ECON 21110
Applied Microeconometrics
Winter 2022
Lecture 5
Regression Discontinuity (RD)

Eyðfríð Juanna Schrøter Joensen

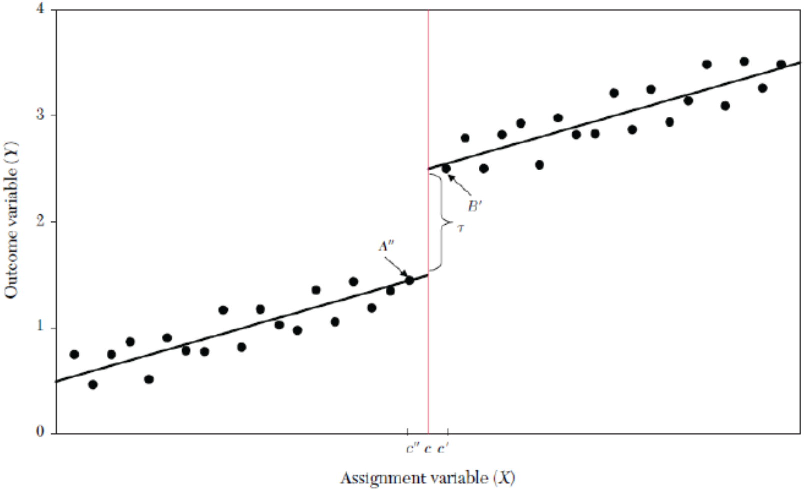
University of Chicago

February 10, 2022

Regression Discontinuity (RD)

- Exploit precise knowledge of the rules determining treatment
- RD design is based on the idea that in a rule-based world, some rules are arbitrary and therefore provide good quasi-experiments
- Two types of RD designs:
  - **Sharp RD**: Treatment is a deterministic function of a covariate X. A *selection-on-observables* story
  - **Fuzzy RD**: Exploits discontinuities in the probability of treatment. An instrumental variables (IV) type of setup
- RD captures a causal effect by distinguishing the treatment, $D_i = 1[X_i \geq X_0]$ a discontinuous function of X_i from a smooth and flexible function $f(X_i)$ that controls for the counterfactual
- Main **identifying assumption**:
  in a sufficiently near neighborhood around the discontinuity X_0, treatment is as good as randomly assigned

Example: Linear RD



Regression Discontinuity (RD)

- An arbitrary jump $\rightarrow$ usually thanks to a quirk in nature or a law
- We are interested in the ones that make very similar people (or firms or i's more generally) get very dissimilar treatment

Regression Discontinuity (RD)
Applications:

- Exam and Financial Aid thresholds:
  - Basically the “top” test-takers get a scholarship. A small difference in test score, X, can mean a discontinuous jump in financial aid (or scholarship) amount
    (Van Der Klaauw, 2002)
- School Class Size:
  - Maimonides Rule $\rightarrow$ No more than 40 kids in a class in Israel. 40 kids in school means 40 kids per class. 41 kids means two classes with 20 and 21
    (Angrist & Lavy, 1999)

Regression Discontinuity (RD)
Applications:

- Union Elections
  - If employers want to unionize, hold election. 50% means the employer does not have to recognize the union, and $(50+1)\%$ means the employer is required to bargain in good faith with the union (DiNardo & Lee, 2004)
- Air Pollution and Home Values
  - The Clean Air Act's National Ambient Air Quality Standards say if the geometric mean concentration of 5 pollutant particulates is 75 micrograms per cubic meter or greater, a county is classified as non-attainment and subject to much more stringent regulation (Chay & Greenstone, 2005)

Combine the “R” and the “D”

- Run a regression based on a situation where you have a discontinuity
- Treat above-the-cutoff and below-the-cutoff like the treatment and control groups from an RCT

Sharp RD design

- Treatment is assumed to be *deterministic* and *discontinuous* function, $f(X_i)$, of a covariate X_i
- For example, a unit will be treated if X_i is above a specific *cutoff/threshold/discontinuity* value, otherwise not:

$$D_i = \begin{cases} 1 & \text{if } X_i \geq X_0 \\ 0 & \text{if } X_i < X_0 \end{cases}$$

- Based on *selection on observables* assumptions:
  - Deterministic: once we know X_i we know D_i
  - Discontinuous: at a certain value, X_0, the probability of treatment D_i jumps discontinuously
- As highlighted by Imbens & Lemieux (2008) there is no value of X_i at which you observe both treatment and control observations $\rightarrow$ the method relies on extrapolation across covariate values
- For this reason we cannot be agnostic about regression functional form $f(.)$ in RD

Sharp RD design
Linear

- Suppose that in addition to the assignment mechanism above, potential outcomes can be described by a linear, constant effects model:

$$\mathbb{E}\left[Y_{0i} \mid X_i\right] = \alpha + \beta X_i$$
$$Y_{1i} = Y_{0i} + \rho$$

- This leads to the regression:

$$Y_i = \alpha + \beta X_i + \rho D_i + \varepsilon_i$$

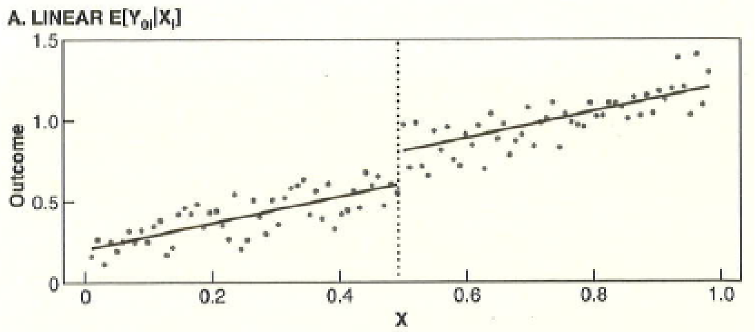
- D_i is now not only correlated with X_i, but it is a deterministic function of X_i
- Estimate causal effects by distinguishing the nonlinear function $D_i = 1$ if $X_i \geq X_0$ from the linear function $f(X_i) = \beta X_i$

Sharp RD design
Key Identifying Assumption

- Key identifying assumption:
  $\mathbb{E}[Y_{0i} \mid X_i]$ and $\mathbb{E}[Y_{1i} \mid X_i]$ are continuous in X_i at X_0
- This means that all other unobserved determinants of Y are continuously related to the running variable X
- This allows us to use average outcomes of units just below the cutoff as a valid counterfactual for units right above the cutoff
- This assumption cannot be directly tested
  - some tests give suggestive evidence whether the assumption is satisfied

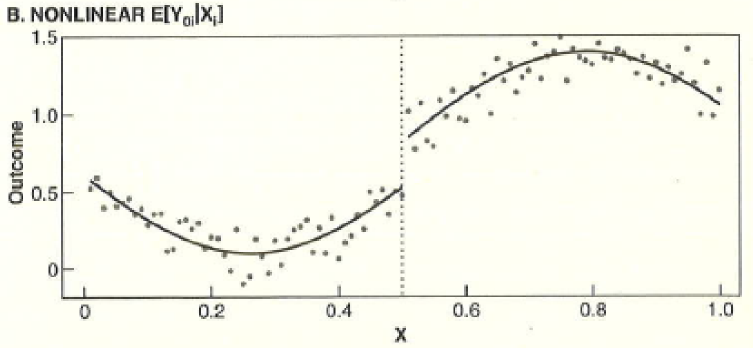
Sharp RD design
Linear

Restrictive assumption: linear function $f(X_i) = \beta X_i$



Sharp RD design
Non-linear

More realistic assumption: $f(X_i)$ is not linear over the entire range of X_i



Sharp RD design
Non-linear

- Suppose the nonlinear relationship is $\mathbb{E}[Y_{0i} \mid X_i] = f(X_i)$ for some reasonably smooth function $f(X_i)$
- In this case we can construct RD estimates by fitting:

$$Y_i = \alpha + f(X_i) + \rho D_i + \varepsilon_i$$

- There are two ways of approximating $f(X_i)$:
  (1) Use a nonparametric kernel method (more on this later)
  (2) Use a k^{th} order polynomial: i.e. estimate

$$Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + ... + \beta_k X_i^k + \rho D_i + \varepsilon_i$$

- Method (1) is more or less becoming the standard but method (2) should lead to very similar results

Sharp RD design
Non-linear, different polynomials on the two sides

- We can generalize the function $f(X_i)$ by allowing the X_i terms to differ on both sides of the threshold by including them both individually and interacting them with D_i
- Allowing different functions on both sides of the discontinuity should be the main results in a RD paper, since otherwise we use values from both sides of the cutoff to estimate the function on each side (Lee & Lemieux, 2010)
- In this case we have:

$$\mathbb{E}\left[Y_{0i} \mid X_i\right] = \alpha + \beta_{01}\widetilde{X}_i + \beta_{02}\widetilde{X}_i^2 + \ldots + \beta_{0k}\widetilde{X}_i^k$$
$$\mathbb{E}\left[Y_{1i} \mid X_i\right] = \alpha + \rho + \beta_{11}\widetilde{X}_i + \beta_{12}\widetilde{X}_i^2 + \ldots + \beta_{1k}\widetilde{X}_i^k$$

  where $\widetilde{X}_i = X_i - X_0$
- Centering at X_0 ensures that the treatment effect at $X_i = X_0$ is the coefficient on D_i in a regression model with interaction terms

Sharp RD design
Non-linear, different polynomials on the two sides

- To derive a regression model that can be used to estimate the causal effect we use the fact that D_i is a deterministic function of X_i:

$$\mathbb{E}[Y_i \mid X_i] = \mathbb{E}[Y_{0i} \mid X_i] + (\mathbb{E}[Y_{1i} \mid X_i] - \mathbb{E}[Y_{0i} \mid X_i])\, D_i$$

- The regression model which you estimate is then:

$$\begin{aligned} Y_i = \alpha &+ \beta_1 \widetilde{X}_i + \beta_2 \widetilde{X}_i^2 + \ldots + \beta_k \widetilde{X}_i^k \\ &+ \beta_1^* \widetilde{X}_i D_i + \beta_2^* \widetilde{X}_i^2 D_i + \ldots + \beta_k^* \widetilde{X}_i^k D_i + \rho D_i + \varepsilon_i \end{aligned}$$

where $\beta_j^* = \beta_{1j} - \beta_{0j}$ for all $j = 1, 2, \ldots, k$

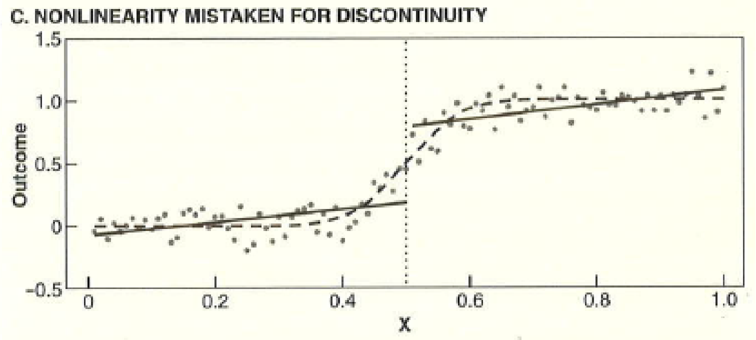
- The treatment effect at X_0 is ρ
- The treatment effect at $X_i - X_0 = c > 0$ is: $\rho + \beta_1^* c + \beta_2^* c^2 + \ldots + \beta_k^* c^k$

Sharp RD design

- Test of identifying assumption?
  - Check what happens with the estimates when the data window gets smaller
  - When a smaller window is chosen estimates get less precise, but we can also remove some of the polynomials to model $f(X_i)$
  - The estimated effect should remain relatively stable as the window (bandwidth) gets smaller

Internal Validity of RD estimates

- The validity of RD estimates depends crucially on the assumption that the polynomials provide and adequate representation of $\mathbb{E}[Y_{0i} \mid X_i]$
- If not what looks like a jump may simply be a non-linearity in $f(X_i)$ that the polynomials have not accounted for



Why not just control for *everything*?

- Can't we just control for everything and use OLS?
  $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$
- Consider the class size example:
  - Get a dataset with different class sizes and the student's test scores
  - Class size is X, test score is Y, run a regression of Y on X
  - Larger class size leads to lower test scores assumption: the error term (containing everything we can't measure, the unobservables) supposed to be uncorrelated with the X's
  - Class size may be correlated with, for example, neighborhood quality
  - Crowded classrooms might be in poorer schools or special needs students might be in small classes

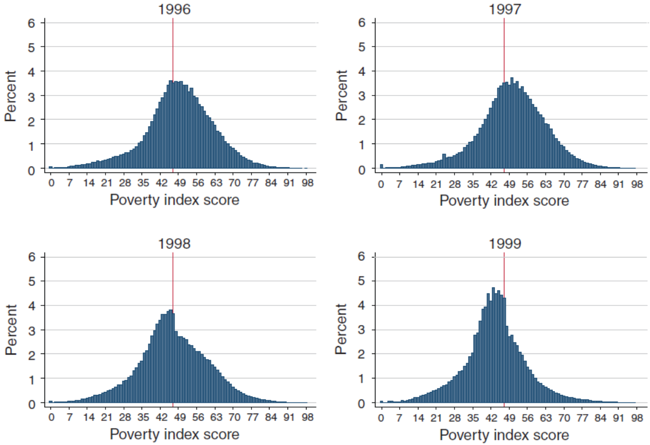
Why not just control for *everything*?

- We can try and control for neighborhood poverty level
  - Does it solve the problem?
  - You will probably not be able to "magically" control for every single variable correlated with the X of interest
  - Find a bandwidth for which these variables are uncorrelated
- Class size is *not* random
  - BUT is a school with 40 kids that different from a school with 41?
  - Right around the cut-off, there is a good chance of as good as random assignment

Random Assignment around the Discontinuity?

- Can test for whether this is the case!
  - Look at the averages of the observables in your below-cutoff group, and the averages of the observables in the above-cutoff group. Are they the same? Hopefully!?
  - Do people know about this cutoff? Are they sorting endogenously? When deciding where to live, did "good" parents look for schools where their kids would be the 41st kid?
  - In addition to checking the observables on either side of the cutoff, we should check the density of the distribution. Is it unusually low/high right around the cutoff?
  - If there is some abnormally large portion of people right around the cutoff, it is quite possible that you do *not* have random assignment

Sorting on the Threshold



Conover & Camacho (2011)
"Manipulation of Social Program Eligibility"

Fuzzy RD design

- Fuzzy RD exploits discontinuities in the probability of treatment conditional on a covariate
- The discontinuity becomes an instrumental variable (IV) for treatment status
- Units with values above a certain threshold value of the underlying variable are *more likely* to be treated than those below
- There is a jump in the probability of treatment at the cut-off value X_0
  $$P(D_i = 1 \mid X_i) = \begin{cases} g_1(X_i) & \text{if } X_i \geq X_0 \\ g_0(X_i) & \text{if } X_i < X_0 \end{cases}$$
  where $g_1(X_0) \neq g_0(X_0)$
- The more they differ the better because then there is more discontinuity to exploit

Fuzzy RD design

- Suppose now that $g_1(X_i) > g_0(X_i)$; i.e. the probability of treatment is higher when $X_i \geq X_0$

$$\mathbb{E}\left[D_i \mid X_i\right] = P(D_i = 1 \mid X_i) = g_0(X_i) + \left[g_1(X_i) - g_0(X_i)\right] T_i$$

where $T_i = 1\left[X_i \geq X_0\right]$

- T_i indicates the point where $\mathbb{E}\left[D_i \mid X_i\right]$ is discontinuous

$$\mathbb{E}\left[D_i \mid X_i\right] = \gamma_0 + \pi T_i + \sum_{j=1}^{k} \gamma_{0j} X_i^j + \sum_{j=1}^{k} \gamma_{0j} X_i^j T_i$$

- As with sharp RD, a good idea to specify a flexible functional form for $f(X_i, T_i)$
- We can use T_i (and all the interactions with T_i) as instrumental variable(s) for D_i

Fuzzy RD design

- The simplest fuzzy RD estimator uses only T_i as an instrument, without the interaction terms
- First-stage:

$$D_i = \gamma_0 + \pi T_i + \sum_{j=1}^{k} \gamma_{0j} X_i^j + \eta_i$$

where π is the first stage effect of T_i (“relevance”)

- Second-stage:

$$Y_i = \alpha_0 + \rho D_i + \sum_{j=1}^{k} \beta_{0j} X_i^j + \varepsilon_i$$

- We estimate this using 2SLS

Fuzzy RD design

- The second-stage model with interaction terms would be:

$$Y_i = \alpha_0 + \rho D_i + \sum_{j=1}^{k} \beta_{0j} \widetilde{X}_i^j + \sum_{j=1}^{k} \beta_{0j}^* \widetilde{X}_i^j D_i + \varepsilon_i$$

  where where $\widetilde{X}$ are not only normalized with respect to X_0, but are also fitted values obtained from the 1st stage regression

- One can therefore use both T_i as well as the interaction terms as instruments for D_i

What does Fuzzy RD Estimate?

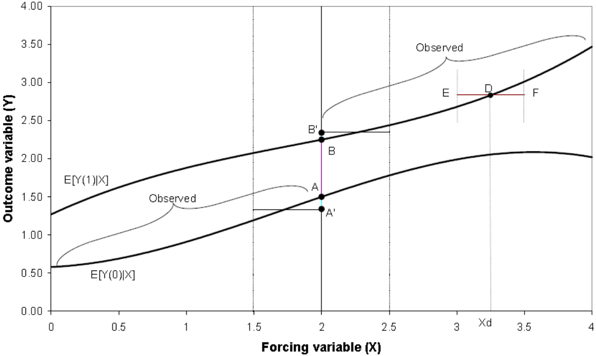
- Needs the same assumptions as in the standard IV framework (Hahn, Todd, and van der Klaauw, 2001)
- As with other binary IVs one then estimates LATE: the average treatment effect of the compliers
- In RD the compliers are those whose treatment status changes as we move the value of X_i from just the left of X_0 to just to the right of X_0

Practical Tips for RDD: Estimating f

- As pointed out before there are essentially two ways of approximating the function, $f(X)$
  1. Kernel regression; i.e. bandwidth and weighting (Imbens & Lemieux, 2008)
  2. Using a polynomial function (Lee, 2008; Lee & Lemieux, 2010)
- There is no right or wrong method, both have advantages and disadvantages

Practical Tips for RDD: The Kernel Method

- Challenge: trying to estimate regressions at the cutoff
- This results in a "boundary problem"



while the "true" effect is AB, with a certain bandwidth a rectangular kernel would estimate the effect as A'B'

- This implies a systematic bias with the kernel method if $f(X)$ is upwards or downwards sloping

Practical Tips for RDD: The Kernel Method
Local Linear Regression

- The standard solution to this problem is to run local linear regression
- In the case drawn above this would substantially reduce the bias
- You simply estimate the following model using local linear regression:

$$Y_i = \alpha + \rho D_i + \beta_1 \widetilde{X}_i + \beta_2^* \widetilde{X}_i D_i + \eta_i$$

  where $\widetilde{X}_i = X_i - X_0$ and $-h < \widetilde{X}_i < h$

- Before we fixed h as the full bandwidth (the entire sample) while compensating this with a higher-order polynomial k
- Here we set the polynomial $k = 1$ and vary the bandwidth

Practical Tips for RDD: The Kernel Method
Local Linear Regression

- But which bandwidth should we choose?
- Trade-off between bias and efficiency
  - Larger h, more data, more precision, but more bias
- Various methods:
  - Cross-validation (Imbens & Lemieux, 2008)
  - Optimal Bandwidth (Imbens & Kalyanaraman, 2008)
    - ★ `rdob.ado` in Stata (download from Guido Imbens' website)
  - Robust Data-Driven Inference (Calonico, Cattaneo & Titiunik, 2013)
    - ★ `rdrobust.ado` in Stata

Practical Tips for RDD: The Polynomial Method

- Alternatively you can estimate the $f(X)$ function including polynomials in X (as we have done!)
- The polynomial method suffers from the problem that you are using data that is far away from the cutoff to estimate the $f(X)$ function
- The equivalent of choosing the right bandwidth for the kernel method is to use the right order of polynomial
- An easy way suggested by Lee & Lemieux (2010) to test whether you have the right polynomial is to estimate the polynomial function and include a full set of bin dummies in the regression
  - Then test the null hypothesis whether all bin dummies are 0
  - Add polynomial terms until you can no longer reject that null hypothesis

Practical Tips for RDD: Estimation

- It is advisable to report results for both estimation types:
  1. Polynomials in X
  2. Local linear regression
- In robustness checks you also want to show that including higher order polynomials does not substantially affect your findings
- You also want to show that your results are not affected if you vary the window around the cutoff (standard errors may go up but hopefully the point estimate does not change)

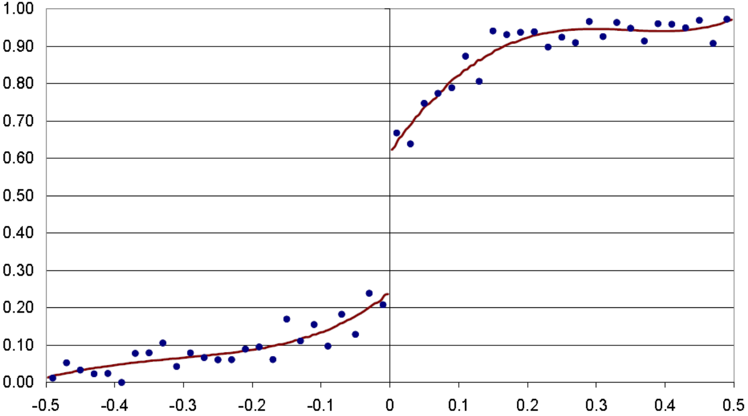
Graphical Analysis in RD Designs

A graphical analysis should be an integral part of any RD study.
You should show the following graphs:

- Outcome by forcing variable, X_i; i.e. the standard graph showing the discontinuity in the outcome variable:
  - Construct bins and average the outcome within bins on both sides of the cutoff
  - Look at different bin sizes when constructing these graphs
  - Plot the forcing variable X_i on the horizontal axis and the average of Y_i for each bin on the vertical axis
  - You may also want to plot a relatively flexible regression line on top of the bin means
  - Inspect whether there is a discontinuity at X_0
  - Inspect whether there are other obvious unexpected discontinuities

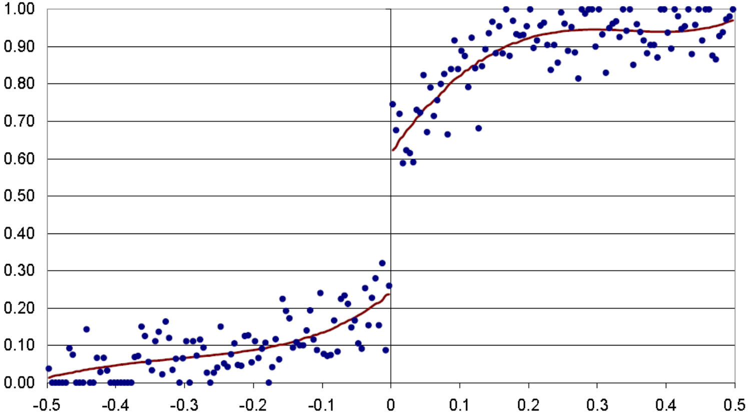
Graphical Analysis in RD Designs
Example: Outcomes by Forcing Variable

(Lee, 2008; Lee & Lemieux, 2010) winning election
bandwidth 0.02 (50 bins)



Graphical Analysis in RD Designs
Example: Outcomes by Forcing Variable - Smaller Bins

(Lee, 2008; Lee & Lemieux, 2010) winning election
bandwidth 0.005 (200 bins)

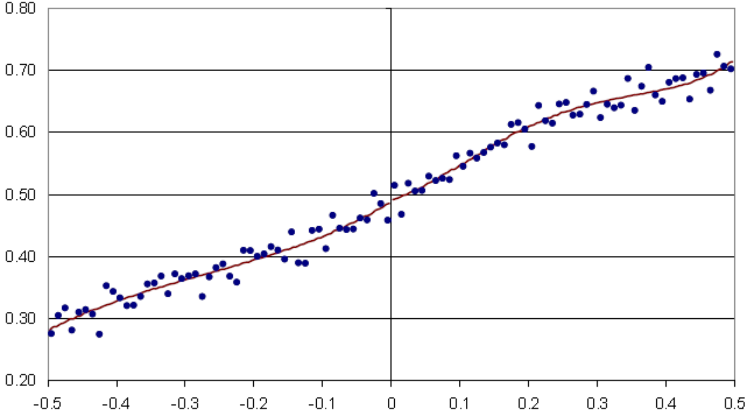


Graphical Analysis in RD Designs

- Probability of treatment by forcing variable if fuzzy RD
  - In a fuzzy RD design you also want to see that the treatment variable jumps at X_0
  - This tells you whether you have a first stage
- Covariates by forcing variable
  - 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 discontinuity one would doubt the identifying assumption

Graphical Analysis in RD Designs
Example: Covariates by Forcing Variable

(Lee, 2008; Lee & Lemieux, 2010) share of vote in prior election



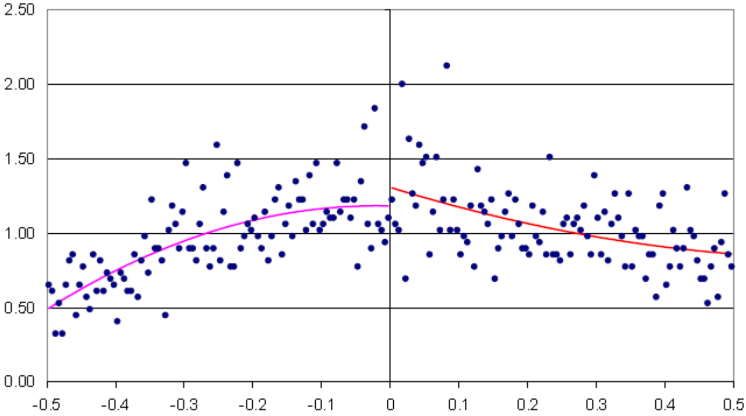
Graphical Analysis in RD Designs

The density of the forcing variable

- One should plot the number of observations in each bin
- This plot allows to investigate whether there is a discontinuity in the distribution of the forcing variable at the threshold
- This would suggest that people can manipulate the forcing variable around the threshold
- This is an indirect test of the identifying assumption that each individual has imprecise control over the assignment variable

Graphical Analysis in RD Designs
Example: Density of the Forcing Variable

(Lee, 2008; Lee & Lemieux, 2010) vote share in previous election



Testing the Validity of the RD Design

- As outlined above the key identifying assumption is that $\mathbb{E}[Y_{0i} \mid X_i]$ and $\mathbb{E}[Y_{1i} \mid X_i]$ are continuous in X_i at X_0
- This implies that each individual has imprecise control over the assignment variable
- It is impossible to test this directly but we can nonetheless get some evidence with the following specification tests

Testing the Validity of the RD Design

- Testing the continuity of the density of the forcing variable, X:
  - McCrary (2008) suggests testing the null hypothesis of continuity of the density of the forcing variable at the discontinuity point
  - In principle one does not need continuity. A discontinuity in the density, however, suggests that there is some manipulation of X around the threshold going on
  - In the first step you partition the assignment variable into bins and calculate frequencies (number of observations) in the bins
  - In the second step you treat those frequency counts as dependent variable in a local linear regression as before

Testing the Validity of the RD Design

- Test involving covariates:
  - Test whether other covariates exhibit a jump at the discontinuity; i.e. re-estimate the RD model with the covariate as the dependent variable
  - This is a type of placebo test
- Testing for jumps at non-discontinuity points:
  - Imbens & Lemieux (2008) suggest to only look at one side of the discontinuity and take the median of the forcing variable in that section and test whether you can find a discontinuity in that part
  - Another type of placebo test

Fuzzy RD Example: Angrist & Lavy (1999)
Maimonides' rule comes to school

- Estimate the effect of class size on test scores
- Extend RD in two ways compared to the discussion above:
  - The causal variable of interest (class size) takes on many values $\rightarrow$ the first stage exploits discontinuities in average class size instead of probabilities of a single treatment
  - They use multiple discontinuities
- Class size in Israeli schools is capped at 40
- Students in a grades with up to 40 students can expect to be in classes as large as 40
- Grades with 41 students are split into two classes (of 20 and 21)
- Grades with 81 students are split into three classes etc...
- **Maimonides' rule** $\rightarrow$ this class size rule was introduced by medieval Talmudic scholar Maimonides

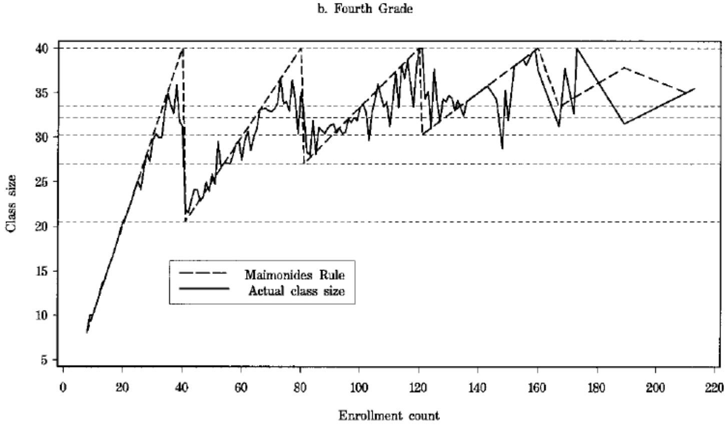
Fuzzy RD Example: Angrist & Lavy (1999)
Maimonides' rule comes to school

- Maimonides' rule:
$$m_{sc} = \frac{e_s}{int\left[\frac{e_s - 1}{40}\right] + 1}$$
where $int\,[X]$ is the integer part of X

- This function has a "sawtooth pattern" with discontinuities at integer multiples of 40

- At the same time, m_{sc}, is an increasing function of e_s making the enrollment variable an important variable to control for in the regression

Fuzzy RD Example: Angrist & Lavy (1999)
Maimonides' rule and Actual Class Size



Rule not followed strictly $\rightarrow$ fuzzy RD

Fuzzy RD Example: Angrist & Lavy (1999)
Maimonides' rule comes to school

- Use the Maimonides rule in a fuzzy RD design, where the outcome equation is:

$$Y_{isc} = \alpha_0 + \alpha_1 d_s + \rho n_{sc} + \beta_1 e_s + \beta_2 e_s^2 + \varepsilon_{isc}$$

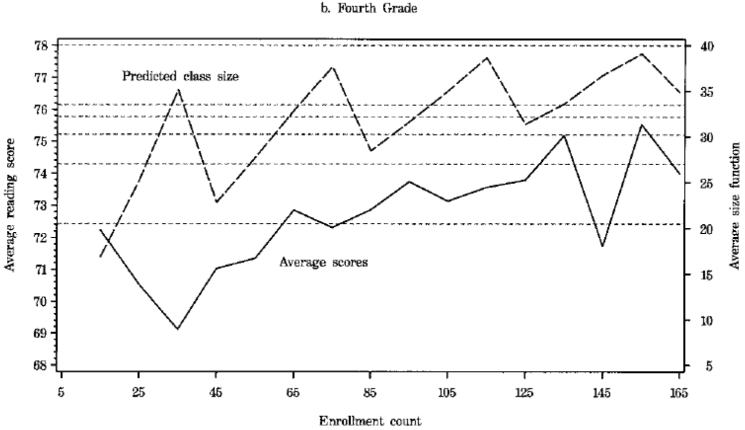
  where Y_{isc} is the test score of student i in school s and class c
  - e_s is enrollment in school s
  - d_s is the percentage of disadvantage students in school s
  - n_{sc} is class size
- The variables relate to the previous description as follows:
  - n_{sc} plays the role of D_i
  - e_s plays the role of X_i
  - m_{sc} plays the role of T_i
- The first stage is:

$$n_{sc} = \gamma_0 + \gamma_1 d_s + \pi m_{sc} + \delta_1 e_s + \delta_2 e_s^2 + \eta_{isc}$$

  where m_{sc} is the function describing Maimonides' rule

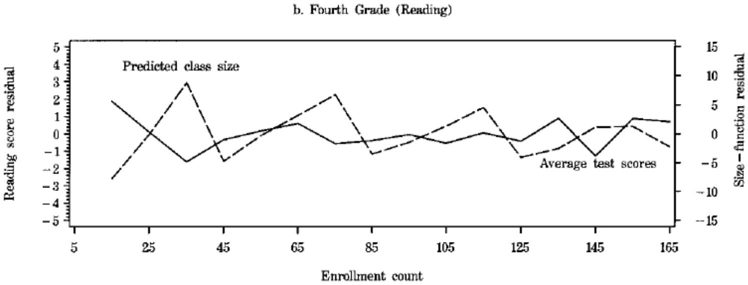
Fuzzy RD Example: Angrist & Lavy (1999)
Reduced Form: Test scores and Maimonides' rule

- Students in schools with more overall enrollment (often in bigger cities) do better on average
- Average test scores are partly a mirror image of predicted class sizes



Fuzzy RD Example: Angrist & Lavy (1999)
Reduced Form: Test scores and Maimonides' rule

- They control for enrolment when they re-draw the relationship between class-size and achievement, because larger schools are often in better-off areas
- Now test-scores are more of a mirror image to predicted class sizes



Fuzzy RD Example: Angrist & Lavy (1999)
Maimonides' rule comes to school

Table 6.2.1: OLS and fuzzy RD estimates of the effects of class size on fifth grade math scores

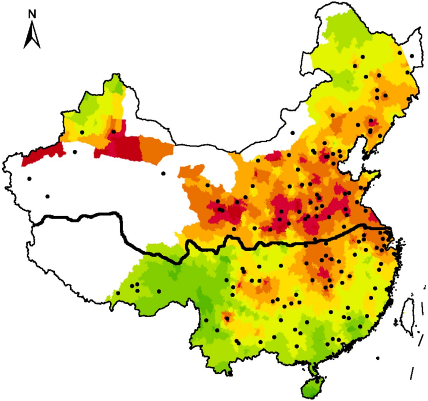
	OLS			2SLS				
				Full sample		Discontinuity samples		
						+/- 5		+/- 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean score		67.3		67.3		67.0		67.0
(s. d.)		(9.6)		(9.6)		(10.2)		(10.6)
Regressors								
Class size	.322	.076	.019	-.230	-.261	-.185	-.443	-.270
	(.039)	(.036)	(.044)	(.092)	(.113)	(.151)	(.236)	(.281)
Percent disadvantaged		-.340	-.332	-.350	-.350	-.459	-.435	
		(.018)	(.018)	(.019)	(.019)	(.049)	(.049)	
Enrollment			.017	.041	.062		.079	
			(.009)	(.012)	(.037		(.036)	
Enrollment squared/100					-.010			
					(.016)			
Segment 1								-12.6
(enrollment 36-45)								(3.80)
Segment 2								-2.89
(enrollment 76-85)								(2.41)
Root MSE	9.36	8.32	8.30	8.40	8.42	8.79	9.10	10.2
R-squared	.048	.249	.252					
N		2,018		2,018		471		302

Notes: Adapted from Angrist and Lavy (1999). The table reports estimates of equation (6.2.6) in the text using class averages. Standard errors, reported in parentheses, are corrected for within-school correlation.

Fuzzy RD: A "Modern" Example
Air Pollution and Life Expectancy

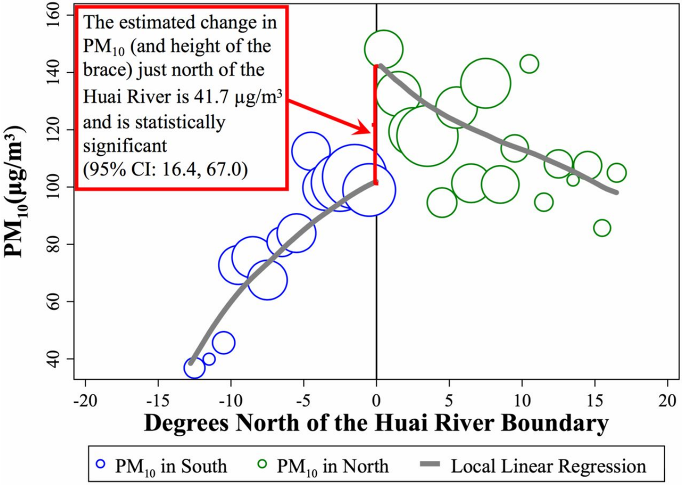
- Ebenstein, Fan, Greenstone, He & Zhou (2017)
  quasi-experimental variation in airborne particulate matter – particulate matter smaller than 10 μm (PM_{10})– generated by China's Huai River Policy
  - free or heavily subsidized coal for indoor heating during the winter to cities North ($T_i = 1$) of the Huai River, but not to those to the South ($T_i = 0$)
- Regression Discontinuity Design (RDD) based on distance ($\widetilde{X} = X_i - X_0$) from the Huai River ($X_0$)
- Find that a 10-$\mu g/m^3$ increase in PM_{10} (D) reduces life expectancy (Y) by 0.64 years: 95% confidence interval = 0.21-1.07
- The estimates imply that bringing all of China into compliance with its Class I standards for PM_{10} would save 3.7 billion life-years

Fuzzy RD: A "Modern" Example
Discontinuity and Data



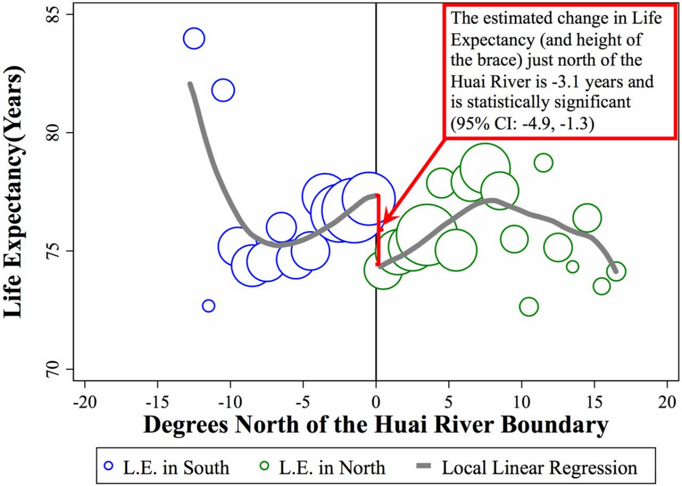
China's Huai River/Qinling Mountain Range winter heating policy line and PM_{10} concentrations. Black dots indicate the DSP locations. Coloring corresponds to interpolated PM_{10} levels at the 12 nearest monitoring stations, where green, yellow, and red indicate areas with relatively low, moderate, and high levels of PM_{10}, respectively. White: no data.

Fuzzy RD: A "Modern" Example
Air Pollution (first stage)



Causal effect of T on D (fitted values, $\widehat{D}$, from local linear regression)

Fuzzy RD: A "Modern" Example
Life Expectancy (reduced form)



Causal effect of T on Y (fitted values, $\widehat{Y}$, from local linear regression)

Fuzzy RD: A "Modern" Example
Air Pollution and Life Expectancy

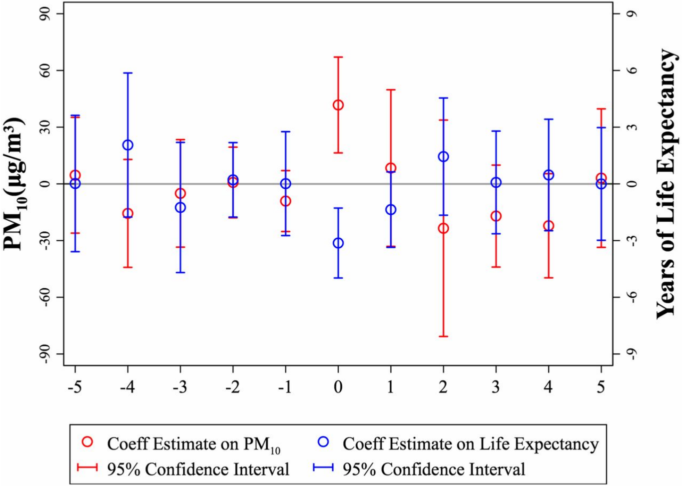
Table 2. RD estimates of the impact of the Huai River Policy

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).
*Significant at 10%.
**Significant at 5%.
***Significant at 1%.

- estimates robust to using parametric and nonparametric estimation methods, different kernel types and bandwidth sizes, and adjustment for a rich set of demographic and behavioral covariates
- the shorter lifespans almost entirely caused by elevated rates of cardiorespiratory mortality, suggesting that PM_{10} is the causal factor

Fuzzy RD: A "Modern" Example
Air Pollution and Life Expectancy



RDD check: "placebo" discontinuities $X_0 \in \{..., -1, 1, ...\}$