Paper for Presentation next week: For next week, I will ask you to present: Black, S., (1999), Do Better Schools Matter? Parental Valuation of Elementary Education, Quarterly Journal of Economics 114, 577-599. Pay close attention not only to the econometrics of the paper, but also to the theory that would allow us to interpret the results.

Intro Comments:

We are going to do Regression Discontinuity today. A lot of the papers on the econometrics are fairly straightforward, the real problem is how to run the estimates.

Regression Discontinuity Theory

Let's start this off with pictures (best way this is done in the literature).

Suppose we are looking at the effect of mergers on prices (my own pet project). We have two treatments $T_{it} \in \{0, 1\}$ where $T_{it} = 1$ is merge and $T_{it} = 0$ is not merge, and the FTC / DOJ uses the following rule to assign treatment of allowing or blocking the merger:

$$T_{it} = \begin{cases} 1 \text{if} C_{it} > \bar{C} \\ 0 \text{if} C_{it} < \bar{C} \end{cases}$$

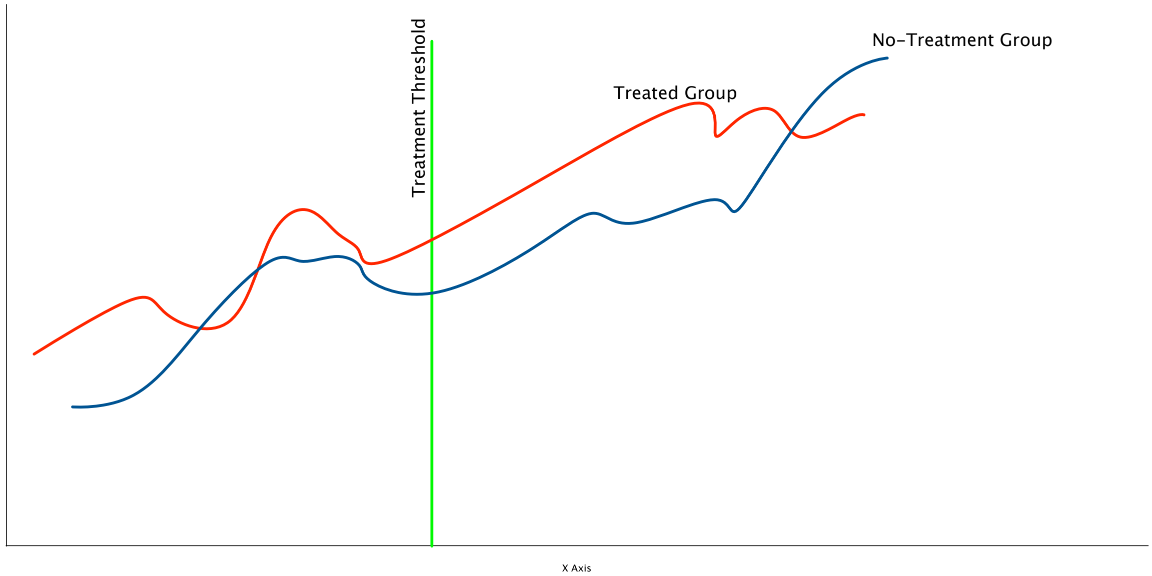
where $\bar{C}$ is the herfindahl at which the merger is challenged.

As well the effect of mergers on prices is:

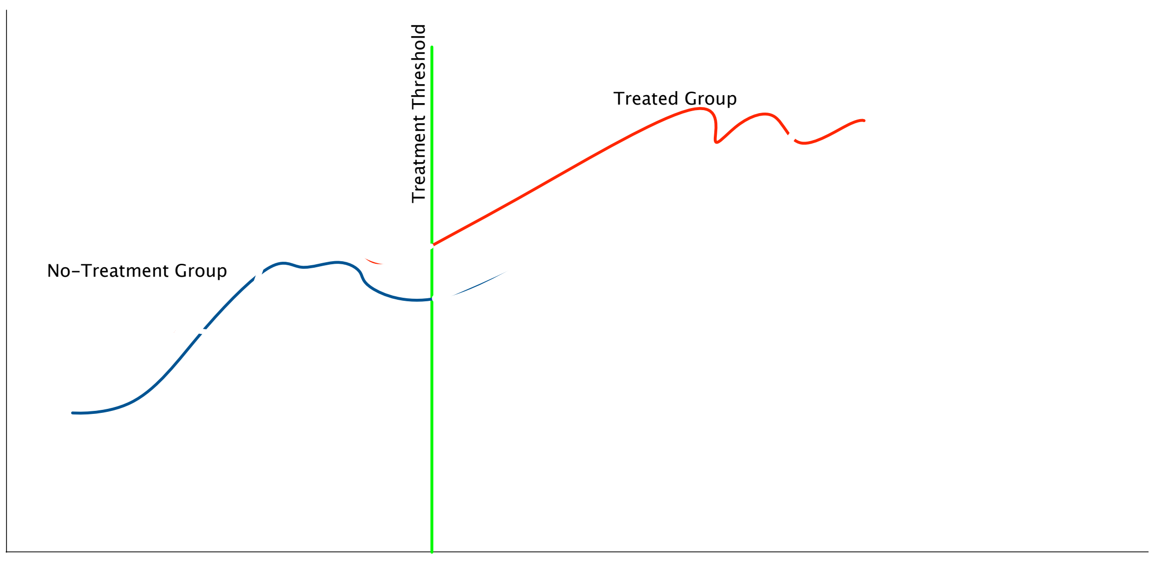
$$Y_{it}(T_{it}) = [X_{it} C_{it}]\beta + \alpha T_{it} + u_{it}$$

and let's assume that $X_{it} \perp C_{it}$ just to make things a bit easier.

The treatment effect looks like:



But we only see the following since there a threshold at which the treatment is applied:



Note the the difference between $Y(T_{it} = 1|X_{it}) - Y(T_{it} = 0|X_{it})$ can be identified at the treatment cutoff $\bar{C}$ by comparing observations very close to each side of the cutoff. Moreover, there may be other variables that also change near the cutoff that generate a problem as well, but as long as they don't change discontinuously, we will be okay (asymptotically).

Now let's look at this in a bit more detail. There are two types of RD designs:

1. Sharp RD.

2. Fuzzy RD.

Sharp RD

Suppose the treatment is assigned as $T_{it} = 1(C_{it} > \bar{C})$, which is know as sharp RD. Then the MTE (treatment effect at the cutoff) is:

$$\begin{aligned} \lim_{c \uparrow \bar{C}} E[Y_{it}|X_{it} = x] - \lim_{c \downarrow \bar{C}} E[Y_{it}|X_{it} = x] = \\ \lim_{c \uparrow \bar{C}} E[Y(1)|X_{it} = x] - \lim_{c \downarrow \bar{C}} E[Y(0)|X_{it} = x] = \\ E[Y(1) - Y(0)|X_{it} = x] = \tau^{SRD}(X_{it}) \end{aligned} \tag{1}$$

In order to use the limit result above we need to assume continuity of the underlying conditional expectations:

Assumptions: Continuity of Conditional Expectations: Assume that $E[Y_{it}|X_{it} = x, T_{it}]$ is continuous in x.

Fuzzy RD

Now let's assume instead that there is a discontinuity in the probability of treatment at a cutoff value:

$$\lim_{c\uparrow C} \Pr[T_{it} = 1|X_{it} = x] \neq \lim_{c\downarrow C} \Pr[T_{it} = 1|X_{it} = x]$$

So there is at least one factor where the probability changes discontinuously at the cutoff C. We need the following assumption to make this work:

Assumptions: Continuity of Conditional Expectations: Assume that $E[Y_{it}|X_{it} = x, T_{it}]$ is continuous in x.

So the treatment effect is:

$$\frac{\lim_{c\uparrow \bar{C}} E[Y_{it}|X_{it} = x] - \lim_{c\downarrow \bar{C}} E[Y_{it}|X_{it} = x]}{\lim_{c\uparrow \bar{C}} E[T_{it} = 1|X_{it} = x] - \lim_{c\downarrow \bar{C}} E[T_{it} = 0|X_{it} = x]}$$
$$= E[Y(1) - Y(0)|X_{it} = x, c = C] = \tau^{FRD}(X_{it}) \quad (2)$$

Lee on House Election Rules

Question: What is the advantage of incumbency in House Elections?

The problem with the regression:

$$Y_{it} = \alpha T_{it} + X_{it}\beta + \epsilon_{it}$$

where Y_{it} is the probability of winning the election, X_{it} are characteristics of the county and of the politicians running for office, and T_{it} is an indicator for having won the election in the last electoral cycle. The issue is that there are factors that effect the probability of winning the election, like having a famous candidate, or being in a district where people drive their Prius'es to pick up food at the farmer's market, or drive pick-up trucks to the NASCAR race. So the assumption of unconfoundness, i.e. $T_{it} \perp \epsilon_{it} | X_{it}$ is hard to swallow without some very high quality X_{it}'s. Since we expect $E[T_{it}\epsilon_{it}] > 0$ things that made the candidate win in the last election are correlated with things that make the candidate win the current election (like a panel data problem). So we would overestimate the advantages of incumbency.

Okay, here is the paper in pictures (the rest is unimportant):

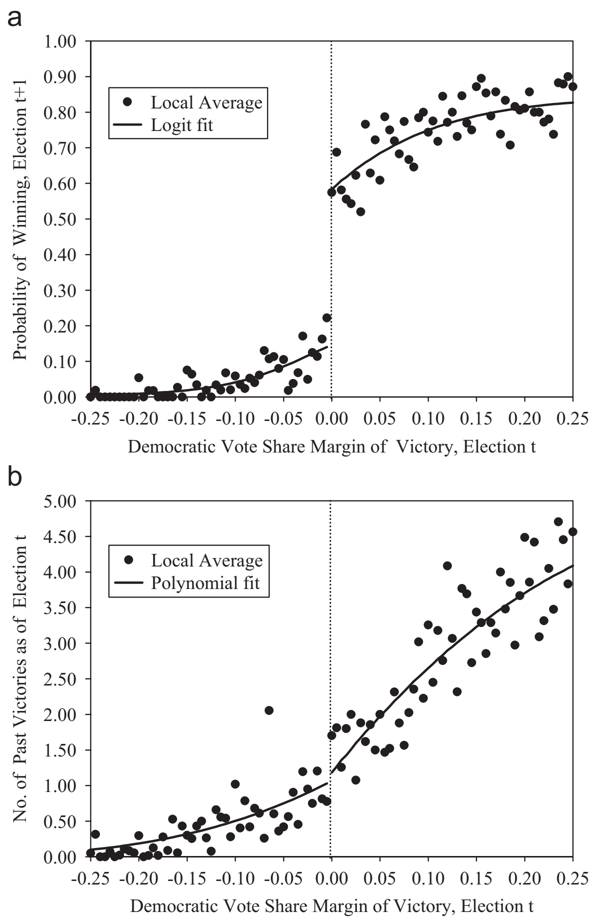


Fig. 2. (a) Candidate's probability of winning election $t+1$, by margin of victory in election t: local averages and parametric fit. (b) Candidate's accumulated number of past election victories, by margin of victory in election t: local averages and parametric fit.

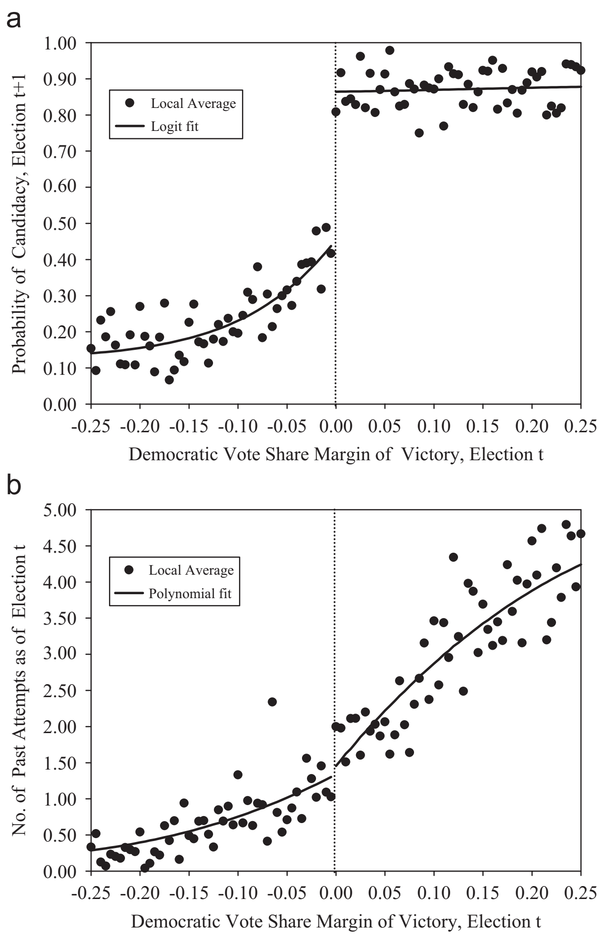


Fig. 3. (a) Candidate's probability of candidacy in election $t+1$, by margin of victory in election t: local averages and parametric fit. (b) Candidate's accumulated number of past election attempts, by margin of victory in election t: local averages and parametric fit.

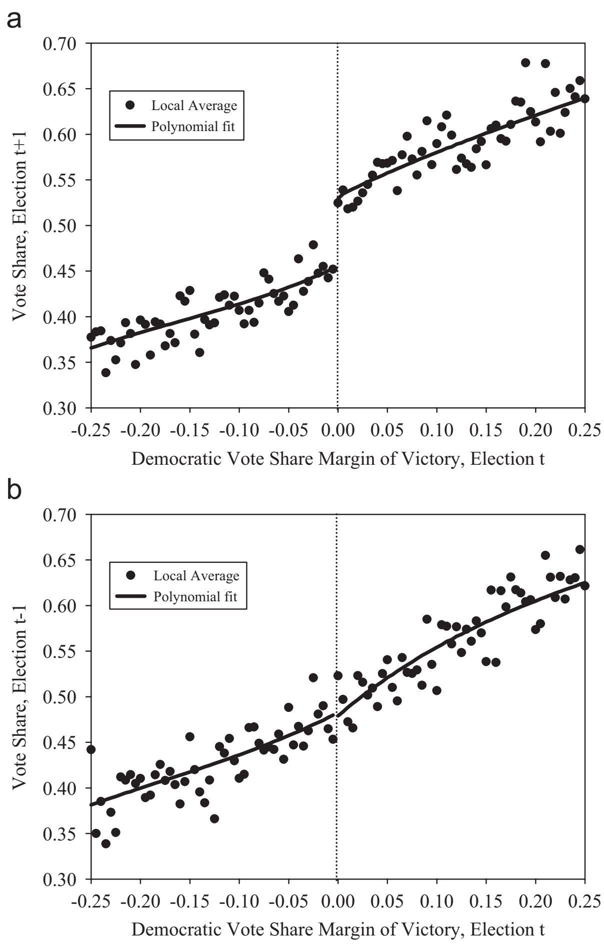


Fig. 4. (a) Democrat party's vote share in election $t+1$, by margin of victory in election t: local averages and parametric fit. (b) Democratic party vote share in election $t-1$, by margin of victory in election t: local averages and parametric fit.

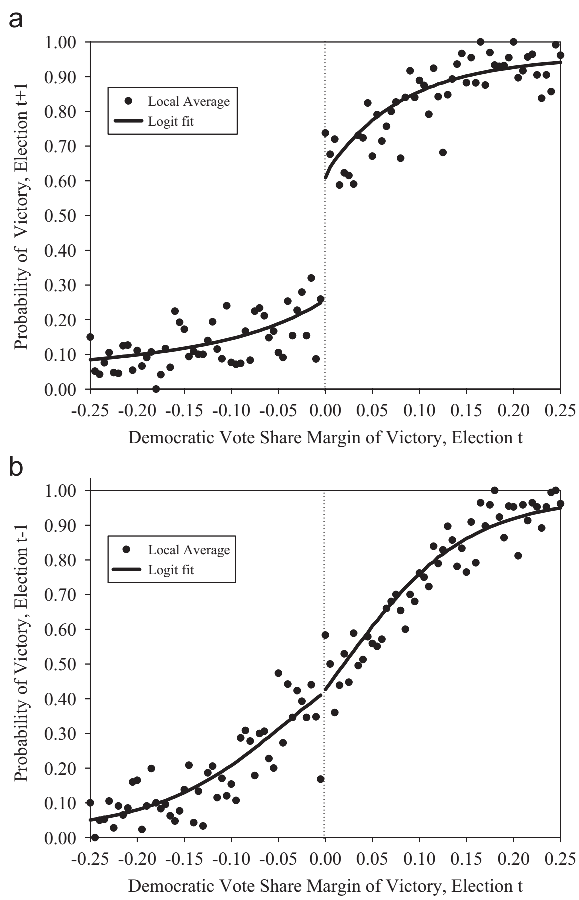


Fig. 5. (a) Democratic party probability victory in election $t+1$, by margin of victory in election t: local averages and parametric fit. (b) Democratic probability of victory in election $t-1$, by margin of victory in election t: local averages and parametric fit.

So the results in the paper are that we still find a large incumbency effect, though smaller than with naive estimator.

There are 2 ways to estimate this effect, using either a kernel regression on the right and left side of the cutoff point, or some local regression on both sides of the cutoff:

1. Kernel Regression

$$MTE = \frac{1}{N_{c^-}} \sum_{\bar{C}-\delta<c<\bar{C}} y_{it} - \frac{1}{N_{c^+}} \sum_{\bar{C}+\delta>c>\bar{C}} y_{it}$$

where we need to choose a bandwidth δ.

2. Bin Regressions:

$$N_k = \sum_{i=1}^{N} 1(x_{it} \in C(k))$$

$$\widehat{y}_{it} = \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i=1}^{N} y_{it} 1(x_{it} \in C(k))$$

where we need to choose a bandwidth δ and we look for a discontinuity in the bin average near $x = c$.

3. Local Regression

$$y_{it} = X_{it}\beta + \alpha 1(c > \bar{C}) + \epsilon_{it}$$

or

$$y_{it} = X_{it}\beta^{+}1(c > \bar{C}) + X_{it}\beta^{-}1(c < \bar{C}) + \alpha 1(c > \bar{C}) + \epsilon_{it}$$

4. Fuzzy RD Local Regression using squared series approximation

$$\left(\widehat{\alpha}_y^-, \widehat{\beta}_y^-\right) = \text{argmin} \sum_{\bar{C}-\delta<c<\bar{C}} \left(y_{it} - \alpha_y^- - \beta_y^-(X_i - c)\right)^2$$

$$\left(\widehat{\alpha}_y^+, \widehat{\beta}_y^+\right) = \text{argmin} \sum_{\bar{C}+\delta>c>\bar{C}} \left(y_{it} - \alpha_y^+ - \beta_y^+(X_i - c)\right)^2$$

and likewise for the probability of treatment:

$$\left(\widehat{\alpha}_T^-, \widehat{\beta}_T^-\right) = \text{argmin} \sum_{\bar{C}-\delta_T<c<\bar{C}} \left(T_{it} - \alpha_T^- - \beta_T^-(X_i - c)\right)^2$$

$$\left(\widehat{\alpha}_T^+, \widehat{\beta}_T^+\right) = \text{argmin} \sum_{\bar{C}+\delta_T>c>\bar{C}} \left(y_{it} - \alpha_T^+ - \beta_T^+(X_i - c)\right)^2$$

So the MTE for the fuzzy design is:

$$\tau^{FRD} = \frac{\widehat{\alpha}_y^+ - \widehat{\alpha}_y^-}{\widehat{\alpha}_T^+ - \widehat{\alpha}_T^-}$$

Note that there is a key question of bandwidth choice δ, δ_T given that there is a bias versus variance tradeoff that needs to be made.

Issues

- While mergers rules state Herfindahl guidelines for challenging mergers, these guidelines are loose in the sense that there are other factors that guide the DOJ/FTC in blocking mergers. I believe that most policies, when examined carefully, have this type of discretion embedded into them. So we need bureaucratic rules which are enforced exactly to make RD work.

- Sometimes we don't know what the cutoff is.

- You need a bunch of data to estimate the effects at the thresholds.

The next graph show the delinquency rate for mortgages with low documentation. The thing that you need to know is that above a FICO score of 620, the loans can be resold to Fannie May and Freddy Mac (government purchasers in the mortgage market). This means that you can identify the effect of moral hazard (the effect of banks having to check an applicant versus not) on delinquency rates.

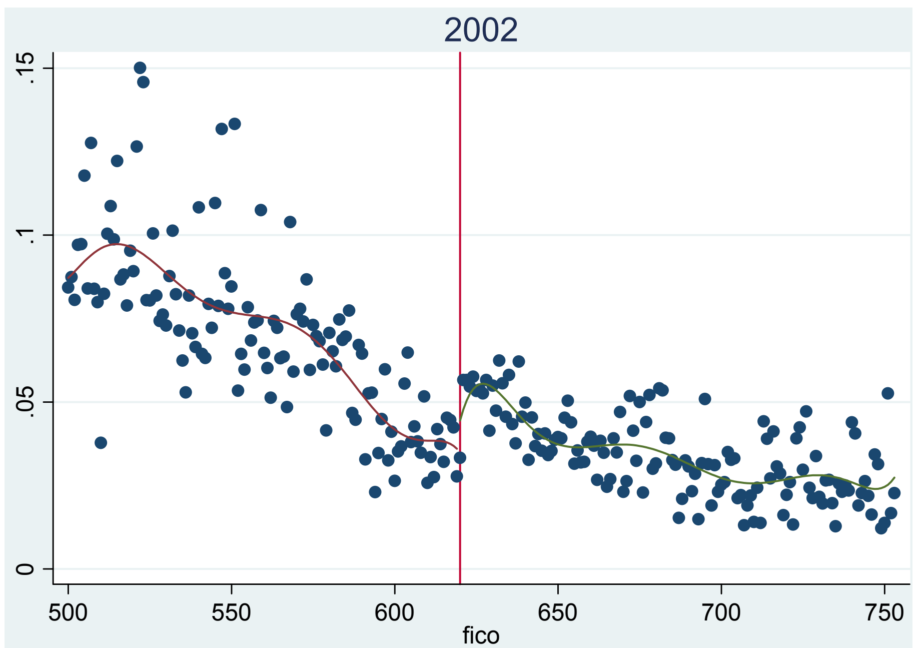


Figure 5B: Annual Delinquencies for Low Documentation Loans in 2002

Figure 5B presents the data for actual percent of low documentation loans that became delinquent in 2002. We plot the dollar weighted fraction of the pool that becomes delinquent for one-point FICO bins between score of 500 and 750. The vertical line denotes the 620 cutoff, and a seventh order polynomial is fit to the data on either side of the threshold. Delinquencies are reported between 10-15 months for loans originated in the year.

Angrist and Lavy on Maimonides's Rule

Question: What is the effect of class size on student achievement? Note that the key question in education economics is the returns of inputs in the education production function, so we need to have some idea if educational investments are worth the cost.

The problem is that class size is correlated with other things, and in Israel the problem is that large classes are correlated with large urban areas with better schools on average.

Idea behind Maimonides Rule (from the Talmudic Scholar Moses Maimonides, one of the more famous ones),

Limit the class size at 40. This means that the average class size at a school with enrollment of 76 has a class size of 37, while a school with 84 students has a class size of 28. So we get plausibly exogenous variation in class size for schools with enrollment near 40, 80, 120 and so on. This is what we are going to use.

TABLE I
UNWEIGHTED DESCRIPTIVE STATISTICS

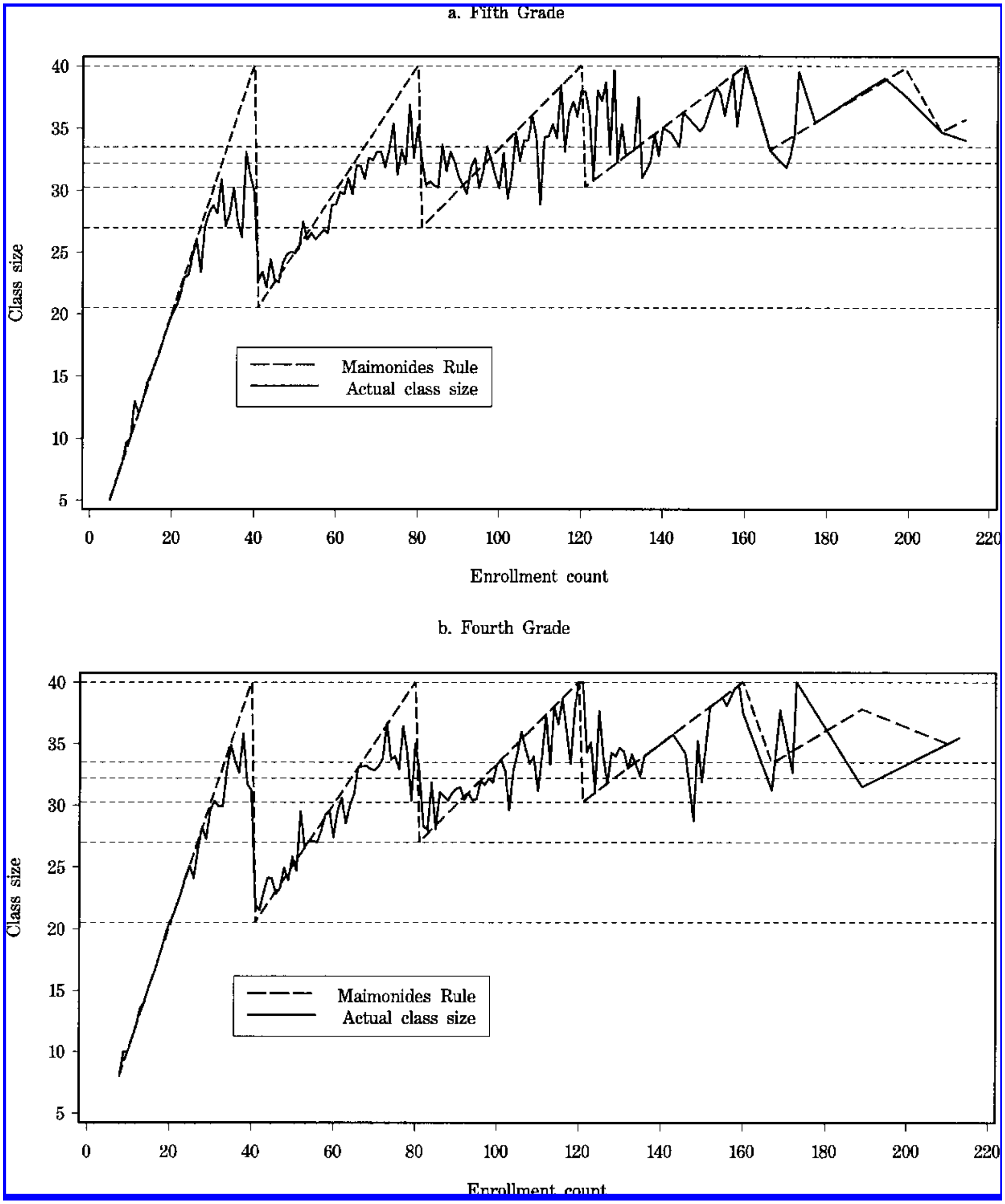
Variable	Mean	S.D.	Quantiles 0.10	0.25	0.50	0.75	0.90
A. Full sample							
5th grade (2019 classes, 1002 schools, tested in 1991)							
Class size	29.9	6.5	21	26	31	35	38
Enrollment	77.7	38.8	31	50	72	100	128
Percent disadvantaged	14.1	13.5	2	4	10	20	35
Reading size	27.3	6.6	19	23	28	32	36
Math size	27.7	6.6	19	23	28	33	36
Average verbal	74.4	7.7	64.2	69.9	75.4	79.8	83.3
Average math	67.3	9.6	54.8	61.1	67.8	74.1	79.4
4th grade (2049 classes, 1013 schools, tested in 1991)							
Class size	30.3	6.3	22	26	31	35	38
Enrollment	78.3	37.7	30	51	74	101	127
Percent disadvantaged	13.8	13.4	2	4	9	19	35
Reading size	27.7	6.5	19	24	28	32	36
Math size	28.1	6.5	19	24	29	33	36
Average verbal	72.5	8.0	62.1	67.7	73.3	78.2	82.0
Average math	68.9	8.8	57.5	63.6	69.3	75.0	79.4
3rd grade (2111 classes, 1011 schools, tested in 1992)							
Class size	30.5	6.2	22	26	31	35	38
Enrollment	79.6	37.3	34	52	74	104	129
Percent disadvantaged	13.8	13.4	2	4	9	19	35
Reading size	24.5	5.4	17	21	25	29	31
Math size	24.7	5.4	18	21	25	29	31
Average verbal	86.3	6.1	78.4	83.0	87.2	90.7	93.1
Average math	84.1	6.8	75.0	80.2	84.7	89.0	91.9

B. +/− 5 Discontinuity sample (enrollment 36–45, 76–85, 116–124)

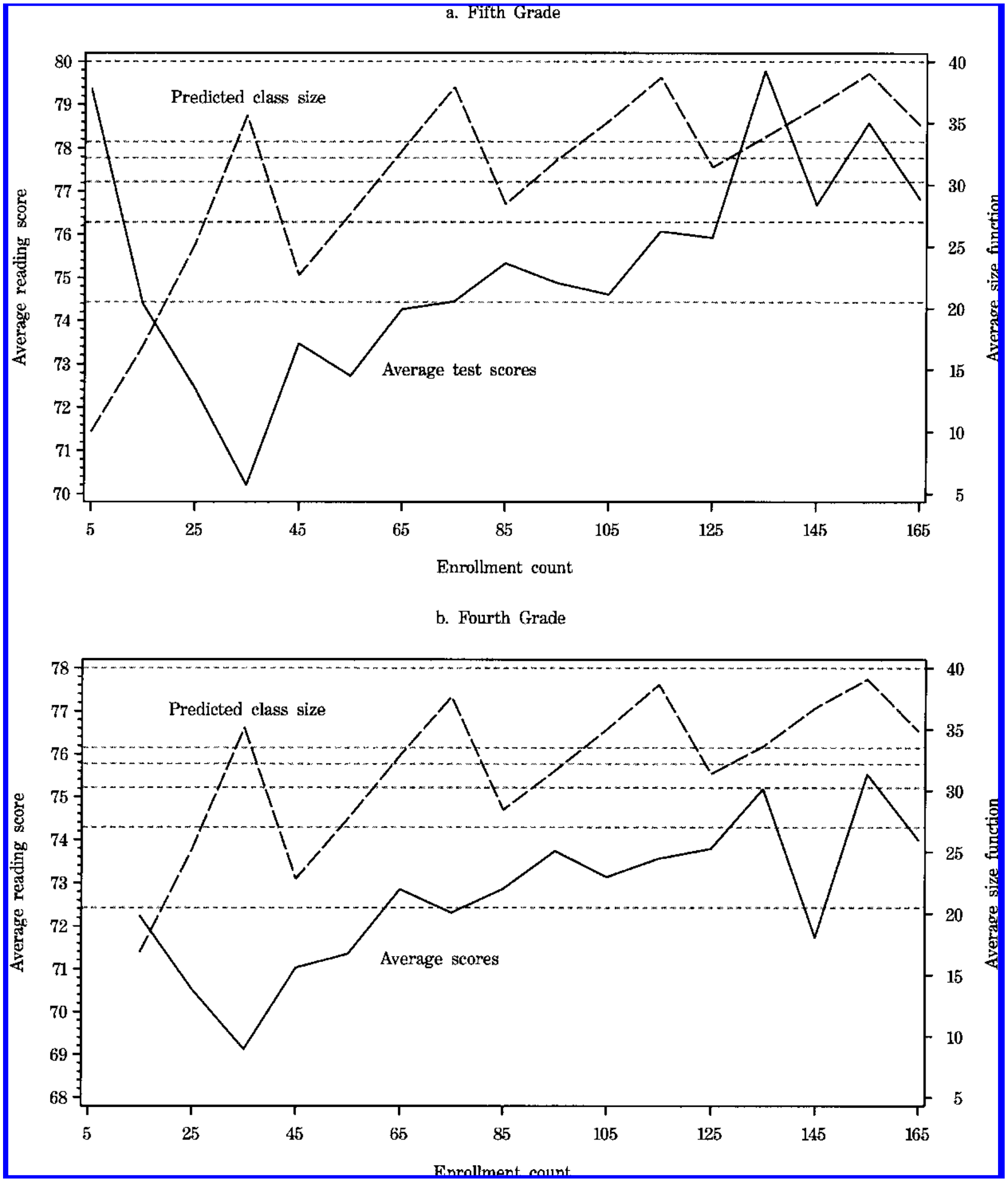
	5th grade		4th grade		3rd grade	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
	(471 classes, 224 schools)		(415 classes, 195 schools		(441 classes, 206 schools)	
Class size	30.8	7.4	31.1	7.2	30.6	7.4
Enrollment	76.4	29.5	78.5	30.0	75.7	28.2
Percent disadvantaged	13.6	13.2	12.9	12.3	14.5	14.6
Reading size	28.1	7.3	28.3	7.7	24.6	6.2
Math size	28.5	7.4	28.7	7.7	24.8	6.3
Average verbal	74.5	8.2	72.5	7.8	86.2	6.3
Average math	67.0	10.2	68.7	9.1	84.2	7.0

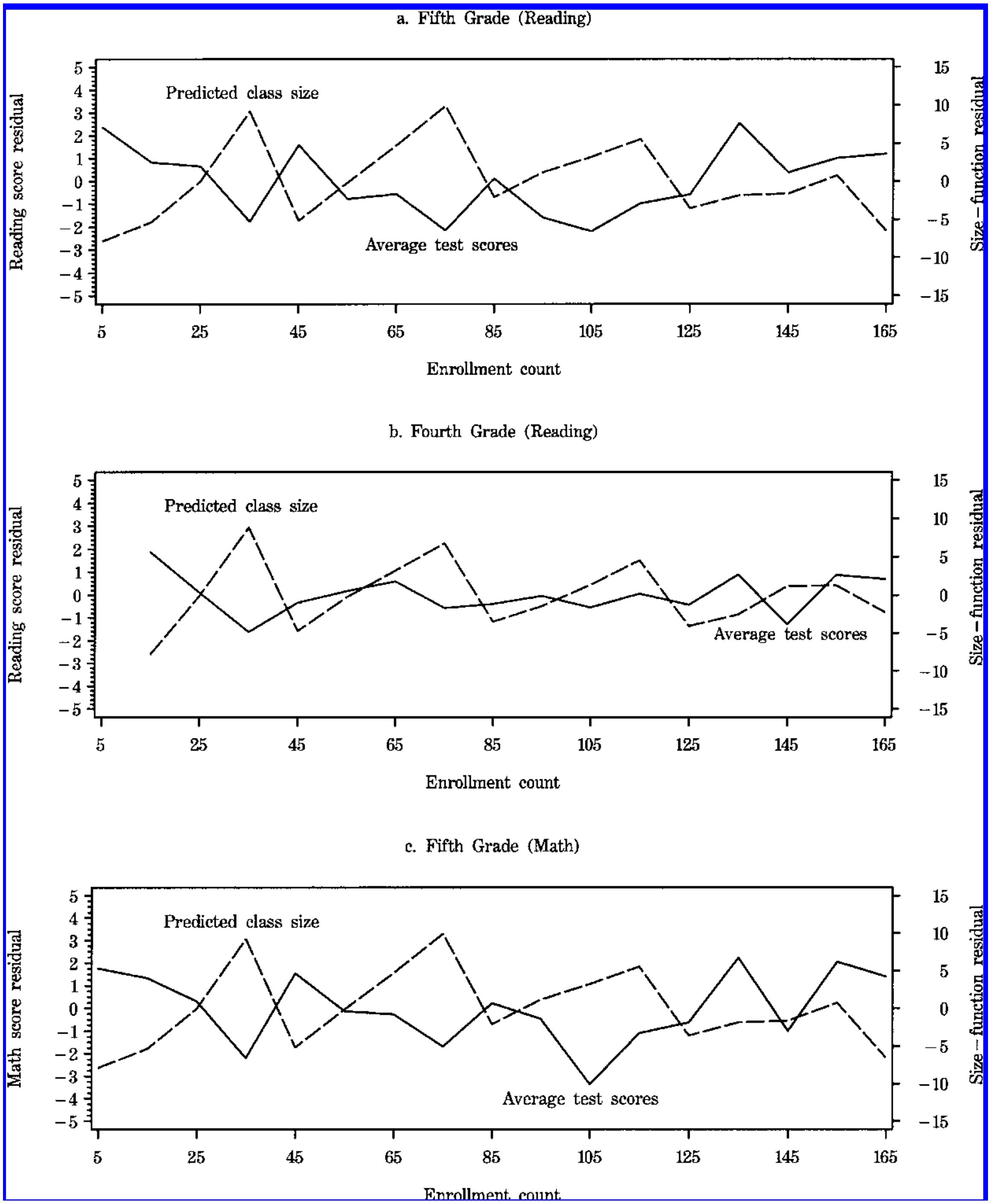
Variable definitions are as follows: Class size = number of students in class in the spring, Enrollment = September grade enrollment, Percent disadvantaged = percent of students in the school from "disadvantaged backgrounds," Reading size = number of students who took the reading test, Math size = number of students who took the math test, Average verbal = average composite reading score in the class, Average math = average composite math score in the class.

The next graph shows the effect of Maimonides rule on class size.



And here are some summary graphs on the effect of class size on educational outcomes. Note that the test scores are increasing with test scores. This effect is mainly due to the changing demographics in larger school districts.




a. Fifth Grade (Reading)
Predicted class size
Average test scores
Reading score residual
Size – function residual
Enrollment count
b. Fourth Grade (Reading)
Predicted class size
Average test scores
Reading score residual
Size – function residual
Enrollment count
c. Fifth Grade (Math)
Predicted class size
Average test scores
Math score residual
Size – function residual
Enrollment count

Now for the IV treatment of RD design. The regression is:

$$y_{it} = \alpha nc_{it} + X_{it}\beta + \epsilon_{it}$$

But we worry that $E[nc_{it}\epsilon_{it}] > 0$ due to the effect of changing demographics in larger schools. As well there are other factors, in particular the PD variable (Percent Disadvantaged) that changes test scores quite a bit.

The instrument is therefore:

$$\widehat{nc}_{it} = mod(e_{it}, 40)$$

which is correlated with nc_{it} (the first stage of $nc_{it} = \gamma_0 + \gamma_1 mod(e_{it}, 40)$, but plausibly uncorrelated with ϵ_{it}.

First the reduced form evidence

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.

The the IV regressions:

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.

TABLE V

2SLS ESTIMATES FOR 1991 (FOURTH 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		72.5			72.5			67.3			68.7	
(s.d.)		(8.0)			(7.8)			(9.6)			(9.1)	
Regressors												
Class size	−.110	−.133	−.074	−.147	−.098	−.150	.049	−.050	−.033	−.098	.095	.023
	(.040)	(.059)	(.067)	(.084)	(.090)	(.128)	(.048)	(.070)	(.081)	(.092)	(.114)	(.160)
Percent disadvantaged	−.346	−.345	−.346		−.354	−.347	−.290	−.284	−.284		−.299	−.290
	(.014)	(.014)	(.014)		(.034)	(.034)	(.017)	(.017)	(.017)		(.042)	(.043)
Enrollment		.005	−.040			.017		−.020	.007			.023
		(.008)	(.024)			(.022)		(.010)	(.029)			(.028)
Enrollment squared/100			.021						.006			
			(.011)						(.014)			
Piecewise linear trend				.100						.130		
				(.026)						(.028)		
Root MSE	6.65	6.66	6.63	8.02	6.64	6.69	7.82	7.82	7.82	8.65	8.23	8.24
N		2049		2001	415			2049		2001	415	

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 2SLS estimates use f_{sc} as an instrument for class size.

TABLE VI

DUMMY-INSTRUMENT RESULTS FOR DISCONTINUITY SAMPLES

	5th grade						4th grade					
	Reading comprehension			Math			Reading comprehension			Math		
	+/− 5 Sample	+/− 3 Sample		+/− 5 Sample	+/− 3 Sample		+/− 5 Sample	+/− 3 Sample		+/− 5 Sample	+/− 3 Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Regressors												
Class size	−.687	−.588	−.451	−.596	−.395	−.270	−.175	−.234	−.380	.018	−.118	−.247
	(.197)	(.198)	(.236)	(.254)	(.254)	(.281)	(.130)	(.157)	(.205)	(.162)	(.202)	(.234)
Percent dis-	−.464	−.452		−.433	−.416		−.350	−.372		−.291	−.323	
advantaged	(.039)	(.045)		(.050)	(.058)		(.034)	(.043)		(.043)	(.055)	
Segment 1	−5.09	−4.54	−10.7	−7.54	−6.94	−12.6	−1.62	−2.67	−6.94	−1.89	−3.57	−7.31
(enrollment 36–45)	(2.40)	(2.59	(3.19)	(3.07)	(3.34)	(3.80)	(1.77)	(2.23)	(2.90)	(2.21)	(2.87)	(3.31)
Segment 2	−1.64	−2.18	−2.96	−1.57	−2.17	−2.89	−1.52	−2.16	−3.83	−1.15	−2.50	−3.96
(enrollment 76–85)	(1.41)	(1.64)	(2.00)	(1.83)	(2.14)	(2.41)	(1.24)	(1.59)	(2.10)	(1.56)	(2.07)	(2.39)
Root MSE	7.46	7.24	8.67	9.41	9.14	10.2	6.72	6.70	8.30	8.25	8.53	9.52
N	471	302		471	302		415	265		415	265	

The table reports results from a sample of classes in schools with enrollment close to points of discontinuity. 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 $1[f_{sc} < 32]$ and interactions with dummies for enrollment segments as instruments for class size. Since there are three segments, there are three instruments. The models include dummies for the first two segments to control for segment main effects.