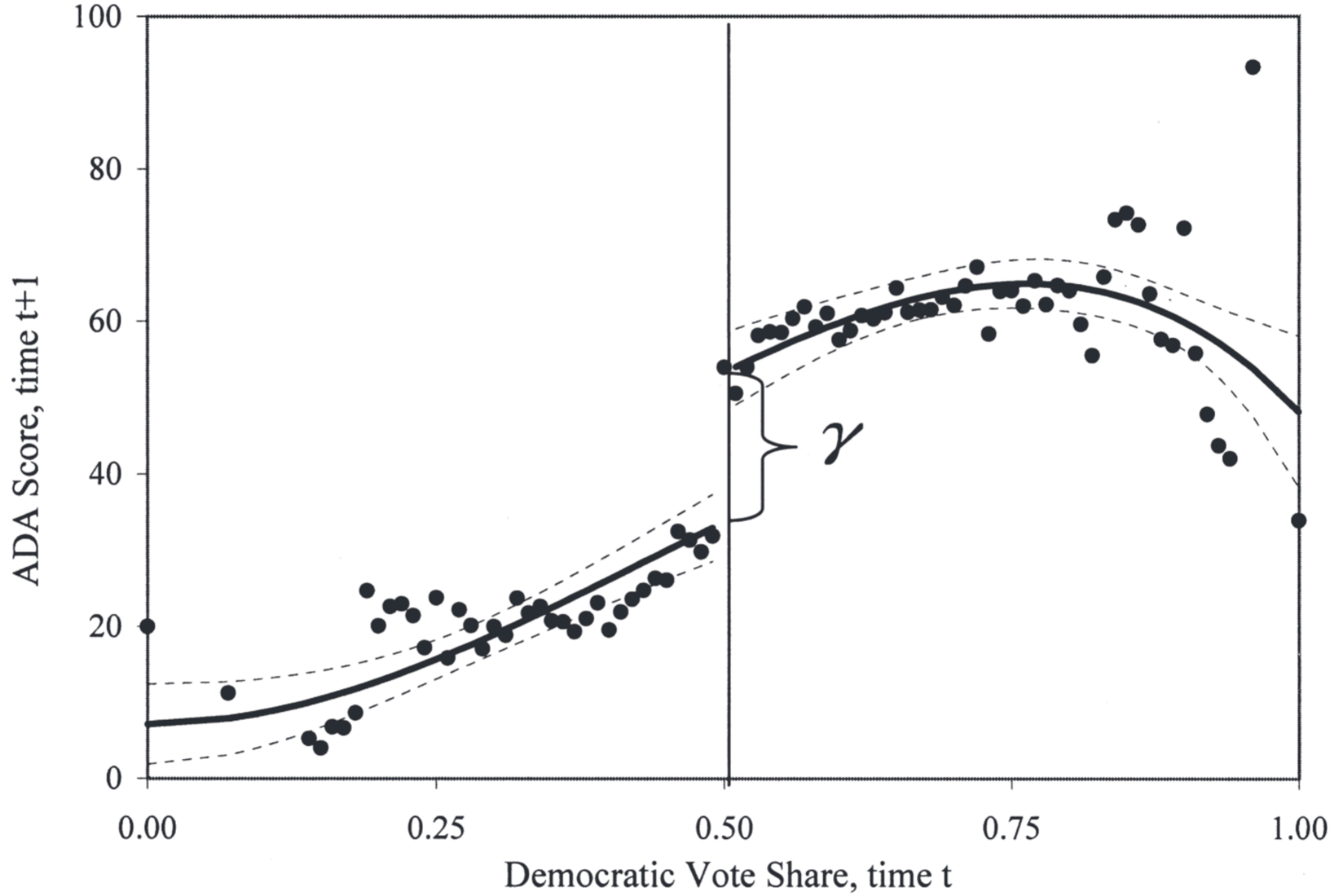


FIGURE I

Total Effect of Initial Win on Future ADA Scores: γ

This figure plots ADA scores after the election at time $t + 1$ against the Democrat vote share, time t. Each circle is the average ADA score within 0.01 intervals of the Democrat vote share. Solid lines are fitted values from fourth-order polynomial regressions on either side of the discontinuity. Dotted lines are pointwise 95 percent confidence intervals. The discontinuity gap estimates

$$\gamma = \underbrace{\pi_0(P^{*D}_{t+1} - P^{*R}_{t+1})}_{\text{“Affect”}} + \underbrace{\pi_1(P^{*D}_{t+1} - P^{*R}_{t+1})}_{\text{“Elect”}}.$$

RDD Example: Close Elections

Lee, Moretti, and Butler (2004) Empirical Framework

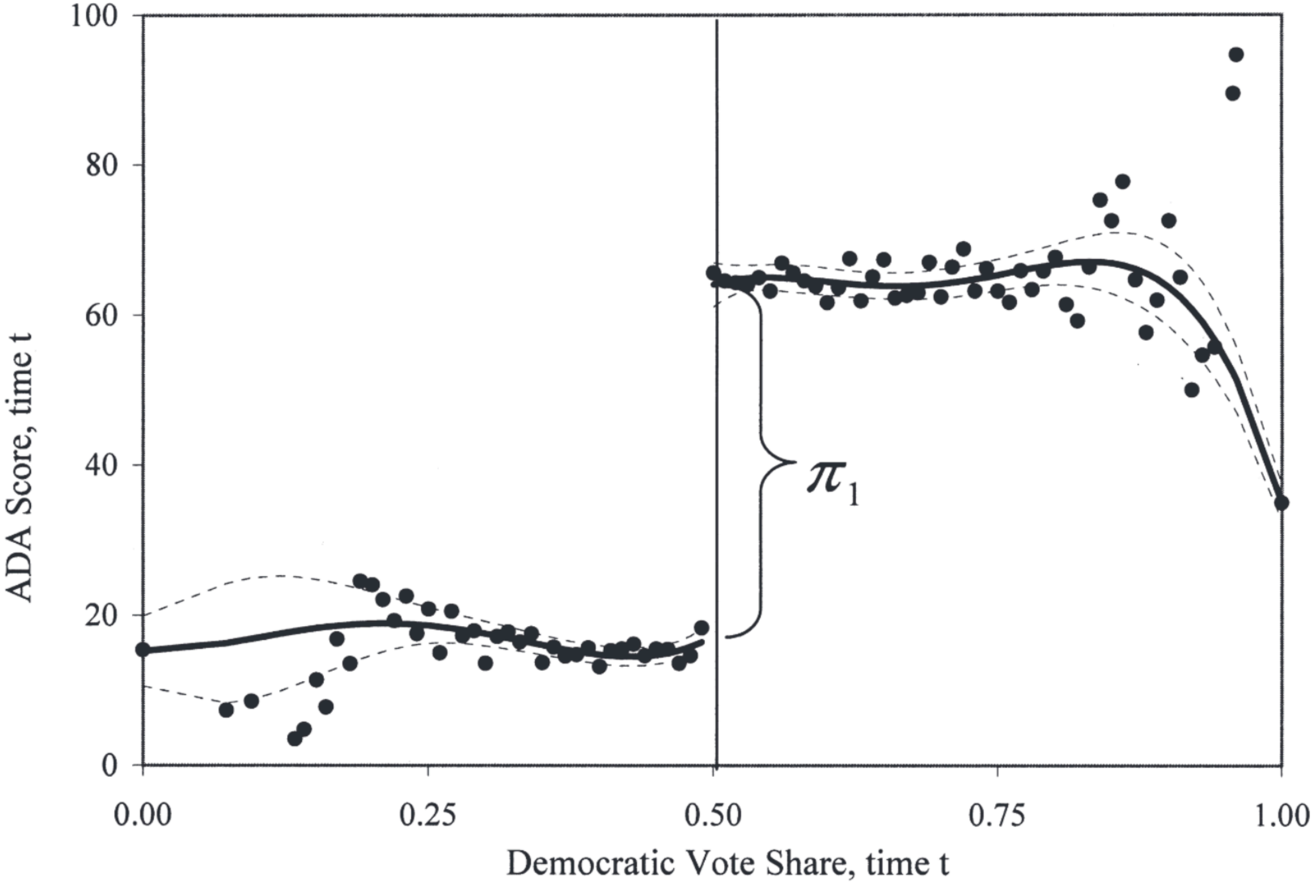


FIGURE IIa
Effect of Party Affiliation: π_1

RDD Example: Close Elections

Lee, Moretti, and Butler (2004) Empirical Framework

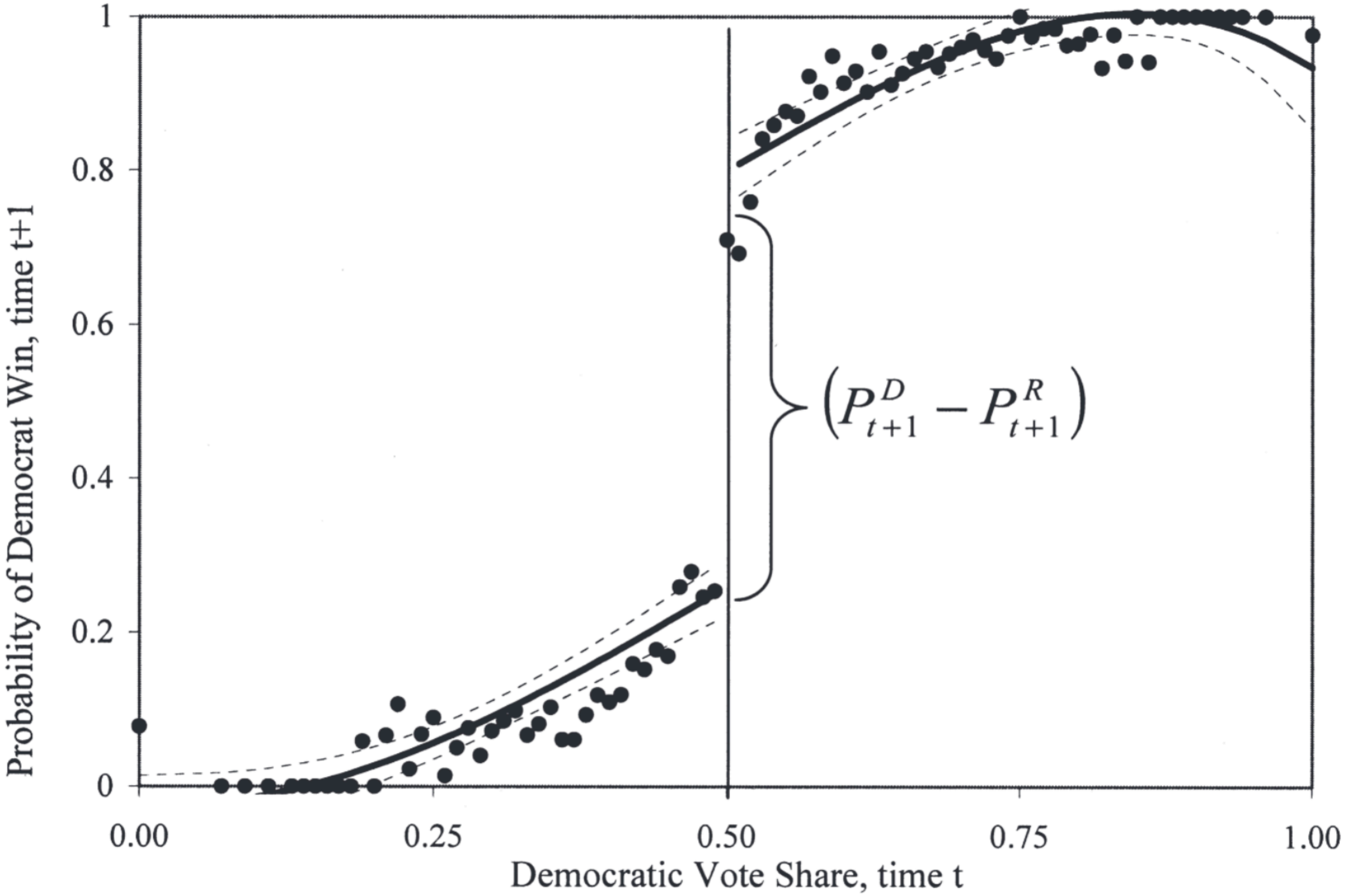


FIGURE IIb

Effect of Initial Win on Winning Next Election: $(P^D_{t+1} - P^R_{t+1})$

Top panel plots ADA scores after the election at time t against the Democrat vote share, time t. Bottom panel plots probability of Democrat victory at $t + 1$ against Democrat vote share, time t. See caption of Figure III for more details.

RDD Example: Close Elections

Lee, Moretti, and Butler (2004)

- Lee, et al. (2004) argue that close elections are "good as random" and isolates the effect of D_t
- Data
  - Roll-call voting records are obtained from Americans for Democratic Action (ADA) linked with House of Representative elections for 1946-1995
- The Running Variable
  - Voting share in an election is the running variable with a cutoff at 50%
- The window is 2% around the cutoff, so 48% to 52% is the sample used

RDD Example: Close Elections

Lee, Moretti, and Butler (2004)

- These are the results γ, π_1, and $P^D_{t+1} - P^R_{t+1}$
- $\pi_0[P^{*D}_{t+1} - P^{*R}_{t+1}]$ is small and insignificant
- This supports the complete divergence theory
- The "elect" component is more important than "affect"

TABLE I

RESULTS BASED ON ADA SCORES—CLOSE ELECTIONS SAMPLE

	Total effect			Elect component	Affect component
	γ	π_1	$(P^D_{t+1} - P^R_{t+1})$	$\pi_1[(P^D_{t+1} - P^R_{t+1})]$	$\pi_0[P^{*D}_{t+1} - P^{*R}_{t+1}]$
Variable	ADA_{t+1}	ADA_t	DEM_{t+1}	(col. (2)*(col. (3))	(col. (1)) − (col. (4))
	(1)	(2)	(3)	(4)	(5)
Estimated gap	21.2	47.6	0.48		
	(1.9)	(1.3)	(0.02)		
				22.84	−1.64
				(2.2)	(2.0)

Standard errors are in parentheses. The unit of observation is a district-congressional session. The sample includes only observations where the Democrat vote share at time t is strictly between 48 percent and 52 percent. The estimated gap is the difference in the average of the relevant variable for observations for which the Democrat vote share at time t is strictly between 50 percent and 52 percent and observations for which the Democrat vote share at time t is strictly between 48 percent and 50 percent. Time t and $t + 1$ refer to congressional sessions. ADA_t is the adjusted ADA voting score. Higher ADA scores correspond to more liberal roll-call voting records. Sample size is 915.

RDD Example: Close Elections

Lee, Moretti, and Butler (2004)

- Stata Exercise to replicate Lee, Moretti, and Butler (2004)

Regression Kink Design

- Sometimes a discontinuity does not describe what happens at a cutoff
- Card, Lee, Pei, and Weber (2015) discuss regression kink design
- Instead of a discontinuous jump at a cutoff, a change in the first derivative (slope) occurs at the cutoff
- Kinks in the slope are often embedded in policy rules
    - We can identify causal effects by exploiting a change in the slope at the cutoff

Regression Kink Design