Lecture 5

Regression Discontinuity Design

Policy Evaluation

Nova SBE - Universidade Nova de Lisboa / IGC

Pedro C. Vicente

http://www.pedrovicente.org/

Introduction

- Discontinuities in program implementation, based on eligibility criteria or other exogenous factors, can be very useful in non-experimental program evaluation
- People above and below the threshold can be compared in terms of outcomes; but, to ensure comparability, the samples across which to compare need to be sufficiently close to the eligibility cutoff
- This way, selection on unobservables close to the eligibility cutoff is minimized
- **Regression Discontinuity Design (RDD)** is similar to IV because it introduces an exogenous variable that is highly correlated with participation

- For example:
  - Grameen Bank's microcredit is targeted to households with landholdings of less than half an acre
  - Pension programs are targeted to populations above a certain age
  - Scholarships are targeted to students with high scores on standardized tests
- By looking at a narrow band of units that are below and above the cutoff point and comparing their outcomes, one can judge the program's impact because the individuals just below and above the threshold are likely to be very similar to each other

Figure 7.1 Outcomes before Program Intervention

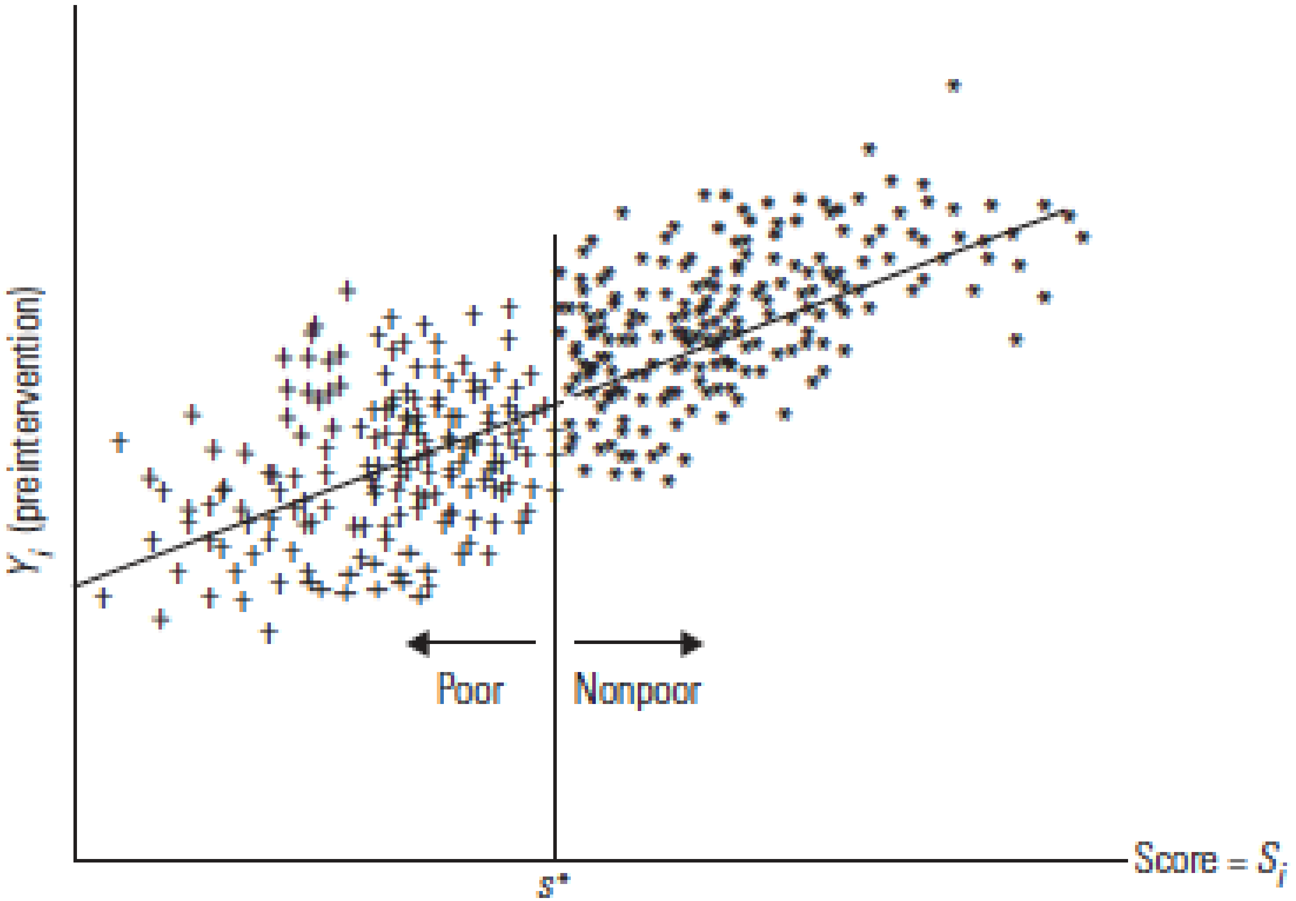
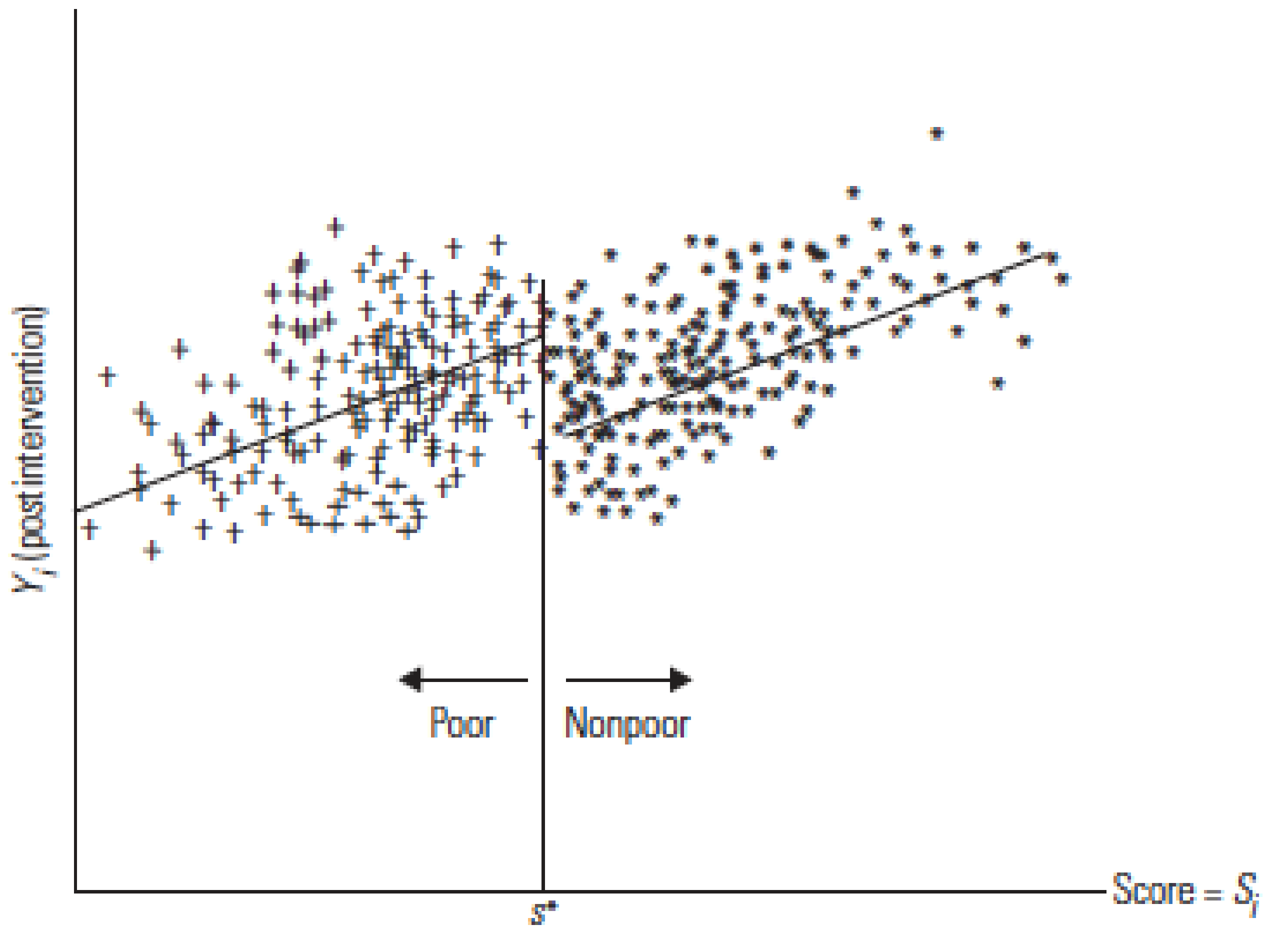


Figure 7.2 Outcomes after Program Intervention



RDD Theory – Sharp RDD

• To model the effect of a particular program on individual outcomes Y_i through an RDD approach, one needs a variable X_i that determines program eligibility (such as age, asset holdings, or the like) with an eligibility cutoff of X^*

• The **sharp RDD** is defined by the following assignment rule:

$$D_i = \begin{cases} 1 \; if \; X_i \geq X^* \\ 0 \; if \; X_i < X^* \end{cases}$$

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

$$E(Y_{0i} \mid X_i) = \alpha + \beta X_i$$

$$Y_{1i} = Y_{0i} + \rho$$

leading to the regression,

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

• The key difference between this regression and others in this course, is that D_i is not only correlated with X_i but also a deterministic function of X_i

• RDD distinguishes the discontinuous function

$$I(X_i \geq X^*)$$

from the smooth (linear) function X_i

• Note that we can be more general and assume a more general model non-linear in X_i

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

• As long as *f(.)* is continuous in a neighborhood of X^*, it should be possible to estimate the model

• For example, modeling *f(.)* with p-th order polynomial, RDD estimates can be constructed from the regression

$$Y_i = a + b_1 X_i + b_2 X_i^2 + ... + b_p X_i^p + r D_i + e_i$$

• A slightly different version of the above also allows different trend functions for both potential outcomes Y_{0i} and Y_{1i}

• The validity of RDD estimates of causal effects based on the above specification relies on whether the polynomial model provides an adequate description of

$$E(Y_{0i} | X_i)$$

• If not, what looks like a jump due to treatment might be an unaccounted for non-linearity in the counterfactual conditional mean function

• Comparisons of average outcomes in a small neighborhood to the left and right of X^* estimate the treatment effect in a way that does not depend on the correct specification of a model for

$$E(Y_{0i} \mid X_i)$$

• Then we have for small positive Δ

$$E(Y_i \mid X^* - \Delta < X_i < X^*) \approx E(Y_{0i} \mid X_i = X^*)$$

$$E(Y_i \mid X^* < X_i < X^* + \Delta) \approx E(Y_{1i} \mid X_i = X^*)$$

so that

$$\lim_{\Delta \to 0} E(Y_i \mid X^* < X_i < X^* + \Delta) - E(Y_i \mid X^* - \Delta < X_i < X^*)$$

$$= E(Y_{1i} - Y_{0i} \mid X_i = X^*)$$

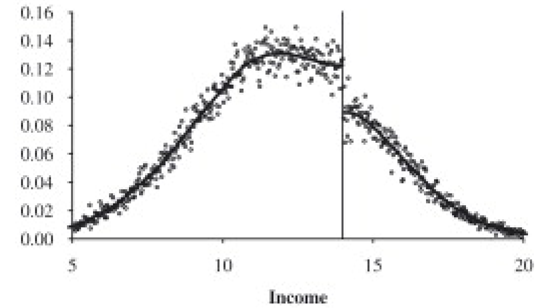
which is our RDD estimate of interest

- The non-parametric approach to RDD requires good estimates of the mean of Y_i in small neighborhoods to the right and left of X^*
- This is difficult because of data availability
- Typical solutions include:
    - Local linear regression (Hahn et al., Econometrica, 2001)
    - Local polynomial regression estimators
- For example, the local linear estimator for

$$E(Y_i \mid X^* < X_i < X^* + \Delta) \equiv Y^+$$

is the constant estimated parameter from (the local polynomial estimator is analogous):

$$(\hat{a}, \hat{b}) \equiv \operatorname*{argmin}_{a,b} \sum_{i=1}^{n} (Y_i - a - b(X_i - X^*))^2 K\left(\frac{X_i - X^*}{h}\right) 1(X_i > X^*)$$



- Three robustness tests particularly useful:
  - The density of the variable determining eligibility around the sharp threshold can help show whether the RDD is valid, i.e., that members of the non-eligible sample do not become participants (namely through misreporting the value of the eligibility variable)
  - As we adjust the interval of analysis around X^* and approach X^* (losing precision of the estimate), and diminish the number of controls, the estimate of the TE should be stable
  - The average values of the pre-determined covariates around the threshold also can provide an indication of RDD validity (they should not vary with the threshold)
- See Lee and Lemieux (JEL, 2010) for many tips on implementation decisions and presentation of results

Fuzzy RDD

- **Fuzzy RDD** exploits discontinuities in the probability or expected value of treatment conditional on a covariate
- This may occur when eligibility rules are not strictly adhered to or when certain geographic areas are targeted but boundaries are not well defined and mobility is common
- There is now a jump in the probability of treatment at X^*

$$P(D_i = 1 | X_i) = \begin{cases} p_1(X_i) \; if \; X_i \geq X^* \\ p_0(X_i) \; if \; X_i < X^* \end{cases} \quad with \; p_1(X_i) \neq p_0(X_i)$$

- Let's assume

$$p_1(X_i) > p_0(X_i)$$

- We can then write

$$E(D_i | X_i) = P(D_i = 1 | X_i) = p_0(X_i) + (p_1(X_i) - p_0(X_i))T_i$$

where

$$T_i = \mathrm{I}(X_i > X^*)$$

- We can then see that T_i can be used as an instrument for D_i

• We can then follow a 2SLS procedure where the first stage is

$$D_i = g_0 + g_1 X_i + g_2 X_i^2 + \dots + g_p X_i^p + p T_i + x_i$$

• We can then substitute the prediction of the above equation in our equation of interest

$$Y_i = a + b_1 X_i + b_2 X_i^2 + \dots + b_p X_i^p + r \hat{D}_i + e_i =$$

$$m + k_1 X_i + k_2 X_i^2 + \dots + k_p X_i^p + r p T_i + V_i$$

with

$$m = a + r g_0, k_j = b_j + r g_j \ \mathit{for}\ j = 1, \dots, p$$

• Just like for the sharp case, it is crucial to distinguish the effect of the discontinuity from the non-linear functions

- The non-parametric version of fuzzy RDD consists of IV estimation in a small neighborhood around the discontinuity
- We know the effect of the instrument on the outcome near the threshold is

$$E(Y_i \mid X^* < X_i < X^* + \Delta) - E(Y_i \mid X^* - \Delta < X_i < X^*) \approx \rho\pi$$

- Similarly for the first stage, we have

$$E(D_i \mid X^* < X_i < X^* + \Delta) - E(D_i \mid X^* - \Delta < X_i < X^*) \approx \pi$$

- These mean we can identify our parameter of interest through

$$\lim_{\Delta \to 0} \frac{E(Y_i \mid X^* < X_i < X^* + \Delta) - E(Y_i \mid X^* - \Delta < X_i < X^*)}{E(D_i \mid X^* < X_i < X^* + \Delta) - E(D_i \mid X^* - \Delta < X_i < X^*)} = \rho$$

Example: Class sizes revisited

- Angrist and Lavy (QJE, 1999)
- The twelfth century rabbinic scholar Maimonides proposed a maximum class size of 40
  - Maimonides' rule (MR) is not the only source of variation in Israeli class sizes, and average class size is generally smaller, but the ceiling of 40 is a real constraint faced in many schools
  - A regression of actual class size on size predicted by MR explains about half the variation in class size
- The class-size function induced by MR is used to construct instrumental variables estimates of effects of class size on test scores
  - This can also be seen as a fuzzy RDD

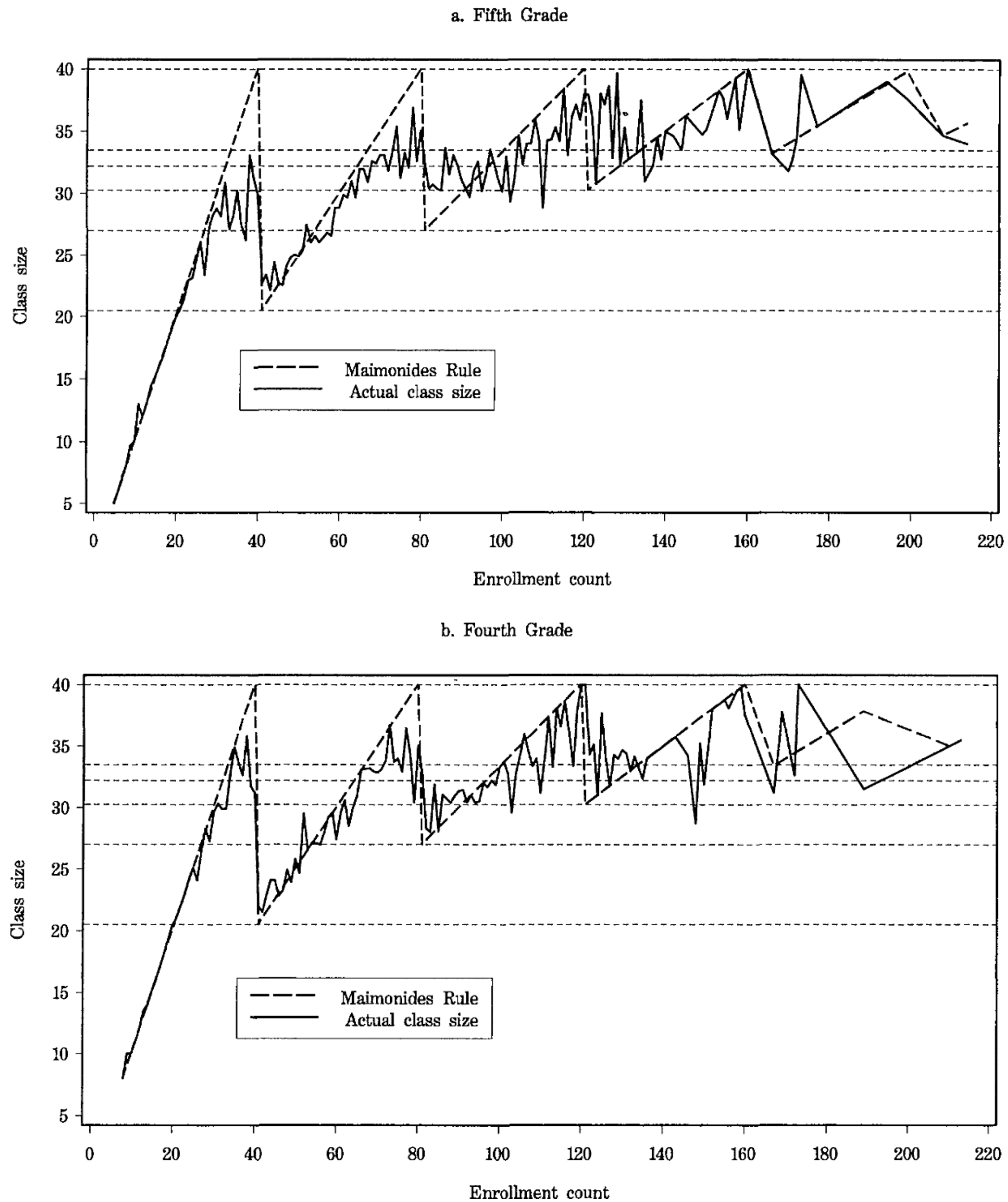


FIGURE I
Class Size in 1991 by Initial Enrollment Count, Actual Average Size and as Predicted by Maimonides' Rule

$$(2) \qquad y_{isc} = X'_s\beta + n_{sc}\alpha + \mu_c + \eta_s + \epsilon_{isc},$$

where y_{isc} is pupil i's score, X_s is a vector of school characteristics, sometimes including functions of enrollment, and n_{sc} is the size of class c in school s. The term μ_c is an i.i.d. random class component, and the term η_s is an i.i.d. random school component. The remaining error component ϵ_{isc} is specific to pupils. The first two error components are introduced to parameterize possible within-school and within-class correlation in scores. The class-size coefficient α is the parameter of primary interest.

$$(4) \qquad n_{sc} = X'_s\pi_0 + f_{sc}\pi_1 + \xi_{sc},$$

- The instrument comes from:

$$f_{sc} = \frac{e_s}{\operatorname{int}\left(\frac{e_s - 1}{40}\right) + 1}$$

TABLE II
OLS Estimates for 1991

	5th Grade						4th Grade					
	Reading comprehension			Math			Reading comprehension			Math		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean score		74.3			67.3			72.5			69.9	
(s.d.)		(8.1)			(9.9)			(8.0)			(8.8)	
Regressors												
Class size	.221	−.031	−.025	.322	.076	.019	0.141	−.053	−.040	.221	.055	.009
	(.031)	(.026)	(.031)	(.039)	(.036)	(.044)	(.033)	(.028)	(.033)	(.036)	(.033)	(.039)
Percent disadvantaged		−.350	−.351		−.340	−.332		−.339	−.341		−.289	−.281
		(.012)	(.013)		(.018)	(.018)		(.013)	(.014)		(.016)	(.016)
Enrollment			−.002			.017			−.004			.014
			(.006)			(.009)			(.007)			(.008)
Root MSE	7.54	6.10	6.10	9.36	8.32	8.30	7.94	6.65	6.65	8.66	7.82	7.81
R^2	.036	.369	.369	.048	.249	.252	.013	.309	.309	.025	.204	.207
N		2,019			2,018			2,049			2,049	

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes.

TABLE IV
2SLS Estimates for 1991 (Fifth Graders)

	Reading comprehension						Math					
	Full sample				+/− 5 Discontinuity sample		Full sample				+/− 5 Discontinuity sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean score	74.4				74.5		67.3				67.0	
(s.d.)	(7.7)				(8.2)		(9.6)				(10.2)	
Regressors												
Class size	−.158	−.275	−.260	−.186	−.410	−.582	−.013	−.230	−.261	−.202	−.185	−.443
	(.040)	(.066)	(.081)	(.104)	(.113)	(.181)	(.056)	(.092)	(.113)	(.131)	(.151)	(.236)
Percent disadvantaged	−.372	−.369	−.369		−.477	−.461	−.355	−.350	−.350		−.459	−.435
	(.014)	(.014)	(.013)		(.037)	(.037)	(.019)	(.019)	(.019)		(.049)	(.049)
Enrollment		.022	.012			.053		.041	.062			.079
		(.009)	(.026)			(.028)		(.012)	(.037)			(.036)
Enrollment squared/100			.005						−.010			
			(.011)						(.016)			
Piecewise linear trend				.136						.193		
				(.032)						(.040)		
Root MSE	6.15	6.23	6.22	7.71	6.79	7.15	8.34	8.40	8.42	9.49	8.79	9.10
N	2019			1961	471		2018			1960	471	

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes. All estimates use f_{sc} as an instrument for class size.

What have we learnt?

- RDD yields an unbiased estimate of treatment effect at the discontinuity, while making use of a known rule in the assignment of a policy
    - Note that compared to randomization, no group of eligible individuals needs to be excluded from the policy
- However, RDD produces a LATE that is not always generalizable (or interesting for the evaluator), it typically relies on a low number of observations (close to the threshold), and the results can be sensitive to functional form
- The typical problem about applying RDD is that program officials often do not know well the eligibility criteria

Problem Set 5 (Optional)

Exercise: Follow the instructions in chapter 16 of Khandker et al (2009). You should produce a do file and a log file, which should be commented to show that you understood the results. These should be emailed to the grader (Matilde Grácio): only one email per group, please.

Note: We should be able to run the do file on a computer given the original datafile and produce the raw log file.

Due date: Friday, March 18.

The Table we show in the next page is taken from Ludwig, Jens, and Douglas Miller (2007), 'Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design', QJE. In that paper, the child mortality effects of the social program 'Head Start' in the U.S. are estimated.

a) Describe in general terms the identification method employed in this paper. Distinguish from fuzzy regression discontinuity.

b) Describe the difference between the main results and the specification tests in the Table.

c) Describe the difference between the non-parametric and the parametric estimators in the Table.

d) What do the bandwidth numbers mean in the Table?

e) What is the advantage the method employed in this paper has over randomization? Explain.

f) Which treatment effect is being estimated in the Table?

TABLE III

REGRESSION DISCONTINUITY ESTIMATES OF THE EFFECT OF HEAD START ASSISTANCE ON MORTALITY

Variable	Control mean	Nonparametric estimator			Parametric: Flexible linear	Parametric: Flexible quadratic
Bandwidth or poverty range		9	18	36	8	16
Number of observations (counties) with nonzero weight		527	961	2,177	484	863
Main results						
Ages 5–9, Head Start-related causes, 1973–1983	3.238	−1.895**	−1.198*	−1.114**	−2.201**	−2.558**
		(0.980)	(0.796)	(0.544)	(1.004)	(1.261)
		[0.036]	[0.081]	[0.027]	[0.022]	[0.021]
Specification checks						
Ages 5–9, injuries, 1973–1983	22.303	0.195	2.426	0.679	−0.164	0.775
		(3.472)	(2.476)	(1.785)	(3.380)	(3.401)
		[0.924]	[0.345]	[0.755]	[0.998]	[0.835]
Ages 5–9, all causes, 1973–1983	40.232	−3.416	0.053	−1.537	−3.896	−2.927
		(4.311)	(3.098)	(2.253)	(4.268)	(4.295)
		[0.415]	[0.982]	[0.558]	[0.317]	[0.505]
Ages 25+, Head Start-related causes, 1973–1983	131.825	2.204	6.016	5.872	2.091	2.574
		(5.719)	(4.349)	(3.338)	(5.581)	(6.415)
		[0.700]	[0.147]	[0.114]	[0.749]	[0.689]
Ages 25+, injuries, 1973–1983	121.191	5.697	7.276*	4.398	2.65	4.276
		(6.527)	(4.531)	(3.249)	(6.206)	(6.059)
		[0.256]	[0.060]	[0.261]	[0.596]	[0.426]
Ages 5–9, Head Start causes, 1959–1964	9.752	−3.327	−1.076	−0.066	−3.754*	−4.869**
		(5.066)	(3.341)	(2.075)	(5.136)	(5.016)
		[0.117]	[0.536]	[0.641]	[0.075]	[0.039]
Whites age 5–9, Head Start-related causes, 1973–1983	2.63	−1.105	−0.865	−0.749	−1.334	−1.746
		(1.056)	(0.862)	(0.618)	(1.061)	(1.332)
		[0.263]	[0.269]	[0.198]	[0.212]	[0.145]
Blacks age 5–9, Head Start-related causes, 1973–1983	4.688	−2.275	−2.719**	−1.589	−1.699	−1.93
		(3.758)	(2.163)	(1.706)	(4.094)	(3.718)
		[0.173]	[0.048]	[0.322]	[0.411]	[0.276]

Outcome of interest is one-year mortality rates per 100,000 Head Start-related causes include deaths due to tuberculosis, other infections, diabetes, nutritional causes, anemias, meningitis, and respiratory causes. Each cell presents a separate estimate of the discontinuity in the outcome measure listed in the left-hand column at OEO's threshold 1960 poverty level for providing counties with grant-writing assistance for Head Start funding. Within each cell the first number represents our point estimate of α from (3), with analytic standard errors in parentheses and percentile-T bootstrapped p-values in square brackets. Nonparametric estimates are based on the locally weighted kernel regression method discussed in Porter [2003], calculated using a triangle kernel. Parametric models give equal weight to observations within the range of the cutoff and model factors that vary with 1960 poverty rates using a linear or quadratic term in 1960 poverty with slopes allowed to differ on both sides of the OEO cutoff for Head Start grant-writing assistance.

*$p < .1$; **$p < .05$, using Percentile-T significance.