Lecture 11: Regression Discontinuity Design

Introduction to Econometrics,Spring 2025

Zhaopeng Qu

Business School,Nanjing University

May 14 2025

1. Review Instrumental Variable
2. Regression Discontinuity Design
3. Identification: Model Specification and Bandwidth Selection
4. Application: Effect of the Minimum Legal Drinking Age (MLDA) on death rates
5. Fuzzy RDD: IV and Application
6. Implement of RDD
7. Case Study: Political hierarchy and regional economic development
8. Summary

Review Instrumental Variable

Review the Basic idea of Causal Inference

- **Selection bias** is a major challenge in estimating causal treatment effects.
- **Randomized Controlled Trial** (RCT) offer the best solution to this problem.
- However, RCTs are not always feasible or ethical.
- Alternative strategies focus on **controlling/balancing** the treatment assignment process:
  - **Selection on Observables**: OLS Regression and Matching
  - **Selection on Unobservables**: IV, RDD, DID, SCM

Instrumental Variable (IV)

- IV relies on two key assumptions:
  - **Relevance**: Instrument correlates with the endogenous variable
  - **Exogeneity**: Instrument is uncorrelated with the error term
- The **two-stage least squares (2SLS)** estimator is used for IV estimation
  - While **biased**, it provides **consistent** estimates

Instrumental Variable (IV)

- **Local Average Treatment Effect (LATE)** represents the average treatment effect for compliers
- IV can be viewed as a weighted OLS regression using **inverse first-stage weights**
- Heterogeneity: The **local average treatment effect (LATE)** is the average treatment effect for the **compliers**.
- Key practical considerations:
  - Establishing instrument **Relevance**: first stage regression is crucial.
  - Addressing **Weak Instruments**: first-stage F-test can help.
  - Proving instrument **Exogeneity**: telling a convincing story with reduced form/placebo test/overidentifying restriction test.

Instrumental Variable (IV)

- While IV is a powerful tool, it faces several inherent limitations that have reduced its popularity in economics:
  - Finding valid instruments is challenging
  - Establishing instrument exogeneity is difficult
  - Interpreting the causal effect can be complex
- These limitations highlight the need for alternative, more robust methods to estimate causal effects.
  - Regression Discontinuity Design (RDD) is one of the most popular alternative methods.
  - It is considered as “the most similar method to RCT” among non-experimental methods.

Regression Discontinuity Design

Main Ideas

- **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.

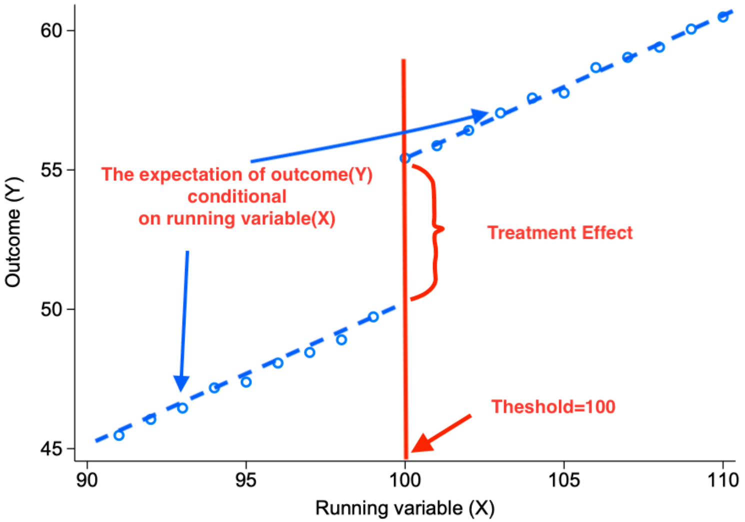
A Motivating Example: Elite University

- Numerous studies have shown that graduates from more **selective** programs or universities earn more than others.
  - e.g Students graduated from **NJU** averagely earn more than those graduated from other ordinary universities like **NUFE**(南京财经大学).
- 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**).
- OLS regression will not give us the right answer for the bias.
- **Question**: Why?

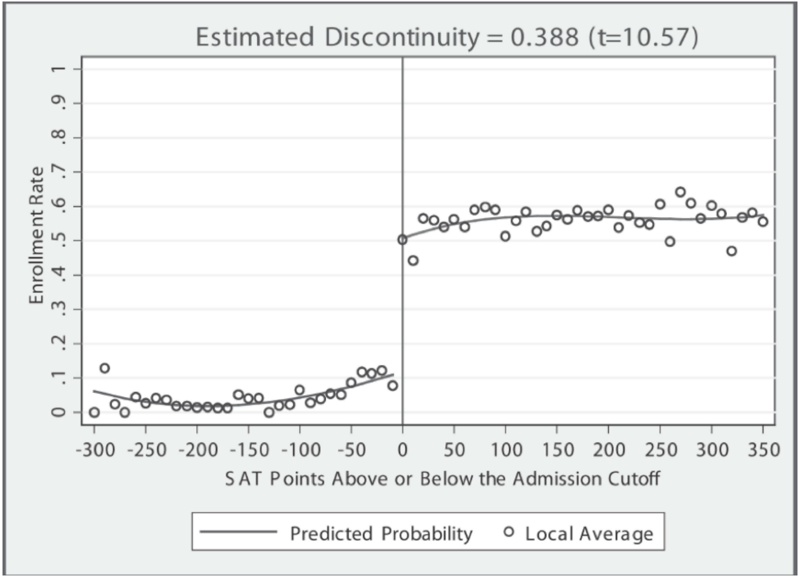
A Motivating Example: Elite University

- But if we could know *National College Entrance Exam Scores*（高考成绩）of all the students. Then we can do something.
- Let us say that the entry cutoff for a score of entrance exam is **600** for NJU.
  - Those with scores **590** or even **599** are unlikely to attend NJU, instead attend NUFE(南京财经大学).
  - Assume that those get *599* or *595* and those get *600* are **essentially identical**, the different scores can be attributed to *some random events*.
- **RDD strategy**:

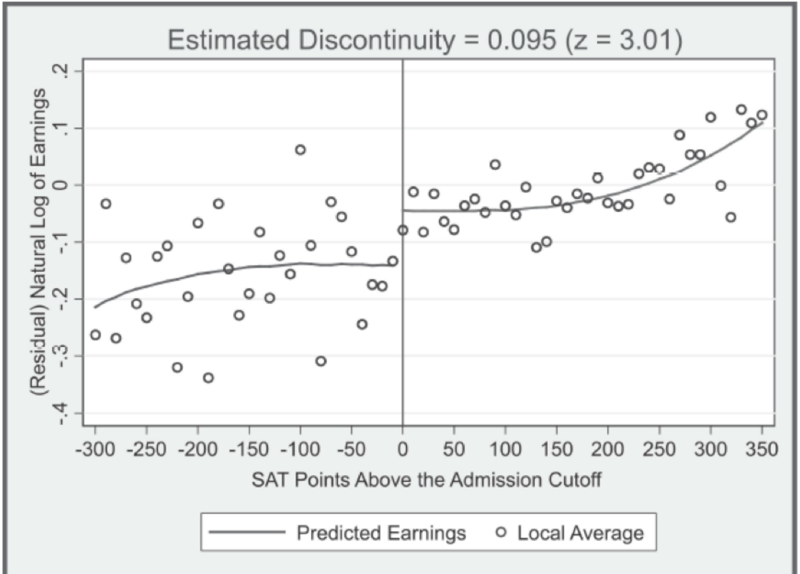
Main Idea of RDD: Outcome



Hoekstra(2009): The flagship state university on Earnings



Hoekstra(2009): The flagship state university on Earnings



More Cases of RDD

- **Academic test scores**: scholarship, prize, higher education admission, certifications of merit.
- **Poverty scores**: means-tested anti-poverty programs.
- **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
- **Geography** in policy: China's Huai River Heating Policy,Spanish's Slavery "Mita" of colonial Peru in sixteen century, and American Air force Bombing in Vietnam War.

RD as Local Randomization

- RD provides "local" randomization if the following assumption holds:
  - Agents have **imperfect** control over the assignment variable X.
- **Assumptions**:

RDD and Potential Outcomes: Notations

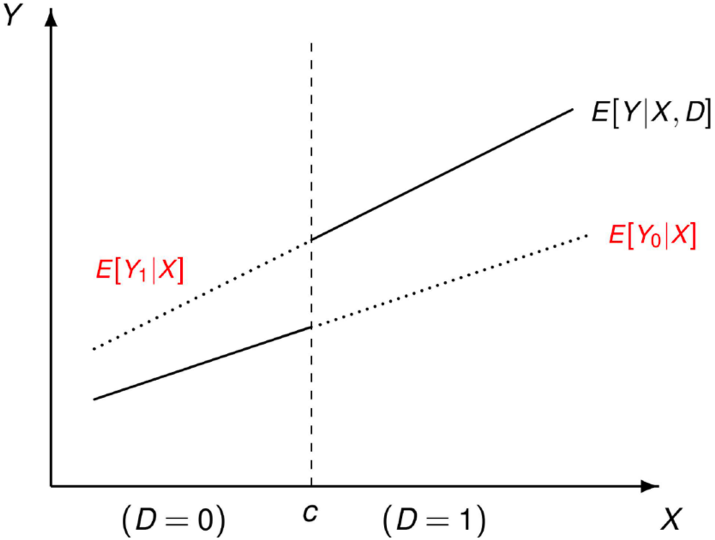
- Treatment
  - **Assignment variable** (running variable): X_i
  - **Threshold** (cutoff) for treatment assignment: c
  - **Treatment variable**: D_i and treatment assignment rule is
- Potential Outcomes
  - Potential outcome for an individual i with treatment, Y_{1i}
  - Potential outcome for an individual i without treatment, Y_{0i}
- Observed Outcomes

Identification for Sharp RDD

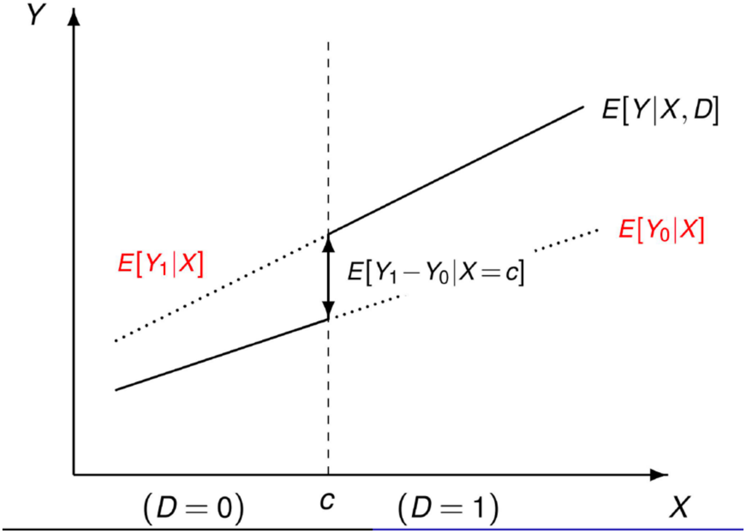
- **Continuity Assumption**

- Which is equivalent to:
  - Assume potential outcomes do not change at cutoff.
  - except treatment assignment, all other unobserved determinants of Y_i are **continuous** at cutoff c.
  - no other confounding factor affects outcomes at cutoff c.
- Then any observed discontinuity in the outcome can be attributed to treatment assignment.
- The treatment effect is identified by the difference in the potential outcomes at the cutoff:

Graphical Interpretation



Graphical Interpretation



Continuity Assumption

- **Continuity** is a natural assumption but could be **violated** if:

Identification: Model Specification and Bandwidth Selection

Basic Parametric RDD specification

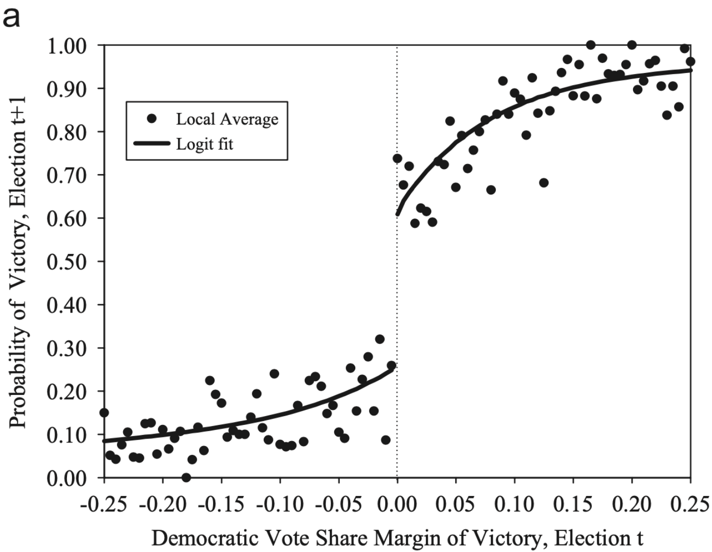
- A simple RD regression is

- 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 **cutoff**
- u_i is the error term including other factors
- **Question**: Which parameter do we care about the most?
- However, this is not enough.
  - Specification and bandwidth selection are very important in RD design.

A Classical Example: Lee(2008)

- Important phenomenon in politics: The incumbency advantage(在任优势)
  - Candidates/parties who won the previous election are **much more likely** to win again.
- Some or all of incumbency advantage could be due to **persistent unobservables**.
  - Position advantage: **name recognition**, **campaign experience**, **networks**, **fundraising** etc.
  - Candidate quality: a **better candidate/party** which is more likely to win last time.
- Lee (2008) uses an RD design to estimate the causal effect of winning US House elections.

Discontinuity for the next election



Basic Parametric RDD specification

- A simple RD regression is

- Y_i is the outcome variable(e.g. The probability of winning the next election)
- D_I is the the treatment variable(eg. Winning the last election)
- X_i is the running variable(e.g. the margin of victory in the last election)
- c is the value of cut-off(e.g. 0)
- u_i is the error term including other factors
- **Question**: Which parameter do we care about the most?
- But Linear function form is not enough.

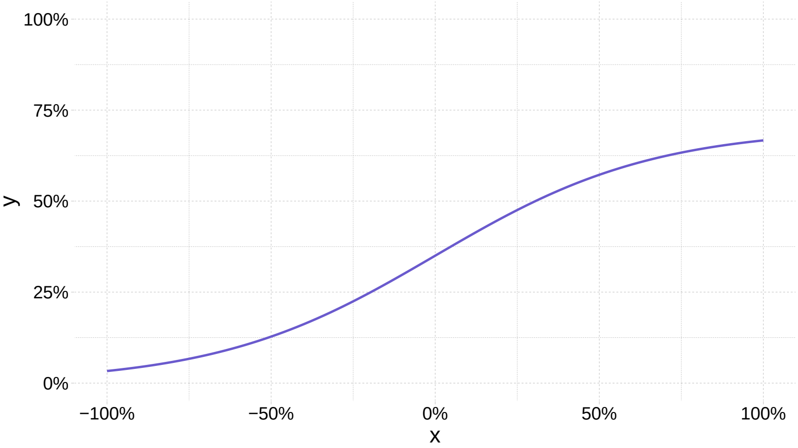
Model Specification and Bandwidth Selection

- Two Keys in specifications in RD :

1. **Specification**: How should we estimate $E\left[\mathrm{Y}_{1i} \mid \mathrm{X}_i\right]$ and $E\left[\mathrm{Y}_{0i} \mid \mathrm{X}_i\right]$?
    - **Parametric**: Estimate treatment effects based on a specific functional form for the outcome and assignment variable relationship.
    - **Nonparametric**: Compare the outcome of treated and untreated observations that lie within specific bandwidth.
2. **Bandwidth Selection**: How much data around the cut-off should we use—*i.e.* the **widows size**
    - **Global**: use all data available.
    - **Local**: only use data with specific bandwidth.

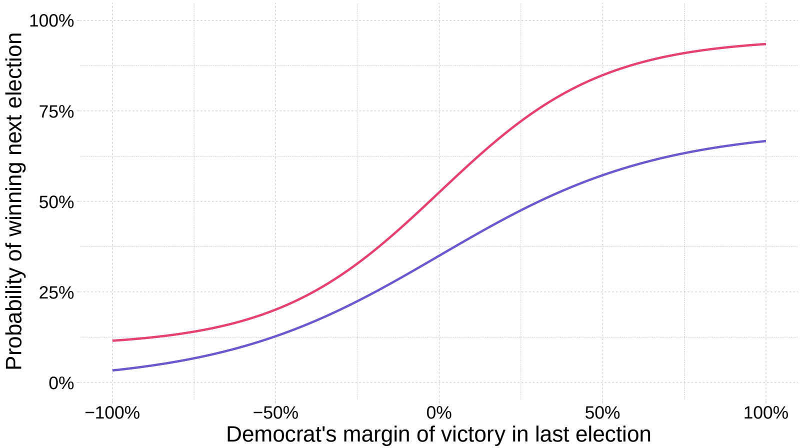
Specification and Bandwidth Selections

- $E\left[\mathbf{Y}_{0i} \mid \mathbf{X}_i\right]$



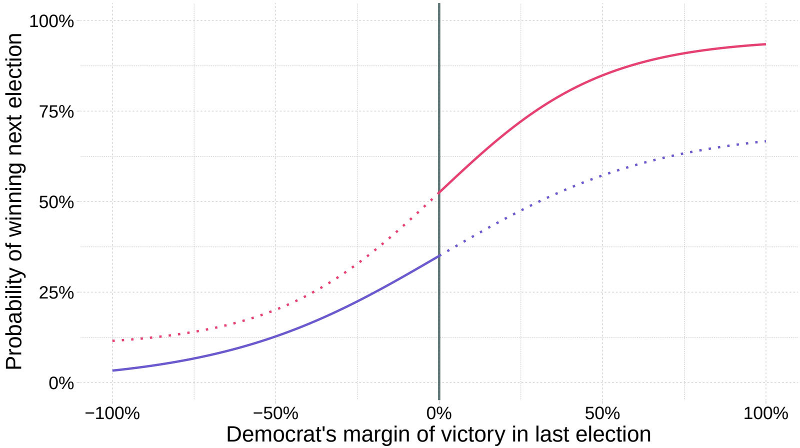
Specification and Bandwidth Selections

- $E\left[\mathbf{Y}_{0i} \mid \mathbf{X}_i\right]$ and $E\left[\mathbf{Y}_{1i} \mid \mathbf{X}_i\right]$



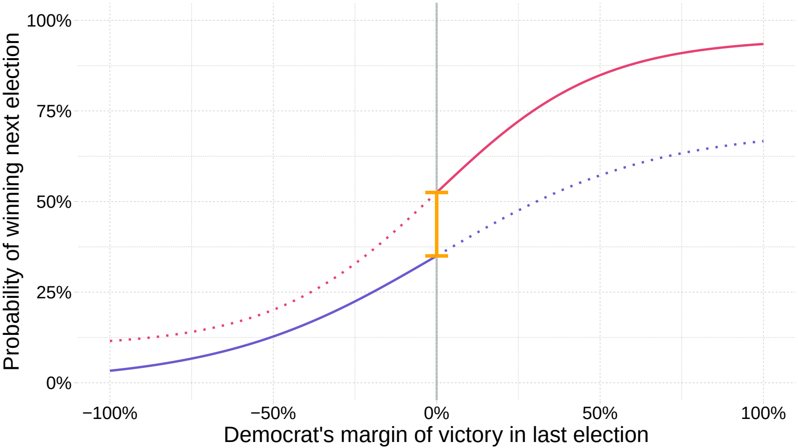
Only one state can be seen

- You only win an election if your margin of victory exceeds zero.



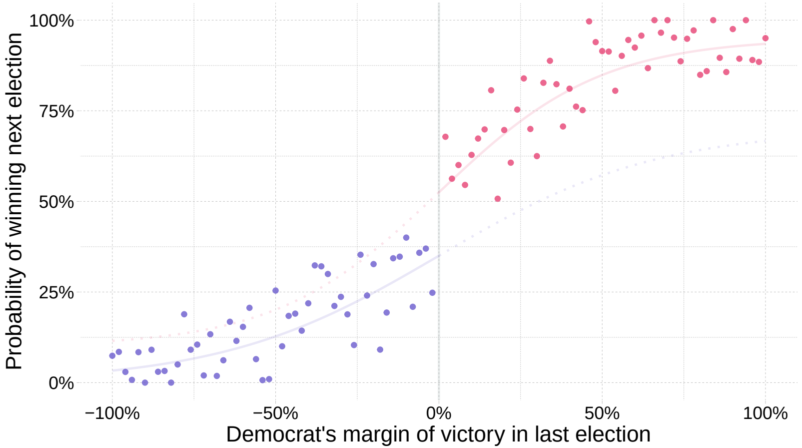
The treatment on the discontinuity

$E\left[\mathbf{Y}_{1i} \mid \mathbf{X}_i\right] - E\left[\mathbf{Y}_{0i} \mid \mathbf{X}_i\right]$ at the discontinuity gives ρ_{SRD}.

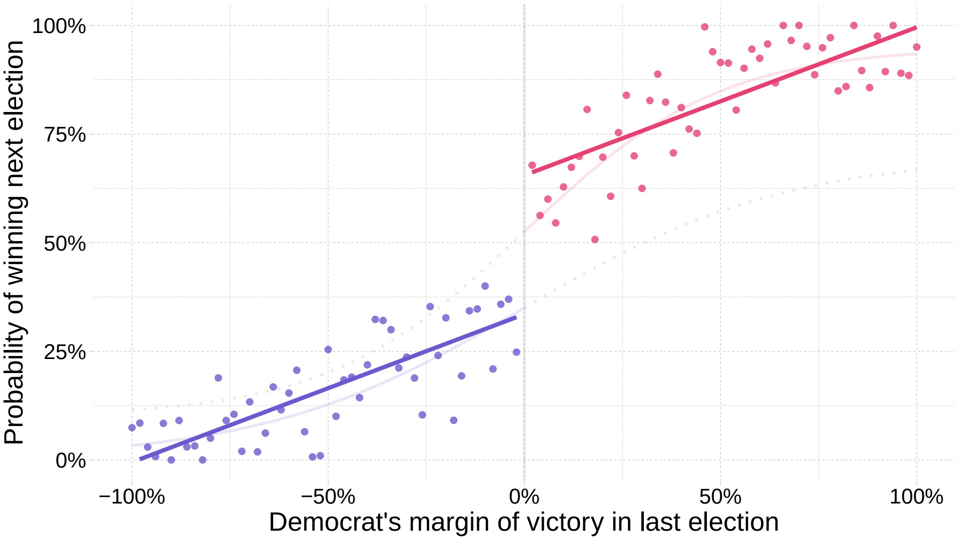


Using data to estimate the treatment effect

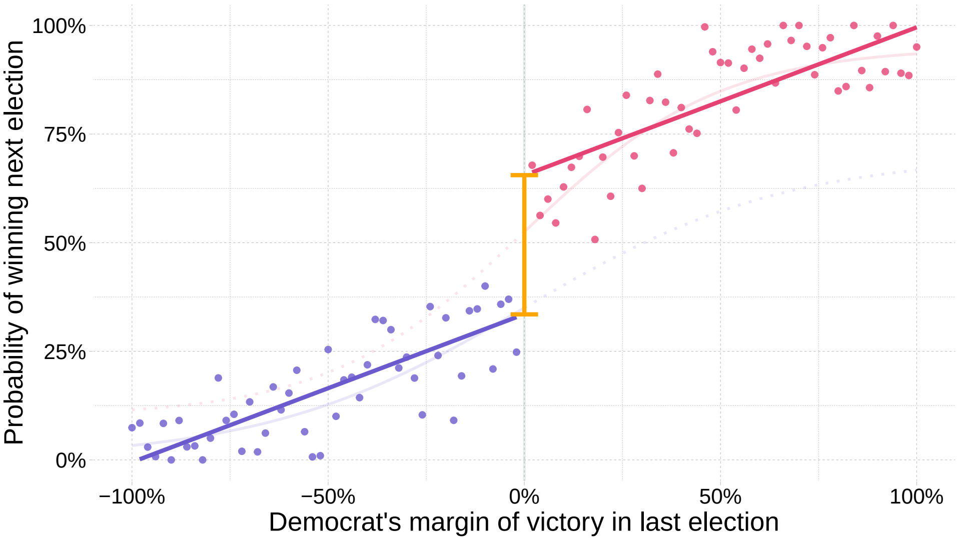
We are going to estimate $E\left[\mathbf{Y}_{1i} \mid \mathbf{X}_i\right]$ and $E\left[\mathbf{Y}_{0i} \mid \mathbf{X}_i\right]$.



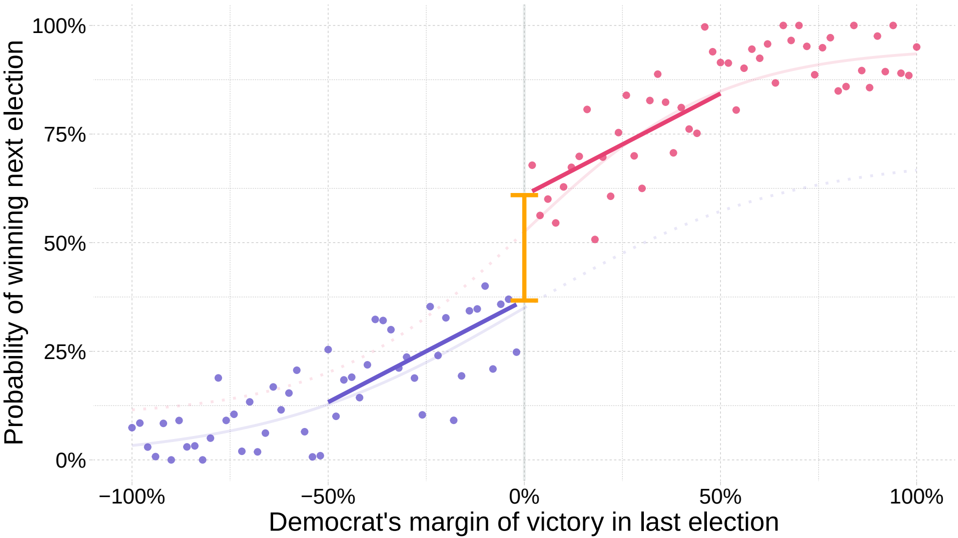
Linear regression with constant slopes (and all data)



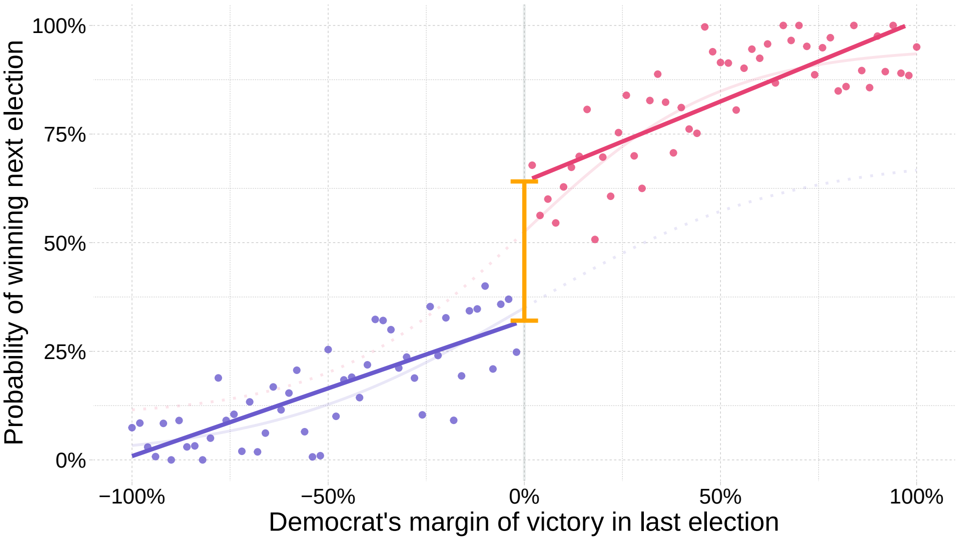
Linear regression with constant slopes (and all data)



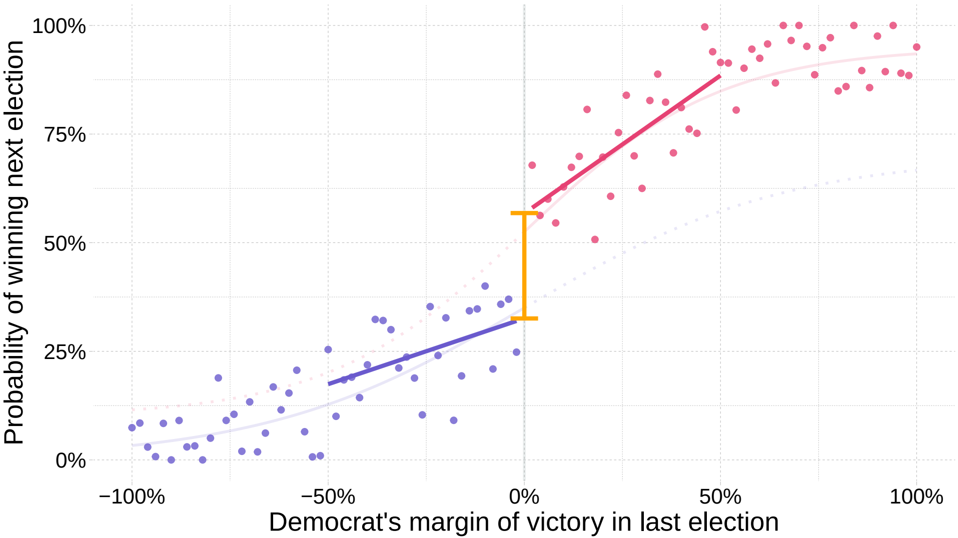
Linear regression with constant slopes; limited to +/- 50%.



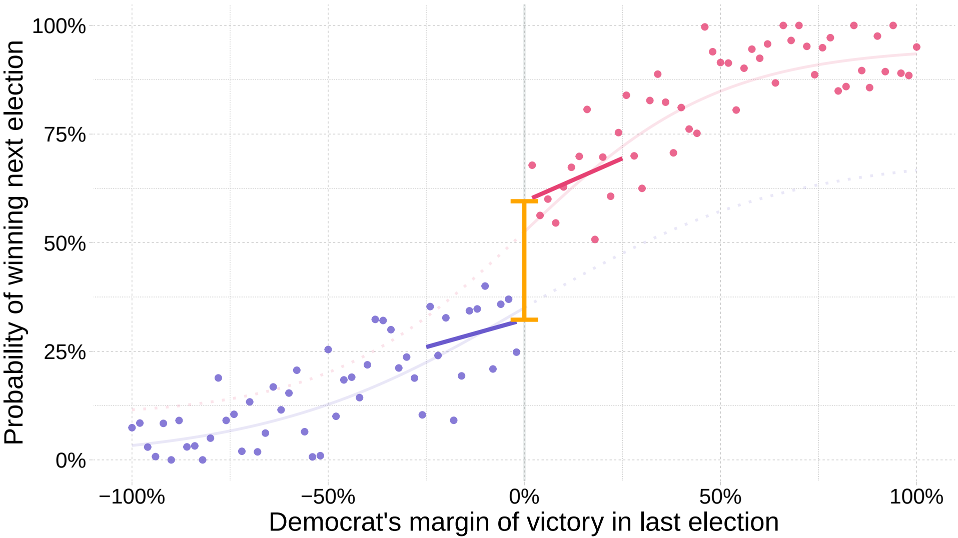
Linear regression with differing slopes (and all data)



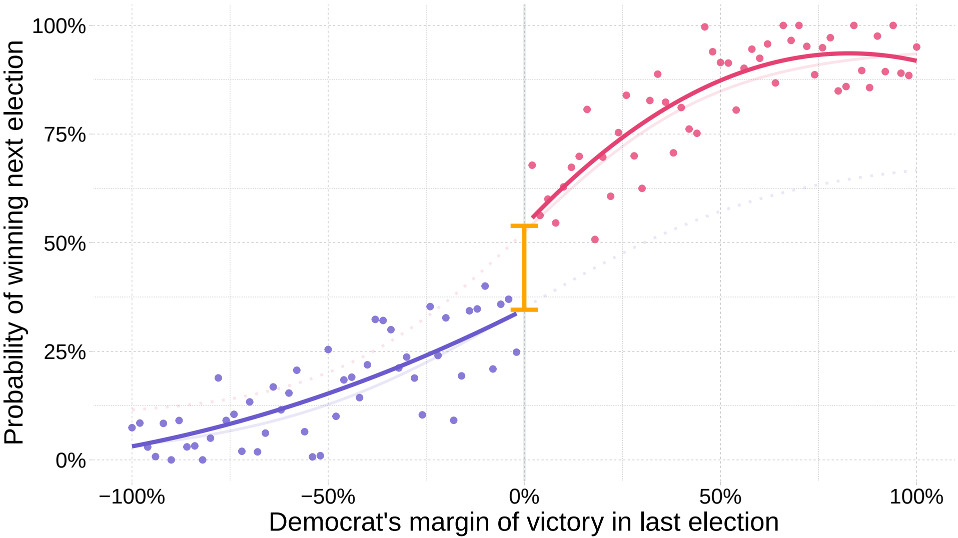
Linear regression with differing slopes; limited to +/- 50%.



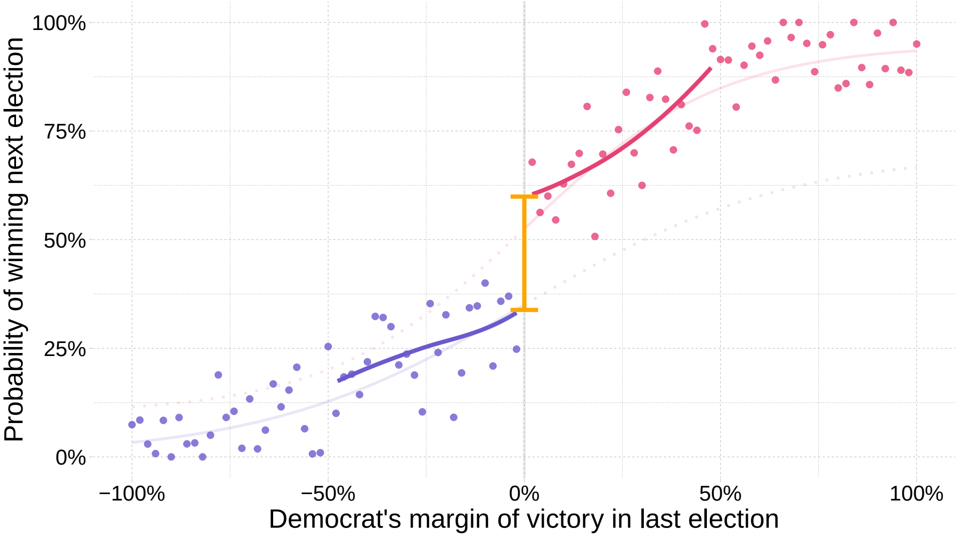
Linear regression with differing slopes; limited to +/- 25%.



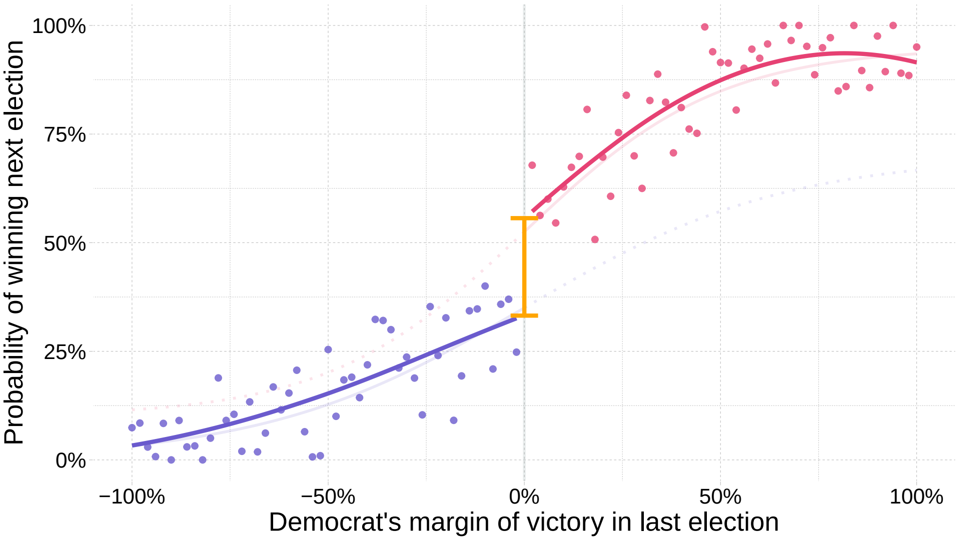
Differing quadratic regressions (all data).



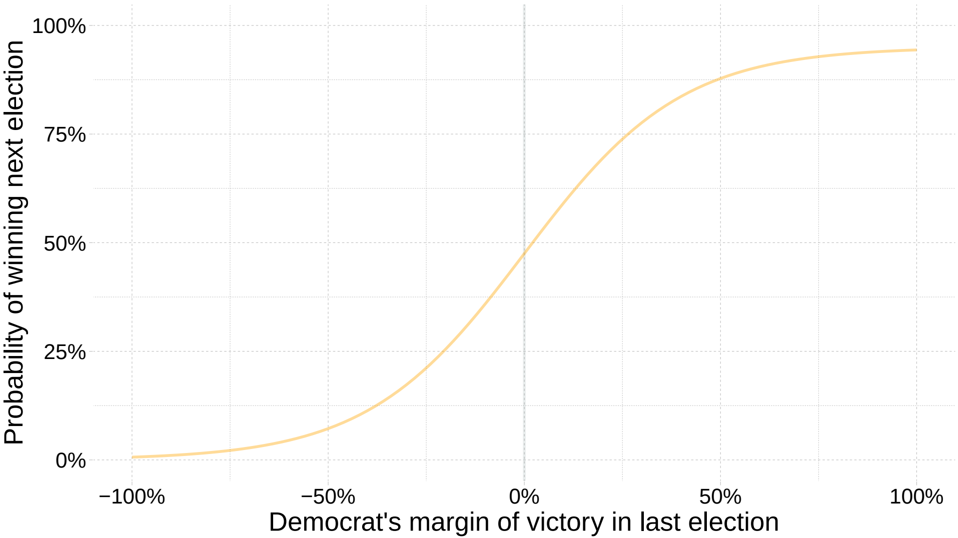
Differing local (LOESS) regressions (limited to +/- 50%).



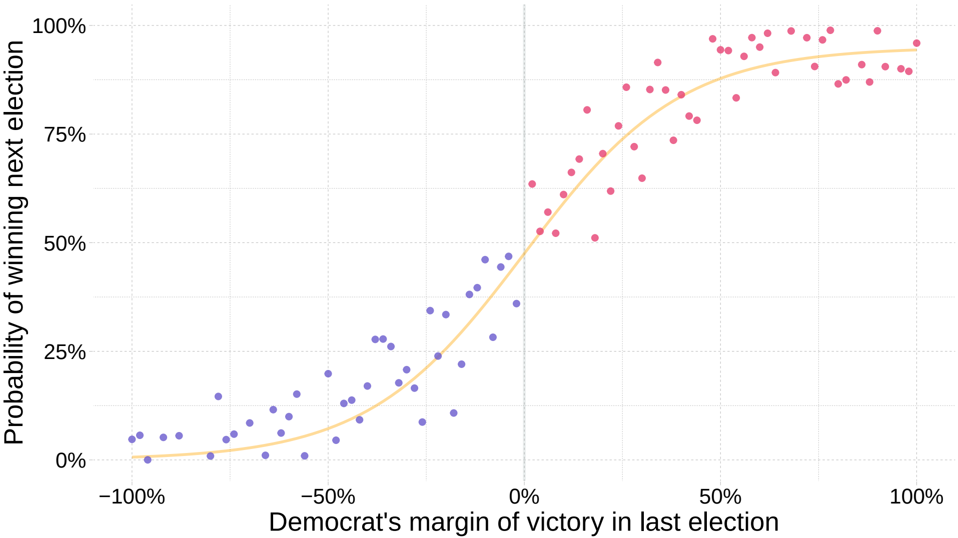
Differing local (LOESS) regressions (all data).



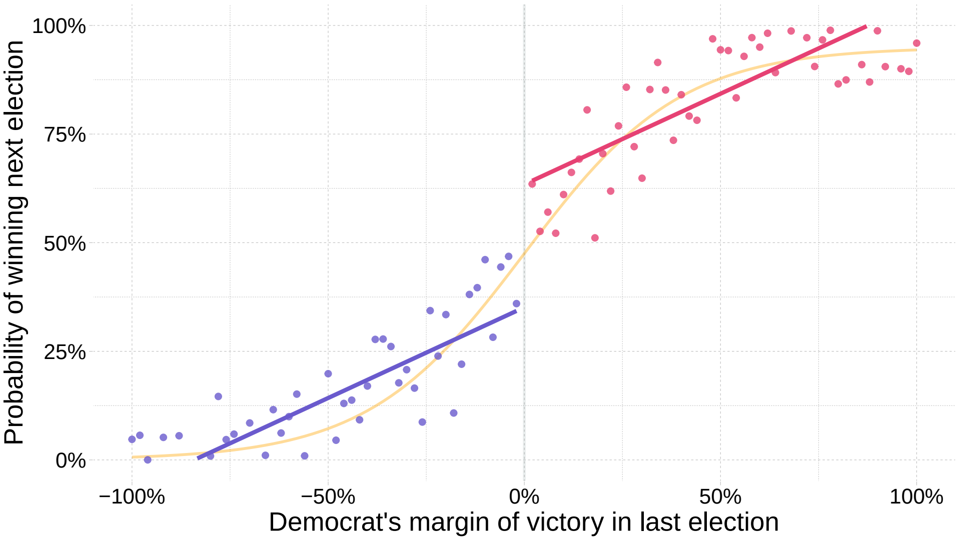
Functional form can be very important



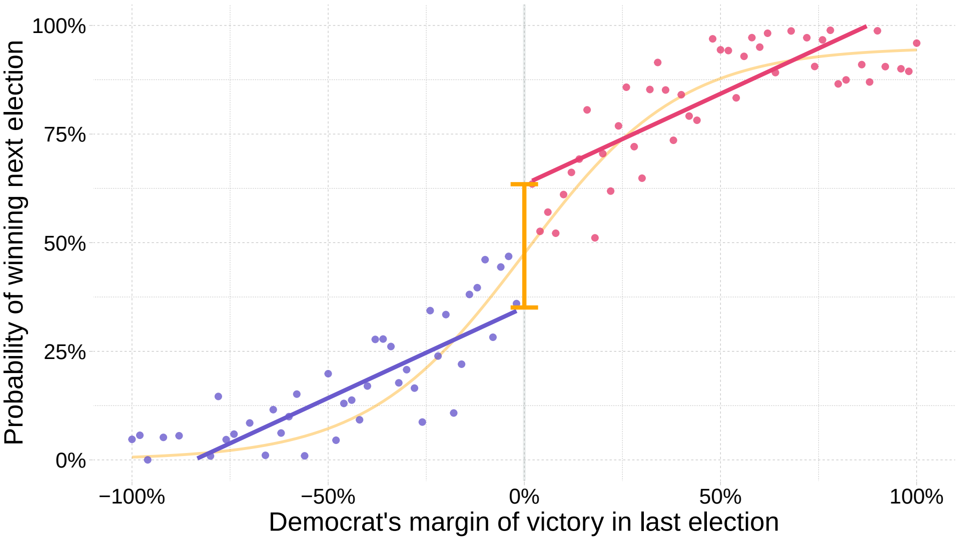
Functional form can be very important



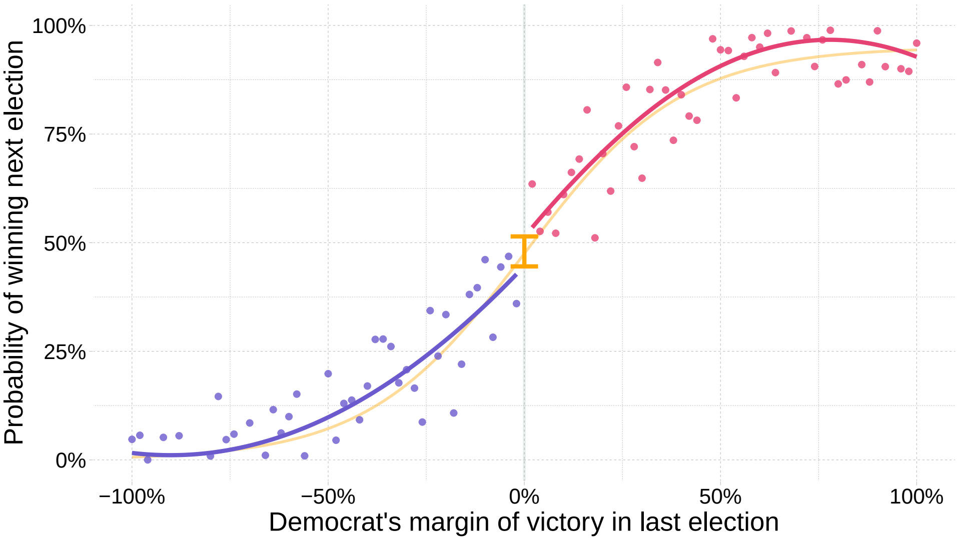
Functional form can be very important



Functional form can be very important



Functional form can be very important



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 method

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

- Simply, we can construct RD estimates by fitting

- Where $f(X_i)$ can be a smooth function of X_i.
- The strong assumption is that $f(X_i)$ is the **same** for all i in *the right* and *the left* side of the cutoff.

Specification in RDD

- Recall **Continuity Assumption**:
  - only for continuity, but no limitions on the functional form of $f(X_i)$
- We could also estimate **two separate regressions** for each side respectively.

- Where Y_i^b is the outcome for the observations below the cutoff, Y_i^a is the outcome for the observations above the cutoff, $f(\cdot)$ and $g(\cdot)$ are continuous functions, and c is the cutoff value.
- Continue Assumption: $f(\cdot)$ and $g(\cdot)$ be any continuous function of $(x_i^{a,b} - c)$, and satisfy

Specification in RDD

- All in one step to estimate the treatment effect:

where D_i is a dummy variable for treated status and ρ is the treatment effect.

  - Then when D_i=0, thus
$$Y_i = \alpha + f(X_i - c) + u_i$$
  - Then when D_i=1,thus
$$Y_i = \alpha + \rho + g(X_i - c) + u_i$$
  - where $g(X_i - c) = f(X_i - c) + h(X_i - c)$
  - $\beta^b = \alpha + \rho$ and $\beta^a = \alpha$

Nonlinear Function forms

- Use a flexible polynomial (pth order polynomial) regression to estimate $f(X_i)$ and $g(X_i)$,thus

$$f(X_i - c) = \beta_1(X_i - c) + \beta_1(X_i - c)^2 + ... + \beta_p(X_i - c)^p$$

- Then we can estimate the following regression:

- How to decide which polynomial to use?
  - start with the **eyeball test**,similar to OLS regression
- Some alternatives
  - F-Test: use F-test in OLS regression to test the order
  - AIC approach: Akaike information criterion (AIC) procedure
  - BIC approach: Bayesian information criterion(BIC) procedure

More Flexible Functional Forms

- Let

$$\begin{aligned} f(X_i - c) &= f(\tilde{X}_i) \\ &= \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + ... + \beta_p \tilde{X}_i^p \\ h(X_i - c) &= h(\tilde{X}_i) \\ &= \beta_1^* \tilde{X}_i + \beta_2^* \tilde{X}_i^2 + ... + \beta_p^* \tilde{X}_i^p \end{aligned}$$

- In a comprehensive case,the regression model which we estimate is then

- The estimated treatment effect at c is still ρ.

Gelman and Imbens (2018)

Gelman and Imbens (2018) on functional form:

- **controlling for global high-order polynomials** is *a flawed approach* with three major problems:
  - it leads to noisy estimates, sensitivity to the degree of the polynomial, and poor coverage of confidence intervals.
- Recommending researchers instead use estimators based on **local linear regression(局部线性回归)** or **quadratic polynomials** or **other smooth functions**.

Nonparametric/Local Approach

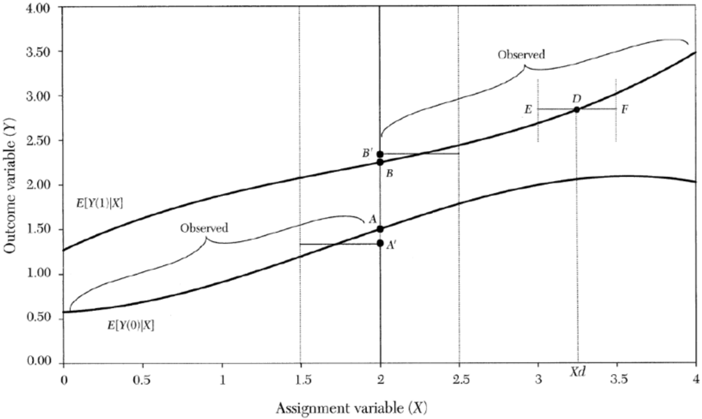
- Recall we can construct RD estimates by fitting

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

- 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:

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

- The most often used nonparametrics method is **local linear polynomial regression**,which is linear smoother within a given interval.
- Thus we estimate the following weighted linear regression within a given window of width h:

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

  - Here we often use some nonparametric functions(such as **kernel**) as the weight, which measures the "distance" to the cut-off.
- *The detail is a little bit beyond the scope of this course. You could refer to Li and Racine(2006) or other nonparametric econometric textbooks.*

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.

How to Choose Bandwidth

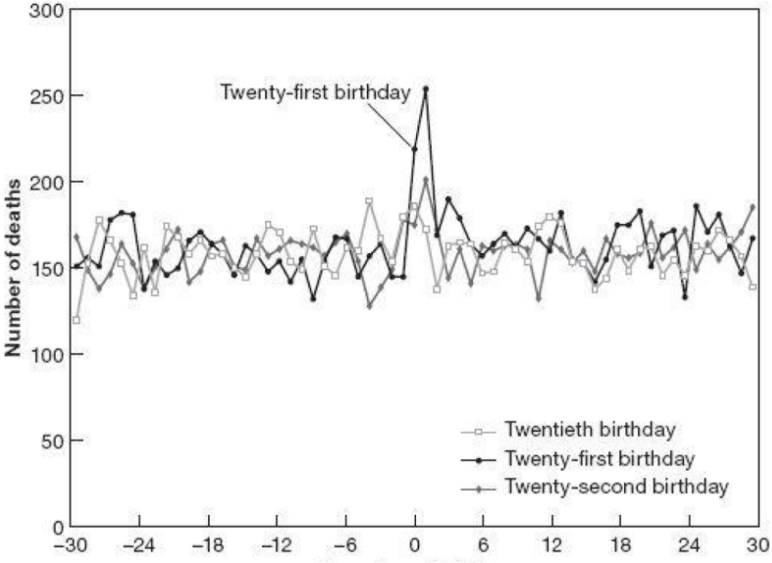
- **Bias/variance trade-off**: Smaller bandwidth reduces bias from using points away from the boundary, but also reduces precision for smaller sample size.
- The optimal bandwidth: use
  - **Cross-Validation Procedure**: Choose the optimal bandwidth h that produces the best fit for the relationship of outcome and assignment variable.
- Usually, we would present the RD estimates by different choices of bandwidth.

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

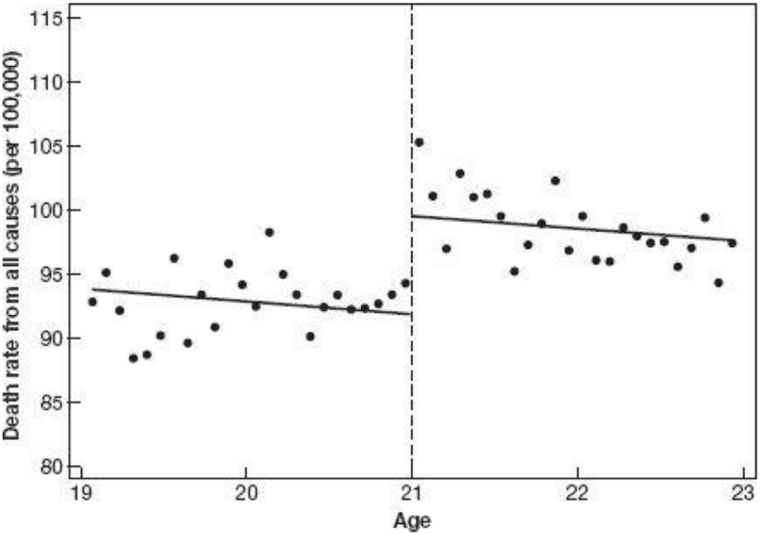
Introduction

- Carpenter and Dobkin (2009): “The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age” American Economic Journal: Applied Economics Vol. 1, No. 1, January 2009 (pp. 164–82)
- **Topic**: Birthdays and Funerals
- In America, the **21st birthday** marks a significant milestone as it represents the legal drinking age.
- Two Competing Views:
  - Some American college presidents have advocated for lowering the minimum legal drinking age (MLDA) back to 18, as it was during the Vietnam era.
  - Proponents argue that legalizing drinking at age 18 would reduce binge drinking and foster more responsible alcohol consumption habits.
  - Opponents maintain that keeping the MLDA at 21 helps prevent harm by restricting youth access to alcohol.
- Which perspective is supported by the evidence?

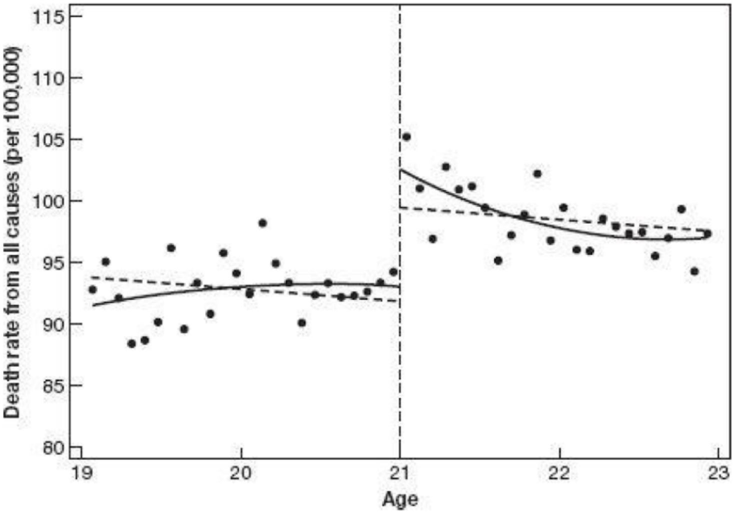
Application: MLDA on death rates



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
- The $f(x_i - 21)$ is the cubic polynomial of $(x_i - 21)$

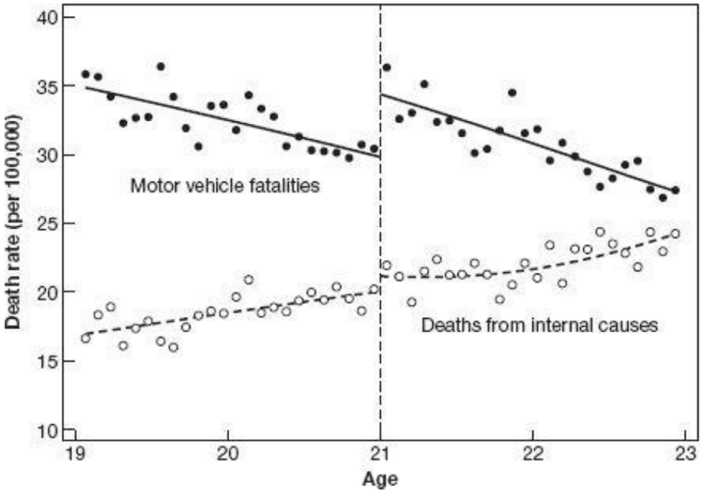
Application: MLDA on death rates

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

Application: MLDA on death rates



Fuzzy RDD: IV and Application

Sharp RDD

- So far, we have assumed that the treatment assignment is **deterministic** at the threshold.
  - Over the cutoff, the treatment assignment is on, thus $D_i = 1$.
  - Under the cutoff, the treatment assignment is off, thus \$D_i=0.
- In a probability framework, the probability of treatment jumps at the threshold
  - Over the cutoff, the probability of treatment is 1, thus $P(D_i = 1|x_i) = 1$.
  - Under the cutoff, the probability of treatment is 0, thus \$P(D_i=1|x_i)=0.
- Thus, in sharp RDD, nobody below the cutoff gets the treatment, everybody above the cutoff gets it.

Fuzzy RDD

- **Fuzzy RDD**: Some individuals *above cutoff* do **NOT** get treatment and some individuals *below cutoff* do receive treatment.
- The probability of treatment is not deterministic at the threshold but a function of X_i,$p_1(X_i)$ and $p_0(X_i)$.

- This creates a research design where the discontinuity serves as an **instrumental variable** for treatment status, rather than directly determining treatment assignment.

Fuzzy RD v.s Sharp RD

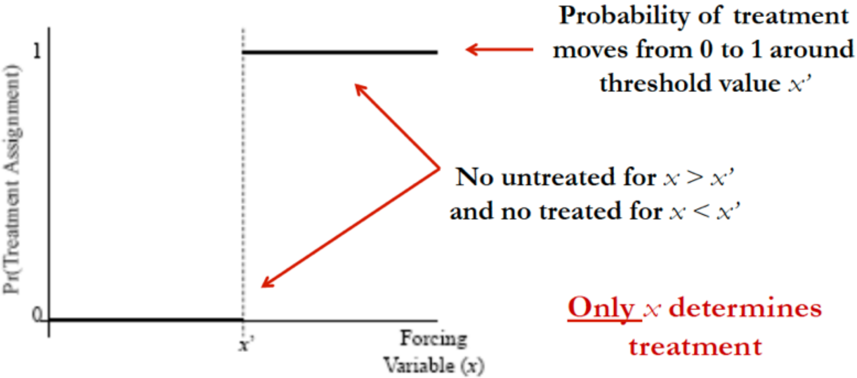
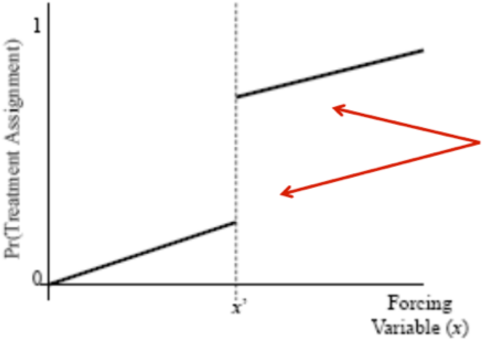


Figure is from Roberts and Whited (2010)

Fuzzy RD v.s Sharp RD



Treatment probability increases at x'

Some untreated for $x > x$' and some treated for $x < x$'

Treatment is <u>not</u> purely driven by x

Figure is from Roberts and Whited (2010)

Identification in Fuzzy RD

- Now besides the treatment varaible, we have an **encourage variable** Z_i, which presents the **eligibility** determined by whether the running variable is above or below the cutoff

- The relationship between the treatment varaible, D_i, and the encourage variable, Z_i,is:

Identification in Fuzzy RD

- Recall in SRD, we estimate

- Then, similar to the SRD, we can estimate the following **First Stage** of FRD regression:

- **Question**:
  1. What is the specification of the **First Stage**?
  2. Which one is the **endogenous varaible**?
  3. Which one is the **instrumetal varaible**?

Identification in Fuzzy RD

- The **second stage** regression takes the form:

- The **reduced form** regression in FRD is specified as:

- Additional covariates can be incorporated into each equation to enhance control.
- The notation and structure remain consistent with standard IV and SRD frameworks.

Fuzzy RDD

- Specification and bandwidth selection remain critical components in FRD, as they are in SRD.
- Two primary approaches for specification and bandwidth selection are available:
    1. **Parametric**/global method
    2. **Nonparametric**/local method
- The validity of the instrumental variable remains a crucial consideration in FRD:
    1. A well-designed RD framework ensures the instrumental variable is enough exogenous.
        - However, the **exclusion restriction** must still be satisfied in FRD.
    2. The relevance of the instrumental variable must be demonstrated.
        - This is essential to avoid the weak instrument problem.

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

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

- Potential bias?

Application: Air pollution in China

- More elegant method: SRD and FRD in Geography
- Natural experiment: “Huai River policy” in China
- Result:
  - 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

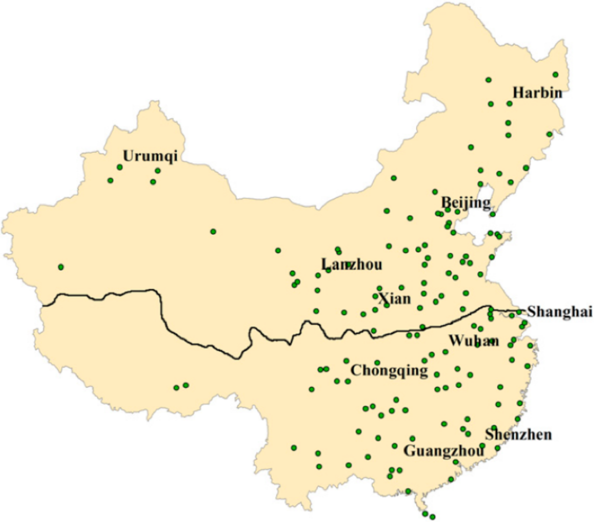


Fig. 1. The cities shown are the locations of the Disease Surveillance Points.

Application: Air pollution in China

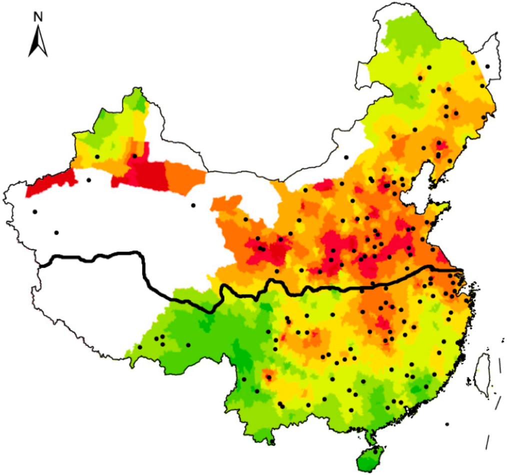


Fig. 1. China's Huai River/Qinling Mountain Range winter heating policy line and PM_{10} concentrations. Black dots indicate the DSP locations. Coloring corresponds to interpolated PM_{10} levels at the 12 nearest monitoring stations,

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

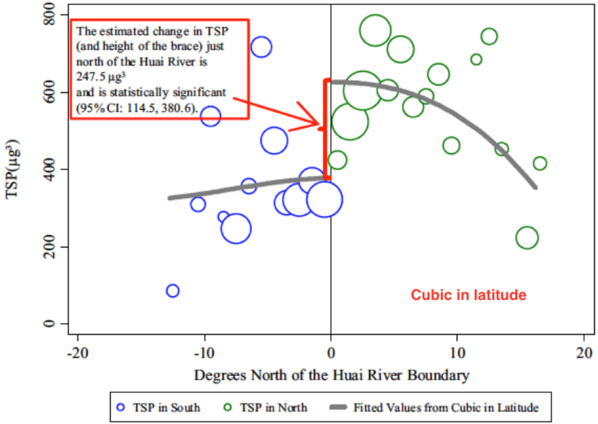


Fig. 2. Each observation (circle) is generated by averaging TSPs across the Disease Surveillance Point locations within a 1° latitude range, weighted by the population at each location. The size of the circle is in proportion to the total population at DSP locations within the 1° latitude range. The plotted line reports the fitted values from a regression of TSPs on a cubic polynomial in latitude using

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

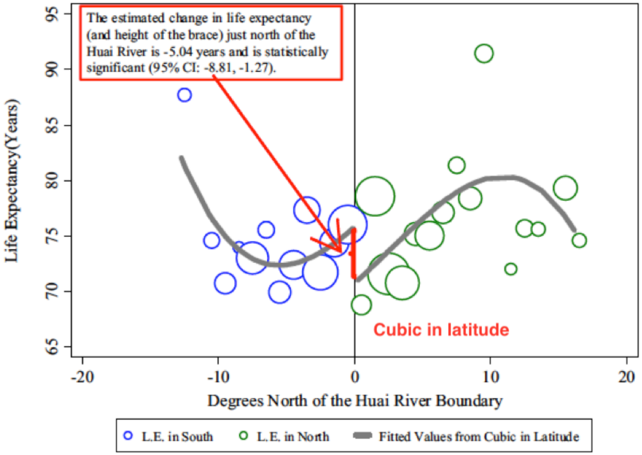


Fig. 3. The plotted line reports the fitted values from a regression of life expectancy on a cubic in latitude using the sample of DSP locations, weighted by the population at each location.

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 Surveillance Points (1991–2000), *China Environment Yearbook* (1981–2000), and World Meteorological Association (1980–2000).

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 Surveillance Points (1991–2000), *China Environment Yearbook* (1981–2000), and World Meteorological Association (1980–2000).

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

- Sharp RDD

$$Y_j = \delta_0 + \delta_1 N_j + \delta_2 f(L_j) + X_j'\phi + u_j$$

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)
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

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

- Fuzzy RDD
  - First Stage:

$$TSP_j = \alpha_0 + \alpha_1 N_j + \alpha_2 f(L_j) + X_j'\kappa + v_j$$

  - Second Stage:

$$Y_j = \beta_0 + \beta_1 \widehat{TSP}_j + \beta_2 f(L_j) + X_j'\gamma + \varepsilon_j$$

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

Dependent variable	(1)	(2)	(3)
Panel 1: Impact of “North” on the listed variable, ordinary least squares			
TSPs, 100 μg/m^3	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)
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

Air pollution in China: Ebenstein et al(2017)

- More accurate measures of pollution particles(PM_{10})
- More accurate measures of mortality from a more recent time period(2004-2012)
- More samples size(eight times than previous one)
- More subtle functional form: Local Linear Regression

Air pollution in China: Ebenstein et al(2017)

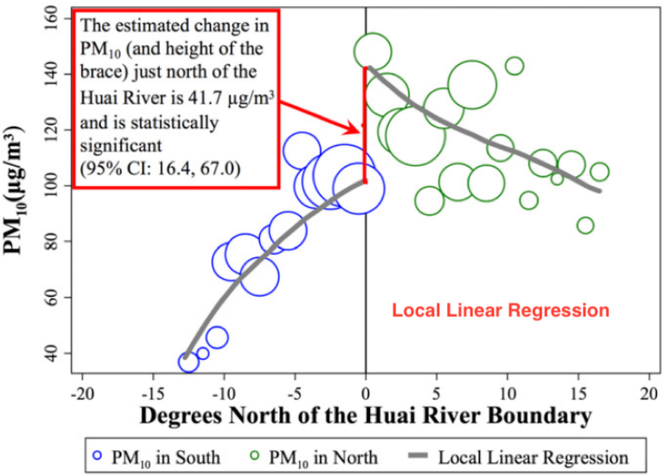


Fig. 2. Fitted values from a local linear regression of PM_{10} exposure on distance from the Huai River estimated separately on each side of the river.

Air pollution in China: Ebenstein et al(2017)

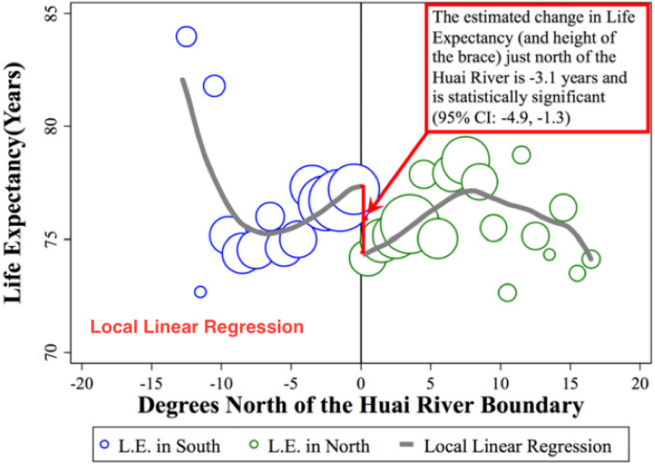


Fig. 3. Fitted values from a local linear regression of life expectancy (L.E.) on distance from the Huai River estimated in the same manner as in Fig. 2.

Air pollution in China: Ebenstein et al(2017)

- **Sharp RD**

$$Y_j = \delta_0 + \delta_1 N_j + f(L_j) + N_j f(L_j) + X_j' \phi + u_j$$

- **Fuzzy RD**
  - **First Stage**

$$PM_j^{10} = \alpha_0 + \alpha_1 N_j + f(L_j) + N_j f(L_j) + X_j' \gamma + u_j$$

  - **Second Stage**

$$Y_j = \beta_0 + \beta_1 \widehat{PM_j^{10}} + f(L_j) + N_j f(L_j) + X_j' \phi + \varepsilon_j$$

Air pollution in China: Ebenstein et al(2017)

Table 2. RD estimates of the impact of the Huai River Policy

Outcome	[1]	[2]	[3]
Pollution and life expectancy			
PM_{10}	27.4*** (9.5)	31.8*** (9.1)	41.7*** (12.9)
Life expectancy at birth, y	−2.4** (1.0)	−2.2* (1.1)	−3.1*** (0.9)
Cause-specific mortality (per 100,000, log)			
Cardiorespiratory	0.30** (0.14)	0.22* (0.13)	0.37*** (0.11)
Noncardiorespiratory	0.06 (0.10)	0.08 (0.09)	0.13 (0.08)
RD type	Polynomial	Polynomial	LLR
Polynomial function	Third	Linear	
Sample	All	5°	

Column [1] reports OLS estimates of the coefficient on a north of the Huai River dummy after controlling for a polynomial in distance from the Huai River interacted with a north dummy using the full sample (n = 154) and the control variables from *SI Appendix*, Table S1. Column [2] reports this estimate for the restricted sample (n = 79) of DSP locations within 5° of the Huai River. Column [3] presents estimates from local linear regression (LLR), with triangular kernel and bandwidth selected by the method proposed by Imbens and Kalyanaraman (14).

Air pollution in China: Ebenstein et al(2017)

Table 2. RD estimates of the impact of the Huai River Policy

Outcome	[1]	[2]	[3]
Pollution and life expectancy			
PM_{10}	27.4*** (9.5)	31.8*** (9.1)	41.7*** (12.9)
Life expectancy at birth, y	−2.4** (1.0)	−2.2* (1.1)	−3.1*** (0.9)
Cause-specific mortality (per 100,000, log)			
Cardiorespiratory	0.30** (0.14)	0.22* (0.13)	0.37*** (0.11)
Noncardiorespiratory	0.06 (0.10)	0.08 (0.09)	0.13 (0.08)
RD type	Polynomial	Polynomial	LLR
Polynomial function	Third	Linear	
Sample	All	5°	

Column [1] reports OLS estimates of the coefficient on a north of the Huai River dummy after controlling for a polynomial in distance from the Huai River interacted with a north dummy using the full sample (n = 154) and the control variables from *SI Appendix*, Table S1. Column [2] reports this estimate for the restricted sample (n = 79) of DSP locations within 5° of the Huai River. Column [3] presents estimates from local linear regression (LLR), with triangular kernel and bandwidth selected by the method proposed by Imbens and Kalyanaraman (14).

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

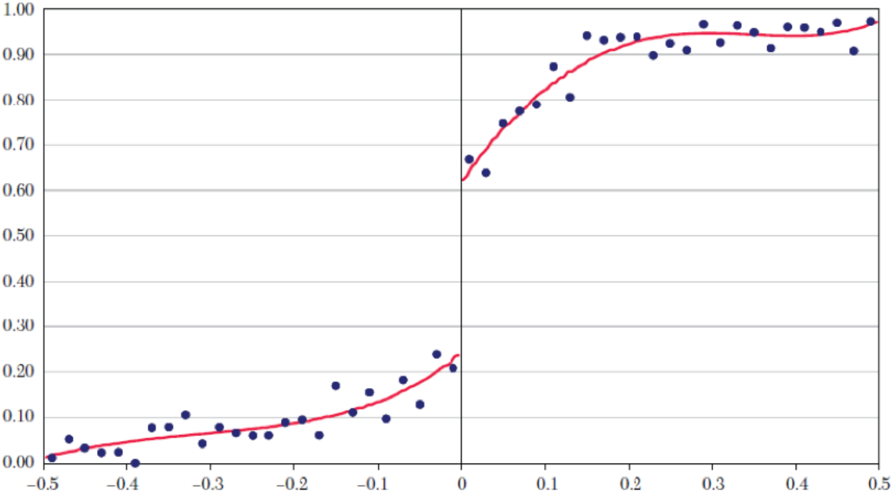
$$[0, 0.5), [0.5, 1), ..., [9.5, 10]$$

- Second,calculate average y in each bin,and plot this above midpoint for each bin.
- Third, plot the forcing 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 unexpected discontinuities.

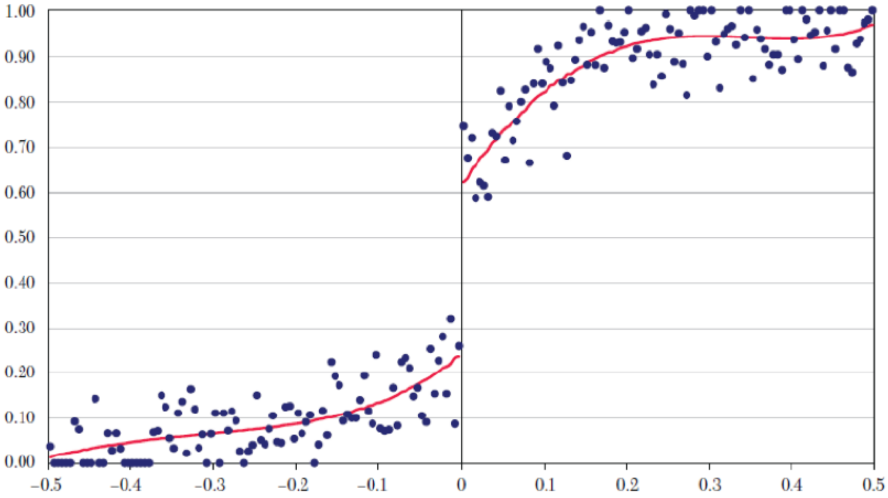
RDD graphical analysis: Select Bin Width

- 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

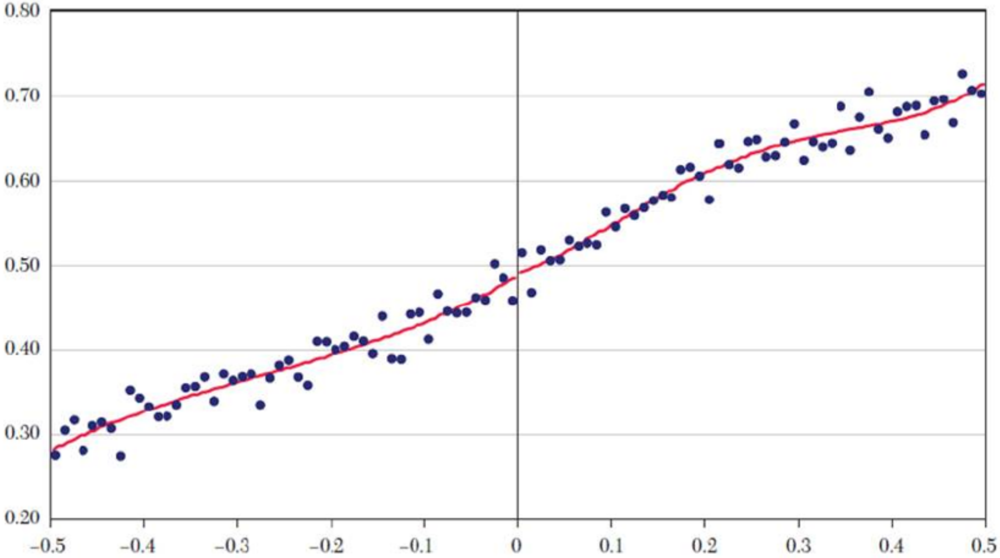
- It is probably advisable to report results for both estimation types:

1. Polynomials in X.
   - In robustness checks you also want to show that including higher order polynomials does not substantially affect your findings.
   - But quadratic(at most Cubic) is enough,higher-order polynomial may hurm and should not be use.(Gelman and Imbens,2019)
2. Local linear regression or other nonparametric estimation
   - 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

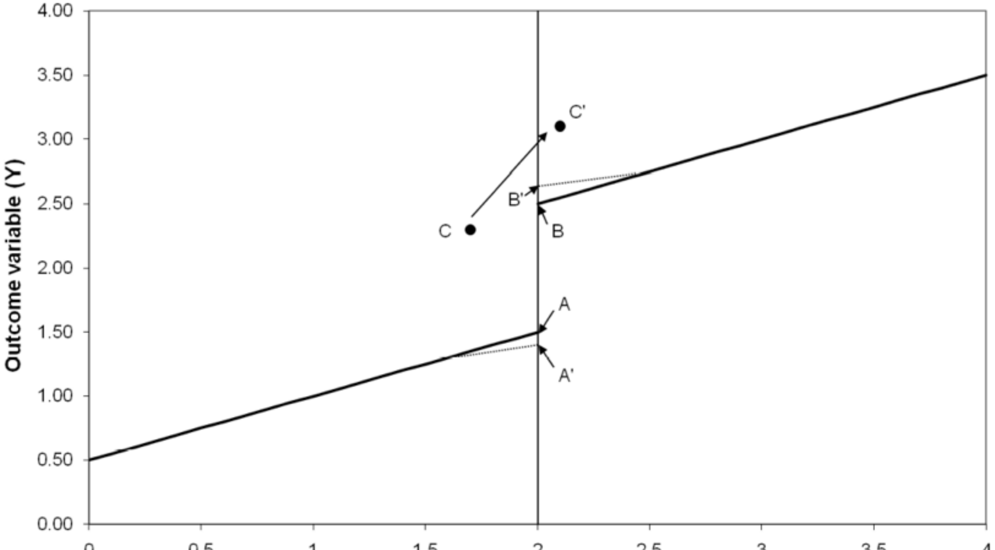


Testing the Validity of the RDD

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

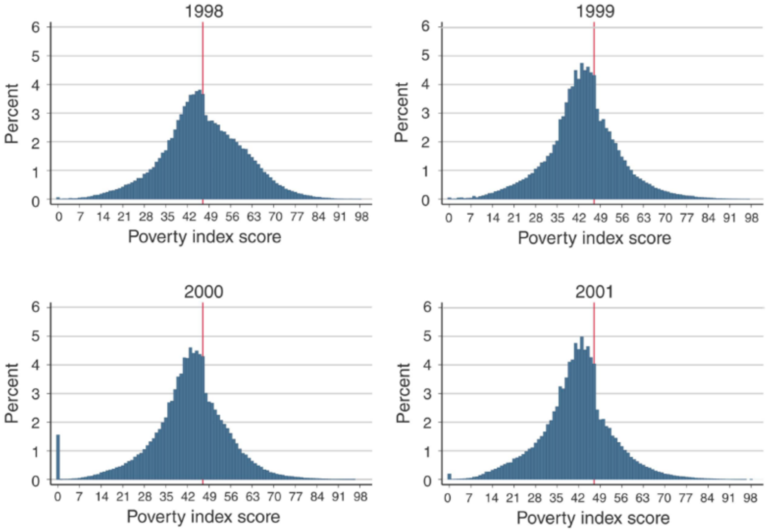


Manipulation of a poverty index in Colombia

- 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.

Manipulation of a poverty index in Colombia

Manipulation of a poverty index in Colombia

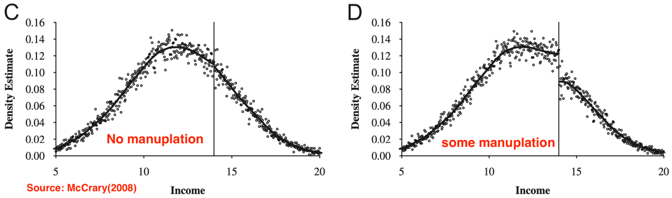


Testing the Validity of the RDD

- Testing for discontinuities in the density of the assignment variable X:
  - Create a histogram showing the number of observations in each bin of the assignment variable
  - Examine whether there is a discontinuity in the distribution of the assignment variable at the threshold
  - A discontinuity in the density indicates potential manipulation of the assignment variable around the threshold

McCrary(2008) Test

- Also a more formal test which is used to check whether units are sorting on the running variable.



Testing the Validity of the RDD

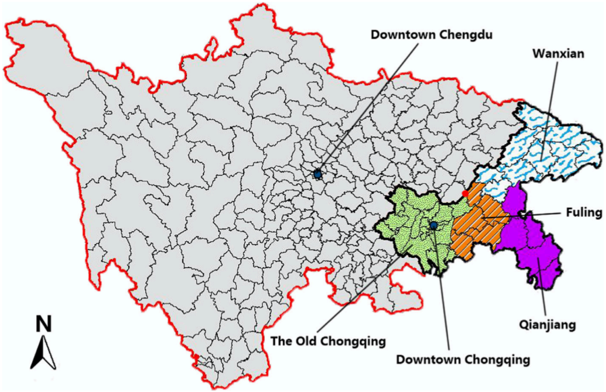
- **Falsification Tests**: testing for jumps at non-discontinuity points
  - If threshold x only existed in certain c or for certain types of observations
  - Make sure no effect in c where there was no discontinuity or for agents where there isn't supposed to be an effect.

Case Study: Political hierarchy and regional economic development

Jia,Liang and Ma(2021)

- *"Political hierarchy and regional economic development: Evidence from a spatial discontinuity in China",Journal of Public Economics Volume 194, February 2021.*
- **Topic**: *Political hierarchy and Regional economic development*
- **Background**: In 1997, the prefecture-level Chongqing city was elevated to a province-level municipality, splitting off from Sichuan province.
  - It consequently gained a substantial increase in decision-making power for administrative, personnel, and fiscal affairs.
- **Question**: Does the promotion of Chongqing to a province-level municipality lead to higher economic growth? if so how much?
- **Empirical Challenge**:
  - OLS and Matching?
  - Panel Data and DID?
  - Geographic RD design by authors.

Chongqing v.s Sichuan



- *Chongqing* contains 30 million people in 43 counties and 933 towns.
- The remaining *Sichuan* contains 85 million people in 20 prefectures, 180 counties,and 4155 towns.

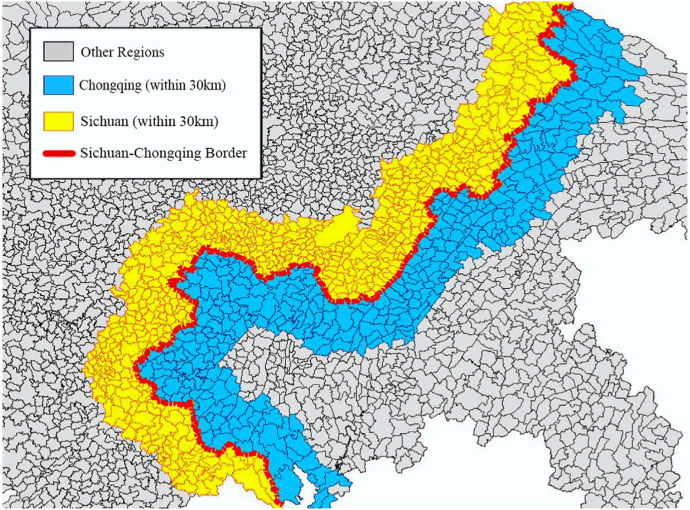
Empirical Strategy

- SRD regression equation is:

$$Y_i = \beta_0 + \beta_1 Chongqing_i + f(L) + \varepsilon_i$$

  - Y_i as outcome variable of interest in town i, thus the economic growth rate.
  - *Chongqing* is a binary indicator.
  - $f(L)$ control for a two-dimensional polynomial in a town centroid's longitude and latitude.
  - β_1 is the coefficient of interest, which captures the Chongqing promotion treatment effect on economic growth.

Geographic RD design: Boundary as Discontinuity



Pretreatment Balance

Table 2
Balance test.

	Chongqing (1)	Sichuan (2)	Mean Difference (s.e.) (3)
	Mean values		
Panel A. town-level variables within 30 km bandwidth			
Light intensity in 1996	0.704	0.764	−0.060 (0.218)
Elevation (meter)	505.530	458.497	47.033 (80.209)
Slope (%)	9.205	8.310	0.894 (2.001)
Distance to Chongqing Downtown (km)	126.794	142.989	−16.196 (25.513)
Distance to Chengdu Downtown (km)	285.987	254.320	31.667 (31.663)
Ethnic minority population share	0.003	0.012	−0.009 (0.011)
Observations	279	467	
Panel B. county-level variables for full sample			
Per capita GDP in 1996 (yuan, in logarithm)	8.098	8.012	0.086 (0.100)
Per capita industrial output in 1996 (yuan, in logarithm)	7.580	7.346	0.235 (0.201)
Per capita fiscal revenue in 1996 (yuan, in logarithm)	4.722	4.713	0.009 (0.099)
Urbanization rate in 1996 (%)	78.495	81.663	−3.168 (3.006)
Observations	43	178	

Notes: *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively. The county-level clustered standard errors are reported in parentheses

Pretreatment Discontinuity

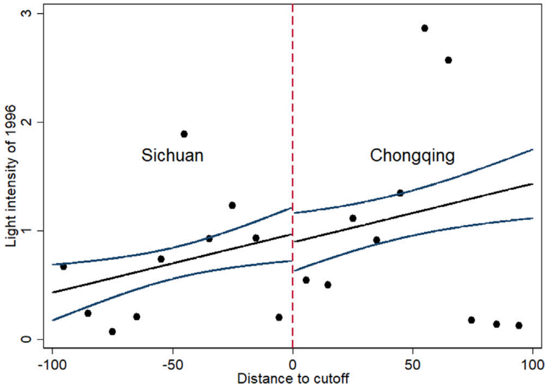


Fig. 3. Balance of initial development level across the border. *Notes:* This figure shows the single dimension RD graphs. The x-axis denotes the distance from a town centroid to the Chongqing–Sichuan border, where negative numbers refer to the control group (Sichuan). The dark dots show growth rates averaged over 10 km wide bins in distance from the border. The black lines fit local linear regressions within 100 km bandwidth on both sides of the boundary and the blue lines denote 95 percent confidence interval.

Baseline Discontinuity

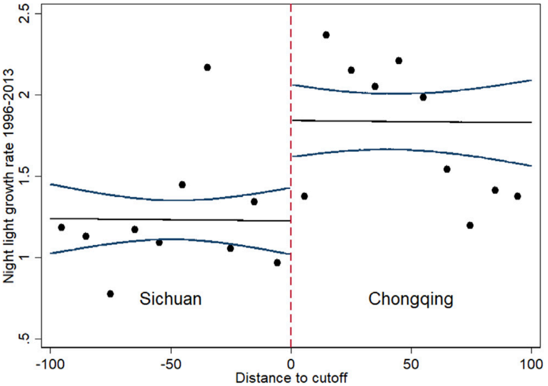


Fig. 5. Discontinuity in growth rate of light intensity from 1996 to 2013. *Notes:* The x-axis denotes the distance from a town centroid to the Chongqing–Sichuan border, where negative numbers refer to the control group (Sichuan). The dark dots show growth rates averaged over 10 km wide bins in distance from the border. The black lines fit local linear regressions within 100 km bandwidth on both sides of the boundary, and the blue lines denote 95 percent confidence interval.

Baseline results

Table 3
Baseline RD results.

Sample within	Dependent variable: light intensity growth from 1996-2013					
	Local linear approach		Local quadratic approach		Global polynomial approach	
	<30 km	<50 km	<30 km	<50 km	Full Sample	Full Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Chongqing	1.038***	1.170***	1.028***	1.199***	1.036***	1.022***
	(0.291)	(0.295)	(0.287)	(0.308)	(0.234)	(0.239)
Polynomial	Linear	Linear	Quadratic	Quadratic	Cubic	Quartic
Observations	746	1,188	746	1,188	5,088	5,088
R-squared	0.104	0.087	0.117	0.094	0.034	0.033

Notes: The dependent variable is $ln(0.01 + LightIntensity_{i,2013}) - ln(0.01 + LightIntensity_{i,1996})$. All regressions include two-dimensional geographic controls. The county-level clustered standard errors are reported in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Parallel Trends and Dynamic effects

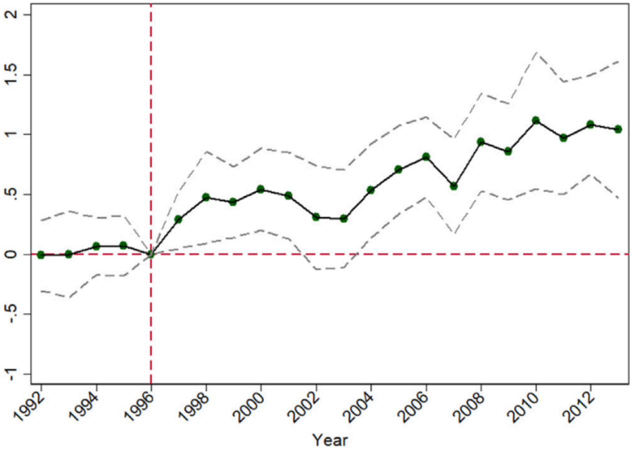


Fig. 4. Dynamics of the effects on light intensity growth. *Notes*: Point estimates are reported under alternative time windows. The basic line is for 1996. The solid line plots the point estimate of a separate estimation of β_1 in Eq. (1) and the dash lines denote 95 percent confidence interval.

Robustness: alternative bandwidths and specifications

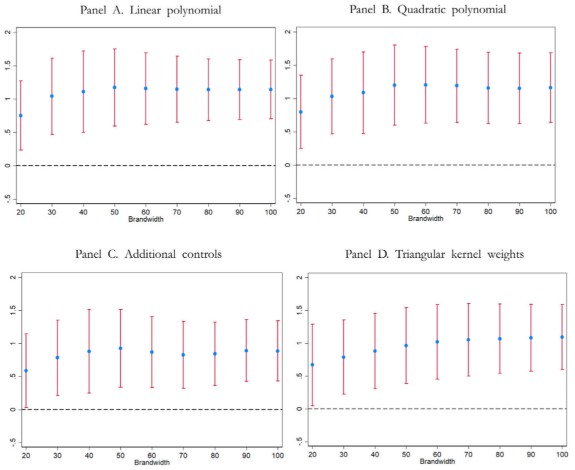


Fig. 6. Robustness to alternative bandwidths and model specifications. *Notes*: Each point plots the point estimate of a separate estimation of β_1 in Eq. (1) along with the 95 percent confidence interval, ranging from 20-km to 100-km bandwidths. Panel A plots estimates using linear polynomials in latitude and longitude. Panel B plots estimates from equivalent regressions but using second-order polynomials in latitude and longitude. Panel C plots estimates additionally controlling for elevation, slope, and a series of segment dummies. Panel D shows results using triangular kernel weights to give higher weight to observations that are closer to the boundary.

Placebo tests

Table 4
Placebo tests.

	Dependent variable: light intensity growth from 1996-2013	
Sample within	Move the true boundary 30 kilometers westward (1)	Move the true boundary 30 kilometers eastward (2)
East of the falsified border	−0.086 (0.249)	0.256 (0.404)
Observations	881	517
R-squared	0.069	0.068

Notes: The dependent variable is $ln(0.01 + LightIntensity_{i,2013}) - ln(0.01 + LightIntensity_{i,1996})$. We set a 30 km bandwidth. All regressions include two-dimensional geographic controls. The county-level clustered standard errors are reported in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Displacement effects

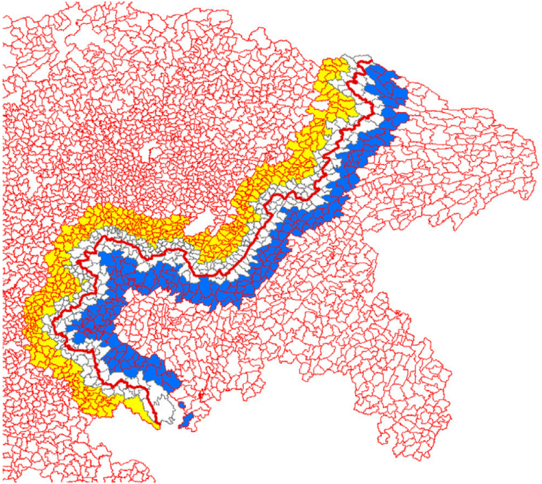


Fig. 8. Spatial exclusion approach. *Notes:* The red line marks the border between Sichuan and Chongqing. The blue and yellow shaded areas are towns in our boundary sample that belong to treated and non-treated areas, respectively. We

Displacement effects

Table 5
Test on displacement effects.

Dependent variable: light intensity growth from 1996-2013						
Sample within	<30 km			<50 km		
	Baseline	Exclude towns within 2*10 km across boundary	Exclude towns within 2*5km across boundary	Baseline	Exclude towns within 2*10 km across boundary	Exclude towns within 2*5km across boundary
	(1)	(2)	(3)	(4)	(5)	(6)
Chongqing	1.038***	1.418***	1.199***	1.170***	1.506***	1.315***
	(0.291)	(0.342)	(0.275)	(0.295)	(0.314)	(0.276)
Observations	746	476	622	1,188	918	1,064
R-squared	0.104	0.126	0.113	0.087	0.095	0.091
Test on equality with the baseline estimate		p = 0.2731	p = 0.5609		p = 0.2886	p = 0.6012

Notes: The dependent variable is $ln(0.01 + LightIntensity_{i,2013}) - ln(0.01 + LightIntensity_{i,1996})$. All regressions include two-dimensional local linear geographic controls. The last row reports the p-value of the Wald test on equality with the baseline estimate. The county-level clustered standard errors are reported in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Summary

RDD in the toolkit of Causal Inference

- RDD is considered the **closest** methodological approach to RCTs for identifying causal treatment effects.
- RDD requires an arbitrary threshold where agents can **partially** influence treatment assignment.

- Two main variants:
  - Sharp RD
  - Fuzzy RD
- Key assumption: Continuity at the threshold

- Methodological considerations:
  - Functional form specification
  - Bandwidth selection
  - Binning strategy
- Practical challenges:
  - Data requirements
  - Computational complexity