

PERFORMANCE PAY AND TEACHERS' EFFORT, PRODUCTIVITY AND GRADING ETHICS

The Online Appendix

August 2008

Difference in Differences Estimates Using the Sample of all Eligible Schools

The advantage of using the RT and the RD samples over the sample of all the eligible schools is the almost perfect balancing that they produce in pre-program characteristics and outcomes. Contrary to this striking resemblance, we see huge treatment-control imbalances in the eligible schools sample in Table 1. However, since some of these differences were of similar size and precision in 2000 and in 2001, allowing for school fixed effects may partly solve the problem. However, there are several concerns remaining that may bias the difference in differences estimates. A first concern is that very large differences in observables may be paralleled in control-treatment differences in time-varying unobservables that might confound the treatment effect as well. Secondly, as many more of the control schools are at the upper part of the distribution of the average pre-program outcomes', mean reversion may upward bias the treatment-effect estimates. Thirdly, including the entire eligible group in the sample implies that the studied sample includes now many more treatment schools with high pre-program average matriculation rates and also a few more schools with low matriculation rates. However the results from quartile analysis imply that schools with higher matriculation rates have smaller treatment effects. Therefore including more high-matriculation schools should lower the estimated average treatment effect but including more low matriculation schools should increase the estimated average treatment. The net effect of these two offsetting changes on the treatment estimates cannot be predicted a priori. All these factors might affect the differences in differences estimates based on the eligible sample but the direction of the bias is ambiguous.¹

Table A2 in this online Appendix presents the results from estimating equations [1] and [2] based on the sample of all eligible schools. All the treatment effect estimates in columns 1 and 7 are positive and significantly different from zero. Adding individual and school controls however lowers most of the point estimates, which is as expected given the much lower background characteristics of the treated

¹ Chay, McEwan and Urquiola (2005) illustrate that measurement error in test score would tend to inflate estimates of treatment effects which would have caused the dif-in-dif estimates to be significantly larger. However, the measurement error discussed in this paper is not the same as the transitorily high or low true test score that is

group. The decline is much larger for the math estimates and as a result the estimated effect on pass rate and the average test score is no longer significantly different from zero. Comparing these DID estimates to the RT estimates shows that the former are much lower in math and not very different in English. Another demonstration of the bias of the difference in difference estimates is reflected in the very different pattern that is revealed regarding the effect of treatment by quartiles. For example, in English the difference in difference estimates are positive and significant in the 4th quartile for all three outcomes while in Table 4 the respective estimates are practically zero.

References

Chay, Kenneth Y. Patrick J. McEwan and Miguel Urquiola, 2005, “The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools”, *The American Economic Review* Vol. 95, No. 4, September 2005, 1237-1258.

considered in Chay et. al. and so experiencing a large negative measurement error, and therefore getting assigned to treatment, does not imply that there will be a mean reversion in the correct test score as shown by Chay et.al.

Table A1 - Parameter Estimates of the RT DID Regressions of Table 4
the Specification with Full Controls, Based on the Full Sample

	Math			English		
	Testing (1)	Passing (2)	Score (3)	Testing (4)	Passing (5)	Score (6)
Treated	0.041 (0.029)	0.087 (0.040)	5.307 (2.739)	0.033 (0.019)	0.039 (0.028)	2.527 (2.040)
One-year lagged school matriculation rate	-0.087 (0.115)	0.093 (0.225)	-3.582 (11.024)	-0.072 (0.095)	0.017 (0.117)	-4.648 (9.109)
Two-years lagged school matriculation rate	-0.007 (0.175)	-0.127 (0.283)	-18.139 (17.830)	-0.054 (0.142)	-0.119 (0.121)	-13.513 (11.667)
Asia-Africa ethnicity	-0.006 (0.010)	0.012 (0.019)	0.230 (1.038)	0.021 (0.011)	0.022 (0.018)	0.961 (0.806)
Immigrant	0.102 (0.033)	0.141 (0.042)	9.808 (2.326)	0.056 (0.028)	0.061 (0.029)	6.463 (1.957)
Gender (1=male)	-0.005 (0.023)	0.006 (0.020)	1.338 (2.258)	-0.025 (0.014)	-0.028 (0.009)	-0.794 (0.852)
Attempted credits	0.027 (0.008)	-0.007 (0.011)	-0.048 (0.786)	0.041 (0.007)	0.013 (0.008)	1.544 (0.451)
Average score in attempted credits	0.007 (0.001)	0.007 (0.001)	0.581 (0.061)	0.007 (0.001)	0.007 (0.001)	0.515 (0.053)
Overall awarded credits	0.004 (0.008)	0.042 (0.011)	3.253 (0.813)	-0.015 (0.007)	0.012 (0.008)	0.716 (0.460)
Awarded credits in the subject	-0.007 (0.014)	-0.078 (0.023)	-2.006 (1.558)	0.017 (0.023)	-0.070 (0.046)	-1.744 (3.631)

Notes:

1. Standard errors in parenthesis are clustered at the school level
2. Observations were weighted with frequency weights in order to have similar number of students in control and treatment schools within each group of schools with close true matriculation rate.
3. Additional categorial controls, of which estimates are not reported here, are father's and mother's education, number of brothers and school dummies.

Table A2 - DID Estimates of the Effect of Teachers' Bonuses on Math and English Outcomes Based on the Sample of all Eligible Schools

	Math						English					
	All quartiles		Estimates by quartile				All quartiles		Estimates by quartile			
	Limited Control (1)	Full Control (2)	1st (3)	2nd (4)	3rd (5)	4th (6)	Limited Control (7)	Full Control (8)	1st (9)	2nd (10)	3rd (11)	4th (12)
Testing rate												
Control group mean		0.835	0.475	0.804	0.934	0.969	0.880		0.583	0.887	0.953	0.975
Treatment effect	0.050 (0.014) [0.020]	0.028 (0.012) [0.017]	0.047 (0.035) [0.047]	0.060 (0.024) [0.034]	0.007 (0.012) [0.017]	0.010 (0.013) [0.017]	0.045 (0.010) [0.015]	0.029 (0.009) [0.013]	0.060 (0.032) [0.044]	0.042 (0.013) [0.018]	0.010 (0.008) [0.011]	0.013 (0.009) [0.012]
Pass rate												
Control group mean		0.736	0.349	0.641	0.834	0.935	0.818		0.481	0.794	0.894	0.954
Treatment effect	0.044 (0.020) [0.028]	0.019 (0.016) [0.022]	0.011 (0.028) [0.039]	0.054 (0.030) [0.041]	0.023 (0.019) [0.027]	-0.015 (0.015) [0.020]	0.058 (0.015) [0.021]	0.040 (0.014) [0.019]	0.054 (0.035) [0.048]	0.065 (0.020) [0.028]	0.003 (0.014) [0.020]	0.024 (0.012) [0.016]
Average score												
Control group mean		62.931	28.547	54.594	71.186	81.791	61.722		37.865	59.959	67.499	75.410
Treatment effect	3.089 (1.530) [2.161]	1.032 (1.272) [1.794]	0.736 (2.157) [2.972]	2.180 (2.212) [3.087]	1.188 (1.437) [2.002]	0.721 (1.386) [1.914]	4.437 (1.033) [1.457]	2.865 (0.946) [1.331]	1.870 (2.242) [3.114]	4.681 (1.495) [2.086]	1.542 (1.130) [1.563]	1.670 (0.756) [1.017]
Passing rate												
Proportion of unconditional effect		-0.005	-0.023	0.004	0.017	-0.024	0.014		0.006	0.027	-0.006	0.011
Average score												
Proportion of unconditional effect		-0.970	-1.974	-1.965	0.693	-0.102	0.861		-1.908	1.805	0.812	0.666
N		24,085	5,993	6,044	6,007	6,041	25,106		6,218	6,326	6,281	6,281

Notes:

- Standard errors in parenthesis are clustered at the school-year-combination level. Standard errors in brackets are clustered at the school level.
- In columns (3)-(6) and (9)-(12) treatment effects vary by quartiles of previous tests score distribution. The estimates are taken from a four separate regressions, one for each of the quartiles.
- Student level controls - in all columns except (1) and (7) - include a set of dummy variables for the number of siblings and father and mother education, the school's (one-year) lagged mean matriculation rate, a dummy for Asia-Africa ethnic background, immigration status, gender dummy, the number of credit units *attempted*, the average score in those attempted units, overall credit units *awarded*, and credit units awarded for the subject in question only. School fixed effects are included in each model.
- In columns (1) and (7), the controls are school fixed effects, school's one-year lagged matriculation rate, and student's attempted credit units
- Control group mean row show the mean for students in control schools in 2001

Table A3 - Teacher's Education And Demographic Characteristic and Balancing Tests Between Treatment and Control Groups

	All interviewd teachers				Eligible schools' teachers				Randomized-treatment schools' teachers			
	Math teachers		English teachers		Math teachers		English teachers		Math teachers		English teachers	
	Sample mean	Treatment-control difference	Sample mean	Treatment-control difference	Sample mean	Treatment-control difference	Sample mean	Treatment-control difference	Sample mean	Treatment-control difference	Sample mean	Treatment-control difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Teacher demographics:												
Age	44.009	0.301 (1.180)	45.309	-0.697 (1.180)	44.972	-0.663 (1.484)	44.265	0.347 (1.450)	45.825	-0.867 (1.701)	45.433	-2.150 (1.965)
Gender (Female=1)	0.598	-0.024 (0.064)	0.814	-0.005 (0.047)	0.563	0.011 (0.075)	0.794	0.014 (0.067)	0.600	0.034 (0.109)	0.833	-0.022 (0.089)
Born in Israel	0.540	-0.054 (0.070)	0.319	0.213 (0.061)	0.394	0.091 (0.093)	0.397	0.135 (0.089)	0.225	0.296 (0.113)	0.233	0.276 (0.103)
Teacher education:												
Teacher certificate	0.011	0.049 (0.018)	0.011	0.025 (0.019)	0.000	0.059 (0.016)	0.000	0.035 (0.018)	0.000	0.028 (0.019)	0.000	0.000 (0.000)
B.A in education	0.042	0.070 (0.027)	0.080	0.012 (0.034)	0.042	0.070 (0.038)	0.132	-0.040 (0.054)	0.050	0.049 (0.060)	0.233	-0.158 (0.088)
B.A	0.402	0.024 (0.053)	0.484	-0.066 (0.054)	0.394	0.032 (0.069)	0.485	-0.067 (0.083)	0.350	0.058 (0.100)	0.367	0.105 (0.141)
M.A	0.524	-0.181 (0.053)	0.410	0.023 (0.051)	0.521	-0.178 (0.072)	0.368	0.065 (0.074)	0.550	-0.170 (0.108)	0.400	0.034 (0.114)
Ph.D	0.016	0.037 (0.022)	0.011	0.011 (0.014)	0.042	0.011 (0.030)	0.015	0.007 (0.019)	0.050	0.020 (0.047)	0.000	0.019 (0.019)
Teaching experience (years)	19.079	-0.139 (1.185)	19.134	-1.375 (1.014)	20.958	-2.017 (1.467)	18.529	-0.771 (1.391)	21.050	-1.275 (1.546)	17.867	-1.150 (1.709)
Education quality:												
Degree from top universities	0.185	0.040 (0.041)	0.138	0.089 (0.058)	0.197	0.028 (0.066)	0.162	0.065 (0.081)	0.250	-0.011 (0.109)	0.167	0.079 (0.126)
Degree from other universities	0.307	0.048 (0.046)	0.330	0.004 (0.060)	0.338	0.017 (0.063)	0.412	-0.078 (0.077)	0.250	0.060 (0.080)	0.367	0.048 (0.135)
Degree from teacher colleges	0.122	-0.045 (0.030)	0.090	-0.020 (0.033)	0.099	-0.022 (0.037)	0.074	-0.003 (0.045)	0.100	-0.001 (0.058)	0.100	-0.100 (0.079)
Degree from overseas universities	0.381	-0.050 (0.052)	0.436	-0.067 (0.054)	0.366	-0.035 (0.075)	0.353	0.016 (0.074)	0.400	-0.062 (0.114)	0.367	-0.027 (0.122)
Number of teachers		358		329		240		209		111		83
Number of schools		109		105		69		64		27		25
Number of treated schools		46		42		46		42		17		15

Notes : Standard errors in parenthesis are clustered at the school level. Top universities: Hebrew University in Jerusalem, Tel-Aviv, Technion and Weizman Institute. Other Universities: Bar-Ilan, Ben Gurion and Haifa university.

Table A4 - RT Sample Breakdown of Schools

Neighborhood	Schools in Treatment Group				Schools in Control Group			
	School Code	Number of Students (n'hood sum)	Weight (weighted num)	True Matriculation Rate	School Code	Number of Students (n'hood sum)	Weight (weighted num)	True Matriculation Rate
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	28	160	1	0.453	73	85	4	0.455
	97	207	1	0.455				
		(367)	(367)			(85)	(340)	
2	11	85	1	0.473	87	124	9	0.470
	13	31	1	0.474				
	30	361	1	0.470				
	57	175	1	0.469				
	93	69	1	0.473				
	94	407	1	0.471				
		(1,128)	(1,128)			(124)	(1,116)	
3	86	362	1	0.486	82	196	2	0.489
		(362)	(362)			(196)	(392)	
4	9	14	3	0.500	8	39	1	0.509
	24	54	3	0.512	56	54	1	0.500
	55	53	3	0.514	71	245	1	0.503
	78	97	3	0.518	74	244	1	0.516
					80	52	1	0.517
		(218)	(654)			(634)	(634)	
5	53	222	2	0.535	33	30	1	0.545
	96	110	2	0.551	42	245	1	0.547
					75	190	1	0.534
					79	70	1	0.542
		(332)	(664)			(535)	(535)	
6	70	71	3	0.569	3	97	1	0.563
					41	96	1	0.577
		(71)	(213)			(193)	(193)	
7	29	74	2	0.707	43	174	1	0.711
		(74)	(148)			(174)	(174)	
8	27	46	6	0.774	34	74	1	0.770
					46	39	1	0.772
					69	182	1	0.772
		(46)	(276)			(295)	(295)	

Notes: Column (1) - Neighborhood - is an identifier of a group with at least one treatment and one control schools that are no more than 0.01 away from each other in terms of their true matriculation rate. The matching of schools is contagious - that is, schools will be assigned to the same "neighborhood" as long as there is a school from the other group (treatment/control) there which is less than 0.01 away. See also the rectangles in Figure 3 in the text and Figure A3 in the online Appendix.

Columns (3) and (7) are based on the 2001 math sample. For other combinations of subject and year, similar tables were produced for weight assignment.

Figure A1: The Relationship Between the Correct and the Erroneously Measured 1999 Matriculation Rate
Sample=97 Schools

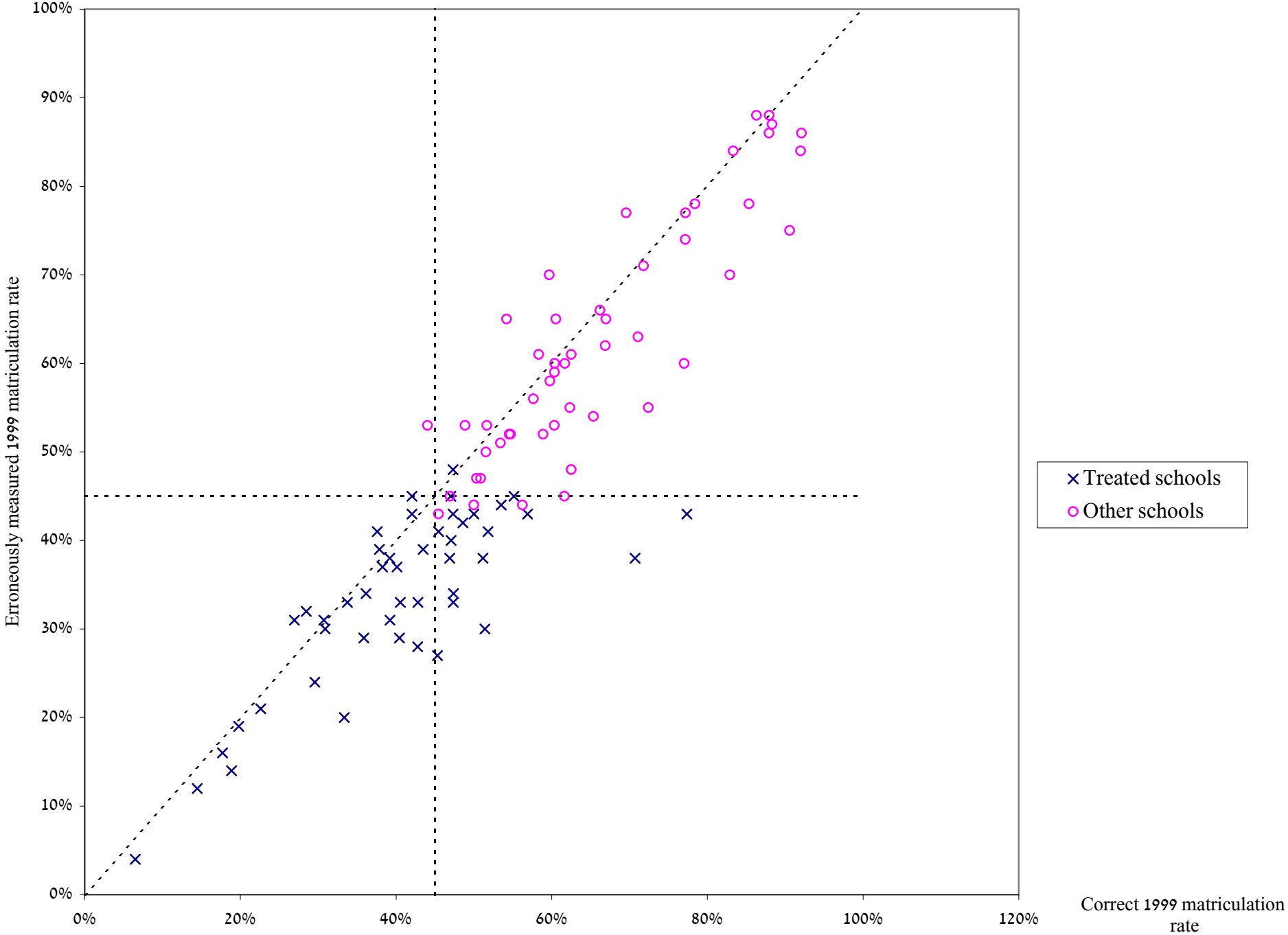


Figure A2: The Correct 1999 Matriculation Rate Versus The Measurement Error
Sample=97 Schools

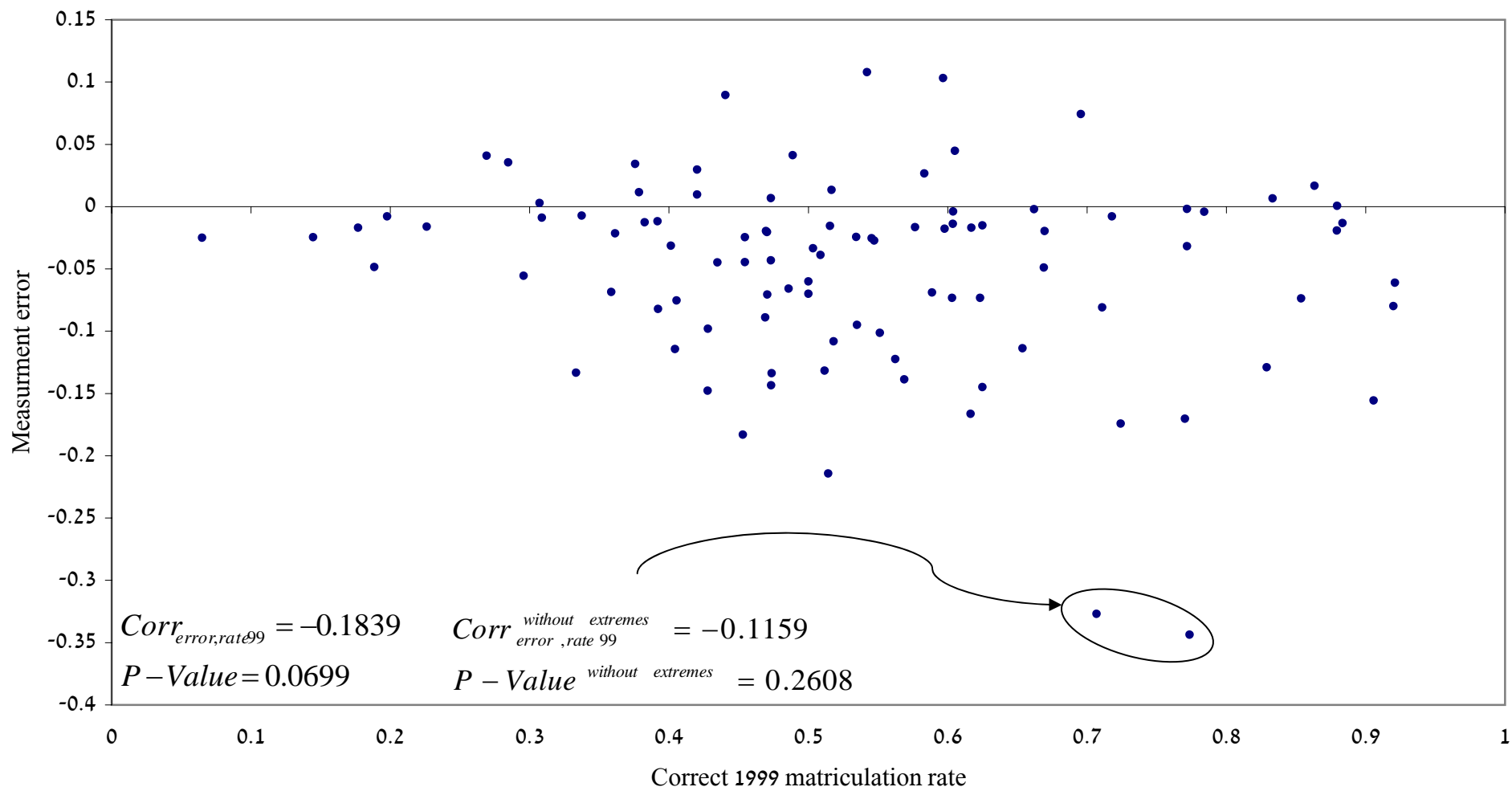


Figure A3: Determining the Sample of Schools That Were Randomly Assigned To Treatment or Control

Elaboration of the Matching Process

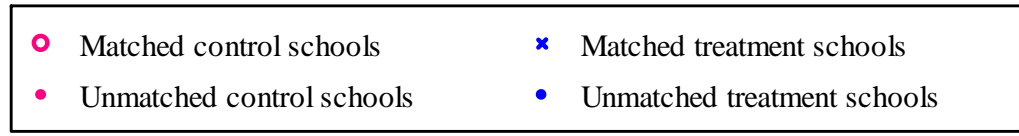
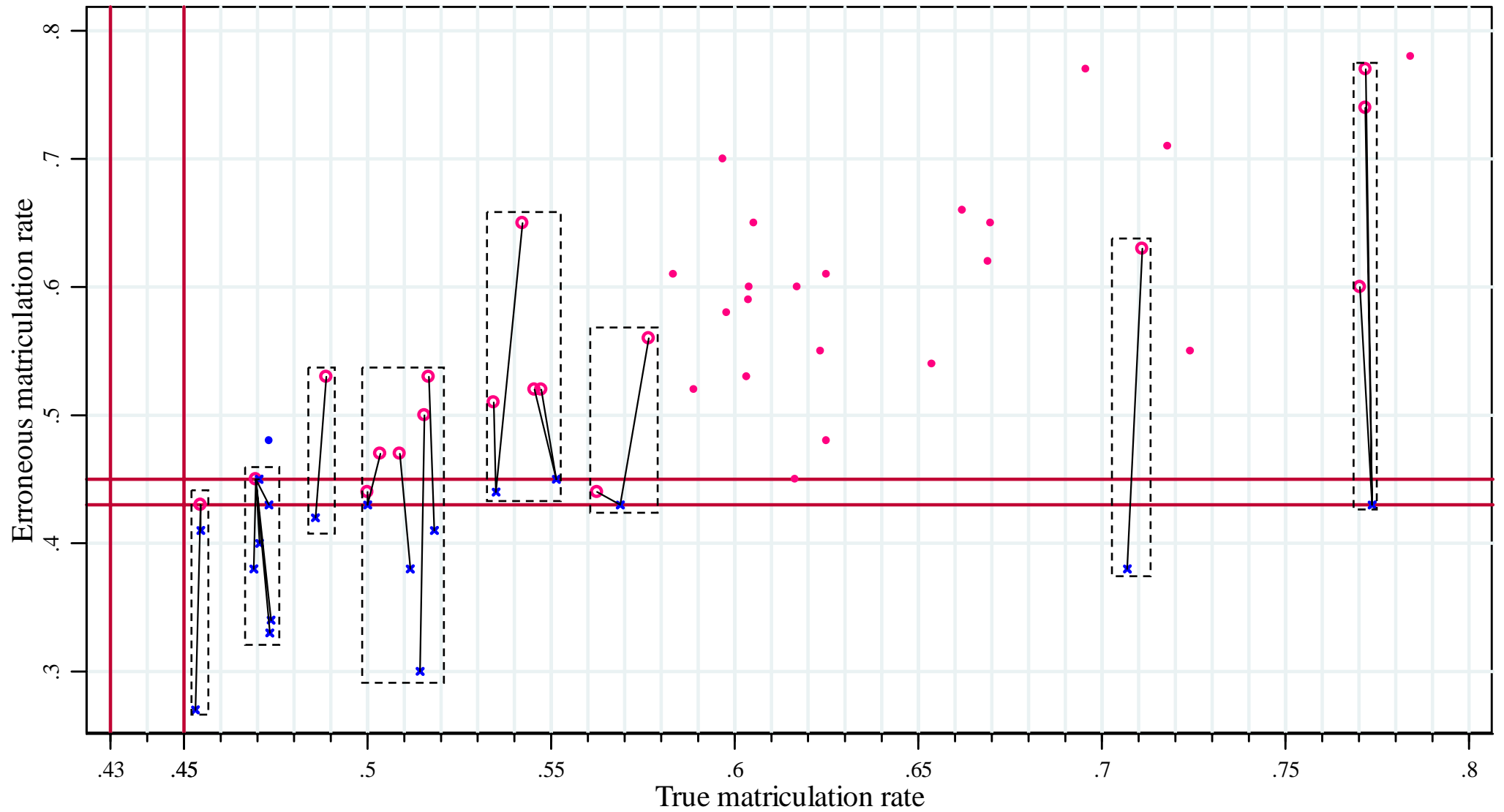


Figure A4: Determining the Discontinuity Sample (Schools Close To the Threshold Value)
Sample=97 Schools

