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

April 23, 2023

Regression Discontinuity Design: Main Idea

Introduction

Selection Bias and RCT

- 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

Main Idea of Regression Discontinuity Design

A Motivating Example

- 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**

Main Idea of Regression Discontinuity Design

A Motivating Example

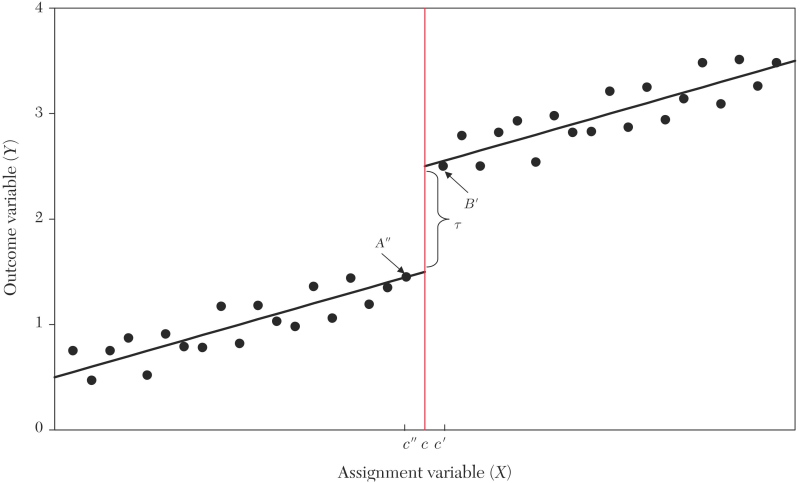
- A great way to answer that question would be to run an experiment:
  - Take students applying both to NTU and NTHU
  - Instead of admitting them the regular way, just flip a coin to decide whether they get into NTU or NTHU
  - Follow them up 10 years later to see whether those admitted to NTU earn more than those admitted to NTHU
- 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 NTHU)

Test Score and Earnings



Source: Lee and Lemieux (2010)

Main Idea of Regression Discontinuity Design

A Motivating Example

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

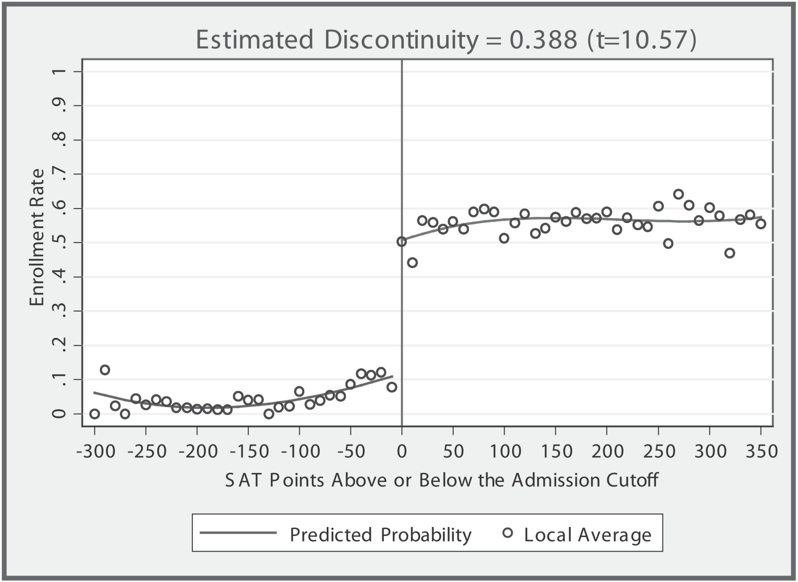
Main Idea of Regression Discontinuity Design

A Motivating Example

- 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

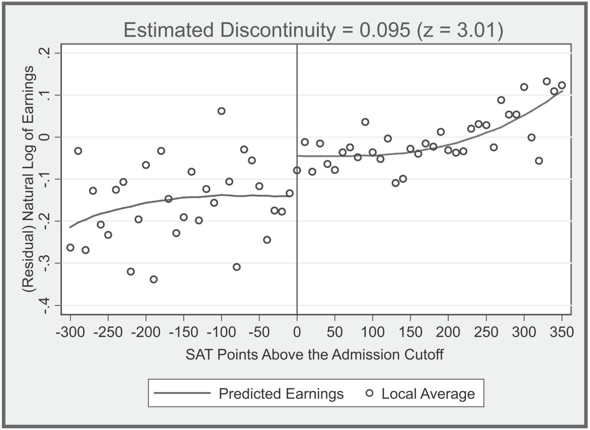
SAT Score and Enrollment

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)



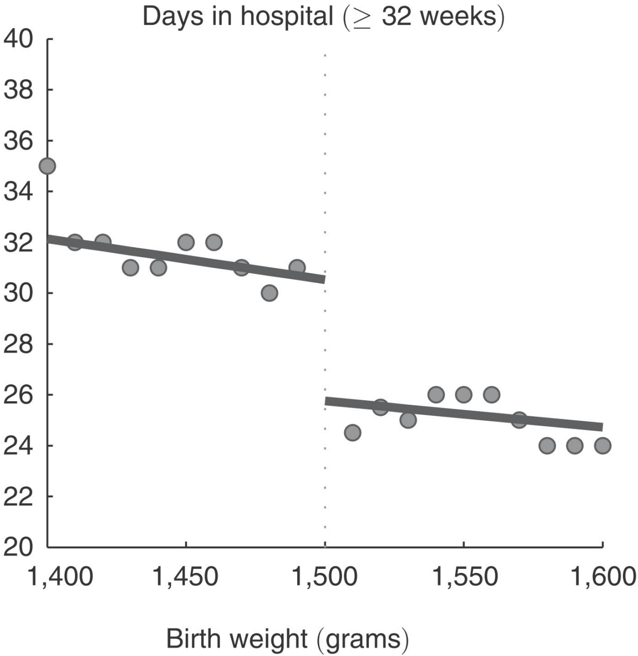
More Examples

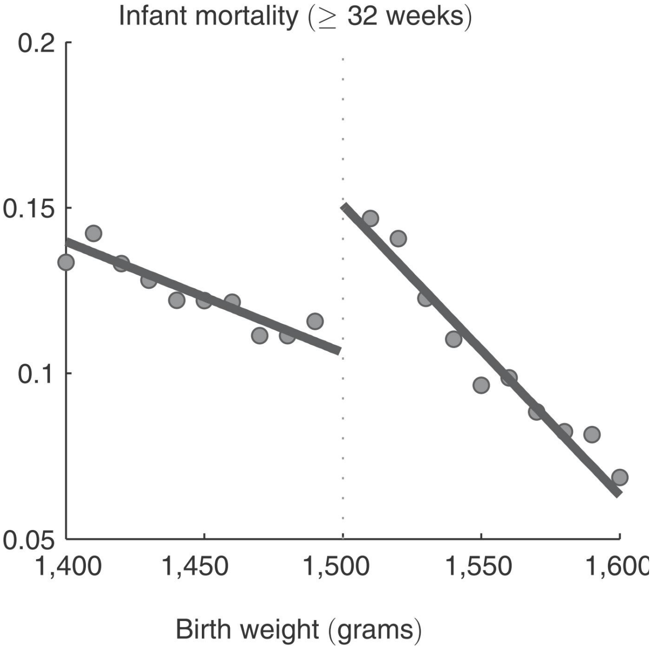
- **Where there is a cutoff there is a RD**

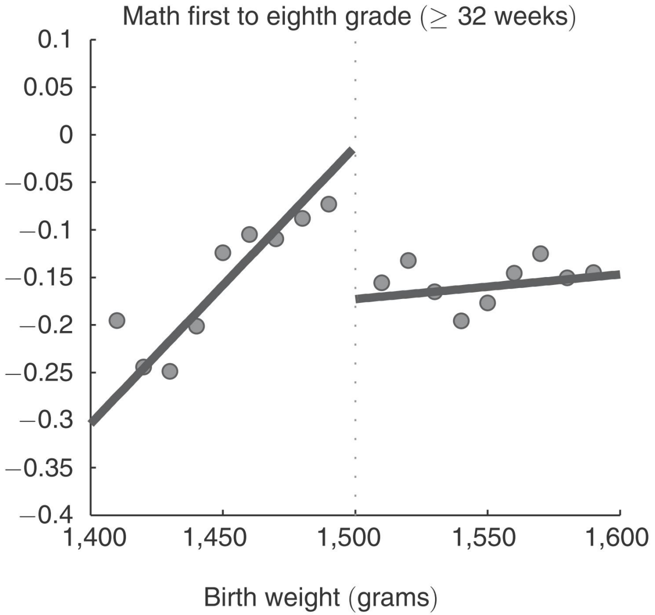
Prashant Bharadwaj, Katrine Vellesen Løken, and Christopher Neilson (2013) "**Early Life Health Interventions and Academic Achievement**" AER

- 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)

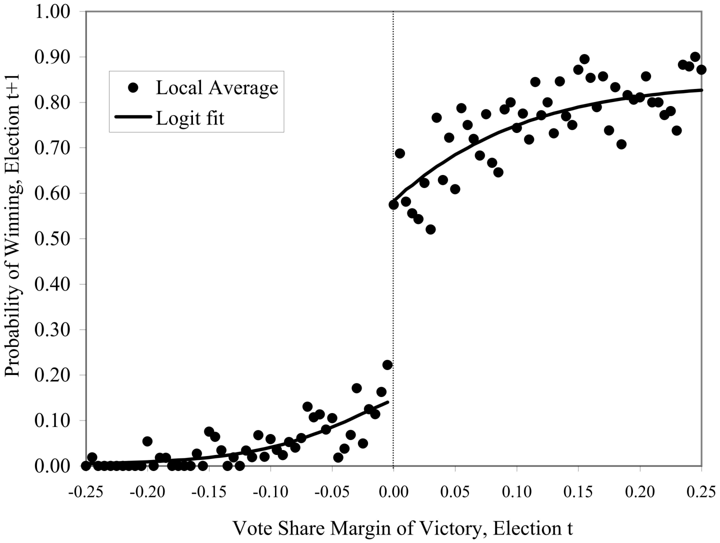

Math first to eighth grade (≥ 32 weeks)
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
−0.25
−0.3
−0.35
−0.4
1,400
1,450
1,500
1,550
1,600
Birth weight (grams)

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**

Labor Economics

Spatial Regression Discontinuity

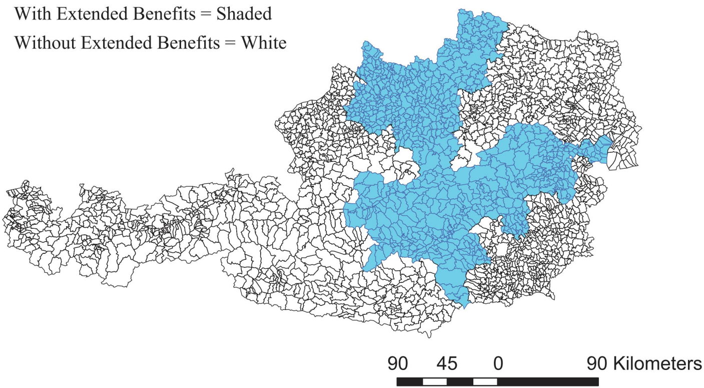
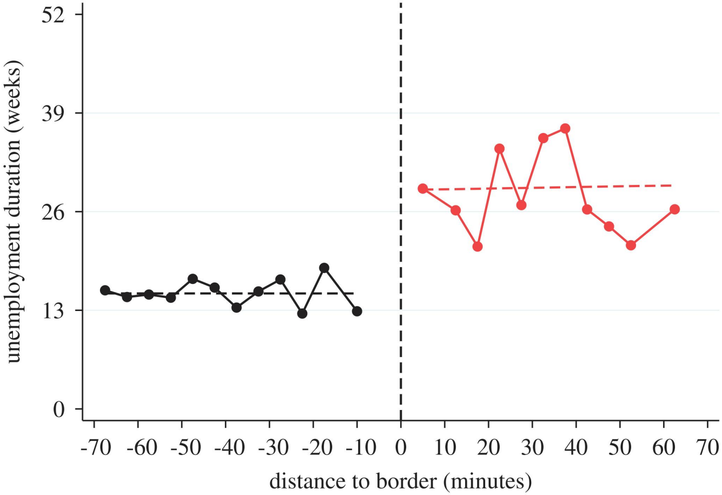


Fig. 1. Regional distribution of REBP.

Labor Economics

Spatial Regression Discontinuity



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. Sharp RDD: nobody below the cutoff gets the "treatment", everybody above the cutoff gets it
    - Everyone follows treatment assignment rule (all are compliers)
    - Local randomized experiment with perfect compliance around cutoff
2. Fuzzy RDD: the probability of getting the treatment jumps discontinuously at the cutoff (NOT jump from 0 to 1)
    - Not everyone follows treatment assignment rule
    - Local randomized experiment with partial compliance around cutoff
    - Using initial assignment as an instrument (IV) for actual treatment

Sharp Regression Discontinuity Design: Potential Outcomes Framework

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

- 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

Sharp RDD and Potential Outcomes

Treatment

- Assignment variable (running variable): $X_i \in \mathbb{R}$
- Threshold (cutoff) for treatment assignment: $c \in \mathbb{R}$
- D_i: a dummy that indicate whether individual i receive treatment or not
- Treatment assignment:

$$D_i = \{X_i \geq c\}$$

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

Sharp RDD and Potential Outcomes

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:

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

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

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:

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

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

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

Deterministic Assumption

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

- Treatment assignment is a deterministic function of the assignment variable X_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:

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

Identification Results for Sharp RDD

Continuity Assumption

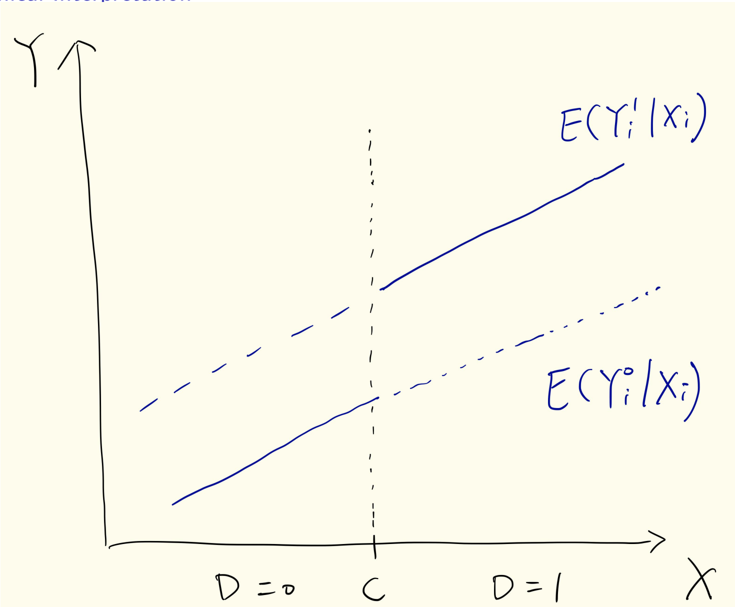
Continuity Assumption

$E[Y_i^1|X_i]$ and $E[Y_i^0|X_i]$ are continuous at $X_i = c$

- 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**

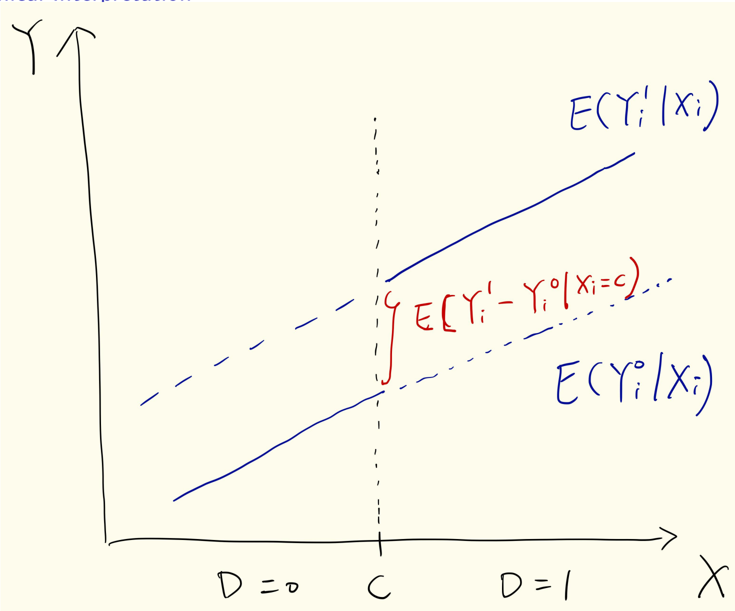
Identification Results for Sharp RDD

Graphical Interpretation



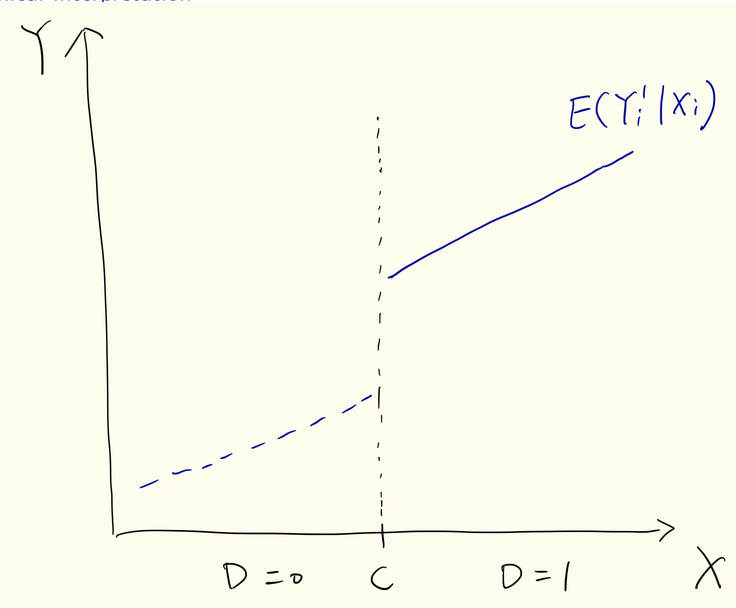
Identification Results for Sharp RDD

Graphical Interpretation



Identification Results for Sharp RDD

Graphical Interpretation



Identification Results for Sharp RDD

Continuity Assumption

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

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

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

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

Identification Results for Sharp RDD

Continuity Assumption

- The continuity assumption and deterministic assumption imply the following:

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

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

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

- 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

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

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

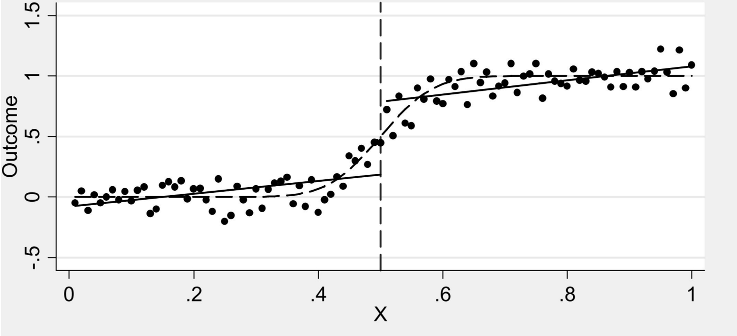
Sharp Regression Discontinuity Design: Estimation

Sharp RDD Estimation

Overview

- There are 2 types of strategies for getting RDD 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

Nonlinear Relationship or Discontinuity in Outcome?



Source: Angrist and Pischke (2009), Mostly Harmless Econometrics, Ch6

Sharp RDD Estimation

Parametric/Global Approach

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

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

- We can get RD estimates by fitting:

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

Sharp RDD Estimation

Parametric/Global Approach

- Use a flexible polynomial (p^{th} order polynomial) regression to estimate $\mathrm{f}(x_i)$

$$Y_i = \alpha + \rho D_i + \underbrace{\beta_1 X_i + \beta_2 X_i^2 + ... + \beta_p X_i^p}_{\mathrm{f}(x_i)} + \eta_i \qquad (1)$$

Sharp RDD Estimation

Parametric/Global Approach

- More general case:
  1. Allow the X_i terms to differ on both sides of the threshold
     - Include X_i both individually and interacting them with D_i
     - By doing this, we can estimate $\mathrm{f}(x_i)$ on each side

Sharp RDD Estimation

Parametric/Global Approach

2 Re-center X_i at c:

- $\widetilde{X}_i = X_i - c$
- This step ensures that the treatment effect at $X_i = c$ is the coefficient on D_i in a regression model with interaction terms
- So we do not have to add values of the D_i interacted with X_i to get the treatment effect at $X_i = c$

Sharp RDD Estimation

Parametric/Global Approach

- Therefore, we usually estimate the following regression model:

$$Y_i = \alpha + \rho D_i + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \ldots + \beta_p \tilde{X}_i^p$$
$$+ \beta_1^* D_i \tilde{X}_i + \beta_2^* D_i \tilde{X}_i^2 + \ldots + \beta_p^* D_i \tilde{X}_i^p + \eta_i \quad (2)$$

  - Note that $\widetilde{X}_i = X_i - c$
  - Equation (1) above is a special case of (2) with $\beta_1^* = \beta_2^* = \beta_p^* = 0$.
  - The treatment effect (ATE) at $X = c$ is ρ

Sharp RDD Estimation

Parametric/Global Approach

- In alcohol example, we estimate the following regression with cubic terms :

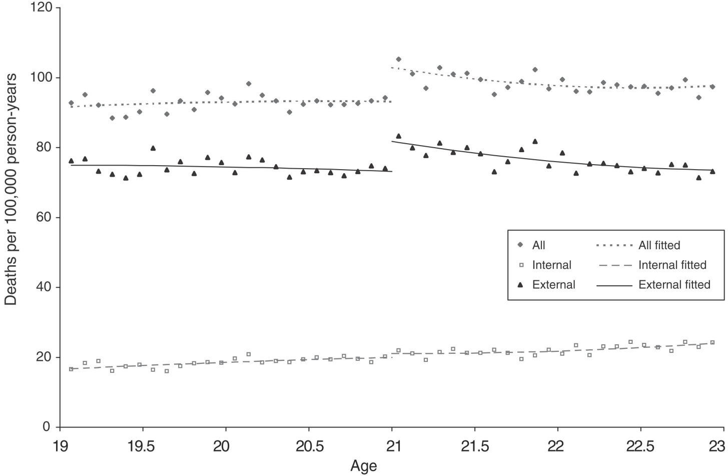
$$
\begin{aligned}
Y_i = \alpha + \rho D_i + \beta_1(X_i - 21) + \beta_2(X_i - 21)^2 + \beta_3(X_i - 21)^3 \\
+ \beta_4 D_i(X_i - 21) + \beta_5 D_i(X_i - 21)^2 + \beta_6 D_i(X_i - 21)^3 + \eta_i
\end{aligned}
$$

- The effect of legal access to alcohol on mortality rate at age 21 is ρ

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

Sharp RDD Estimation

How to Select Polynomial Order

1. F-Test Approach
2. AIC Approach

How to Select Polynomial Order

F-Test Approach

To implement F-Test, one can complete the following steps:

1. Create a set of indicator variables for $K-2$ of the bins used to graphically depict the data
2. Exclude any two of the bins to avoid having a model that is collinear
3. Add the set of bin dummies B_k to the polynomial regression and jointly test the significance of the bin dummies

$$Y_i = \alpha + \rho D_i + \beta_1 (X_i - c) + \beta_2 D_i (X_i - c) + \sum_{k=2}^{K-1} \phi_k B_k + \varepsilon_i$$

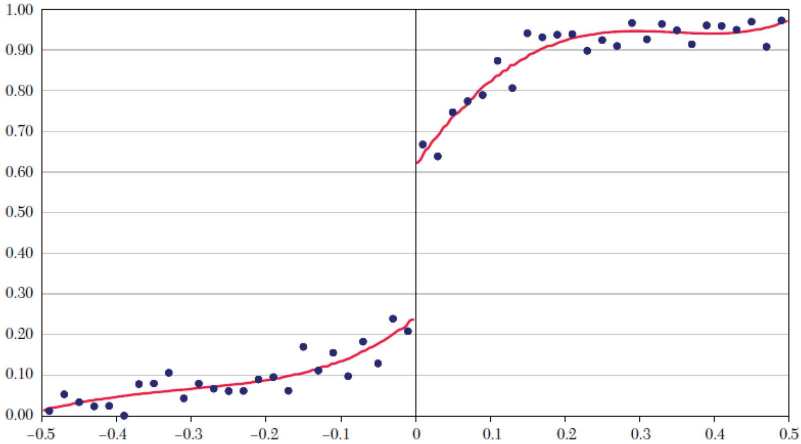
How to Select Polynomial Order

F-Test Approach

To implement F-Test, one can complete the following steps:

4. Test the null hypothesis that $\phi_2 = \phi_3 = ... = \phi_{K-1} = 0$
5. In terms of specification choice procedure, the idea is to add a higher order term to the polynomial until the bin dummies are no longer jointly significant

Bins



Source: Lee and Lemieux (2010)

How to Select Polynomial Order

AIC Approach

- Another strategy that can be used is the Akaike information criterion (AIC) procedure.
- Conceptually, AIC describes the trade-off between the **goodness of fit** of the model and the **simplicity** of the model

How to Select Polynomial Order

AIC Approach

- These two terms move in opposite directions as the model becomes more complex:
- The estimated residual variance $\widehat{\sigma_b^2}$ should decrease with more complex models, but the number of parameters p used increases
- In a regression context, the AIC is given by

$$AIC = N \ln(\widehat{\sigma_b^2}) + 2p$$

where $\widehat{\sigma_b^2}$ is the estimated residual variance based on a model with p parameter, and p is the number of parameters in the regression model including the intercept

How to Select Polynomial Order

AIC Approach

- The set of models are then ranked according to their AIC values, and the model with the **smallest AIC value** is deemed the optimal model among the set of candidates

2 Nonparametric/local method:

- Use the observations around cutoff
- Compare the outcome of treated and untreated observations that lie **within specific bandwidth**

Sharp RDD Estimation

Nonparametric/Local Approach

- Nonparametric approach does NOT specify particular functional form of the outcome and the assignment variable
- Instead, it uses only data within a small neighborhood (known as **bandwidth**) to estimate the discontinuity in outcomes at the cutoff:
  1. Compare means in the two bins adjacent to the cutoff (treatment v.s. control groups)
  2. Local linear regression

Sharp RDD Estimation

Nonparametric/Local Approach

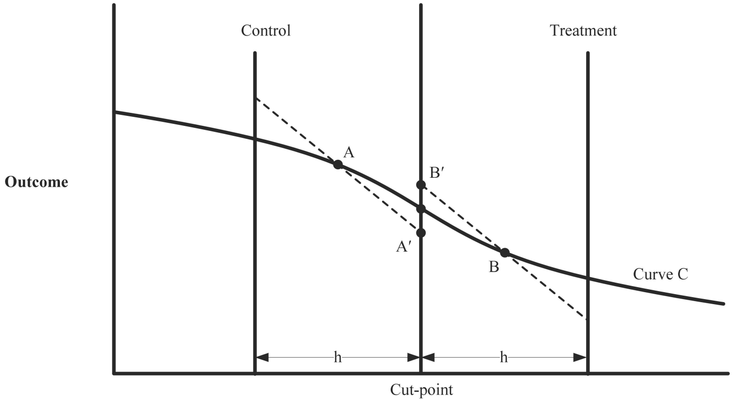
- However, comparing means in the two bins adjacent to the cutoff is generally biased in the neighborhood of the cutoff
  - The RD estimate is $\overline{AB}$
  - This estimate has a **bias**

Sharp RDD Estimation

Nonparametric/Local Approach

Figure 5

Boundary Bias from Comparison of Means vs. Local Linear Regression (Given Zero Treatment Effect)



Source: Robin Jacob, Pei Zhu, Marie-Andrée Somers, and Howard Bloom (2016)

Sharp RDD Estimation

Nonparametric/Local Approach

- The standard solution to reduce this bias is to run **local linear regression**

- We simply estimate the following linear regression **within a given window of width** h **around the cutoff**:

$$Y_i = \alpha + \rho D_i + \beta_1 \tilde{X}_i + \beta_1^* D_i \tilde{X}_i + \eta_i$$

  where $\tilde{X}_i = X_i - c$

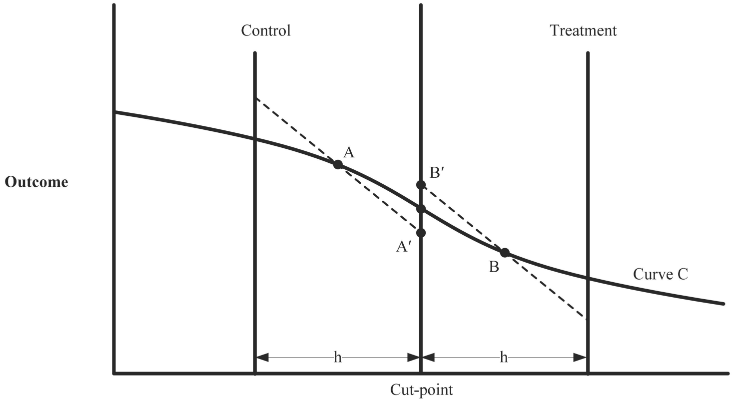
- By using local linear regression, we can substantially reduce the bias
  - The RD estimate is $\overline{B'A'}$
  - This estimate results in less bias

Sharp RDD Estimation

Nonparametric/Local Approach

Figure 5

Boundary Bias from Comparison of Means vs. Local Linear Regression (Given Zero Treatment Effect)



Source: Robin Jacob, Pei Zhu, Marie-Andrée Somers, and Howard Bloom (2016)

Sharp RDD Estimation

Nonparametric/Local Approach

- 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

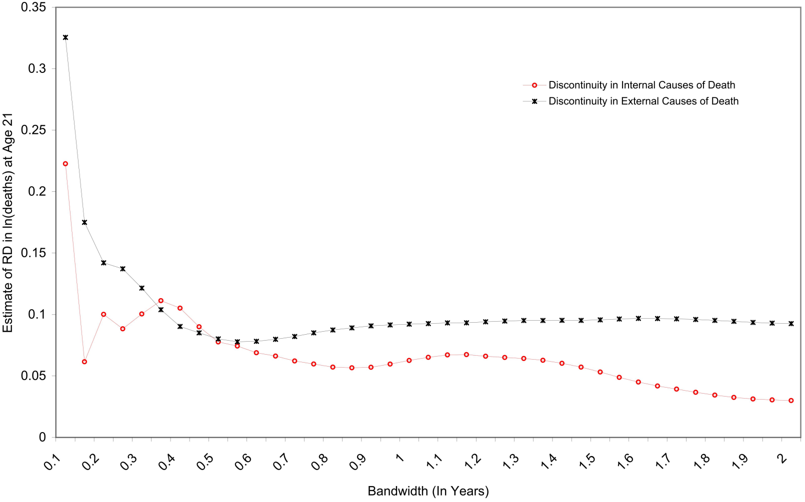
Nonparametric/Local Approach

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

$$Y_i \;=\; \alpha + \rho D_i + \beta_1(X_i - 21) + \beta_2 D_i(X_i - 21) + \eta_i$$

- $h = 180$
- The effect of legal access to alcohol on mortality rate at age 21 is ρ
- Usually, we would present the RD estimates by different choices of bandwidth

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



Sharp RDD Estimation

How to Choose Bandwidth

1. **Cross-Validation Procedure**:
   - Choose bandwidth h that produces the best fit for the relationship of outcome and assignment variable
2. **Plug-In Procedure**:
   - Solve for the optimal bandwidth formula in terms of minimizing mean square error
   - Plug the parameters into formula to get optimal bandwidth
   - See, e.g. Imbens and Kalyanaraman (2009), Calonico, Cattaneo, and Titiunik (2014)

Regression Discontinuity Design: Examine Validity

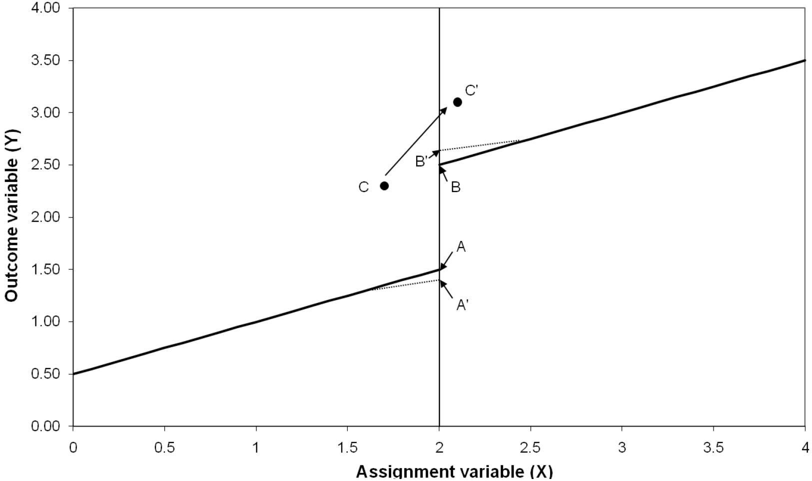
1. **Examine “Discontinuity” in Nonoutcome Variables**
   - Construct a similar graph to the one before but using a covariate as the “outcome”
   - There should be **NO jump in other covariates**
   - If the covariates would jump at the cutoff one would doubt the identifying assumption

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

- Individuals may invalidate the **continuity assumption** if they strategically manipulate assignment variable X 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 X at the cutoff**
- That is, “bunching” to the right or to the left of the cutoff

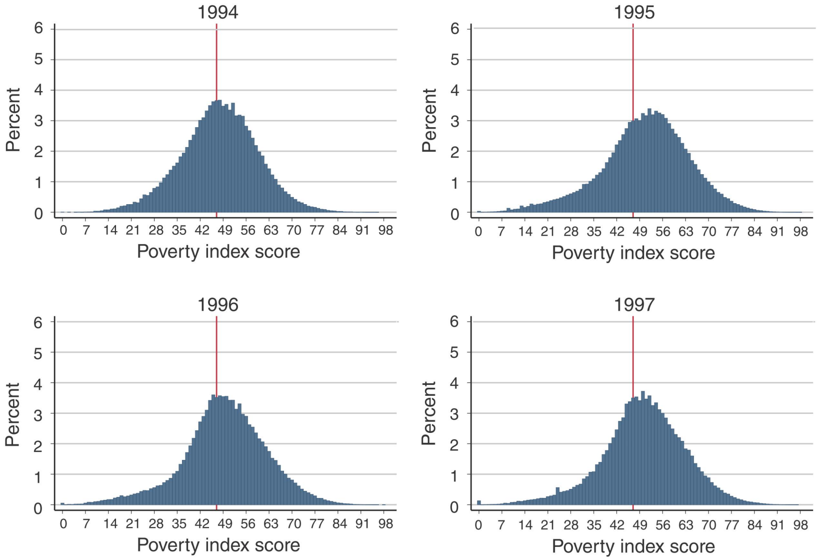
Sorting Behavior
Example

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

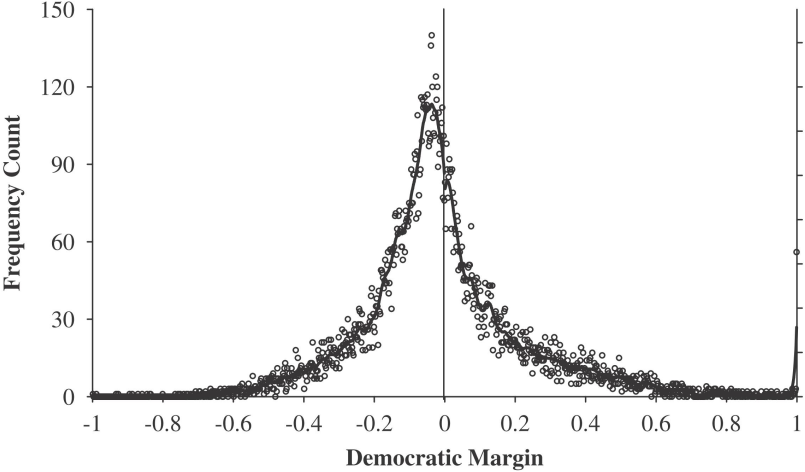
- Plot the number of observations in each bin of assignment variable
- Investigate whether there is a **discontinuity in the distribution of the assignment variable** at the threshold
- A discontinuity in the density suggests that people might manipulate the assignment variable around the threshold

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 midpoint rating values of each of the bins 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



Regression Discontinuity Design: STATA Example

Empirical Example 1: Lee, Moretti, and Butler (2004)

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

- They test two different theories of the role of elections in policy formation
- 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**

Empirical Example 1: Lee, Moretti, and Butler (2004)

Identification Strategy

- Test the above issue using sharp RDD
- Compare a district where the Democrat candidate barely lost at time t (49.5%) with a district where the Democrat candidate barely won at time t (50.5%)
- The close election generated random assignment of who wins in t

Empirical Example 1: Lee, Moretti, and Butler (2004)

Identification Strategy

- Thus, the congressman from a district that Democrat candidate barely won should have similar congressional voting bebavior as one from a district that Democrat candidate barely lose
- The observed difference in voting behavior represents a credible estimate of the **average policy differences between the two parties**

Empirical Example 1: Lee, Moretti, and Butler (2004)

Identification Strategy

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

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

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

Empirical Example 1: Lee, Moretti, and Butler (2004)

STATA Implementation

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

Empirical Example 1: Lee, Moretti, and Butler (2004)

Step 1: Graphical Analysis

- 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]
```

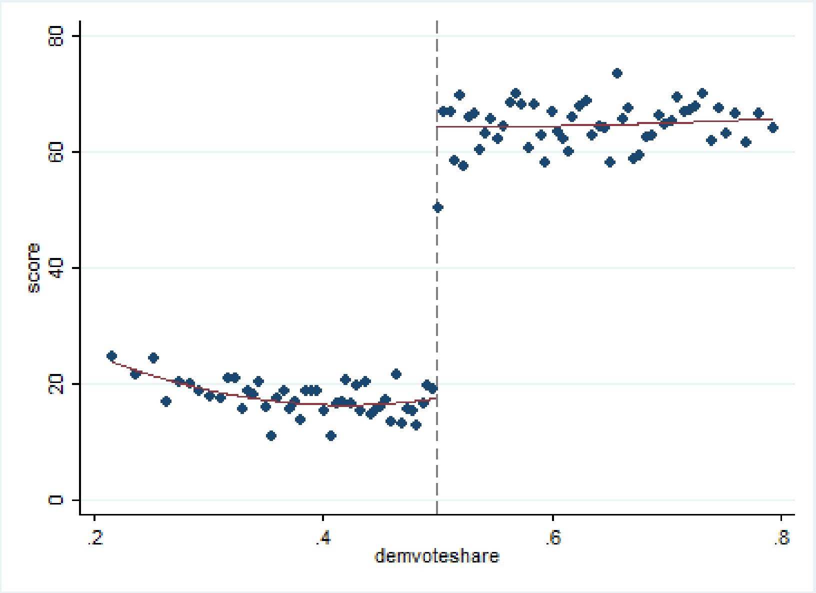
- Example:

```
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



Empirical Example 1: Lee, Moretti, and Butler (2004)

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

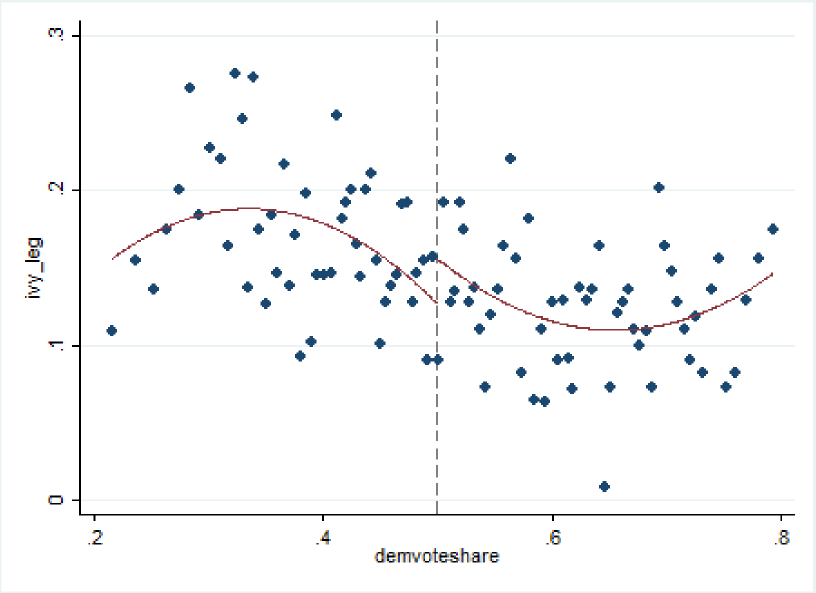
STATA Command: binscatter

- Example:

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

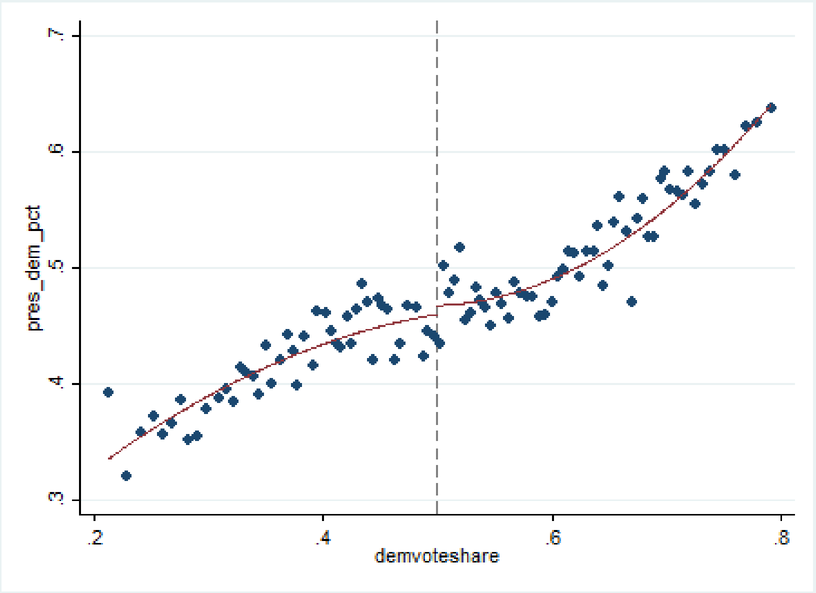
Covariates by assignment variable

Ivy league graduate by Democrat vote share



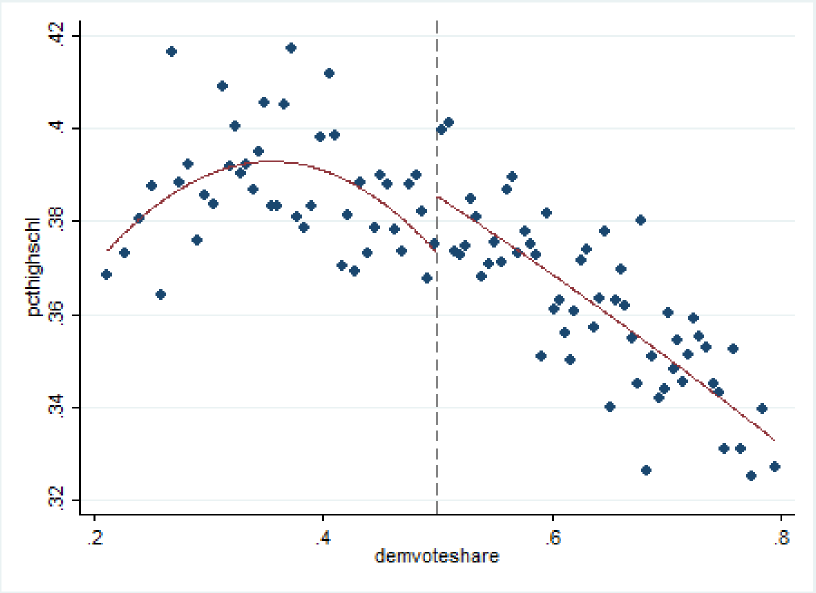
Covariates by assignment variable

Presidential vote share by Democrat vote share



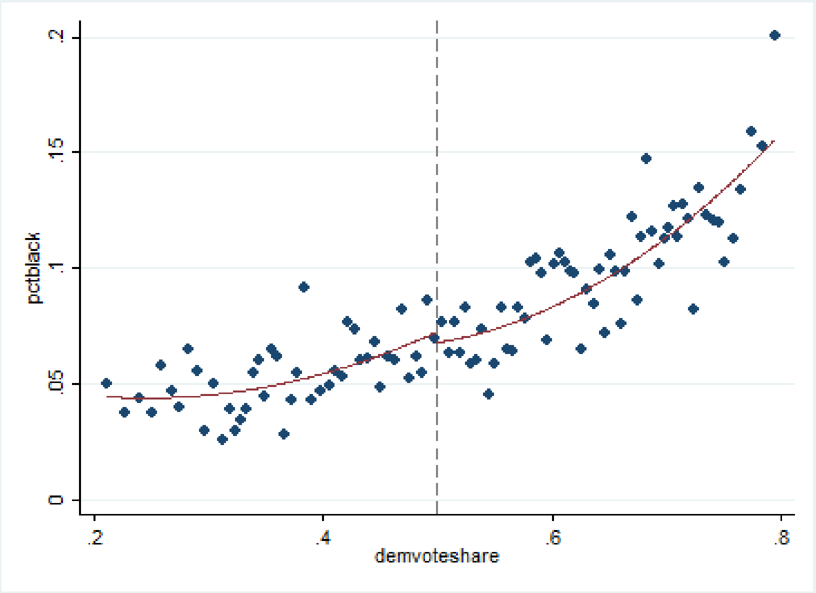
Covariates by assignment variable

High school ratio by Democrat vote share



Covariates by assignment variable

African American ratio by Democrat vote share



Empirical Example 1: Lee, Moretti, and Butler (2004)

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:

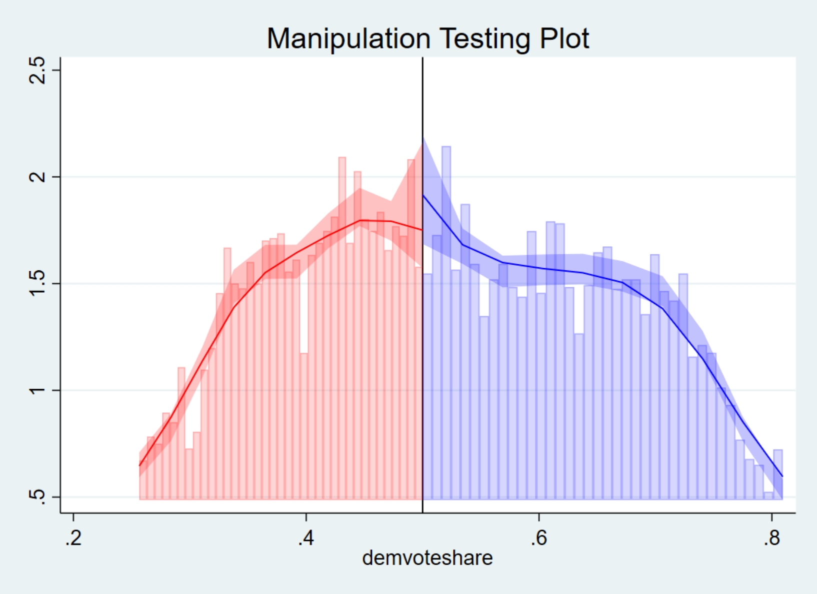
```
rddensity runvar [if] [in] [, options]
```

- 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



Empirical Example 1: Lee, Moretti, and Butler (2004)

Step 3: Preparation for Estimation – STATA

- 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
gen x3_c = x_c^3
```

- Generate some Polynomials interacted with D_i

```
gen d_x_c = democrat*x_c
gen d_x2_c = democrat*x2_c
gen d_x3_c = democrat*x3_c
```

Empirical Example 1: Lee, Moretti, and Butler (2004)

Step 4: Estimation – STATA

- Parametric Approach using all available data

```
reg score democrat x_c-x5_c d_x_c-d_x5_c,cl(id2)
```

- Nonparametric Approach using sample around cutoff (local linear regression)
- use “rdrobust”

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

- Use different choices of bandwidth and order of polynomial to investigate the robustness of estimates

Robustness Check

Bandwidth Choice and Order of Polynomial

Table 2b: RD estimates of the effect of winning the previous election on probability of winning the next election

Bandwidth:	1.00	0.50	0.25	0.15	0.10	0.05	0.04	0.03	0.02	0.01
Polynomial of order:										
Zero	0.814	0.777	0.687	0.604	0.550	0.479	0.428	0.423	0.459	0.533
	(0.007)	(0.009)	(0.013)	(0.018)	(0.023)	(0.035)	(0.040)	(0.047)	(0.058)	(0.082)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.011]	[0.201]	[0.852]	[0.640]	[0.479]	
One	0.689	0.566	0.457	0.409	0.378	0.378	0.472	0.524	0.567	0.453
	(0.011)	(0.016)	(0.026)	(0.036)	(0.047)	(0.073)	(0.083)	(0.099)	(0.116)	(0.157)
	[0.000]	[0.000]	[0.126]	[0.269]	[0.336]	[0.155]	[0.400]	[0.243]	[0.125]	
Two	0.526	0.440	0.375	0.391	0.450	0.607	0.586	0.589	0.440	0.225
	(0.016)	(0.023)	(0.039)	(0.055)	(0.072)	(0.110)	(0.124)	(0.144)	(0.177)	(0.246)
	[0.075]	[0.145]	[0.253]	[0.192]	[0.245]	[0.485]	[0.367]	[0.191]	[0.134]	
Three	0.452	0.370	0.408	0.435	0.472	0.566	0.547	0.412	0.266	0.172
	(0.021)	(0.031)	(0.052)	(0.075)	(0.096)	(0.143)	(0.166)	(0.198)	(0.247)	(0.349)
	[0.818]	[0.277]	[0.295]	[0.115]	[0.138]	[0.536]	[0.401]	[0.234]	[0.304]	
Four	0.385	0.375	0.424	0.529	0.604	0.453	0.331	0.134	0.050	0.168
	(0.026)	(0.039)	(0.066)	(0.093)	(0.119)	(0.183)	(0.214)	(0.254)	(0.316)	(0.351)
	[0.965]	[0.200]	[0.200]	[0.173]	[0.292]	[0.593]	[0.507]	[0.150]	[0.244]	
Optimal order of the polynomial	4	3	2	1	1	2	0	0	0	1
Observations	6558	4900	2763	1765	1209	610	483	355	231	106

Regression Discontinuity Design: My Own Researches

My Own Researches

Apply RDD to National Health Insurance in Taiwan

Hsing-Wen Han, Hsien-Ming Lien and Tzu-Ting Yang (2017) "**Patient Cost-Sharing and Health Care Utilization in Early Childhood: Evidence from a Regression Discontinuity Design**" Conditional Accept at AEJ: Economic Policy

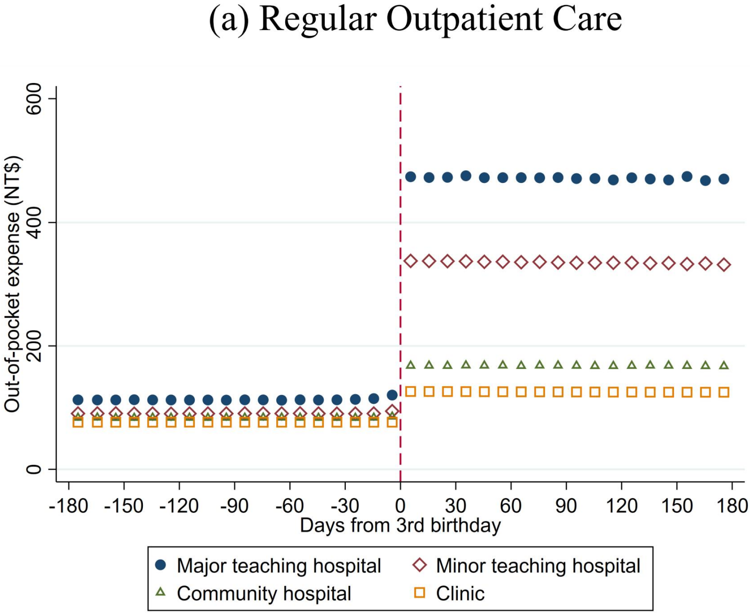
- The effect of copayment (部分負擔) on children's healthcare utilization and provider choice (院所選擇)
- **Selection bias**: those who are charged higher copayment could be due to their bad health condition

My Own Researches

Apply RDD to National Health Insurance in Taiwan

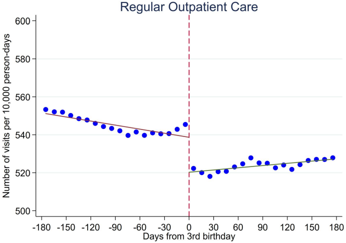
- **RDD solution**: In Taiwan, all children under 3 are completely free, both inpatient (住院) and outpatient (門/急診) services, from copayments (coinsurance) of Taiwanese national health insurance
- Compare the healthcare utilization for children **just before and after their 3rd birthday**
- Children's health conditions just before the 3rd birthday should be very similar to those just after the 3rd birthday

Coapyment and Age

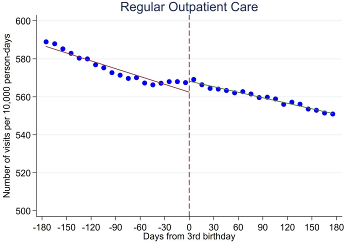


Number of Visit and Age

(c) Number of Visits per 10,000 Person-Days: 2005–2008

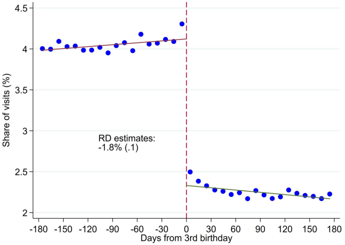


(d) Number of Visits per 10,000 Person-Days: 1997–2001

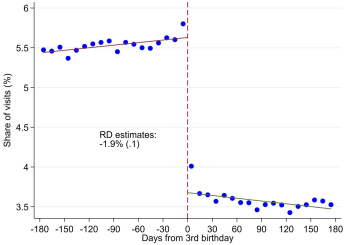


Number of Visit and Age by Different Healthcare Providers

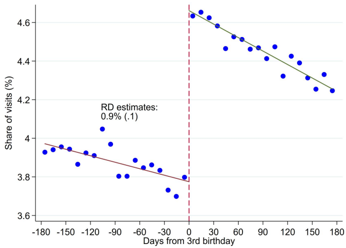
(a) Major Teaching Hospital



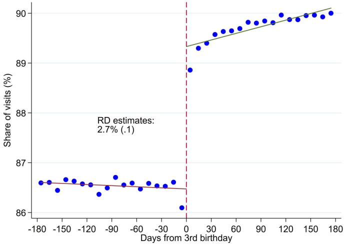
(b) Minor Teaching Hospital



(c) Community Hospital



(d) Clinic



My Own Researches

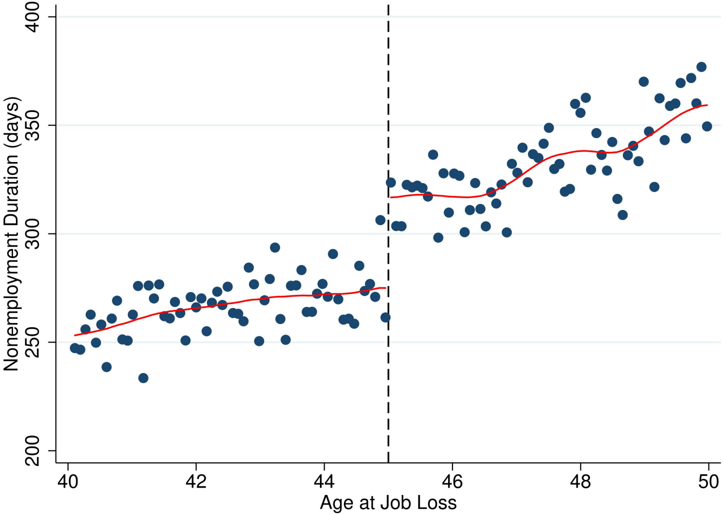
Apply RDD to Unemployment Insurance in Taiwan

Po-Chun Huang and Tzu-Ting Yang (2017) "**An Evaluation of Optimal Unemployment Insurance Using Two Natural Experiments**" Working Paper

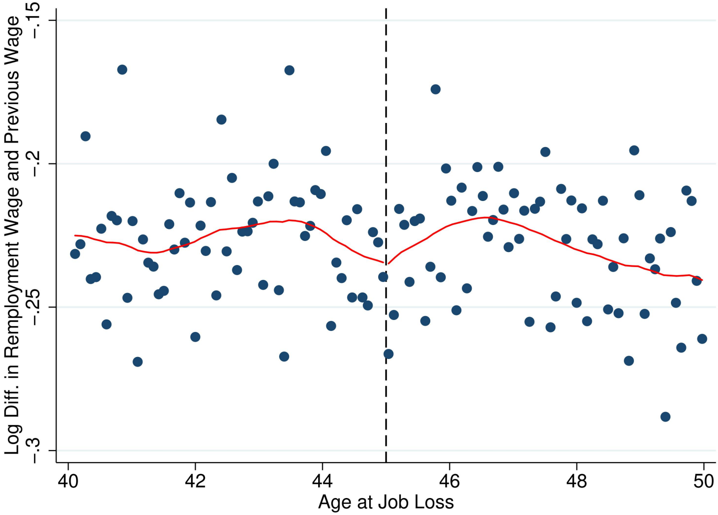
- The effect of extended unemployment insurance (UI) benefit (延長失業給付) on middle age worker's employment (unemployment duration)
- Key parameter for evaluating optimal UI
- **Selection bias**: those who are eligible for more UI benefits usually do not have good working potential

- **RDD solution**: In Taiwan, workers aged at least 45 when leaving their job involuntarily are eligible for 9 months of benefits, rather than 6 months for those under 45
- Compare the unemployment duration for those **lose their job just before and after their 45th birthday**
- These workers' employment potential should be quite similar around age cutoff

Unemployment Duration and Age at Job Loss



Re-employed Wage and Age at Job Loss

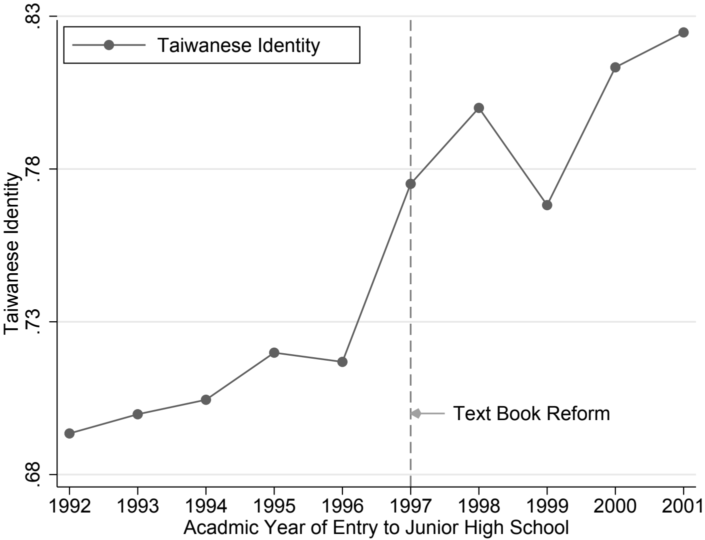


Wei-Lin Chen, Ming-Jen Lin and Tzu-Ting Yang (2017)
"**Curriculum and National Identity: Evidence from the 1997 Textbook Reform in Taiwan**" Working Paper

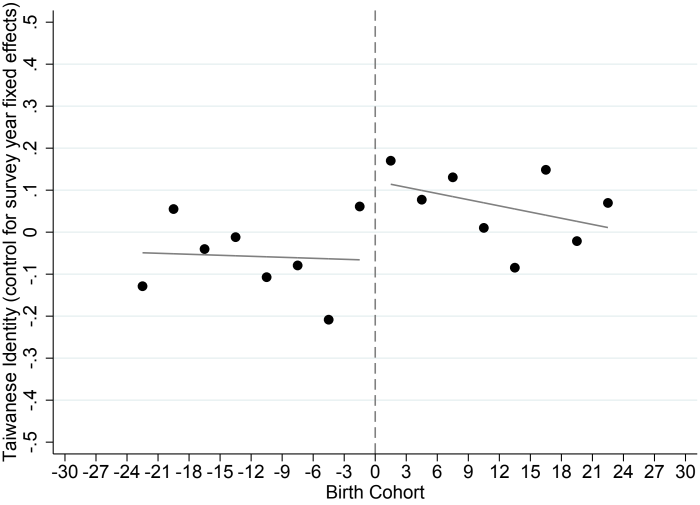
- The effect of curriculum for junior high school (中學教科書) on individuals' national identity (Taiwanese identity)
- This issue arises many debates in Taiwan
- **Selection bias**: Government could change the content of textbook based on social trend

- **RDD solution**: In 1997, Taiwanese government implement a new curriculum (認識台灣) for the students who attend junior high school after September 1997
- That is, those who were born after September 1984 had to read new textbook, which focused on Taiwanese history, geography, and society.
- Those who were born before September 1984 would read old textbook, which exclusively focused on China
- Compare the national identity of those born right before and those born after September 1984

Taiwanese Identity and School Entry Year



Taiwanese Identity and Birth Quarter



Concluding Remark

Checklist for RDD

- Graphing the raw data
  - Graphs for treatment and outcome by assignment variable
  - Density of the assignment variable
- Estimating the regression
  - Polynomial models
  - Local linear regressions and choice of bandwidth
- Testing the validity of the RD design
  - Discontinuity in the density of assignment variable
  - Testing whether covariates are balanced
- Robustness check: bandwidth choice and order of polynomial

Concluding Remark

Some References

- More recent methodological developments can be found in:
  - **Regression Discontinuity Designs: Theory and Applications**, Advances in Econometrics, volume 38
  - Edited by Matias D. Cattaneo and Juan Carlos Escanciano

- Review Papers
  - Christopher Skovron and Rocio Titiunik (2015), **A Practical Guide to Regression Discontinuity Designs in Political Science**, Working Paper
  - David S. Lee and Thomas Lemieux (2014), **Regression Discontinuity Designs in Social Sciences**, The SAGE Handbook of Regression Analysis and Causal Inference
  - David S. Lee and Thomas Lemieux (2010), **Regression Discontinuity Designs in Economics**, Journal of Economic Literature