Regression Discontinuity Design

Yona Rubinstein

July 2016

Outline

- The basic idea of **R**egression **D**iscontinuity **D**esign (*RDD*).
- **Sharp** versus **Fuzzy** *discontinuities*.
- **Estimating** regression discontinuity.
- **Interpretation** and **Internal validity**.
- **Heterogeneous** effects and **external** validity.
- **Applications** from **recent** literature.

Regression Discontinuity Design (RDD) in Brief

- This method was developed to estimate treatment effects in non-experimental settings.
- The basic idea of *Regression Discontinuity Design* (**RDD**) is the following:
- Observations / subjects (e.g. individuals, firms, governments, economies etc.) are 'treated' based on a known cutoff rule.
- There is an observed variable, X, that assigns subjects to treatment.
- Subjects are **treated** if $X \geq C$ and subjects **are not treated** if $X < C$, where C is the *cutoff value*.
- This cutoff rule should generate discontinuity in outcomes (Y) – if indeed treatment matters. X might have a direct effect on the outcome of interest (Y). Yet in the absence of treatment the relationship between X and Y should be "smooth".
- We can **evaluate** treatment effects at the **margin** by estimating the relationship between Y and X around the *cutoff point*.

Applications

- **Finance**: credit constraints (Keys, et al, 2010).
- **Political Economy**: incumbency effects (Lee, 2008). Do voters influence government policies? (Lee, Moretti and Buttler, 2004)
- **Health**: the effect of medical treatment on infants (Almond et al., 2010); eligibility for Medicare (Card, Dobkin and Maestas, 2008); health insurance (Anderson, 2012).
- **Information**: Yelp ratings effect (Anderson and Magruder, 2012; Lucas, 2012).
- **Education**: class size (Angrist and Lavy, 1999), demand for schooling (Black, 1999), returns to schooling (Thistlethwaite and Campbell, 1960; Abdulkadiroglu et al. 2014)
- **Environment**: air quality and housing prices (Chay and Greenstone, 2005)

Pros and Cons: Credible yet Limited

Pros:

- **Internal validity**: some key identifying assumptions can be empirically verified; specifically the absence of other discontinuities
- **Easy to estimate** (like RTC)
- **Credible** causal estimates of treatment effects.
- **Transparency**: treatment and outcomes can be illustrated using graphical methods

Cons:

- Treatment effects are local (*LATE*)
- *Limits* **external validity**.

RDD and Randomized Treatment-Control (RTC) Setting

- **Assignment** to treatment and control **is not** *random*!
- Subjects **cannot precisely** *manipulate* the **assignment variable** X, that is whether they are *treated* or not. It is a function of X.
- A **consequence** of this is that the *variation in treatment* **near the threshold** is *randomized* as though from a randomized experiment.
- Subjects that are *just below* the threshold are **similar/comparable** to subjects *just above* the X threshold.
- The RD designs can be analyzed—and tested—like *randomized experiments*.
- Therefore RDD is a way of estimating treatment effects in a *nonexperimental* setting where treatment is determined by whether an observed "assignment" variable.

RDD as a Nonparametric Differences- in-Differences

- Has similar flavor to Diff-in-Diff experiment setting.
- Graphical presentation of an RD design using raw data is helpful and informative.
- A graph provides a sense of whether the "jump" in the outcome of interest at the cutoff is unusually large compared to bumps away from the cutoff.
- Plot outcome (Y) against the observed "*assignment variable*" (X). If treatment matters we should observe **sharp**, **discontinuous** *change* in Y at the cutoff value of C.
- At the very same time we *should not observe* and observe **sharp**, **discontinuous** *change* in any explanatory variable at the cutoff value of $X = C$.
- The **first change** in the *first diff*; the **second change** is the *second diff*.
- Note that treatment only depends on X we never observe subjects with the **same** X for a treatment and a control ("sharp RDD").

RDD Terminology

- The variable X is called the "*assignment variable*" or the "*forcing variable*".
- The *cutoff rule* can be a **single** variable or **multiple variables**; for simplicity of illustrations, we'll discuss the single variable setting.
- We label the cutoff value – the threshold – with the letter C.
- Outcomes: Y^0 and Y^1 are the **potential outcomes of interest** *with* and *without treatment*.

Two Types of RDD Settings

- *Sharp RDD*: **Assignment** to **treatment** occurs through **known** and *deterministic* **decision rule** that **depends** only on X; i.e. if $X \geq C$. Subject i is treated with probability 1. If $X_i < C$ and not treated otherwise:

$$D_i = \left\{ \begin{array}{lll} 1 & if & X_i \geq C \\ 0 & if & X_i < C \end{array} \right\}. \quad (1)$$

- *Fuzzy RDD*: The **probability** of being *treated* "**jumps**" at $X_i = C$:

$$\Pr\left(D_i = 1|X\right) = \left\{ \begin{array}{lll} g^1\left(X\right) & if & X_i \geq C \\ g^0\left(X\right) & if & X_i < C \end{array} \right\}, \quad (2)$$

where $g^1\left(X = C\right) > g^1\left(X = C\right)$.

RDD Settings and Compliance

Let Z_i be an indicator variable that equals 1 if $X_i \geq C$ and 0 otherwise:

$$Z_i = 1\left(X_i \geq C\right). \tag{3}$$

1. *Sharp RDD*: **Assignment** to **treatment depends** only on Z_i:

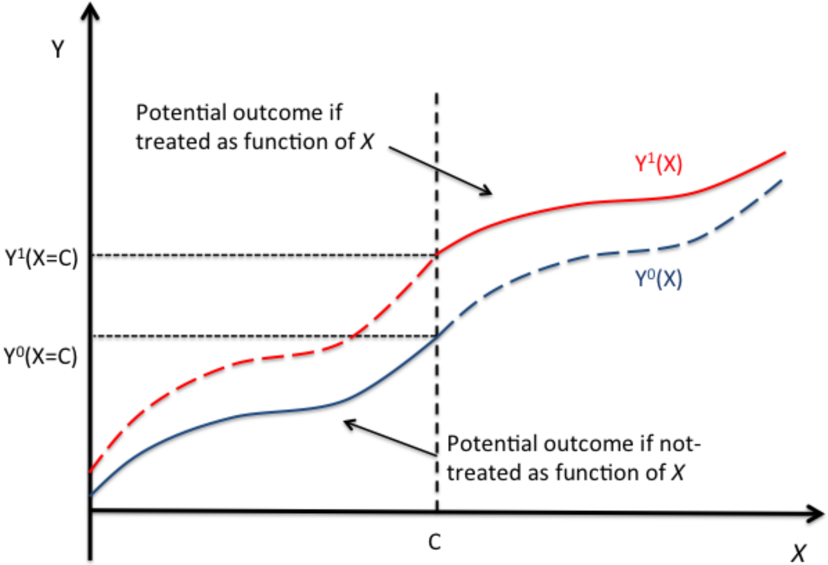
$$D_i = \left\{ \begin{array}{lll} 1 & if & Z_i = 1 \\ 0 & if & Z_i = 0 \end{array} \right\}. \tag{4}$$

2. *Fuzzy RDD*: **Assignment** to **treatment depends** on Z_i, on X and on "*unobservables*":

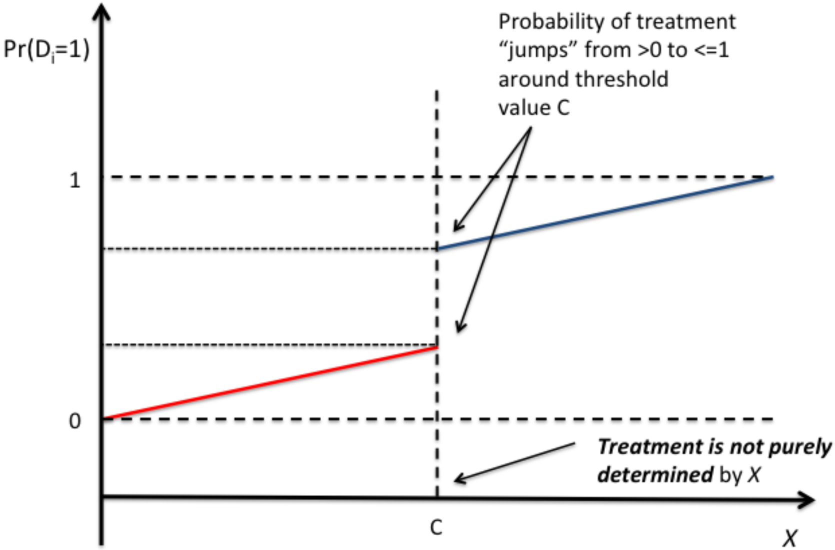
$$\Pr\left(D_i = 1|X\right) = \left\{ \begin{array}{lll} E\left(D_i = 1|X, Z = 1\right) & if & Z_i = 1 \\ E\left(D_i = 1|X, Z = 0\right) & if & Z_i = 0 \end{array} \right\}. \tag{5}$$

- *Sharp RDD* is a particular case of the *Fuzzy RDD* with **perfect compliance**. We'll discuss that later.

Discontinuity in Treatment: Sharp RDD Visually

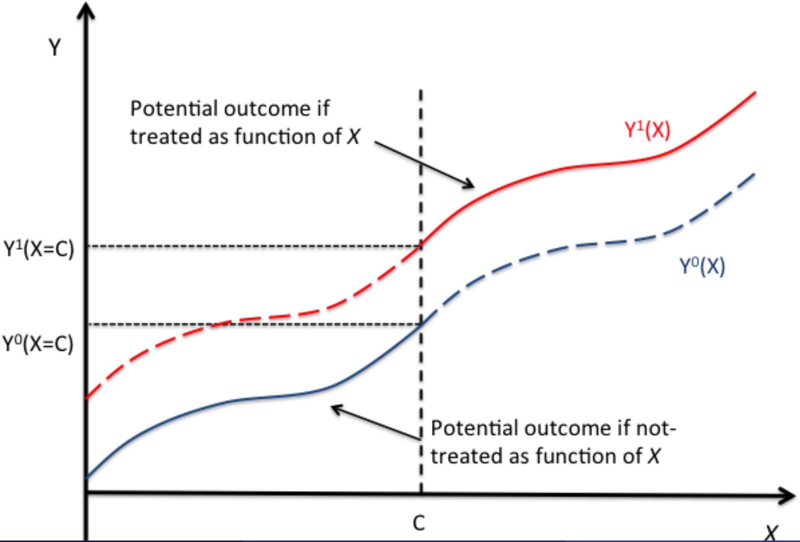


Discontinuity in Treatment: Fuzzy RDD Visually



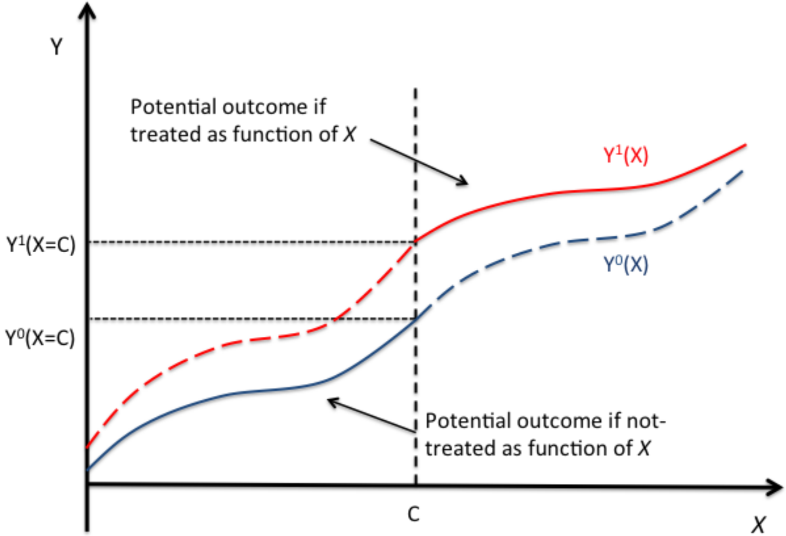
Continuity in Potential Outcomes

$Y^0(X)$ and $Y^0(X)$ continuous at threshold C; <u>*common effect*</u>.



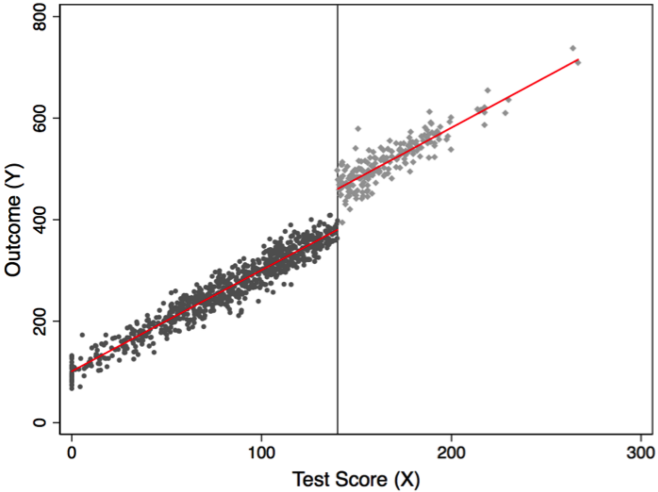
Continuity in Potential Outcomes

$Y^0(X)$ and $Y^0(X)$ continuous at threshold C; *heterogeneous effects*.



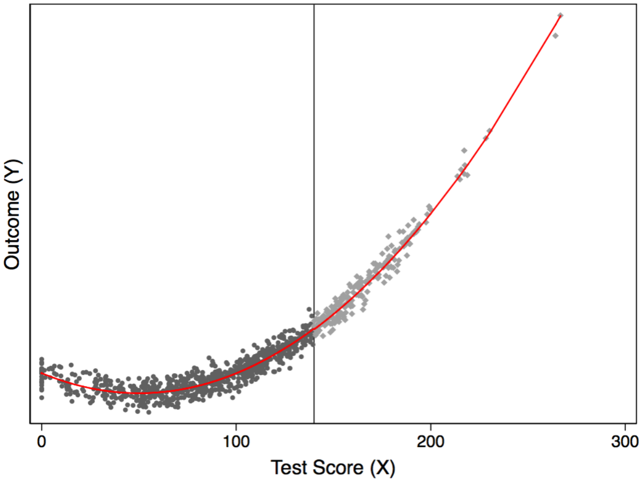
Sharp RDD: Graphical Example

- Simulated data with STATA: $C = 140$; "**gen y = 100 + 80*T + 2*x + rnormal(0, 20)**"



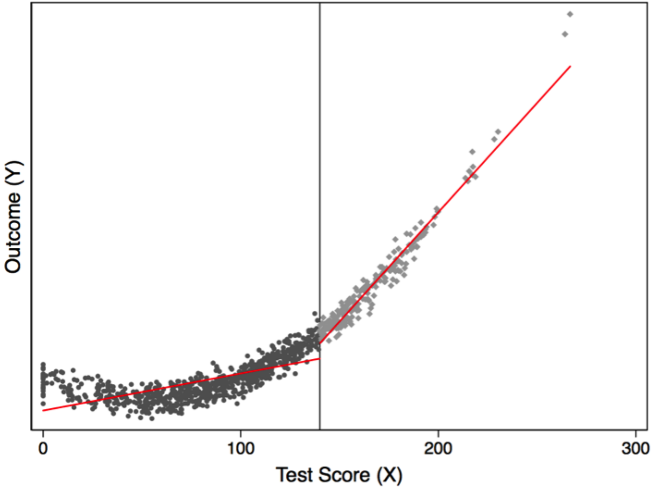
Nonlinear Effect of X on Y, Graphical Example

- Simulated data with STATA: $C = 140$; "**gen y = 10000 + 0*T - 100*x +x2 + rnormal(0, 1000)**"



Nonlinear Effect of X on Y and Miss-specified Sharp RDD:

- What if nonlinear? Could result in a biased treatment effect if one assumes a linear model?



Sharp RDD: Estimation

Two practical strategies to implement a RDD.

1. **First**, *globally* using **all data**: **control** for the direct effect of X on Y in a very **general and rigorous form**.
2. **Second**, *locally* using data in a **narrow window** / **margin** around **threshold**: control for the local direct effect of X on Y

Estimating Sharp RDD using All Data

- Specify the *potential outcomes* models as a **flexible function** of the forcing variable X.
- *Potential outcomes* in the **absence** of *treatment*:

$$Y_i^0 = \beta_0 + f^0 (X_i - C) + U_i^0. \quad (6)$$

- *Potential outcomes* **with** *treatment*:

$$Y_i^1 = \beta_0 + \beta_D + f^1 (X_i - C) + U_i^1. \quad (7)$$

- $f^0 (X_i)$ and $f^1 (X_i)$ are general continuous functions of X where:

$$f^0 (0) = f^1 (0) = 0. \quad (8)$$

- The **average treatment effect** for subjects with $X = C$ (LATE) is:

$$\beta_D = Y^1_{X_i=C} = C - Y^0_{X_i=C}. \quad (9)$$

Sharp RDD using All Data in Practice

- Practically we can estimate the local treatment effect using a fully interacted – switching regression – model:

$$Y_i = (1 - D_i)\, Y_i^0 + D_i Y_i^0 \tag{10}$$

- Where:

$$D_i = 1\,(X_i \geq C)\,. \tag{11}$$

- By introducing (6) and (7) into the **switching regression model** above we obtain:

$$Y_i = \beta_0 + \beta_D D_i + f^0\,(X_i - C) + f\,(X_i - C)\, D_i + U_i \tag{12}$$

- where:

$$f\,(X_i - C) = f^1\,(X_i - C) - f^0\,(X_i - C)$$

- β_D is the average treatment effect for subjects with $X_i = C$ – the **Local Average Treatment Effect** at $(X = C)$.

in Practice (cont.)

- A common practical practice is to approximate model $f\left(X_i - C\right)$ with a pth order polynomial, specifically:

$$\begin{aligned} f^0\left(X_i - C\right) &= f\left(x_i\right) = \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 \ldots + \beta_p x_i^p \qquad (13) \\ f^1\left(X_i - C\right) &= g\left(x_i\right) = \gamma_1 x_i + \gamma_2 x_i^2 + \gamma_3 x_i^3 + \gamma_4 x_i^4 \ldots + \gamma_p x_i^p, \end{aligned}$$

- Note that $x_i = X_i - C$.
- We estimate the model in (12) using the following specification:

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_D D_i + \qquad (14) \\ & \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 \ldots + \beta_p x_i^p + \\ & \delta_1 x_i D_i + \delta_2 x_i^2 D_i + \delta_3 x_i^3 D_i + \delta_4 x_i^4 D_i \ldots + \delta_p x_i^p D_i + \\ & U_i, \end{aligned}$$

- Note that $\delta_j = \gamma_j - \beta_j$.

Estimating Sharp RDD using Windows

- Estimate, this time, using observations with X_i such that:

$$C - \Delta \leq X_i \leq C + \Delta$$

- Specifically, estimate the following model using the sub-sample Δ above:

$$Y_{i\Delta} = \beta_0 + \beta_D D_i + \tilde{f}^0(x_i) + \tilde{f}(x_i) D_i + U_{i\Delta} \tag{15}$$

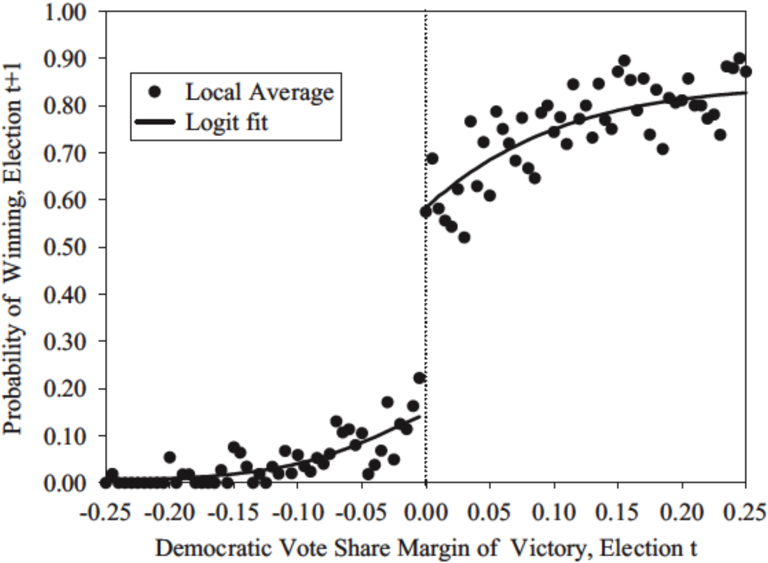
- $\tilde{f}^0(x_i)$ and $\tilde{f}(x_i)$ are general continuous functions of X; might be a bit less flexible than the comparable $f^0(x_i)$ and $f(x_i)$.
- Note that as $\Delta \to 0$ the estimated $\beta_D^{OLS} \to \beta_D$

$$\lim_{\Delta \to 0} E[Y_i | C \leq X_i \leq C + \Delta] - E[Y_i | C - \Delta \leq X_i \leq C] = E[(Y_i^1 - Y_i^0) | X_i = C] = \beta_D.$$

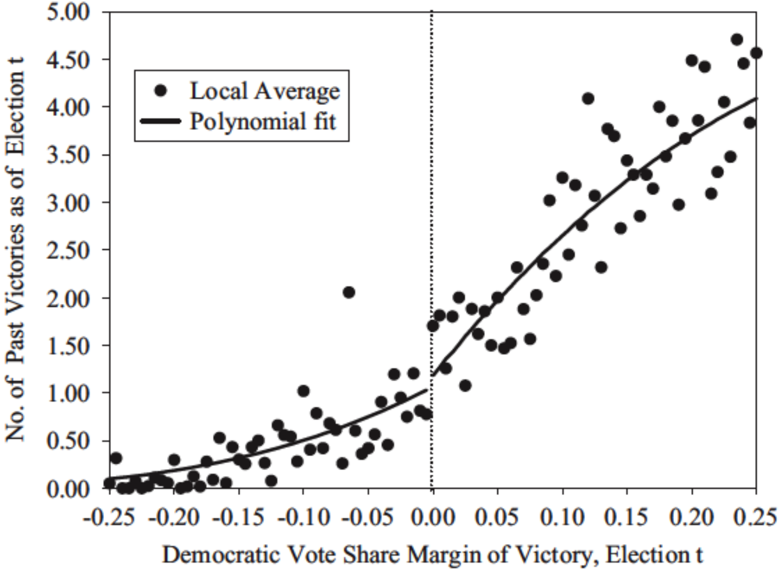
Sharp RDD: Winning the Next Election, Lee 2008

- Lee (2008) uses a Sharp Regression Discontinuity Design to estimate the causal effect of incumbency in U.S. House elections.
- The population sample of contains more than six thousands elections over the 1946–98 period.
- The assignment variable in this setting –X– is the fraction of votes awarded to Democrats in the previous election.
- The cutoff rule, C,is 50%
- When the fraction exceeds 50%, that is when $X \geq 0.5$ a Democrat is elected and the party becomes the incumbent party in the next election.
- Both the share of votes and the probability of winning the next election are considered as outcome variables.

Sharp RDD: Probability of Winning Election t+1, Lee 2008



Sharp RDD: Probability of Winning Election t-1, Lee 2008



Fuzzy RDD

- The causal model - represented by the **switching regression model** (12) – is the same as in the Sharp RD Design:

$$Y_i = \beta_0 + \beta_D D_i + f^0(x_i) + f(x_i) D_i + U_i \quad (16)$$

- Yet, unlike the Sharp RDD, **treatment** itself (D_i) is correlated with whether the subject is above or below the cutoff (Z_i) but not solely determined by that:

$$\begin{aligned} Z_i &= 1(x_i \geq 0), \\ D_i &\neq Z_i. \end{aligned}$$

- Hence, we can write the probability of treatment as:

$$\Pr(D_i = 1|X) = \alpha_0 + \beta_Z Z_i + g(x_i) \quad (17)$$

- Note that the Sharp RDD is a special case where:

$$\alpha_0 = 0, \beta_Z = 1, g(x_i) = 0.$$

Fuzzy RDD as an IV

- Hence, the *cutoff rule* serves as an *instrumental variable*.
- Assuming that Y is a smooth and continuos function of X, then the *cutoff rule* reflected in Z can serve as an instrumental variable since it fulfills the two necessary conditions:

1. *Relevancy*: Z *affects* the probability of treatment (D).
2. *Exclusion Restriction*: Z **does not have** a *direct effect* on Y **conditional** on X.

- We use 2SLS to estimate the parameter of interest β_D.

1. First stage: since $D_i = \Pr(D_i = 1|X) = \alpha_0 + \beta_Z Z_i + g(x_i) + V_i$

$$D_i = \alpha_0 + \beta_Z Z_i + g(x_i) + V_i. \tag{18}$$

2. Second stage:

$$Y_i = \beta_0 + \beta_D \hat{D}_i + f(x_i) + W_i. \tag{19}$$

Fuzzy RDD: an Example from Angrist and Lavy (1999)

- Angrist and Lavy (1999) study the **impact** of class size on educational attainment.
- Angrist and Lavy exploit an old Talmudic rule –."*Maimonides' Rule*" – that classes should be split if they have more than 40 students in Israel. Specifically:

1. A school with 40 students has only one class of 40.
2. A school with 41 students has two classes. with 21 and 20 students.

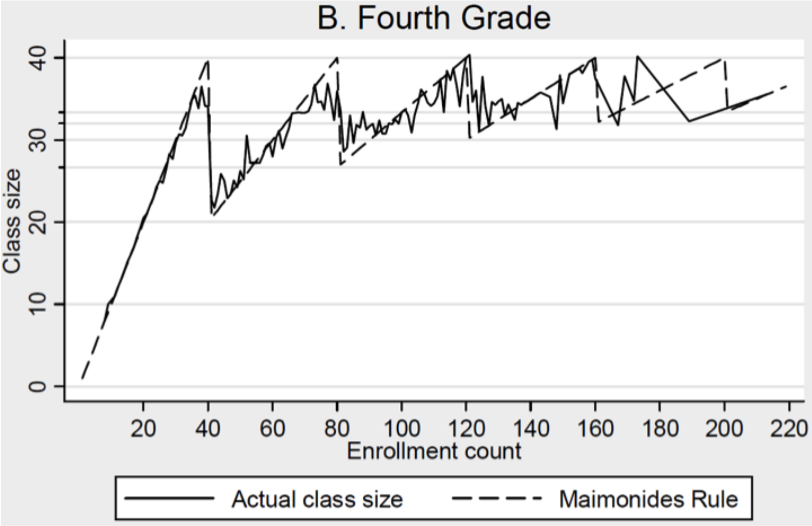
- This rule implies that class C's size at school S, M_{SC}, as a function of enrollment E_S is:

$$M_{SC} = \frac{E_S}{int\left[\frac{E_S - 1}{40}\right] + 1},$$

where $int\,[A]$ is the integer part of the real number A.

The Maimonides' Rule as an Instrumental Variable

Note, that the rule was NOT followed strictly! Yet, a very clever IV.



OLS and Fuzzy RD Estimates

	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

Fuzzy Local RDD

- Similarly to the sharp RDD, we can zoom in around the threshold and get *non-parametric* estimates.
- Specifically, we restrict the sample to the following range of X:

$$C - \Delta \leq X_i \leq C + \Delta$$

- Then we estimate the model using 2SLS.

1. The *reduced form* estimate is:

$$E\left[Y_i | C \leq X_i \leq C + \Delta\right] - E\left[Y_i | C - \Delta \leq X_i \leq C\right] \tag{20}$$

2. The *first stage* estimate is:

$$E\left[D_i | C \leq X_i \leq C + \Delta\right] - E\left[D_i | C - \Delta \leq X_i \leq C\right] \tag{21}$$

- The Wald estimator is:

$$\lim_{\Delta \to 0} \frac{E\left[Y_i | C \leq X_i \leq C + \Delta\right] - E\left[Y_i | C - \Delta \leq X_i \leq C\right]}{E\left[D_i | C \leq X_i \leq C + \Delta\right] - E\left[D_i | C - \Delta \leq X_i \leq C\right]} = \beta_D \tag{22}$$

Parametric or Non-Parametric? Bias versus Noise

- Similarly to the sharp RDD, we can "**zoom in**" *around* the *threshold* and get **non-parametric** *estimates*.
- When would parametric or non-parametric or window size matter?
  - Small effect
  - Relationship between Y and X different away from cutoff,
  - Functional form not well captured by polynomials (or other functional form).
- What do we trade-off?
- *Zoom In*: consistent yet noisy estimate. smaller window can be subject to greater noise, **robust** to *functional form* assumptions.
- *Zoom Out*: efficient, yet biased estimate; contamination by X or "*unobservables*" the functional form **fails** to *condition out*.

RDD: Some Concerns

1. **First concern**: functional form and latent variables around cutoff
2. **Second concern**: subject manipulate X around the cutoff. Note that if subjects can't perfectly manipulate it, then there will still be randomness in treatment.

- **Addressing 1st concern**: *plot RD pictures* for other variables that we observe, but that **shouldn't be affected** by treatment (D), see Lee (2008)
- **Addressing 2nd concern**: test whether subjects are *bunching* around the threshold by looking for **discontinuities** in the *distribution of characteristics* that **shouldn't be affected** by the treatment (gender?, age?, etc.).
- **Additional checks**: falsification test

Take Home Message

- *RDD* allows to **identify** *causal effects* using **observational data**.
- Makes use of *treatment assignment* that **isn't random**, but where process follows some **known** and **arbitrary** *cutoff rule*.
- Very useful with large samples; allows for **nonparametric** estimation and assessment of **internal validity**.
- Very common scenario in practice, and estimator likely to be of increasing use.
- If *treatment effect* is **heterogeneous** *estimators'* interpretation is **LATE**