9) Fuzzy Regression Discontinuity Design

Vitor Kamada

July 2018

Reference

Tables, Graphics, and Figures from:

Cattaneo et al. (2018). **A Practical Introduction to Regression Discontinuity Designs: Volume I**

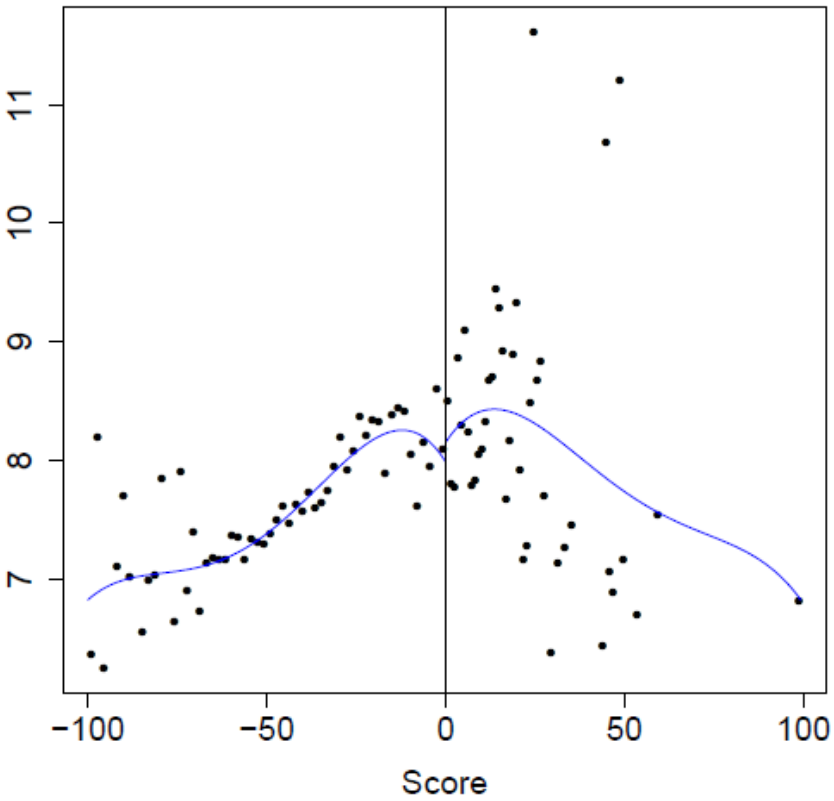
Angrist and Pischke (2008). **Mostly Harmless Econometrics: Ch 6**

Descriptive Statistics for Meyersson (2014)

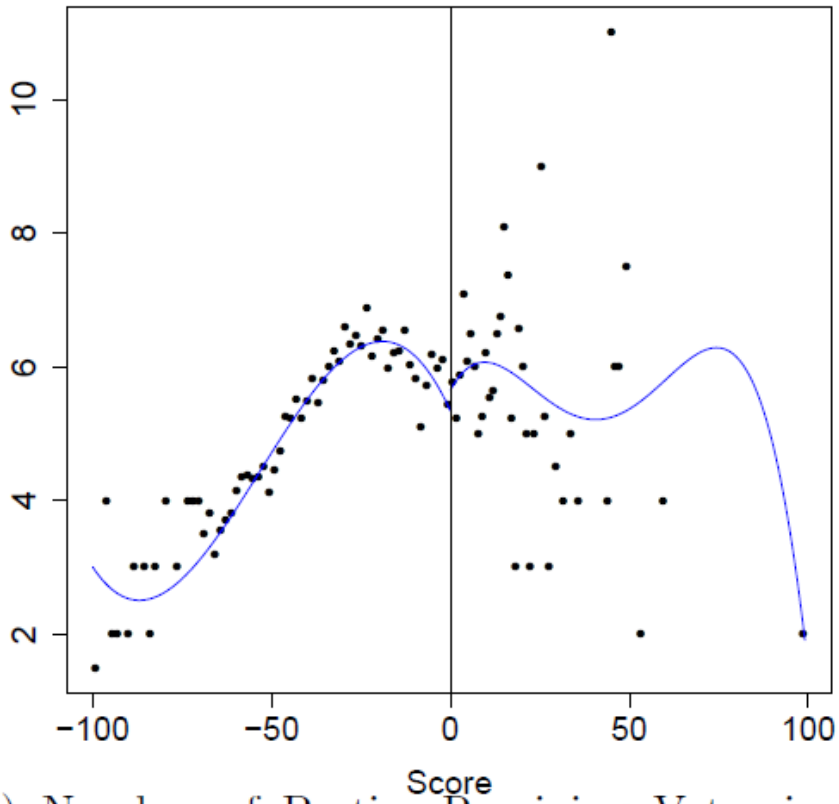
Variable	Mean	Median	Std. Dev.	Min.	Max.
Y	16.306	15.523	9.584	0.000	68.038
X	-28.141	-31.426	22.115	-100.000	99.051
T	0.120	0.000	0.325	0.000	1.000
Percentage of men aged 15-20 with high school education	19.238	18.724	7.737	0.000	68.307
Islamic percentage of votes in 1994	13.872	7.029	15.385	0.000	99.526
Number of parties receiving votes 1994	5.541	5.000	2.192	1.000	14.000
Log population in 1994	7.840	7.479	1.188	5.493	15.338
Percentage of population below 19 in 2000	40.511	39.721	8.297	6.544	68.764
Percentage of population above 60 in 2000	9.222	8.461	3.960	1.665	27.225
Gender ratio in 2000	107.325	103.209	25.293	74.987	1033.636
Household size in 2000	5.835	5.274	2.360	2.823	33.634
District center	0.345	0.000	0.475	0.000	1.000
Province center	0.023	0.000	0.149	0.000	1.000
Sub-metro center	0.022	0.000	0.146	0.000	1.000

Note: the number of observations for all variables is 2,629

Predetermined Covariates I

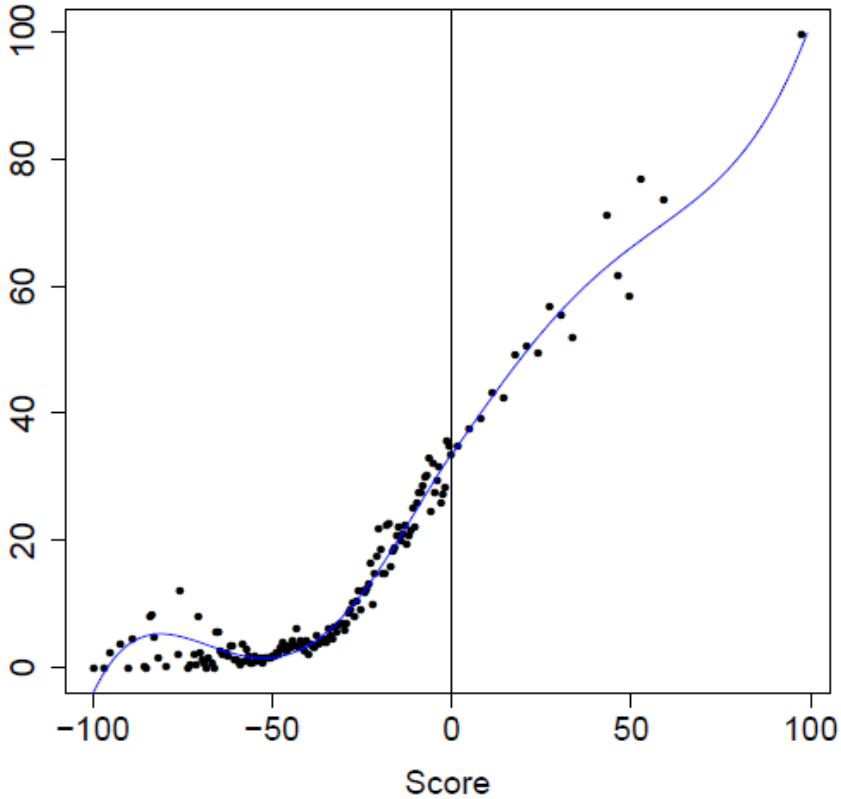


(a) Log Population in 1994

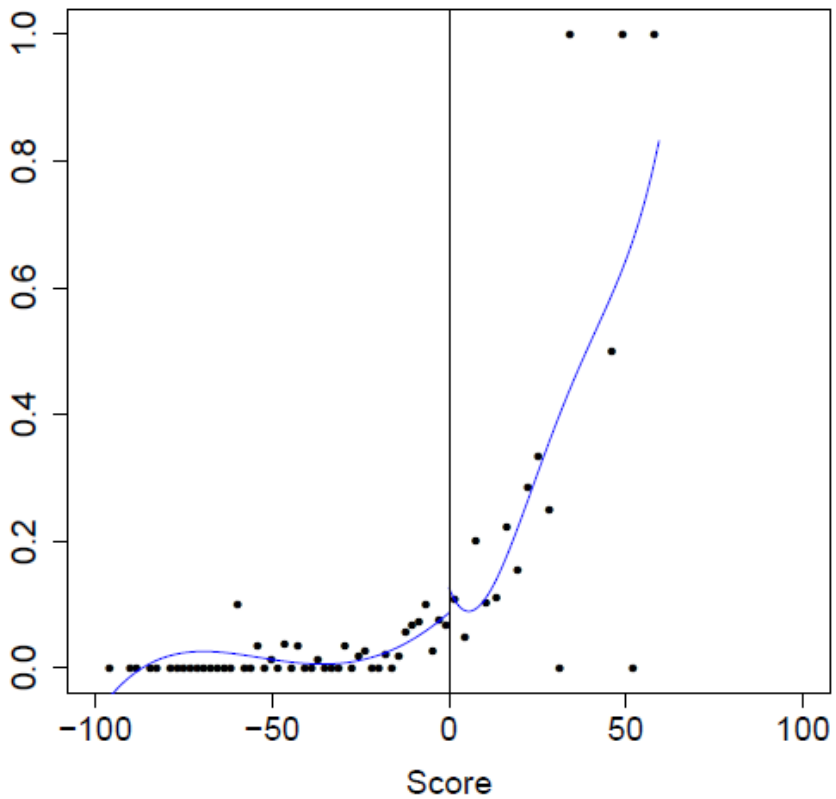


(b) Number of Parties Receiving Votes in 1994

Predetermined Covariates II



(c) Islamic Vote Percentage in 1994

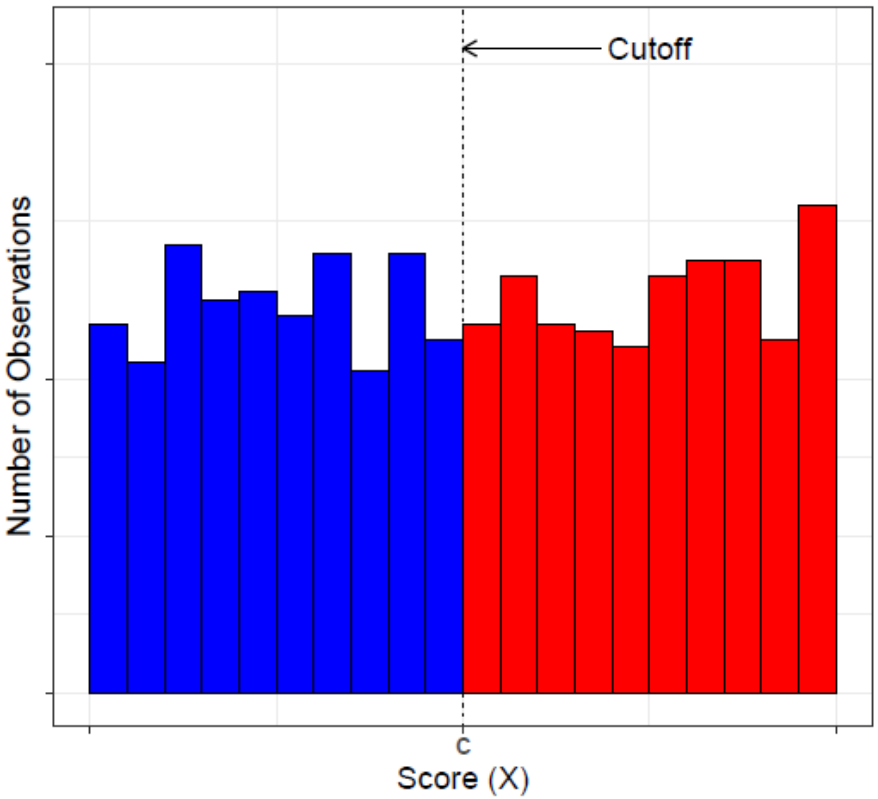


(d) Islamic Mayor in 1989

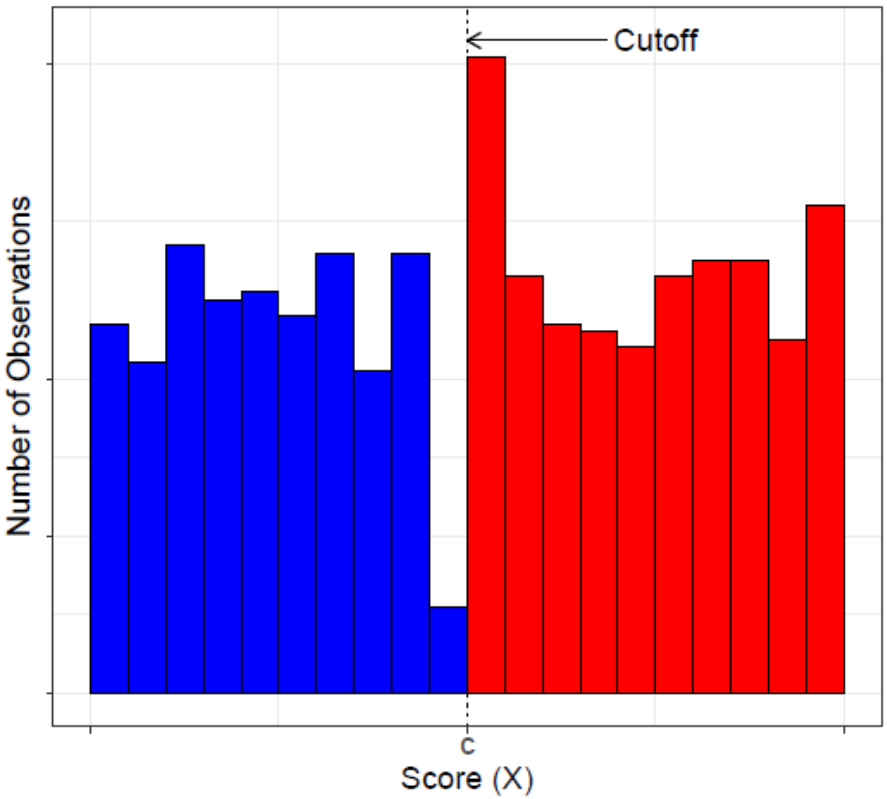
Formal Continuity-Based Analysis for Covariates

Variable	MSE-Optimal Bandwidth	RD Estimator	Robust Inference p-value	Robust Inference Conf. Int.	Eff. Number Observations
Percentage of men aged 15-20 with high school education	12.055	1.561	0.358	[-1.757, 4.862]	590
Islamic Mayor in 1989	11.782	0.053	0.333	[-0.077, 0.228]	418
Islamic percentage of votes in 1994	13.940	0.603	0.711	[-2.794, 4.095]	668
Number of parties receiving votes 1994	12.166	-0.168	0.668	[-1.357, 0.869]	596
Log population in 1994	13.319	0.012	0.999	[-0.644, 0.645]	633
District center	13.033	-0.067	0.462	[-0.285, 0.130]	624
Province center	11.556	0.029	0.609	[-0.064, 0.109]	574
Sub-metro center	10.360	-0.016	0.572	[-0.114, 0.063]	513
Metro center	13.621	0.008	0.723	[-0.047, 0.068]	642

Two Hypothetical RD Examples

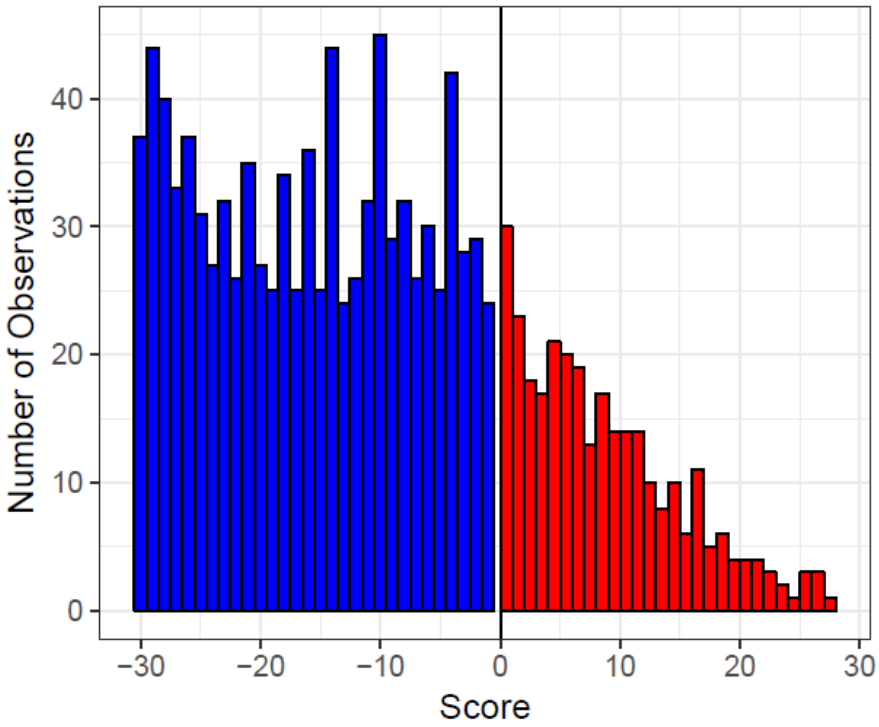


(a) No Sorting

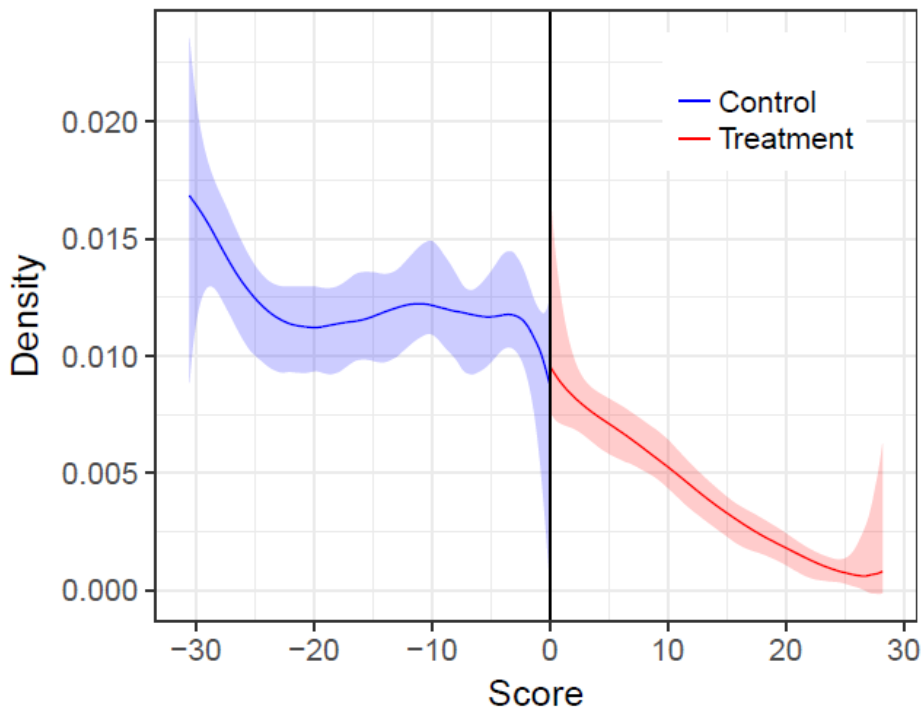


(b) Sorting

Histogram and Estimated Density of the Score



(a) Histogram



(b) Estimated Density

Density of Running Variable

out = rddensity(X); summary(out)

```
RD Manipulation Test using local polynomial density estimation.

Number of obs =       2629
Model =               unrestricted
Kernel =              triangular
BW method =           comb
VCE method =          jackknife

Cutoff c = 0          Left of c          Right of c
Number of obs         2314               315
Eff. Number of obs    965                301
Order est. (p)        2                  2
Order bias (q)        3                  3
BW est. (h)           30.54              28.285

Method                T                  P > |T|
Robust                -1.394             0.1633
```

$|X_i| < 0.3$ are excluded

out = rdrobust(Y[abs(X) >= 0.3], X[abs(X) >= 0.3])

```
Number of Obs.                 2616
BW type                       mserd
Kernel                    Triangular
VCE method                        NN

Number of Obs.                2307         309
Eff. Number of Obs.            482         248
Order est. (p)                   1           1
Order bias  (p)                  2           2
BW est. (h)                 16.043      16.043
BW bias (b)                 27.520      27.520
rho (h/b)                    0.583       0.583

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional     3.414     1.517     2.251     0.024     [0.441 , 6.387]
        Robust         -         -     1.923     0.055     [-0.067 , 6.965]
```

Donut-Hole Approach

Donut-Hole Radius	MSE-Optimal Bandwidth	RD Estimator	Robust Inference p-value	Robust Inference Conf. Int.	Number of Observations	Excluded Obs. Left	Excluded Obs. Right
0.00	17.239	3.020	0.076	[-0.309, 6.276]	795	0	0
0.10	17.954	3.081	0.064	[-0.175, 6.298]	815	1	1
0.20	16.621	3.337	0.052	[-0.033, 6.759]	765	5	4
0.30	16.043	3.414	0.055	[-0.067, 6.965]	730	7	6
0.40	17.164	3.286	0.050	[-0.001, 6.601]	774	9	9
0.50	15.422	3.745	0.028	[0.408, 7.292]	697	13	14

Fuzzy RD

$$P(D_i = 1|x_i) = \begin{cases} g_1(x_i) & if \quad x_i \geq x_0 \\ g_0(x_i) & if \quad x_i < x_0 \end{cases}$$

$$g_1(x_0) \neq g_0(x_0)$$

$$E[D_i|x_i] = P(D_i = 1|x_i)$$

$$g_0(x_i) + [g_1(x_i) - g_0(x_i)]T_i$$

$$T_i = 1\{x_i \geq x_0\}$$

Fuzzy RD is IV

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + ... + \beta_p x_i^p + \rho D_i + \eta_i$$

$$D_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + ... + \gamma_p x_i^p + \pi T_i + \xi_{1i}$$

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

IVs: $T_i, T_i\tilde{x}_i, T_i\tilde{x}_i^2, ..., T_i\tilde{x}_i^p$

The Nonparametric Version of Fuzzy RD

$$E[y_i|x_0 \leq x_i < x_0 + \Delta] - E[y_i|x_0 - \Delta \leq x_i < x_0] \cong \rho\pi$$

$$E[D_i|x_0 \leq x_i < x_0 + \Delta] - E[D_i|x_0 - \Delta \leq x_i < x_0] \cong \pi$$

$$\lim_{\Delta \to 0} \frac{E[y_i|x_0 \leq x_i < x_0+\Delta] - E[y_i|x_0-\Delta \leq x_i < x_0]}{E[D_i|x_0 \leq x_i < x_0+\Delta] - E[D_i|x_0-\Delta \leq x_i < x_0]} = \rho$$

Angrist and Lavy (1999)

$$Y_{isc} = \alpha_0 + \alpha_1 d_s + \beta_1 e_s + \beta_2 e_s^2 + ... + \beta_p e_s^p + \rho n_{sc} + \eta_{isc}$$

Y_{isc}: student i's test score in school s and class c

n_{sc}: size of class

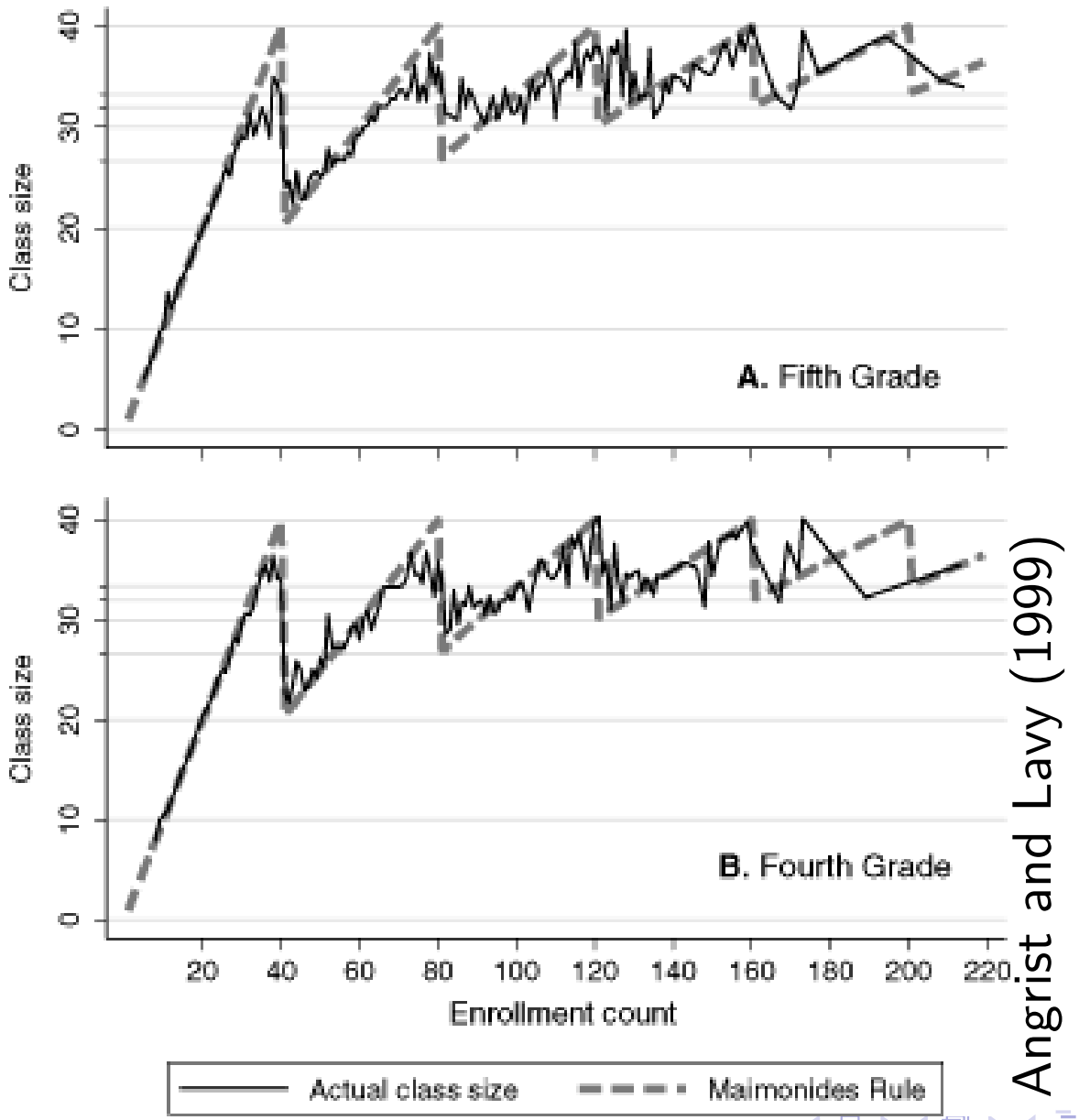
e_s: enrollment

d_s : proportion of disadvantaged students

m_{sc}: Maimonides' rule

$$m_{sc} = \frac{e_s}{int[\frac{(e_s - 1)}{40}] + 1}$$

Maimonides' Rule



Angrist and Lavy (1999)

OLS and Fuzzy RD Estimates of the Effect 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
(SD)		(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 (enrollment 38–43)								−12.6
								(3.80)
Segment 2 (enrollment 78–83)								−2.89
								(2.41)
R^2	.048	.249	.252					
Number of classes		2,018		2,018		471		302

Angrist and Lavy (1999)