Lecture 7:Regression Discontinuity Design

Micro-Econometrics,Fall 2017

Zhaopeng Qu

Nanjing University

12/4/2018

Causal Inference and Regression Discontinuity Design

Introduction

- Social science (Economics) theories always ask causal question
- In general, a typical causal question is: The effect of a treatment(D) on an outcome(Y)
  - Outcome(Y): A variable that we are interested in
  - Treatment(D): A variable that has the (causal) effect on the outcome of our interest
- 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 Experiment** (RCT).

Experimental Idea

- 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
- 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

Causal Inference

- 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"
  - Instrumental Variable Method
  - Regression Discontinuity Design(RDD)

Main Idea of Regression Discontinuity Design

- Regression Discontinuity Design (RDD) exploits the facts that:
  - Some rules are arbitrary and generate a discontinuity in treatment assignment.
  - The treatment assignment is determined based on whether a unit exceeds some threshold on a variable (**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.

A Motivating Example: Elite University

- A large number of studies have shown that graduates from more selective programs or schools earn more than others.
  - e.g Students graduated from NJU earn more than those graduated from other ordinary university.
- 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(**Selection Bias**)

A Motivating Example: Elite University

- But say that the entry cutoff for a score of entrance exam is 400 at NJU.
- Those with scores 399 or even 395 are unlikely to attend NJU, instead attend NUFE(南京财经大学).
- 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 NJU) and the 399 (admitted at NUFE)

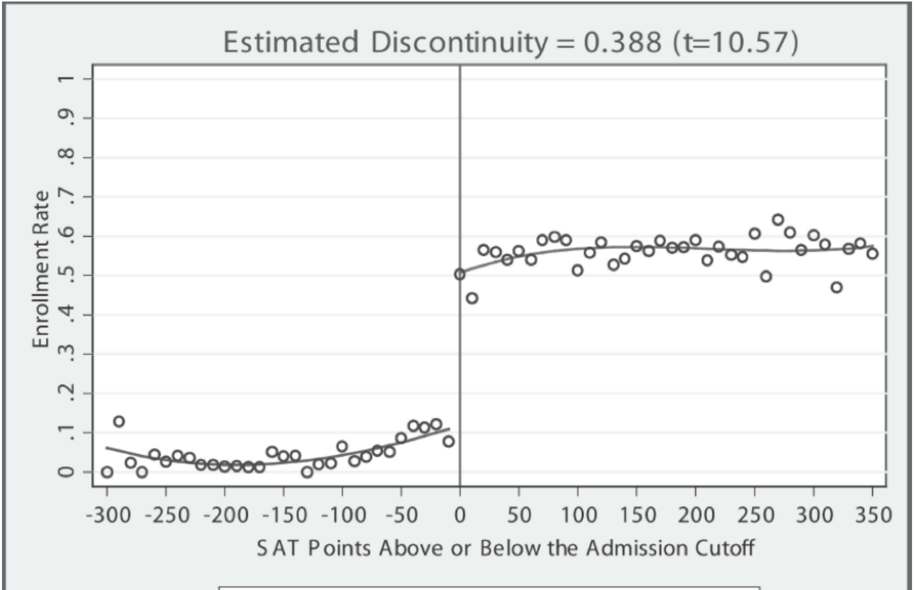
A Motivating Example: Elite University

- 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.

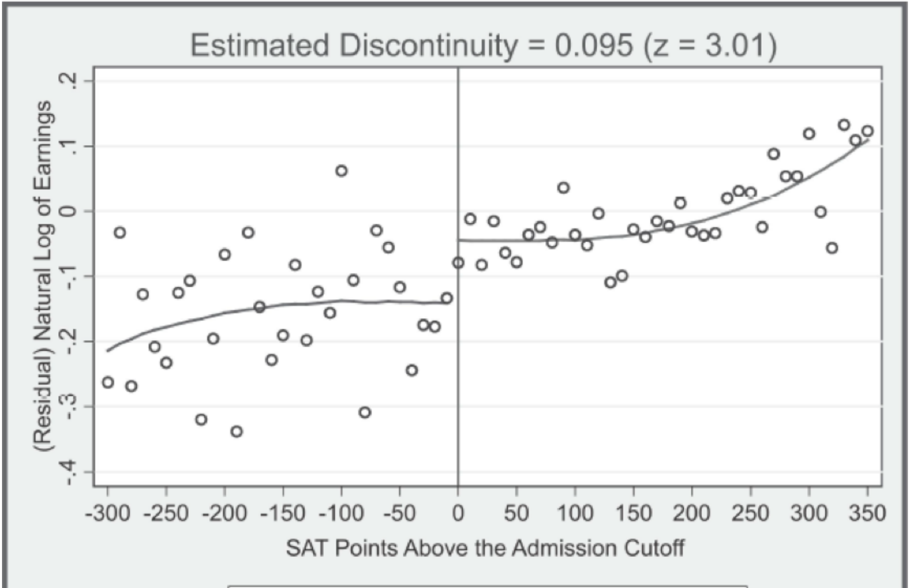
A Motivating Example: Elite University

- For the sake of simplicity, we just focuses on the SAT score.
- 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



SAT Score and Earnings



More Cases of RDD

- Academic test scores: scholarship, prize, higher education admission, certifications of merit.
- Poverty scores: (proxy-) means-tested anti-poverty programs(generally: any program targeting that features rounding or cutoffs)
- Land area: fertilizer program, debt relief initiative for owners of plots below a certain area
- Date: age cutoffs for pensions, dates of birth for starting school with different cohorts, date of loan to determine eligibility for debt relief.
- Elections: fraction that voted for a candidate of a particular party
- Graphically in policy: "China's Huai River policy",Spanish's Salvery "Mita" of colonial Peru in sixteen century.

RD as local randomization

- RD provides “local” randomization if the following assumption holds:
  - Agents have **imperfect** control over the assignment variable X.
- Intuition: the randomness guarantees that the potential outcome curves are smooth (e.g continuous) around the cutoff point.
- There is an element of randomness to whether a given individual is treated.
- Why would it be a problem if agents had perfect control over X?

RDD: Theory and Application

RDD and Potential Outcomes

- Treatment
  - assignment variable (running variable):X_i
  - Threshold (cutoff) for treatment assignment:c
  - Treatment variable: D_i and treatment assignment rule is

$$D_i = 1\, if\; X_i \geq c\; and\; D_i = 0\, if\; X_i < c$$

RDD and Potential Outcomes

- Potential Outcomes
  - Potential outcome for an individual i with treatment,Y_{1i}
  - Potential outcome for an individual i without treatment,Y_{0i}
- Observed Outcomes

$$Y_{1i}\, if\ D_i = 1(X_i \geq c)\, and\, Y_{0i}\, if\ D_i = 0(X_i < c)$$

Identification for Sharp RDD

- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment.
- We can use sharp RDD to investigate the behavior of the outcome around the threshold

$$\alpha_{SRD} = lim_{\varepsilon \to 0} E[Y_i|X_i = c + \varepsilon] - lim_{\varepsilon \to 0} E[Y_i|X_i = c - \varepsilon]$$

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 for actual treatment.

Identification for Sharp RDD

- **Deterministic Assumption**

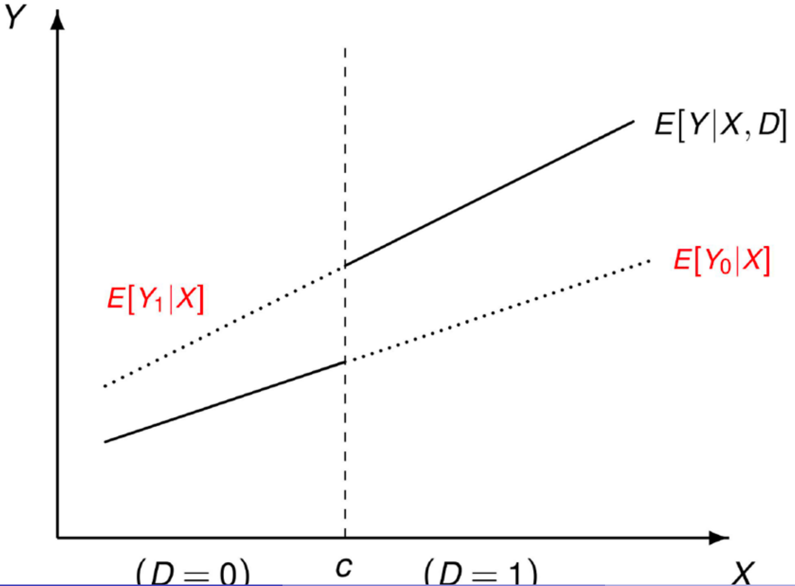
$$D_i = 1(X_i \geq c)$$

- Treatment assignment is a deterministic function of the assignment variable X_i and the threshold c.

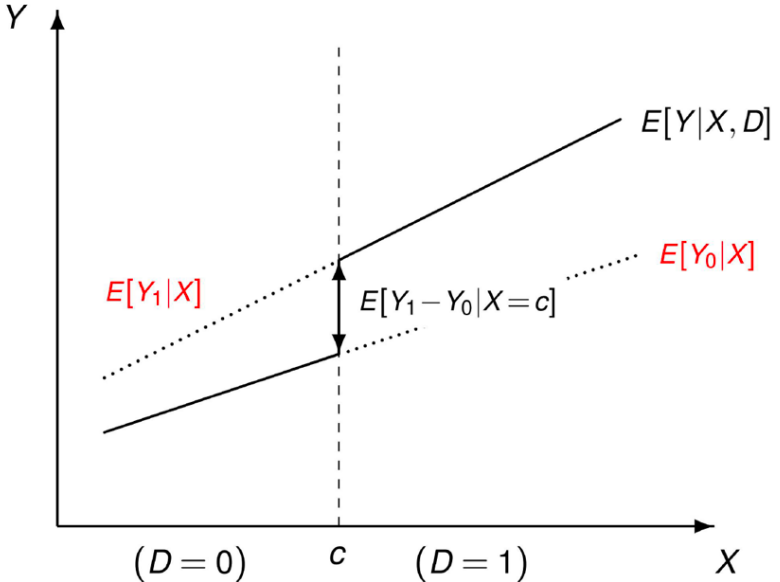
Identification for Sharp RDD

- **Continuity Assumption**
  - $E[Y_{1i}|X_i]$ and $E[Y_{0i}|X_i]$ are continuous at $X_i = c$
- Assume potential outcomes do not change at cutoff.
- 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.

Graphical Interpretation



Graphical Interpretation



Continuity Assumption

- Continuity is a natural assumption but could be **violated** if:
  1. There are differences between the individuals who are just below and above the cutoff that are NOT explained by the treatment.
     - The same cutoff is used to assign some other treatment.
     - Other factors also change at cutoff.
  2. Individuals can **fully manipulate** the running variable in order to gain access to the treatment or to avoid it.

Sharp RDD specification

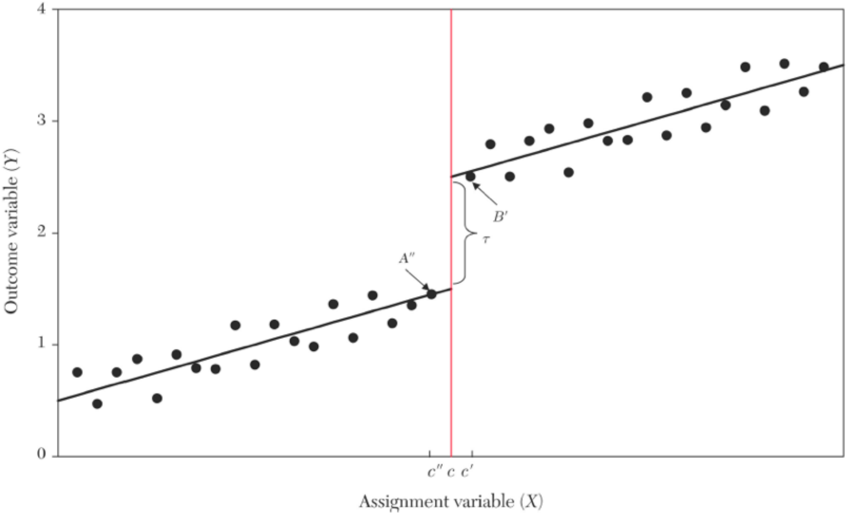
- A simple RD regression is

$$Y_i = \alpha + \rho D_i + \gamma(X_i - c) + u_i$$

  - Y_i is the outcome variable
  - D_I is the the treatment variable(indepent variable)
  - X_i is the running variable
  - c is the value of cut-off
  - u_i is the error term including other factors

- **Qustion**: Which parameter do we care about the most?

Linear Specification

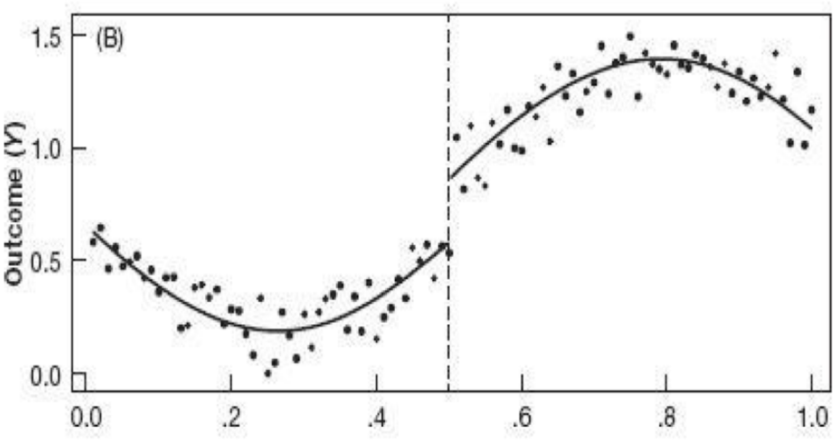


Internal Validity of RD Estimates

- the validity of RD estimates depends crucially on the assumption that the polynomials provide an adequate representation of $E[Y_{0i}|X]$
- If not what looks like a jump may simply be a non-linear in $f(X_i)$ that the polynomilas have not accounted for.

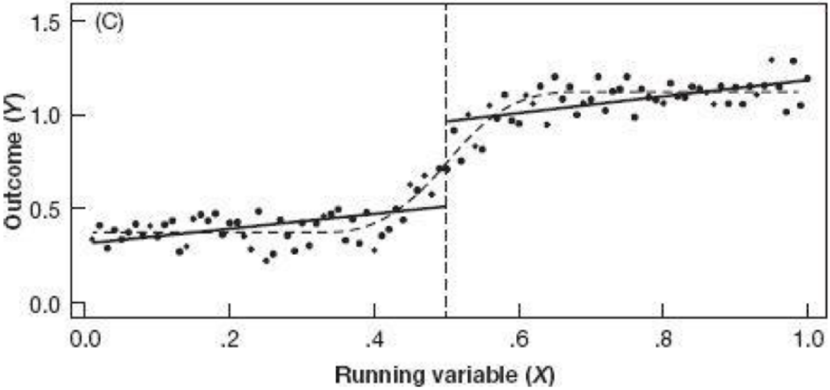
Nonlinear Case

- What if the Conditional Expection Function is **nonlinear**?



Nonlinear Case

- The function form is very important in RDD.



Specification in RDD

- Suppose that in addition to the assignment mechanism above, potential outcomes can be described by some reasonably smooth function $f(X_i)$

$$E[Y_{i0}|x_i] = \alpha + f(x_i)$$
$$Y_{i1} = Y_{i0} + \rho$$

- Simplely, we can construct RD estimates by fitting

$$Y_i = \alpha + \rho D_i + f(x_i) + u_i$$

Specification in RDD

- More generally,we could also estimate two separate regressions

$$Y_i^b = \beta^b + f(x_i^b - c) + u_i^b$$
$$Y_i^a = \beta^a + g(x_i^a - c) + u_i^a$$

- Continue Assumption: $f()$ and $g()$ be any continuous function of$(x_i^{a,b} - c)$, and satisfy

$$f(0) = g(0) = 0$$

- We estimate equationa using only data above c and only data below c.
- Then the treatment effet is $\rho = \beta^b - \beta^a$

Specification in RDD

- Can do all in one step; just use all the data at once and estimate:

$$Y_i = \alpha + \rho D_i + f(x_i - c) + D_i \times g(x_i - c) + u_i$$

- What if dropping $D_i \times g(x_i - c)$?
- Answer: assume the same functional form above and below c.

Sharp RDD Estimation

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

Parametric/Global Approach

- In a simple case: a flexible polynomial (p_{th} order polynomial) regression to estimate $f(x_i)$ and $g(x_i)$

$$Y_i = \alpha + \rho D_i + \beta_1 X_i + \beta_1 X_i^2 + ... + \beta_p X_i^P + \eta_i$$

- How to decide which polynomial to use?
  - start with the **eyeball test**,similar to OLS regression

Parametric/Global Approach

- In a comprehensive case:
- Let

$$
\begin{aligned}
f(x_i - c) &= f(\tilde{x}_i) \\
&= \beta_1 \tilde{x}_i + \beta_1 \tilde{x}_i^2 + ... + \beta_p \tilde{x}_i^p
\end{aligned}
$$

Parametric/Global Approach

- The regression model which we estimate is then

$$Y_i = \alpha + \rho D_i + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + ... + \beta_{0p}\tilde{x}_i^p \\ +\beta_1^* D_i\tilde{x}_i + \beta_2^* D_i\tilde{x}_i^2 + ... + \beta_p^* D_i\tilde{x}_i^p + u_i$$

- Where $\beta_1^* = \beta_{11} - \beta_{01}$
- The treatment effect at c is ρ

How to Select Select Polynomial Order

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 Select Polynomial Order

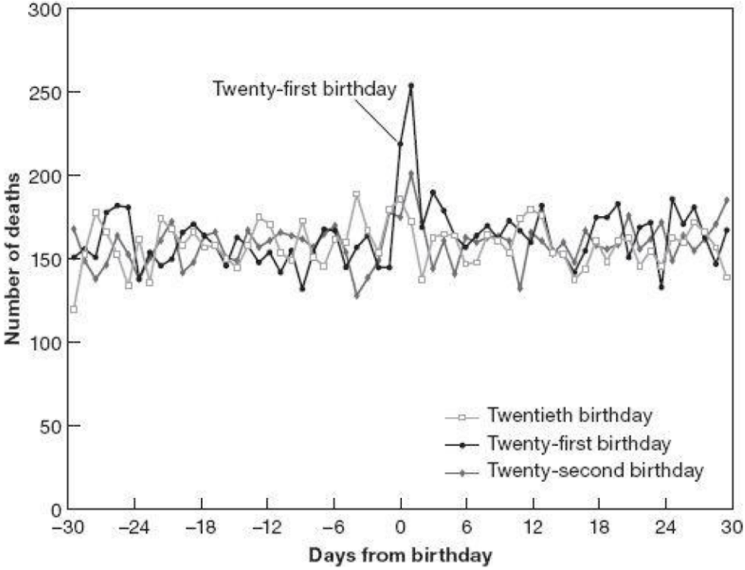
To implement F-Test, one can complete the following steps:

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

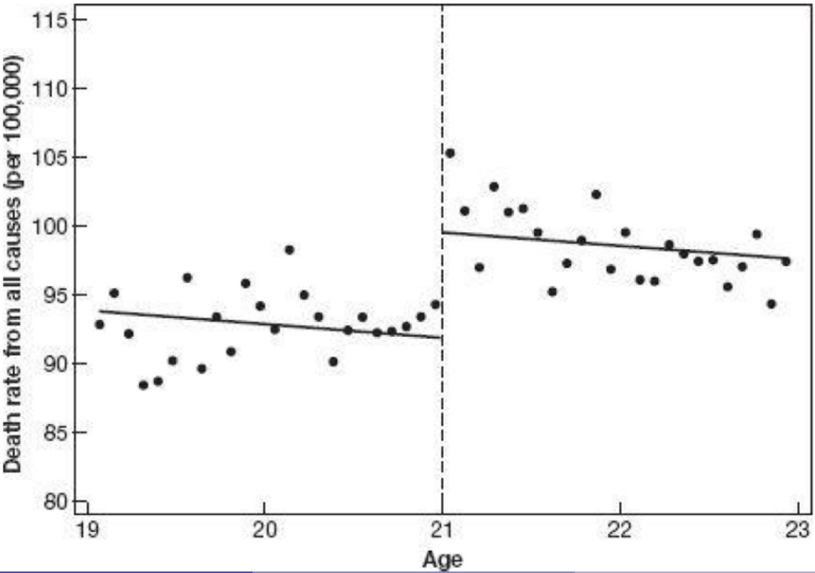
Application: Effect of the Minimum Legal Drinking Age (MLDA) on death rates

- Carpenter and Dobkin (2009)
- Topic: Birthdays and Funerals
- In American, **21th birthday** is an very important milestone. Because over-21s can drink legally.
- Two Views:
  - A group of American college presidents have lobbied states to return the minimum legal drinking age (MLDA) to the Vietnamera threshold of 18.
  - They believe that legal drinking at age 18 discourages binge drinking and promotes a culture of mature alcohol consumption.
  - MLDA at 21 reduces youth access to alcohol, thereby preventing some harm.
- Which one is right?

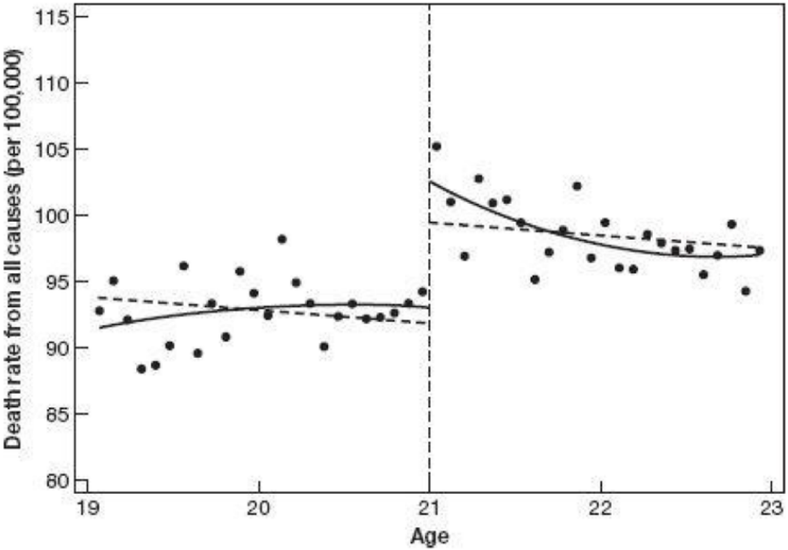
Application: MLDA on death rates



Application: MLDA on death rates



Application: MLDA on death rates



Application: MLDA on death rates:

- The cut off is age 21, so estimate the following regression with cubic terms

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

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

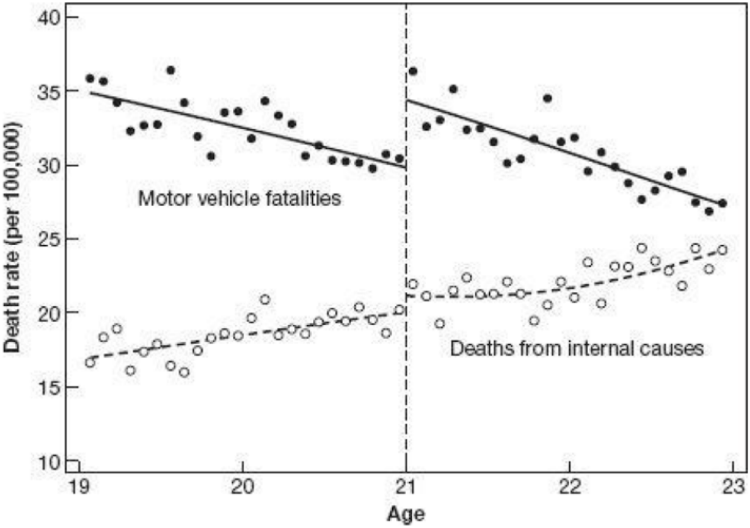
Application: MLDA on death rates

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,

Application: MLDA on death rates



Nonparametric/Local Approach

- Recall we can construct RD estimates by fitting

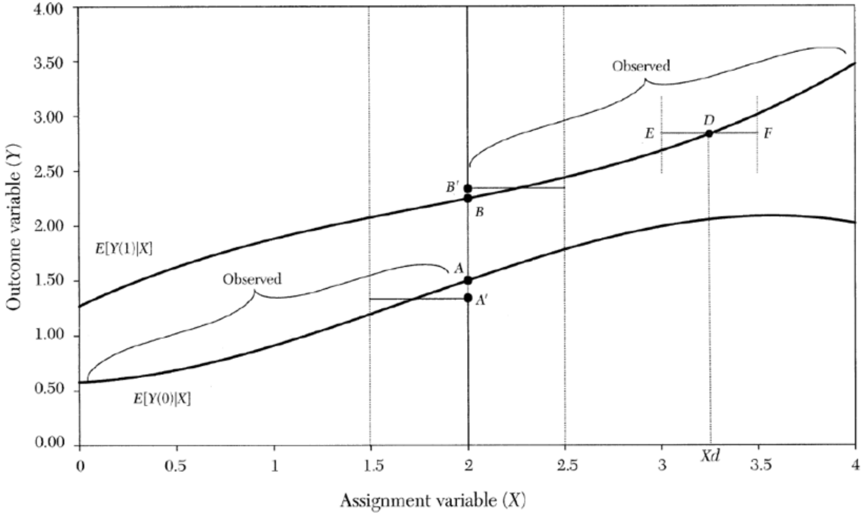
$$Y_i = \alpha + \rho D_i + f(x_i) + u_i$$

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

Nonparametric/Local Approach

- However, comparing means in the two bins adjacent to the cutoff is generally **biased** in the neighborhood of the cutoff.
- This is called **boundary bias**.

Nonparametric/Local Approach:boundary bias



Nonparametric/Local Approach:boundary bias

- The main challenge of nonparametric approach is to **choose a bandwidth**.
- There is essentially a trade-off between bias and precision
- Use a larger bandwidth:
- Get more **precise** treatment effect estimates since more data points are used in the regression.
- But the linear specification is less likely to be accurate and the estimated treatment effect could be biased.

Nonparametric/Local Approach

- The standard solution to reduce the boundary bias is to run **local linear regression**.
- Local linear regressions is a nonparametric method which is linear smoother within a given bandwidth (window) of width h around the threshold.
- Usually, we would present the RD estimates by different choices of bandwidth.

Fuzzy RD

- In sharp RDD not the treatment assignment but **the probability of treatment** jumps at the threshold.
  - **Sharp RDD**:
    - the probability of treatment jumps at the threshold from 0 to 1.
    - Nobody below the cutoff gets the "treatment", everybody above the cutoff gets it.

Fuzzy RD

- Treatment Assignment:

$$P(D_i = 1|x_i) = p_1(X_i) \; if \; x_i \geq c, the \; probability \; assign \; to \; treatment \; grou$$
$$P(D_i = 1|x_i) = p_0(X_i) \; if \; x_i < c, the \; probability \; assign \; to \; control \; group$$

- **Fuzzy RDD**: Some individuals *above cutoff* do **NOT** get treatment and some individuals *below cutoff* do receive treatment.
- The result is a research design where the discontinuity becomes an **instrumental variable** for treatment status instead of deterministically switching treatment on or off.

Fuzzy RD v.s Sharp RD

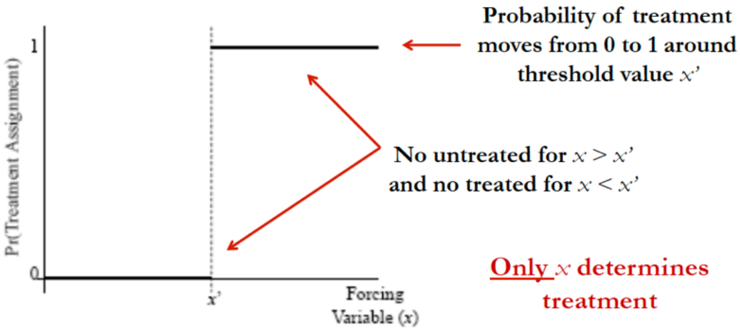


Figure is from Roberts and Whited (2010)

Fuzzy RD v.s Sharp RD

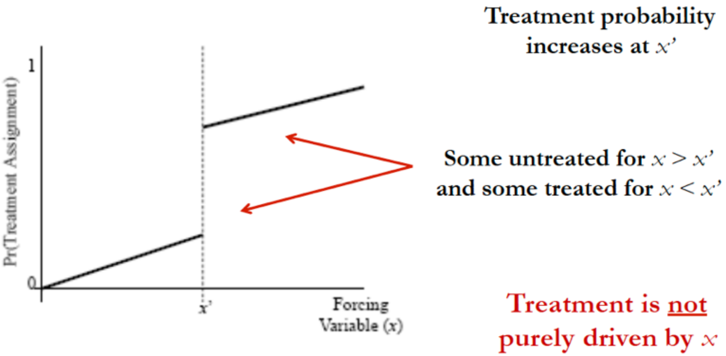


Figure is from Roberts and Whited (2010)

Identification in Fuzzy RD

- Encourage Variable:

$$Z_i = 1 \; if \; assign \; to \; treatment \; group$$
$$Z_i = 0 \; if \; assign \; to \; control \; group$$

- The relationship between the probability of treatment and X_i

$$P(D_i = 1|x_i) = p_0(x_i) + [p_1(x_i) - p_0(x_i)]Z_i$$

Identification in Fuzzy RD

- Recall in SRD, we estimate

$$Y_i = \alpha + \rho D_i + f(x_i - c) + D_i \times g(x_i - c) + u_i$$

- First stage of RD regression:

$$P(D_i = 1|x_i) = \alpha_1 + \phi Z_i + f(x_i - c) + Z_i \times g(x_i - c) + \eta_{1i}$$

- Recall IV terminology: Which one is
  - **endogenous varaible**?
  - **instrumetal varaible**?

Identification in Fuzzy RD

- The **second stage** regression is

$$Y_i = \alpha_2 + \delta \hat{D}_i + f(x_i - c) + \hat{D}_i \times g(x_i - c) + \eta_{2i}$$

- The **reduced form** regression in FRD is

$$Y_i = \alpha_3 + \beta Z_i + f(x_i - c) + Z_i \times g(x_i - c) + \eta_{3i}$$

Fuzzy RD

- Still 2 types of strategies for correctly specifying the functional form in a FRD:
  1. **Parametric**/global method:
  2. **Nonparametric**/local method

Application: Air pollution in China

- Chen et al(2013),"Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy",PNSA,vol.110,no.32.
- Ebenstein et al(2017),"New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy",PNSA,vol.114,no.39.
- Topic: Air pollution and Health
- A Simple OLS regression

$$Health_i = \beta_0 + \beta_1 Air\ pollution + \gamma X_i + u_i$$

- Potential bias?

Application: Air pollution in China

- More elegant Method: SRD and FRD in Geography
- Natural experiment: “Huai River policy” in China
- Result:
  - the results indicate that life expectancies(预期寿命) are about **5.5** year lower in the north owing to an increased incidence of cardiorespiratory(心肺) mortality.
  - the PM_10 is the causal factor to shorten lifespans and an additional 10 $\mu g/m^3$ PM10 reduces life expectancy by **0.86** years.

Application: Air pollution in China

Application: Air pollution in China

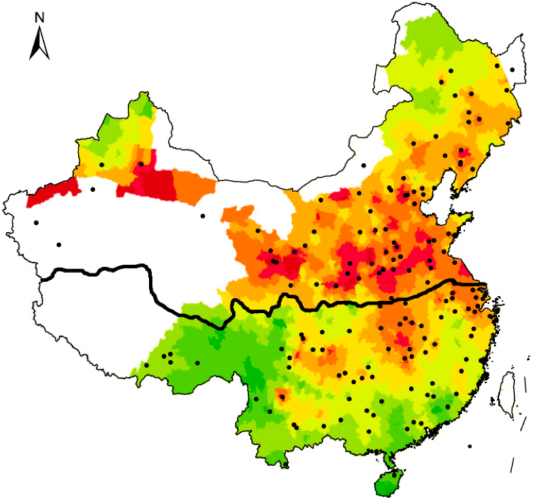


Fig. 1. China's Huai River/Qinling Mountain Range winter heating policy line and PM_{10} concentrations. Black dots indicate the DSP locations. Coloring corresponds to interpolated PM_{10} levels at the 12 nearest monitoring stations

Application: Air pollution in China:Chen et al(2013)

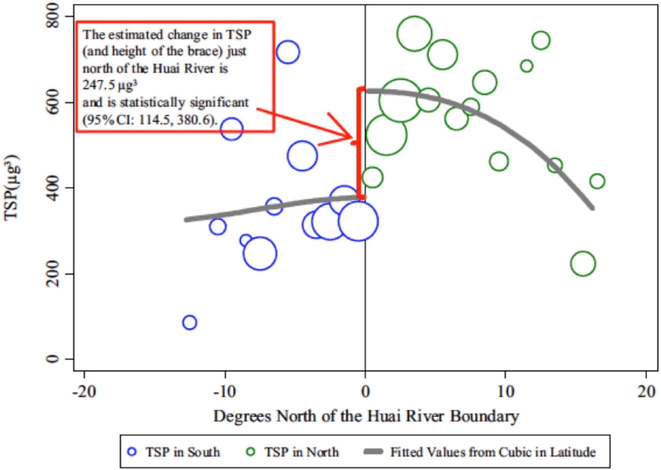


Fig. 2. Each observation (circle) is generated by averaging TSPs across the Disease Surveillance Point locations within a 1° latitude range, weighted by the population at each location. The size of the circle is in proportion to the total

Application: Air pollution in China:Chen et al(2013)

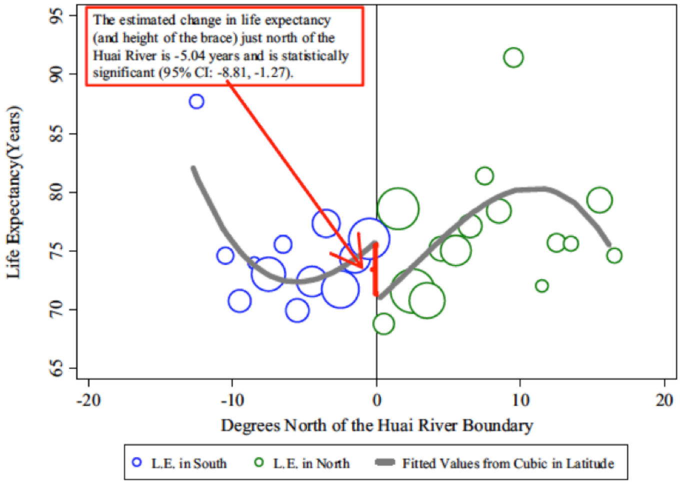


Fig. 3. The plotted line reports the fitted values from a regression of life expectancy on a cubic in latitude using the sample of DSP locations, weighted by the population at each location.

Application: Air pollution in China:Chen et al(2013)

Table 2. Impact of TSPs (100 μg/m^3) on health outcomes using conventional strategy (ordinary least squares)

Dependent variable	(1)	(2)
ln(All cause mortality rate)	0.03* (0.01)	0.03** (0.01)
ln(Cardiorespiratory mortality rate)	0.04** (0.02)	0.04** (0.02)
ln(Noncardiorespiratory mortality rate)	0.01 (0.02)	0.01 (0.02)
Life expectancy, y	−0.54** (0.26)	−0.52** (0.23)
Climate controls	No	Yes
Census and DSP controls	No	Yes

$n = 125$. Each cell in the table represents the coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses. The cardiorespiratory illnesses are heart disease, stroke, lung cancer and other respiratory illnesses. The noncardiorespiratory-related illnesses are violence, cancers other than lung, and all other causes. Models in column (2) include demographic controls and climate controls reported in Table 1. Regressions are weighted by the population at the DSP location. *Significant at 10%. **significant at 5%. ***significant at 1%. Sources: China Disease

Application: Air pollution in China:Chen et al(2013)

Table 3. Using the Huai River policy to estimate the impact of TSPs (100 μg/m³) on health outcomes

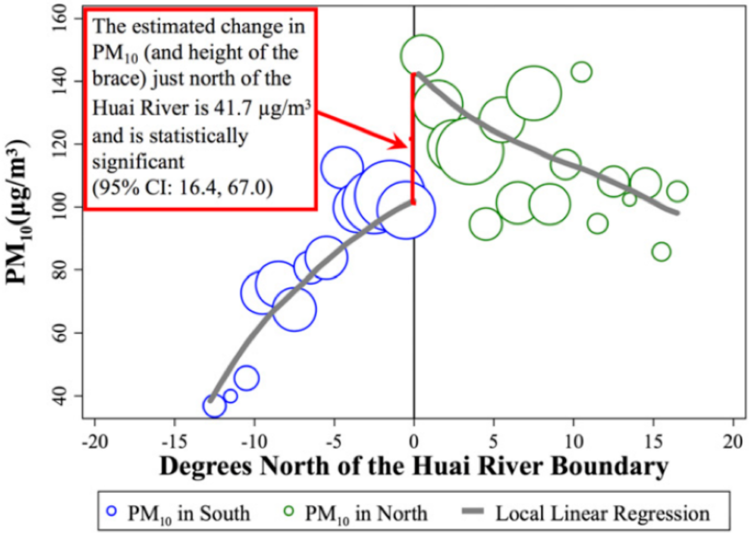
Dependent variable	(1)	(2)	(3)
Panel 1: Impact of "North" on the listed variable, ordinary least squares			
TSPs, 100 μg/m³	2.48*** (0.65)	1.84*** (0.63)	2.17*** (0.66)
ln(All cause mortality rate)	0.22* (0.13)	0.26* (0.13)	0.30* (0.15)
ln(Cardiorespiratory mortality rate)	0.37** (0.16)	0.38** (0.16)	0.50*** (0.19)
ln(Noncardiorespiratory mortality rate)	0.00 (0.13)	0.08 (0.13)	0.00 (0.13)
Life expectancy, y	−5.04** (2.47)	−5.52** (2.39)	−5.30* (2.85)

Application: Air pollution in China:Chen et al(2013)

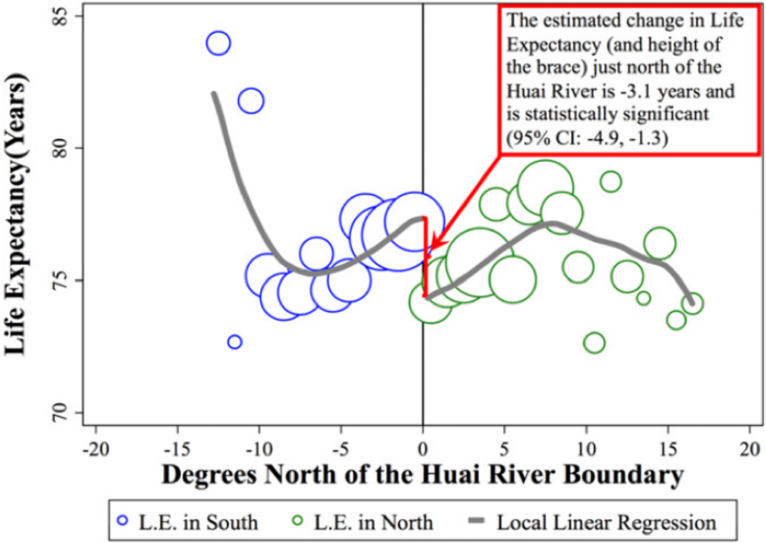
Panel 2: Impact of TSPs on the listed variable, two-stage least squares			
ln(All cause mortality rate)	0.09* (0.05)	0.14** (0.07)	0.14* (0.08)
ln(Cardiorespiratory mortality rate)	0.15** (0.06)	0.21** (0.09)	0.23** (0.10)
ln(Noncardiorespiratory mortality rate)	0.00 (0.05)	0.04 (0.07)	0.00 (0.06)
Life expectancy, y	−2.04** (0.92)	−3.00** (1.33)	−2.44 (1.50)
Climate controls	No	Yes	Yes
Census and DSP controls	No	Yes	Yes
Polynomial in latitude	Cubic	Cubic	Linear
Only DSP locations within 5° latitude	No	No	Yes

The sample in columns (1) and (2) includes all DSP locations ($n = 125$) and in column (3) is restricted to DSP locations within 5° latitude of the Huai River boundary ($n = 69$). Each cell in the table represents the coefficient from a separate regression, and heteroskedastic-consistent SEs are reported in parentheses. Models in column (1) include a cubic in latitude. Models in column (2) additionally include demographic and climate controls reported in Table 1. Models in column (3) are estimated with a linear control for latitude. Regressions are weighted by the population at the DSP location. *Significant at 10%, **significant at 5%, ***significant at 1%. Sources: China Disease Surveillance Points (1991–2000), *China Environment Yearbook* (1981–2000), and World Meteorological Association (1980–2000).

Air pollution in China: Ebenstein et al(2017)



Air pollution in China: Ebenstein et al(2017)



Air pollution in China: Ebenstein et al(2017)

Table 3. Comparing OLS and RD estimates of PM's impact on health outcomes

Outcome	[1]	[2]	[3]
Life expectancy at birth, y	−0.27*** (0.09)	−0.86* (0.51)	−0.64*** (0.22)
Cardiorespiratory (per 100,000, log)	0.02*** (0.01)	0.11* (0.06)	0.08*** (0.03)
Estimation method	OLS	IV	IV
RD type		Polynomial	LLR

Implement of RDD

Three Steps

1. Graph the data for visual inspection
2. Estimate the treatment effect using regression methods
3. Run checks on assumptions underlying research design

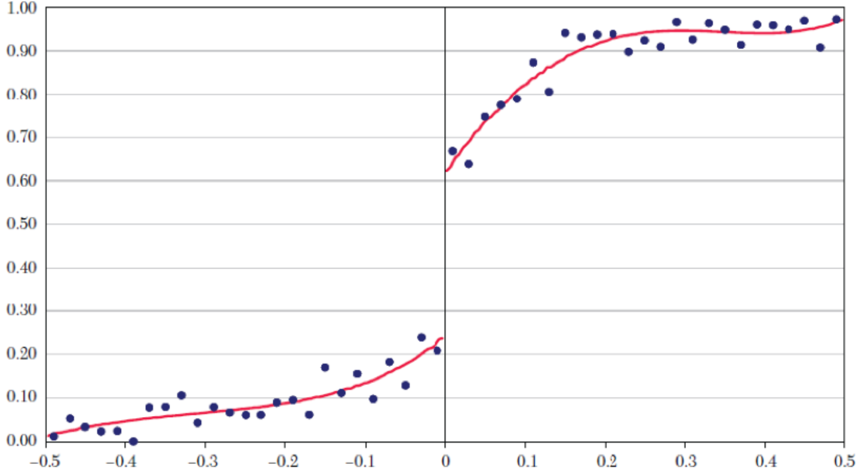
RDD graphical analysis

- First,divide X into bins, making sure no bin contains c as an interior point
  - if x ranges between 0 and 10 and $c = 5$, then you could construct 10 bins:
$$[0, 1), [1, 2), ..., [9, 10]$$
  - if $c = 4.5$, you may use 20 bins, such as
$$[0, 0.5), [0.5, 1), ..., [9.5, 10]$$
- Second,calculate average y in each bin,and plot this above midpoint for each bin.
- Third, plot the forceing variable X_i on the horizontal axis and the average of Y_i for each bin on the vertical axis.(Note: You may look at different bin sizes)
- Fourth,plot predicted line of Y_i from a flexible regression
- Fifth,inspect whether there is a discontinuity at c and there are other

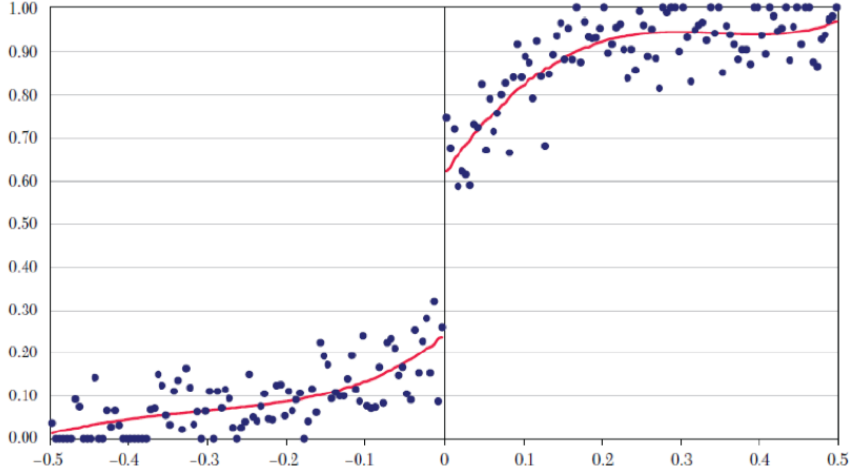
RDD graphical analysis: Select Bin Width

- What is optimal # of bins (i.e. bin width)?
- Choice of bin width is subjective because of tradeoff between precision and bias
    - By including more data points in each average, wider bins give us more precise estimate.
    - But, wider bins might be biased if $E[y|x]$ is not constant within each of the wide bins.
- Sometimes software can help us.

Graphical Analysis in RD Designs: different bin size



Graphical Analysis in RD Designs: different bin size



Estimate the treatment effect using regression methods

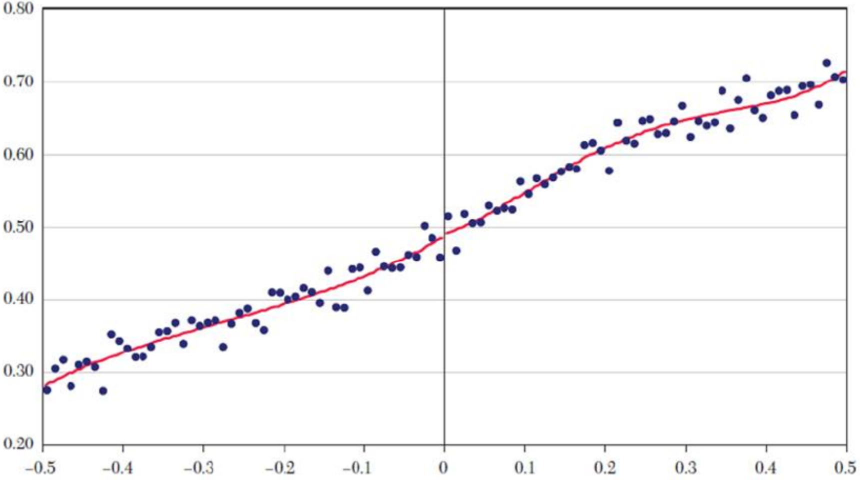
- It is probably advisable to report results for both estimation types:

1. Polynomials in X.
    - In robustness checks you also want to show that including higher order polynomials does not substantially affect your findings.
2. Local linear regression.
    - Your results are not affected if you vary the window(bandwidth)around the cutoff.
    - Standard errors may go up but hopefully the point estimate does not change.

Testing the Validity of the RDD

1. Test involving covariates(Nonoutcome Variable):
    - Test whether other covariates exhibit a jump at the discontinuity. (Just re-estimate the RD model with the covariate as the dependent variable).
    - Construct a similar graph to the one before but using a covariate as the "outcome".
    - There should be no jump in other covariates

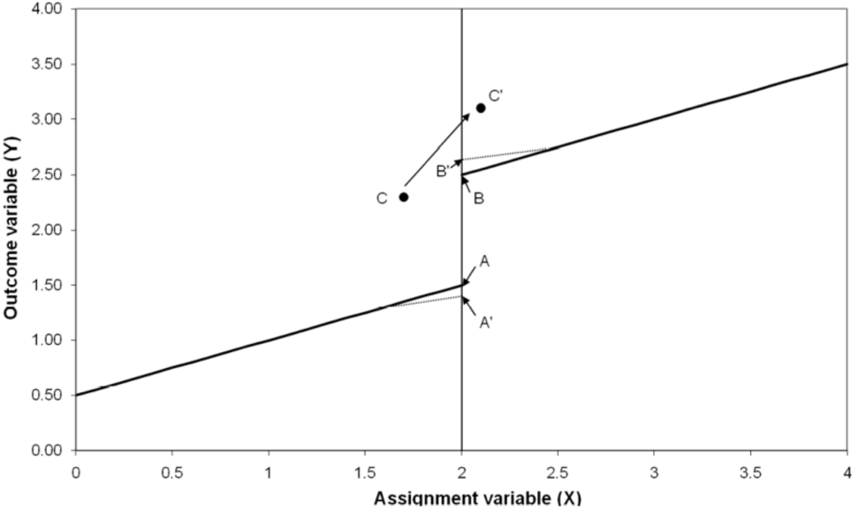
Graphical:Example Covariates by Forcing Variable



Testing the Validity of the RDD

2. Test sorting behavior
   - Individuals may invalidate the continuity assumption if they strategically **manipulate assignment variable X** to be just above or below the cutoff
   - Recall a key assumption of RD is that agents can **NOT perfect** control over the assignment variable X.
   - That is, people just above and just below the cutoff are no longer comparable.

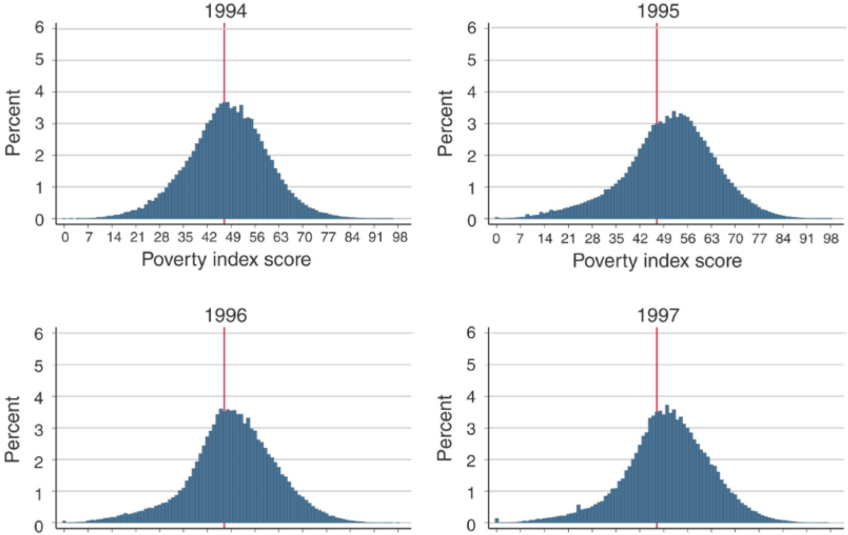
Sorting behavior



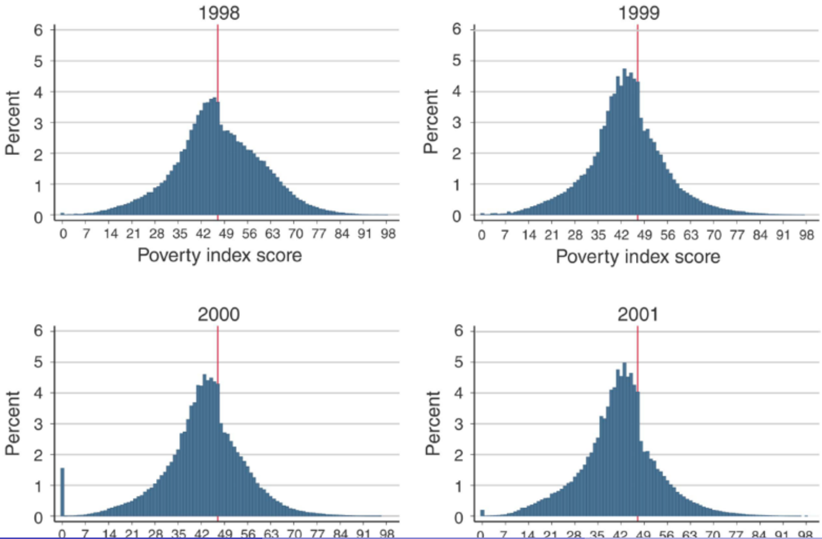
Sorting behavior: Manipulation of a poverty index in Colombia

- Adriana Camacho and Emily Conover (2011) “Manipulation of Social Program Eligibility” AEJ: Economic Policy
- 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: Manipulation of a poverty index in Colombia



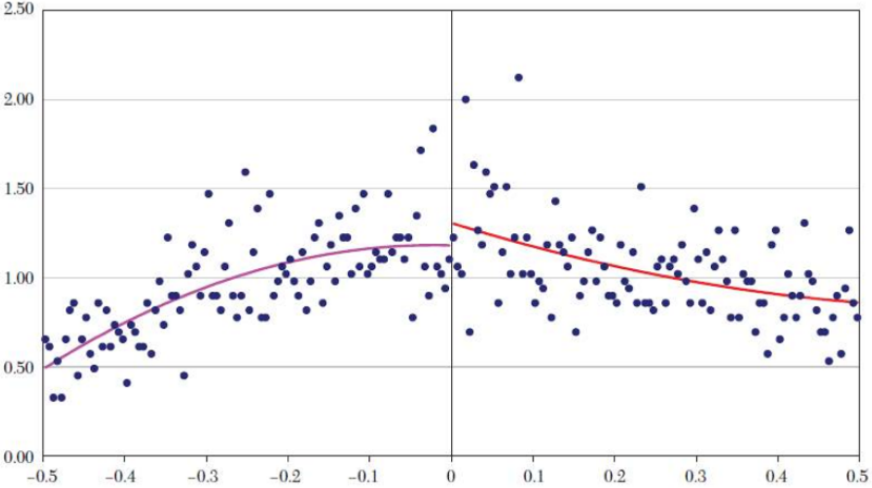
Sorting behavior: Manipulation of a poverty index in Colombia



Testing the Validity of the RDD

- Testing the continuity of the density of assignment variable X: Balance tests
  - 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.
- Also a more formal test: McCrary(2008) test.

Graphical: Density of the Forcing Variable



Testing the Validity of the RDD

- Falsification Tests: testing for jumps at non-discontinuity points
- If threshold x only existed in certain c or for certain types of observations...
- Make sure no effect in c where there was no discontinuity or for agents where there isn't supposed to be an effect.

In a Summary

RDD is a wonderful method in the toolkit of causal inference

- It is so called the nearest method to RCT which identify causal effect of treatment on outcome.
- RDD needs a arbitratry cut-off and agent can **imperfect** manuplate the treatment.
- Two types
  - Sharp RD
  - Fuzzy RD