Class Size and the Regression Discontinuity Design: The Case of Public Schools

Danny Cohen Zada
Mark Gradstein
Ehud Reuven

DISCUSSION PAPER SERIES

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Class Size and the
Regression Discontinuity Design:
The Case of Public Schools

Danny Cohen Zada
Ben Gurion University

Mark Gradstein
*Ben Gurion University
and IZA*

Ehud Reuven
Ben Gurion University

ABSTRACT

Class Size and the Regression Discontinuity Design:
The Case of Public Schools

Using a rich individual-level dataset on secondary public schools in Israel, we find strong evidence for discontinuities in the relationship between enrollment and household characteristics at cutoff points induced by a maximum class size rule. Our findings extend existing work that documents such discontinuities only among private schools (Urquiola and Verhoogen, 2009). These discontinuities violate the assumptions underlying the regression discontinuity design, which are crucial for identification. Consequently, IV estimates of class size effects are likely to be seriously biased. Potential manipulation of the treatment assignment rule by public schools warrants caution in applying a regression discontinuity design to estimate class size effects and indicates that institutional context is crucial for its scope of applicability.

Corresponding author:

Mark Gradstein
Department of Economics
Ben-Gurion University of the Negev
P.O. Box 653
Beer-Sheva 84105
Israel
E-mail: grade@bgu.ac.il

1. Introduction

The unresolved debate over the importance of class size as a determinant of student performance remains a key issue in the economics of education, and it feeds into the ongoing public debate about the value of class size reduction.[1] Studies dealing with this empirical question are particularly preoccupied with the problem of non-random placements of students between and within schools. To address this issue, they often adopt a quasi-experimental approach that exploits the presence of instrumental variables to disentangle the causal effect of class size from the effect of sorting. Angrist and Lavy, 1999, and Hoxby, 2000, are seminal papers which apply regression discontinuity (RD) design to estimate the causal effect of class size on student achievements, the former in Israeli elementary schools, and the latter in public schools in Connecticut.[2] Numerous studies similarly implemented this approach in related contexts more recently (Urquiola 2006, Woessmann 2005, among others).

These studies differ significantly in regard to the estimates of class size effects, varying from no effects to significant and substantial ones. This is true even with papers employing similar methodology; for example, Angrist and Lavy, 1999, and Urquiola, 2006, in their different contexts, point out robust positive effects of reduction in class size on achievements, while Hoxby, 2000, and Urquiola and Verhoogen, 2009, find none, all employing a very similar RD design.

Indeed, it is now recognized that the RD is an application-dependent tool, whose ultimate success may depend on underlying institutional details, see Lee and Lemieux, 2009. In a closely related paper, Urquiola and Verhoogen (2009) illustrate neatly this

---

[1] Several important studies indicate no significant effect of class size on student outcomes (Hoxby 2000, Hanushek 1986, Woessmann and West 2006) while others demonstrate that smaller class sizes enhance student outcomes (e.g., Angrist and Lavy 1999, Krueger 1999, 2003, Sander 2003).

[2] Hoxby, 2000, also uses an additional identification strategy that does not rely on discontinuities, to confirm the findings.

issue for the case of discontinuities in class size in private schools. It turns out that Chilean private schools, seeking to maximize profits, manipulate enrollment to avoid adding an additional classroom, and households sort themselves across schools in response to this situation. Urquiola and Verhoogen's, 2009, empirical evidence demonstrates that this behavior invalidates the RD design.[3] More specifically, their data suggest that students after the cutoffs (smaller class sizes) consistently come from stronger backgrounds, which means that RD estimates overestimate the effect of class size on student outcomes. However, they argue that the continuity assumptions underlying the design are not likely to be violated in public school systems, where students are required to attend local schools that cannot control enrollments. Indeed, this argument is very much likely to be true in the context, for example, of public schools in New England, as shown convincingly in Hoxby, 2000. It is also likely to hold true in the context of elementary public schools in Israel, studied in Angrist and Lavy, 1999 (see below); but, as will be made the case in this paper, this argument fails in secondary public schools, there and possibly elsewhere.

To make our case, we first model a mechanism that provide incentives to public schools administrators who seek to maximize their student outcomes under a given per-student funding budget constraint, to create smaller classes for weaker students and thus to manipulate the treatment assignment rule in this direction. The mechanism is quite general and is likely to exist in many public school systems. We also discuss institutional characteristics that make such manipulation possible and show that it is more likely to occur in the context of secondary schools than in elementary schools in Israel. Then, we use a rich individual-level dataset on students in Israeli high schools to examine whether the predictions of our model are borne out in the data.

---

[3] More generally, the evolving literature on RD design has shown that when the treatment assignment rule is public knowledge, agents can potentially manipulate the rule, thereby invalidating the continuity assumptions required for identification (Lee 2008, McCrary 2008).

The incentive presented in our model to create smaller classes for weaker students while larger classes for stronger students is commonplace in many countries as is shown in West and Woessmann (2006). Examining the sorting of students into classes of different sizes for 18 national education systems, this paper finds that in all countries, with the exception of the US, there exist strong compensatory patterns of sorting within and especially between schools. We, therefore, argue that since school administrators have a strong incentive to allocate students in a compensatory manner, depending on a context, it may be naïve to assume that they react mechanically to the treatment assignment rule instead of manipulating it to achieve their goals.

In Israel, secondary public schools face a maximum class size rule of 38 students, which induces a class size function of enrollment with discrete jumps at enrollments that are multiples of the class-size cap. These jumps can be used to identify the causal effect of class size on student outcomes only if the relationship between enrollment and household characteristics is continuous at these jumps (Hahn et al., 2001). Implementing several tests we find strong evidence that discontinuities exist in the relationship between household characteristics and enrollment, which invalidates the continuity assumptions that are crucial for identification. Moreover, in stark contrast to Urquiola and Verhoogen (2009) but consistent with our model and with West and Woessmann's (2006) findings of compensatory patterns of student sorting in public school system across countries, we find that students after the cutoffs (smaller class sizes) come from weaker backgrounds. This, in turn, means that RD estimates substantially understate the effect of class size on student outcomes. Using the methodology of Altonji et al. (2005), we further show that the IV bias induced by these discontinuities is likely to be quite strong. In contrast, implementing the same tests on the Angrist and Lavy's dataset, we find no such discontinuities for elementary

schools, thus reinforcing support for their findings. This leads us to conclude that the RD design can be very much context sensitive and its application success depends on institutional details. Without paying close attention to these details, research based on the RD design can yield misleading results.

The paper proceeds as follows. The next section describes the institutional structure of public schooling in Israel and presents simple analytics of students' allocation in public schools. It is followed, in Section 3, by the description of the data and design. Section 4 contains our main empirical findings, and Section 5 concludes.

2. Institutional and theoretical background

We first provide institutional background about public schooling in Israel and then we present a simple model that rationalizes students' assignment in public schools.

2.1 The structure of public schools in Israel

Public schools in Israel are financed and (tightly) regulated by the Ministry of Education. Curriculum is highly centralized, and teacher salaries – the main spending component - are determined through centralized bargaining. Schooling is divided into levels, elementary (grades 1-6), middle school (graded 7-9), and high school (grades 10-12). The secondary public school system (the middle school and the high school levels) has three tracks, general (academic), technological and agricultural. In the general academic track, students complete a full academic program that leads to matriculation. Matriculation exams are designed and delivered by the Ministry of Education in 10-12th grades. The technological track offers students both academic and vocational subjects. The range of technological subjects is wide and varies among

schools. The subjects include design, computer skills, industrial automation, electronics, and telecommunications. Some technological studies culminate in a matriculation exam and others in only a practical completion certificate. The public school system includes four school types, comprehensive schools (which include all three tracks), technological/vocational schools, general schools (academic track), and agricultural schools. Only a third of the students in the general track are enrolled in general schools. In secondary schools, financing is by a voucher per student, whereas elementary schools also get a subsidy per class. A maximal class size is dictated by the Ministry of Education and stands at 40 students (the, so called, "Maimonides' rule") in elementary schooling and at 38 students in secondary schooling.

School choice within the public school system is virtually non-existent in Israel, students being educated within the school district of their residence (municipality). Typically, however, a school district has a number of schools within its jurisdiction, and assignment of students to a particular school within the district may be flexible, especially at the high school level. Specifically, while at the elementary school level responsibility for students' assignment rests with the Ministry of Education, and students typically attend a school geographically closest to their residence, at the high school level this responsibility rests with the school district, and school capacity and teachers' availability are among the considerations determining students' assignment.[4]

2.2 Analytical background

To inform our empirical analysis, it is helpful to understand the constraints and the objectives facing public school administrators. In this section, we first model a

---

[4] We are grateful to Mr. David Lifschitz, the Chief of the Education Division of the City Council of Beer Sheva, for very helpful information and an illuminating discussion on which this account is based.

mechanism that provides incentives to school administrators to manipulate the class size rule, using essential features of the system, such as per student subsidy and limited mobility across school districts; then we discuss institutional characteristics that make such manipulation possible. We should state from the outset that, while our model assumptions capture essential characteristics of public school systems, so that its predictions in regard to the incentives of school administrators are fairly generic, their implementation hinges upon a particular institutional setting.

Consider a problem faced by administrators of a school district. Let N denote the district's enrollment; s - subsidy per student; n – student per class; $k = N/n$ is the number of classes (we will ignore issues of indivisibility); q – a student's academic propensity or quality; F – a fixed cost per class (that consists of a teacher's salary, a rent of a classroom etc.). We let $c(n, q)$ denote the cost of attaining a desired level of schooling in a class populated by n student with quality q, where c_n, $c_{nn} > 0$, $c_q < 0$, $c_n(0, q) > 0$.

Note that the number of students in a district is treated as exogenously given. This reflects the relatively low household mobility in Israel and implies – given the budgeting formula – that the district's revenues are constant. It then follows that a plausible objective of a district's school administrator is cost minimization required to attain a desired level of academic standard, in other words,

$$Min_n \; C(n) = (N/n)[F + c(n, q)]$$

The model is inspired by Lazear, 2001, where a framework for analyzing optimal class sizes in the context of profit maximizing schools is constructed, but differs from it in several essential features. The most important of these is the source

of revenue and the objective function of school authorities, which are here supposed to reflect the nature of public schools.

The first order condition is:

$$dC(n)/dn = -(N/n^2)[F + c(n, q)] + (N/n)c_n(n, q) = 0$$

and our assumptions guarantee that the second order condition, $d^2C(n)/dn^2 > 0$, holds. Totally differentiating the first order condition we obtain that $\partial n/\partial F = (N/n^2)/(SOC) > 0$, implying that the optimal class size increases in the fixed cost component. Also,

$$\partial n/\partial q = (c_q/n - c_{nq})/(SOC)$$

which is positive, indicating a positive effect of student quality on class size, if $c_{nq} < c_q/n$. It, in particular, holds when $c_{nq} = 0$, so that the cost effect of student quality is independent of class size. Alternatively, suppose that $c(n, q) = n^{\alpha} q^{\beta}$, $\alpha > 1$, $\beta < 0$. Then $c_{nq} = -\alpha\beta \, n^{\alpha-1} q^{\beta-1} < c_q/n = -\beta \, n^{\alpha-1} q^{\beta-1}$, so that the condition holds again.

This latter result indicates that, optimally, under plausible assumptions low quality students will be found in smaller classes. This result obviously hinges upon the assumed financing scheme under which a school is reimbursed on a per student basis.

Differentiating twice, we further obtain under our assumptions that $\partial^2 n/\partial q \partial F < 0$, implying that the relationship between student quality and class size is moderated through the fixed cost component. Now, subsidization per class – as takes place in elementary schools, but not in secondary schools in Israel – decreases the fixed cost component F, as viewed by school administrators, thus resulting in a smaller effect of

student quality on class size. This may explain why manipulation of class size is relatively more attractive in secondary than in elementary schools.

Further, a school district operating a number of schools can affect class size through the number of students it assigns to various schools (as the enrollment number N applies to a school district as a whole). For example, it would be likely to assign a number of students just under the cut-off point to a school located in an affluent neighborhood with stronger students, thereby creating larger classes for these stronger students, while assigning a number of students just over the cut-off point to a school located in a poorer neighborhood with weaker students, to create smaller classes for weaker students.[5] This mechanism does not necessarily imply stacking since local authorities may assign schools a number of students not exactly at one of the cutoff points but rather before them or after them. In addition, this mechanism is also consistent with our finding that manipulation of the treatment assignment rule applies in secondary schooling but not in elementary schooling, as in secondary schooling the motivation of school district's administration to report better measured results (matriculation rates) is higher. Further, only in the context of high schools have local authorities in Israel the flexibility to shift students, based on physical capacity and teaching needs considerations.

3 Data and design

Our data contain administrative records collected by the Israel Ministry of Education for the year 2003 on all 12th grade public school students in Israel living in localities

---

[5] The motivation of schools and local authorities to assign weaker students to smaller classes in order to increase the matriculation success rate is reflected in some simple OLS regressions we performed of actual class size on mother's education, father's education and the disadvantaged index of the school (not presented in the paper but available from the authors upon request). The regressions indicate that weaker students are generally placed in smaller classes.

with a population higher than ten thousand residents (amounting to 64,429 students from 517 schools).[6] Each record provides data on whether the student is eligible for a matriculation diploma,[7] along with detailed family background including gender, father's years of schooling, mother's years of schooling, number of siblings, ethnicity, a dummy variable indicating immigrant status and dummies for the mother's and father's continent of birth. In addition, the record includes several variables on the class and school of each student such as number of students in the class, percent females in the class, disadvantaged index of the school in 2003, disadvantaged index of the school for the year 2008-2009 and enrollment level of the school.[8,9] Finally, information is also available on the socioeconomic status of the city in which the student resides.

Our database includes information on the track of 64,337 students. It indicates that 39,580 students study in the general track, 24,255 in the technological track and 502 in the agricultural track. In our database, of the 39,580 students who study in the general track only 12,242 are in general schools. We focus our analysis on students from general schools where the maximum class size rule is relevant but ignore technological and comprehensive schools where it is not relevant as these schools

---

[6] The statistical abstract in Israel (2004, Table 8.19) reports that the number of pupils in 12th grade public schools in 2003 was 72, 771 (58,971 in non-religious state schools plus 13,800 in religious state schools). The discrepancy between this figure and the number of observations in our dataset (64, 429 – 89% of the entire number) is due to the fact that 8,342 pupils are omitted for reasons of confidentiality since they come from localities with population of less than 10,000.

[7] A matriculation diploma is a necessary condition for the continuation of academic studies in a university or college. Many employers also make it a condition for employment. The diploma is granted to students who pass a series of state-administered exams.

[8] Ideally, we would prefer to use class size as measured at the beginning of high school (9th grade). However, since we do not have this information, we use class size measured in 12th grade.

[9] The disadvantaged index was created by the Ministry of Education and is based on the students' family background. A high index reflects a low socioeconomic status. The disadvantaged index of schools in 2003 contained missing values for more than 50% of the pupils. Fortunately for us, based on registration records, a new disadvantaged index was recently calculated for the school year 2008-2009 (beginning in September 2008), which includes almost all schools in Israel. Assuming that the socioeconomic status of the schools did not change substantially in this period, we can use the 2008-2009 index to control for the socioeconomic status of students. Indeed, the correlation between the 2003 and 2008-2009 indices is 0.77.

regularly open additional classes before their enrollment level reach the class-size cap in order to separate students from different tracks. Obviously, there is no need to test whether the treatment assignment rule is manipulated in schools where the maximum class size rule is irrelevant. Our sample includes all 12th grade pupils who are enrolled in general schools in localities with a population over 10,000, in sum 12,242 pupils from 167 schools. Summary statistics are reported in Table 1.[10] Comparing the summary statistics of the students in our sample with those of the general 12th grade population we find that the former have somewhat stronger background characteristics. For example, while in our sample the father's years of schooling is 13.7, in the general population it is only 12.79. Similarly, the mother's years of schooling in our sample is 13.68, compared to only 12.93 in the general population. The disadvantaged index of the school also indicates that the students in our sample have a higher socio-economic status. In our sample the disadvantage index is 3.97, while in the general population it is 5.55. General schools are generally smaller in size relative to other high schools as they include only one track. In our sample, the mean enrollment level of schools is 145 while in the general population it is 212. Although the students in our sample differ slightly from the general population, it is valid for

---

[10] For several covariates we had missing values (see Table 1). Thus, in regressions where these covariates appear on the right-hand side so as not to end up with a selective sample we completed the data for some of these covariates, in which cases we also added in the regressions indicators for missing values of these covariates. More specifically, in cases where the number of siblings was not reported, we assigned the pupil the average number of siblings in the class. Similarly, instead of separately controlling for each of the parent's years of schooling we controlled for parents' average years of schooling and in cases where one of the parent's years of schooling was missing we assigned this variable the years of schooling of the other parent. In cases where the father's or mother's continent of birth was missing we just added a dummy indicator for missing values of these covariates. We obtained a sample of 11,754 pupils for which we have all the covariates, losing about 4% of sample due to missing data in several school-level covariates. As identification is at the school level, and because only a very few observations were missing in these covariates, we preferred not to complete them. Note that when those household characteristics appear on the left-hand side of the regression we use the raw data before completion.

our purpose. The reason for this is that, as mentioned before, only in general schools the maximum class size rule is relevant.[11]

We follow Angrist and Lavy (1999), a class size function induced by a maximum class size rule is used to construct IV estimates of class size effects on student outcomes among elementary schools in Israel. Assuming that schools divide their grade enrollment into classes of equal size and also that schools open additional classes only when triggered by the maximum class size rule, Angrist and Lavy (1999) obtained a non-monotonic class size function of enrollment. Since class size is a function of enrollment, and enrollment is likely to be correlated with factors that are related to student outcomes, they included smooth functions of enrollment - a second order polynomial function of enrollment, and a continuous piecewise linear trend - in their covariate set in order to extract out of the instrument any potential relationship between enrollment and student outcomes. The validity of their design relies on two assumptions, that the smooth functions of enrollment fully control for any relationship between class size and enrollment (in which case identification is achieved merely from the discontinuities that exist at the cut-off points); and that discontinuities do not exist at the cut-off points of the maximum class size rule in any of the variables that are related to student outcomes.

There are strong reasons to believe that these assumptions are much less likely to hold in secondary than in elementary public schools. In elementary schools, pupils are younger, and are restricted to attend schools near their homes. Consequently, rules are much stricter regarding exclusion of pupils from local elementary schools which weakens their ability to control enrollment. In secondary schools, manipulation of the maximum class size rule is more likely to occur, also because the responsibility for

---

11 Hence, the results may not be generalizeable to comprehensive schools or technological schools.

students' assignment rests with local authorities, not with the Ministry of Education. This is the case both institutionally and theoretically, as argued above; now we proceed with empirical results.

4. Results

4.1. Instrument's validity

We, first, test whether when controlling for the smooth functions of enrollment the instrument is indeed not correlated with characteristics that are likely to affect student outcomes. To this end, we estimate each of our covariates on the expected class size (the instrument) and one of the smooth functions of enrollment (either the polynomial function or the continuous piecewise linear trend). Column 1 in Table 2 reports the estimated effect of the expected class size on each of our covariates when the smooth function included in the estimation is a piecewise linear trend and Column 2 for when the smooth function is a second order polynomial of enrollment. The results indicate that both smooth functions fail to extract from the instrument the continuous relationship with enrollment. Column 1 shows that when controlling for the second-order polynomial function of enrollment, the instrument is significantly correlated at the 5% level with father's years of schooling, mother's years of schooling, father's birthplace and school type. In addition, it is significantly correlated at the 10% level with number of siblings and disadvantaged index of the school in 2008-2009. Similarly, Column 2 shows that when controlling for a piecewise linear trend, the instrument is significantly correlated at the 5% level with dummies for father birthplace and school type, and at the 10% level with number of siblings. In this case, just as the instrument is correlated with our observed covariates it may be also

correlated with unobserved covariates that affect student outcomes, thus invalidating the RD design.

Next, we follow Urquiola and Verhoogen (2009) and implement the density discontinuity test suggested by McCrary (2008). It is noteworthy that unlike them, in our dataset we do not find any evidence for stacking at enrollments that are multiples of the maximum class size rule. The histogram of 12th grade enrollments among public schools in Israel presented in Figure 1 does not show any pattern in which the number of schools just at the cut-off points is much larger than the number of schools before or after the cut-offs. In addition, the figure indicates that we must concentrate our analysis on the two first cut-off points since there are not enough schools in the third, fourth or fifth cut-offs.

Although we do not find evidence for stacking, we do find strong evidence for discontinuities at the cut-off points between enrollment and household characteristics. Figure 2 presents enrollment-cell means of mother's years of schooling, father's years of schooling and disadvantaged index of the school for the year 2008-2009, along with fitted values obtained from a weighted kernel regression within each enrollment segment. This figure shows that students after the cut-off points come from families with less educated parents and schools with a higher disadvantaged index (lower socioeconomic status). This evidence indicates a pattern in which predetermined covariates change discontinuously around the cut-off points, and that students after the cut-off points come from weaker backgrounds.

Like Urquiola and Verhoogen (2009) we also report regression results for each of the mentioned household characteristics on dummies on whether enrollments are above each of the cut-off points and on piecewise linear splines for enrollment (see Table 3). The results for mother's years of schooling and father's years of schooling

14

indicate that the jumps at all three first enrollment cut-offs are substantial and significant. In addition, the jump at the first cutoff point is significant for the disadvantaged index of the school. Taken together, these results strongly confirm Figure 2 that the continuity assumptions are violated.

Finally, we focus more narrowly around the discontinuities, following a local linear specification, as used in McCrary, 2008, and others, which allows for linear trends to the left and to the right of the cutoff point. For example, to test whether a discontinuity exists at the first cutoff point between enrollment and mother's education we estimate the following specification

$$ME_{ics} = b0 + b1 * ERC_s + b2 * (ERC_s * AC_s) + b3 * AC_s + U_{ics} \ ,$$

where *ME* is mother education, *ERC* (enrollment relative to the cutoff) is actual enrollment minus 38, and *AC* (after the cutoff) equals one if *ERC*>0 and zero otherwise. In this specification, *b1* captures the slope of the line to the left of the potential discontinuity, *b2* captures the slope to the right, and *b3* estimates the discontinuity. Table 4 reports the results. It indicates that discontinuities exist at the first cutoff point between enrollment and all households' characteristics, while at the second cutoff point they exist only between enrollment and mother's education.

4.2 How large is the IV bias?
Before discussing how large the bias of the IV estimates is likely to be, we first present OLS and IV estimates of the class size effect on eligibility for matriculation diploma. Table 5 presents naive estimates of the effect of class size on eligibility for a

matriculation diploma both with and without control variables.[12] The results indicate that class size always has a strong significant positive effect on student outcome. However, this effect is smaller when we control for our set of covariates (Columns 3 and 4).

Table 6 reports IV estimates of a specification that controls for a piecewise linear trend. The standard errors are clustered by enrollment levels, as is appropriate when the assignment variable is discrete (Lee and Card, 2008).[13] The first stage estimations indicate that the instrument has a significant positive effect on actual class size, and the f-statistic on the excluded instruments pass the minimal critical value required by Stock and Yogo (2005) in order for the instrument not to be considered weak. IV estimates without controls indicate that class size has a positive effect though insignificant on eligibility for a matriculation diploma. When control variables are included in the estimation the estimated class size effect falls from 0.02 to 0.005. Although both effects are insignificant they differ substantially in size.

Table 7 presents IV estimates when we control for a second-order polynomial function of enrollment. The results again indicate that the instrument is not weak and the estimated class size effects on outcome are quite similar to those obtained with the piecewise linear trend. As previously, the effect of class size on the eligibility for a

---

[12] We are aware that our outcome measure, eligibility for a matriculation diploma, ignores quality. It also suffers from the fact that while we estimate the effect of the homeclass size on eligibility for a matriculation diploma, about half of the subjects on which the students are tested in the matriculation exams are not studied in the homeclass. To address this latter issue we also estimated class size effects on test scores in specific subjects that are studied in the homeclass and found them to be insignificant (the results of these estimations are available from the authors upon request).

[13] As enrollment categories are supersets of schools (i.e., multiple schools may have the same enrollment but multiple enrollments in the same school is not feasible), clustering on enrollment is considered a more conservative strategy than clustering at school level. We choose to follow a conservative line and to cluster the standard errors by enrollment. In addition, to show that our results are not sensitive to the type of clustering we reran all the estimations with school clusters as well and found that all the findings were qualitatively very similar. These results are available from the authors upon request.

matriculation diploma is quite sensitive to the inclusion of control covariates in the estimation.

Next, we use the methodology of Altonji et al. (2005) to assess the size of the IV bias induced by the discontinuities between enrollment and household characteristics. The calculation of the bias is based on the assumption that the amount of selection on unobservables exactly equals the amount of selection on observables. Alternatively, we also calculate the bias under the assumption that the amount of selection on unobservables equals a specific proportion of the amount of selection on observables. In our study the bias is calculated somewhat differently from that in Altonji et al. (2005) because of two reasons. First, the smooth functions of enrollment are not treated as regular observables but rather as part of the identification strategy. Thus, the correlations between the Maimonides' rule and the smooth functions of enrollment do not affect the calculated size of the bias. That is, we assume that a bias is likely to exist only if controlling for the smooth functions of enrollment, the instrument is still correlated with our other observables. Second, the bias formula in Altonji et al. (2005) is written for the case in which the instrument is an indicator variable while in our case the instrument is a scalar. Because of these two differences we try to be formal in describing how we calculate the bias.[14]

Let the outcome Y_i (eligibility for matriculation diploma) be determined by

$$(1) \qquad Y_{ics} = \alpha \cdot CS_{cs} + \delta_0 \cdot EN_s + \delta_1 \cdot EN^2{}_s + X'_{ics}\,\gamma + \varepsilon_{ics}$$

---

where *CS* is class size, *EN* is enrollment, *X* is a set of exogenous control covariates and γ is determined such that $\text{cov}(X_{ics}, \varepsilon_{ics}) = 0$. Obviously, class size is endogenous and thus correlates with ε_{ics}. Let class size be determined by

$$(2) \qquad CS_{cs} = X_{ics}{}'\beta + \lambda \cdot MR_s + \eta_1 \cdot EN + \eta_2 \cdot EN^2 + u_{ics}$$

where the Maimonides' rule (our instrument) is assumed to not directly influence the eligibility for a matriculation diploma but to possibly be correlated with ε_{ics}. Finally, we define v_{ics} as the residuals from an OLS estimation of MR_s on X_{ics}. It is well known that the asymptotic bias in an IV estimate of α equals[15]

$$(3) \qquad \text{IV Bias} = Cov(v_{ics}, \varepsilon_{ics})/[\lambda \cdot Var(v_{ics})] = Cov(MR_{cs}, \varepsilon_{ics})/[\lambda \cdot Var(v_{ics})]$$

Assuming that the amount of selection on unobservables equals a proportion *p* of the amount of selection on observables we obtain that

$$(4) \qquad \frac{Cov(MR_s, \varepsilon_{ics})}{Var(\varepsilon_{ics})} = p \cdot \frac{Cov(MR_s, X'_{ics}\gamma)}{Var(X'_{ics}\gamma)}$$

Substituting (4) into (3) we obtain the IV bias as a function of *p*

$$(5)\ \text{IV Bias} = p \cdot \frac{Cov(MR_s, X'_{ics}\gamma) \cdot Var(\varepsilon_{ics})}{Var(X'_{ics}\gamma) \cdot [\lambda \cdot Var(v_{ics})]}$$

---

[15] See Altonji et al. (2005, page 805).

It is then straight forward that the correct class size effect is the IV estimated effect minus the IV bias. The corrected class size effects for different values of p are reported in Table 8, where column 1 reports the corrected class size effects for when the smooth function included in the estimation is the second order polynomial function of enrollment, and Column 2 reports for when the smooth function is the piecewise linear trend. Column 1 shows that even for an amount of selection on unobservables that is $1/10^{th}$ as strong as the amount of selection on observables, the entire positive effect would be eliminated. Similarly, if the amount of selection on observables is 50% as strong as selection on observables, class size has a substantial negative effect on the eligibility for a matriculation diploma. Specifically, reducing class size by ten pupils increases the probability of achieving a matriculation diploma by 15 percentage points. The results in Column 2 are even more sensitive to p. If selection on unobservables is only 20% as strong as selection on observables then class size has a quite substantial negative effect on eligibility for matriculation diploma. Taken together, it seems that the bias induced by the discontinuities between enrollment and household characteristics is not likely to be negligible.

4.3. Elementary schools

To illustrate the importance of institutional details for the validity of the RD design, we now apply the above analysis on Israeli elementary public schools first studied in Angrist and Lavy's (1999). To this end we employ four variables used there, percent disadvantaged, grade enrollment, piecewise linear trend, and the Maimonides' rule (I should mention that these are all the variables that they used).[16] To begin, we confirm that among Israeli elementary public schools (where, as mentioned earlier, rules

---

[16] A full description of the dataset in available at: http://econ-www.mit.edu/faculty/angrist/data/anglavy99.

regarding exclusions of pupils from local schools are relatively stricter) manipulation of the treatment assignment rule is not likely to occur. One issue is whether, when controlling for the smooth functions of enrollment, the expected class size according to the Maimonides' rule (our instrument) is indeed not correlated with percent disadvantaged. We estimate percent disadvantaged on the expected class size and a smooth function of enrollment. Columns 1 and 2 in Table 9 report the estimated effect when the smooth function included in the estimation is a piecewise linear trend and when the smooth function is a second order polynomial of enrollment, respectively. The results indicate that, indeed, the instrument is not correlated with percent disadvantaged. Next, we perform a similar test but use the specification of Urquiola and Verhoogen (2009) and estimate percent disadvantaged on dummies for whether enrollments are above each of the cut-off points as well as on piecewise linear splines for enrollment (see Table 10). The results, again, indicate that the jumps at all five enrollment cut-offs are insignificant.

In addition, we also report regressions of percent disadvantaged only for narrow intervals around the first three cut-offs. As for secondary schooling, we follow a local linear specification which allows for linear trends to the left and to the right of the cutoff point. Table 11 reports the results. It shows that there is no discontinuity in the relationship between enrollment and percent disadvantaged. To sum up, our analysis suggests that the RD design among Israeli elementary public schools is valid, offering further indirect support to Angrist and Lavy's, 1999, findings.

5. Conclusion

In this paper, guided by a mechanism that can explain the discontinuities in the relationship between household characteristics and enrollment, we demonstrate existence of such discontinuities in the context of secondary public schools in Israel. Specifically, household characteristics jump around cut-off points induced by a maximum class size rule, where students after the cut-off points consistently come from weaker backgrounds. This is consistent with the cross national evidence in West and Woessmann (2006) that public school systems typically exhibit strong compensatory patterns of student sorting. This, in turn, invalidates the RD design and may cause researchers to mistakenly conclude that class size has no causal effect on student outcomes even if it actually has a negative effect. Furthermore, we show that the IV bias induced by the discontinuities may be quite substantial. Yet, we found that these results do not apply to elementary schools. Our findings extend those of Urquiola and Verhoogen (2009) who detect discontinuities in Chilean private schools. We further show that discontinuities may exist even when there is no evidence for stacking of schools at enrollments that are multiples of the class size cap, indicating that even under this condition the RD design may not be valid. Our main general conclusion is that the ultimate success of the RD design hinges upon the institutional details of the context in which it is applied.

References

Altonji, J., Elder, T., and Taber, C. (2005). "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schools," *Journal of Human Resources*, *40*, 791–821.

Angrist, J. D. and Lavy, V. (1999). "Using Maimonides' rule to estimate the effect of class size on scholastic achievement," *Quarterly Journal of Economics* 114(2), 533–575.

Hahn, J., Todd, P. E. and van der Klaauw,W. (2001). "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica* 69(1), 201– 209.

Hanushek, E., 1986, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature, 49,* 1141-1177.

Hoxby, C. M. (2000). "The effects of class size on student achievement: New evidence from population variation," *Quarterly Journal of Economics* 115(4), 1239–1285.

Krueger, A.B., 1999, "Experimental estimates of education production functions," *Quarterly Journal of Economics* 114, 497–532.

Krueger, A.B., 2003, "Economic considerations and class size," *Economic Journal* 113 (2003), F34–F63.

Lazear, E. P. 2001. "Educational Production." *Quarterly Journal of Economics*, 116(3): 777–803.

Lee, D. S. (2008). "Randomized experiments from non-random selection in U.S. house elections," *Journal of Econometrics* 142(2), 675–697.

Lee, D. S. and Card, D. (2008). "Regression discontinuity inference with specification error," *Journal of Econometrics* 142(2), 655–674.

Lee, D. S. and T. Lemieux, 2009, "Regression discontinuity designs in economics," NBER Working Paper 14723.

McCrary, J. (2008). "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics* 142(2), 698–714.

Sander, W. (1993). "Expenditures and student achievement in Illinois: New evidence," *Journal of Public Economics* 52, 403–416.

Stock, J. and Yogo, M. (2005). "Testing for weak instruments in linear IV regressions," in D. Andrews and J. Stock (Eds.), *Identification and inference for econometric models* (pp 80-108). Cambridge/New-York: Cambridge University Press.

Urquiola, M. S. (2006). "Identifying class size effects in developing countries: Evidence from rural schools in Bolivia," *Review of Economics and Statistics* 88(1), 171-177.

Urquiola, M. S. and Verhoogen, E. (2009). "Class size caps, sorting, and the regression discontinuity design," *American Economic Review*, 99(1), 179-215.

West, M.R. and L. Woessmann, 2006, "Which school systems sort weaker students into smaller classes? International evidence", *European Journal of Political Economy* 22, 944-968.

Woessmann, **L. and M.R. West**, 2006, "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS", *European Economic Review* 50 (3), 695-736.

Woessmann, L. (2005). "Educational production in Europe," *Economic Policy* (43), 445-493.

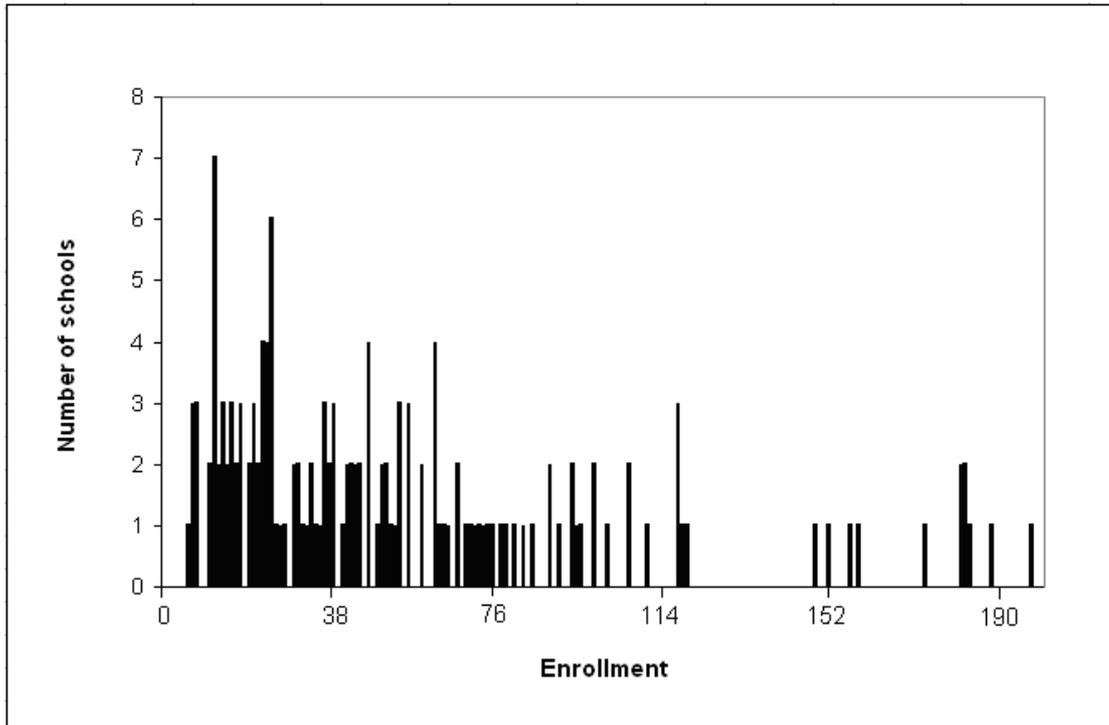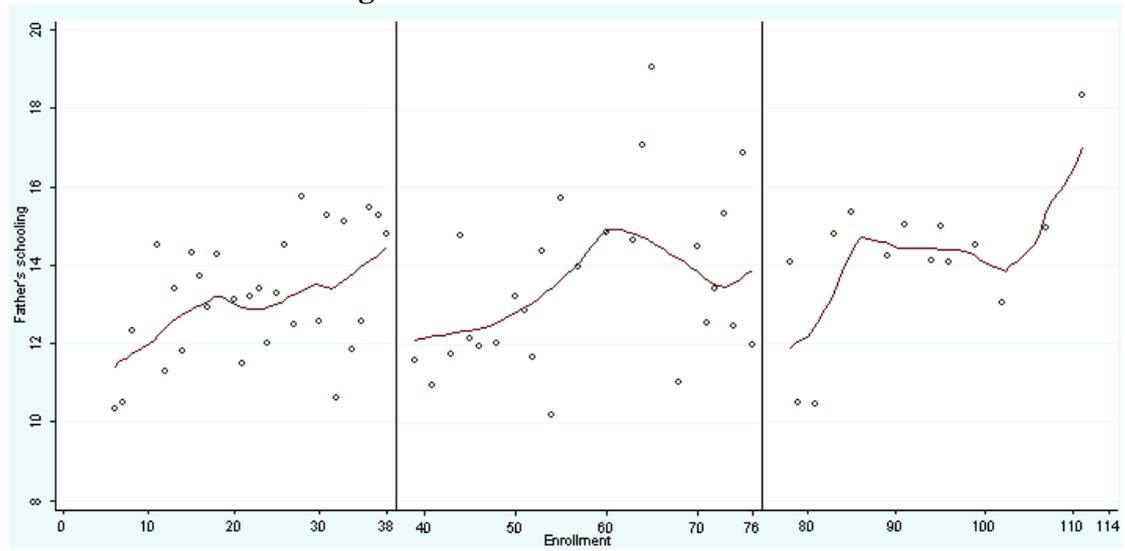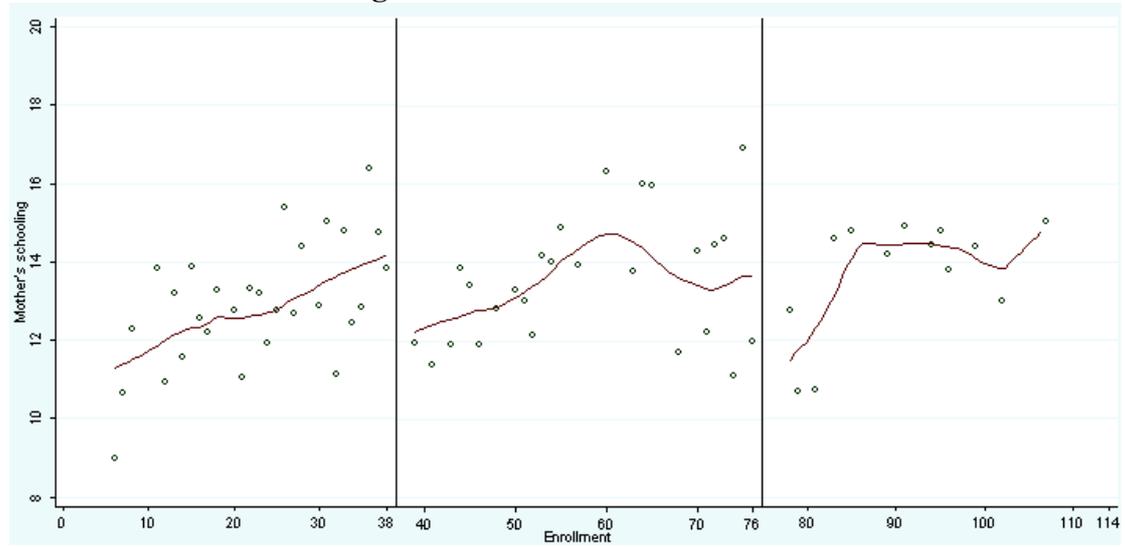**Figure 1. 12th grade enrollments among public schools in Israel, 2003**

Figure 2. Student characteristics and enrollment in secondary public schools, 2003

Panel A: Father's schooling

Panel B: Mother's schooling

Panel C: Disadvantaged index 2008-2009
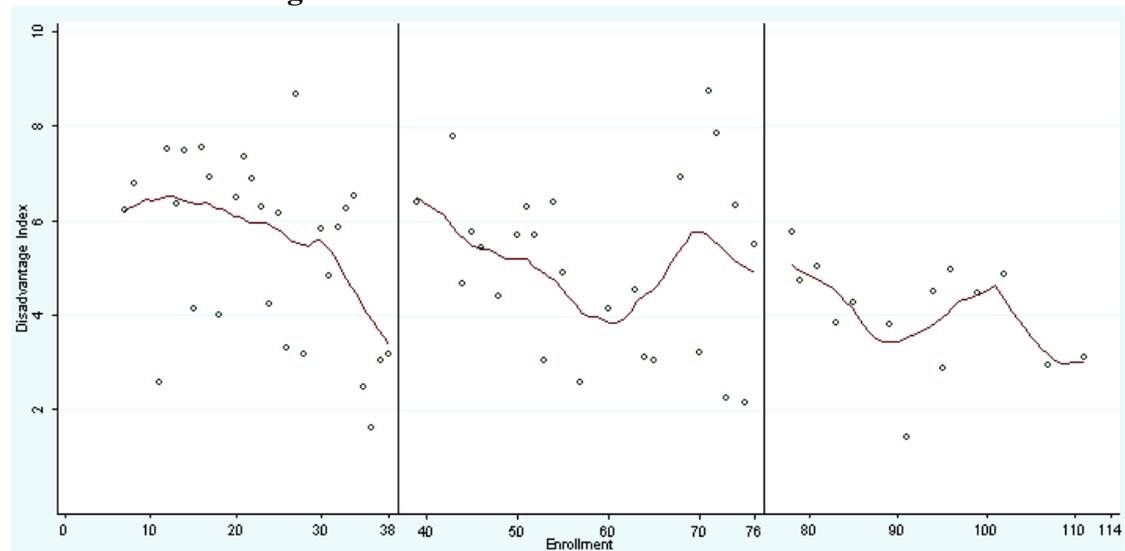
Table 1. Summary statistics			
Variable	Mean	Standard Deviation	Number of Observations
Eligibility	0.690	0.463	12242
Gender (1=female)	0.521	0.500	12242
Father's years of schooling	13.703	3.380	10199
Mother's years of schooling	13.677	3.060	10059
Number of siblings	1.281	1.459	12097
Disadvantaged index 2003	3.966	2.792	5650
Disadvantaged index 2008-2009	3.710	2.017	11792
Socioeconomic index	0.524	0.603	11002
Actual class size	30.178	8.254	12228
Enrollment	144.910	95.927	12228
Expected average class size	31.790	5.728	12228
Proportion of females in the class	0.523	0.294	12204
New Immigrant (1=Yes)	0.163	0.369	12242
School type (1 = Religious)	0.307	0.461	12116
Father's birthplace: America-Europe (1=Yes)	0.225	0.418	10702
Father's birthplace: Asia-Africa (1=Yes)	0.191	0.393	10702
Father's birthplace: Other location (1=Yes)	0.001	0.039	10702
Mother's birthplace: America-Europe (1=Yes)	0.238	0.426	11107
Mother's birthplace: Asia-Africa (1=Yes)	0.149	0.356	11107
Mother's birthplace: Other location (1=Yes)	0.001	0.033	11107

Table 2. Univariate regressions of each of our covariates on the Maimonides' rule and the smooth functions of enrollment			
	(1)	(2)	
Specification	Second order polynomial function of enrollment	Piecewise linear trend	Number of observations
gender (1=female)	0.001 (0.005)	0.002 (0.006)	12,228
Father's years of schooling	0.091** (0.042)	0.069 (0.049)	10,196
Mother's years of schooling	0.081** (0.040)	0.059 (0.047)	10,052
Number of siblings	0.023* (0.013)	0.024* (0.014)	12,084
Disadvantaged index 2003	-0.063 (0.081)	-0.085 (0.086)	5,650
Disadvantaged index 2008-2009	-0.070* (0.039)	-0.061 (0.039)	11,778
Socioeconomic index	0.007 (0.013)	0.002 (0.013)	10,988
Proportion of females in the class	8E-04 (0.005)	0.002 (0.006)	12,204
New immigrant (1=Yes)	-0.001 (0.003)	-0.002 (0.003)	12,228
Father's birthplace: America-Europe (1=Yes)	0.003 (0.002)	0.002 (0.003)	10,692
Father's birthplace: Asia-Africa (1=Yes)	-0.005** (0.002)	-0.005** (0.002)	10,692
Father's birthplace: Other location (1=Yes)	3E-04** (1E-04)	3E-04** (1E-04)	10,692
Mother's birthplace: America-Europe (1=Yes)	0.001 (0.002)	5E-04 (0.003)	11,096
Mother's birthplace: Asia-Africa (1=Yes)	-0.002 (0.002)	-0.002 (0.002)	11,096
Mother's birthplace: Other location (1=Yes)	9E-05 (1E-04)	9E-05 (1E-04)	11,096
School type (1 = Religious)	0.018** (0.007)	0.019** (0.007)	12,102

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Notes: Standard errors clustered at the enrollment level are in parentheses.

Table 3. Behavior of selected variables at the cut-off points

Dependent variable	OLS Father's years of schooling	OLS Mother's years of schooling	OLS Disadvantaged index 2008-2009
Constant	11.255*** (0.628)	10.743*** (0.556)	8.170*** (0.867)
1{enroll≥39}	-2.481*** (0.705)	-2.023*** (0.564)	1.596* (0.808)
1{enroll≥77}	-2.158* (1.284)	-2.061* (1.090)	0.158 (0.873)
1{enroll≥115}	-2.627** (1.036)	-2.532*** (0.918)	0.851 (1.352)
1{enroll≥153}	3.412 (3.309)	2.962 (3.238)	1.278 (1.781)
1{enroll≥191}	0.410 (1.249)	0.328 (1.294)	3.180** (1.424)
1{enroll≥229}	1.672 (1.606)	1.207 (1.712)	3.938** (1.672)
1{enroll≥267}	-1.534* (0.851)	-1.191 (0.802)	-0.268 (0.980)
enrollment	0.085*** (0.025)	0.095*** (0.022)	-0.111*** (0.034)
(enroll-39)*1{enroll≥39}	-0.005 (0.042)	-0.033 (0.037)	0.087* (0.046)
(enroll-77)*1{enroll≥77}	-0.010 (0.056)	0.013 (0.048)	-0.020 (0.036)
(enroll-115)*1{enroll≥115}	-0.118 (0.105)	-0.122 (0.100)	0.001 (0.047)
(enroll-153)*1{enroll≥153}	0.002 (0.098)	0.009 (0.095)	0.032 (0.069)
(enroll-191)*1{enroll≥191}	0.027 (0.069)	0.033 (0.072)	-0.195** (0.089)
(enroll-229)*1{enroll≥229}	0.041 (0.074)	0.018 (0.078)	0.194** (0.077)
(enroll-267)*1{enroll≥267}	-0.004 (0.037)	-3E-04 (0.037)	0.023 (0.028)
Number of clusters	96	95	94
Number of observations	10,196	10,052	11,778

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Notes: Standard errors clustered at the enrollment level are in parentheses.

Table 4. Discontinuities in household characteristics at each cut-off point				
Sample	**Explanatory Variables**	Dependent Variable: **Father's years of schooling**	Dependent Variable: **Mother's years of schooling**	Dependent Variable: **Disadvantaged index 2009**
34≤enroll≤43	(enroll-38)	0.768** (0.238)	0.325 (0.313)	-0.724 (0.378)
	(enroll-38) *1{enroll≥39}	-0.747** (0.246)	-0.357 (0.317)	1.077** (0.378)
	1{enroll≥39}	-4.068*** (0.588)	-2.784*** (0.635)	3.680*** (0.581)
72≤enroll≤81	(enroll-76)	1.080 (0.637)	1.329 (0.808)	-0.463 (0.694)
	(enroll-76) *1{enroll≥77}	-2.120** (0.758)	-1.929* (0.841)	0.278 (0.705)
	1{enroll≥77}	-1.474 (2.595)	-3.943* (1.789)	1.867 (1.605)

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Notes: Standard errors clustered at the enrollment level are in parentheses.

Table 5. Naive estimations

	Without controls		With controls	
	OLS	Probit	OLS	Probit
Dependent variable	Eligibility	Eligibility	Eligibility	Eligibility
Constant	0.321*** (0.098)	-0.522* (0.278)	0.176* (0.090)	-1.101*** (0.290)
Class size	0.013*** (0.003)	0.036*** (0.010)	0.010*** (0.002)	0.031*** (0.007)
Number of clusters	94	94	94	94
Number of observations	11,754	11,754	11,754	11,754

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Notes: Dependent variable is a dummy indicating eligibility for a matriculation diploma. Individual controls include: gender, parents' average years of schooling, number of siblings, ethnicity, a dummy variable indicating immigrant status, dummies indicating the mother's and father's continent of birth, and indicators for missing values of number of siblings, parents' average years of schooling, mother's continent of birth and father's continent of birth. Class and school controls include: percent females in the class and disadvantaged index of the school for the year 2008-2009. Standard errors clustered at the enrollment level are in parentheses.

	Without controls			With controls		
	(1)	(2)	(3)	(4)	(5)	(6)
	First stage	Reduced form	2SLS	First stage	Reduced form	2SLS
Dependent variable	Class size	Eligibility	Eligibility	Class size	Eligibility	Eligibility
Constant	13.242*** (2.724)	0.211* (0.115)	-0.073 (0.351)	17.331*** (3.990)	0.044 (0.149)	-0.044 (0.350)
Expected average class size	0.326*** (0.125)	0.007 (0.005)		0.309** (0.121)	0.002 (0.004)	
Class size			0.0214 (0.017)			0.005 (0.013)
Piecewise linear trend	0.100*** (0.026)	0.004*** (0.001)	0.002 (0.003)	0.064** (0.031)	0.005*** (0.001)	0.005*** (0.002)
Stock-Yogo test (2005) (F-Statistic)	6.85			6.57		
Range of critical values for Stock-Yogo test (2005)	5.53-16.38			5.53-16.38		
Number of clusters	94	94	94	94	94	94
Number of observations	11,754	11,754	11,754	11,754	11,754	11,754

Table heading: **Table 6. IV estimates controlling for a piecewise linear trend**

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Notes: Dependent variable is a dummy indicating eligibility for a matriculation diploma. Individual controls include: gender, parents' average years of schooling, number of siblings, ethnicity, a dummy variable indicating immigrant status, dummies indicating the mother's and father's continent of birth, and indicators for missing values of number of siblings, parents' average years of schooling, mother's continent of birth and father's continent of birth. Class and school controls include: percent females in the class and disadvantaged index of the school for the year 2008-2009. Standard errors clustered at the enrollment level are in parentheses.

	Table 7. IV estimates controlling for a second order polynomial function of enrollment					
	Without controls			With controls		
	(1)	(2)	(3)	(4)	(5)	(6)
	First stage	Reduced form	2SLS	First stage	Reduced form	2SLS
Dependent variable	Class size	Eligibility	Eligibility	Class size	Eligibility	Eligibility
Constant	13.681*** (2.877)	0.244** (0.120)	-0.045 (0.346)	17.551*** (3.895)	0.069 (0.146)	0.008 (0.336)
Expected average class size	0.340*** (0.126)	0.007 (0.005)		0.320*** (0.122)	0.001 (0.004)	
Class size			0.021 (0.016)			0.003 (0.012)
Enrollment	0.073*** (0.026)	0.003*** (0.001)	0.001 (0.002)	0.050* (0.026)	0.004*** (0.001)	0.004*** (0.001)
Enrollment squared	-0.152** (0.074)	-0.005*** (0.002)	-0.002 (0.004)	-0.113 (0.071)	-0.007*** (0.002)	-0.007** (0.003)
Stock-Yogo test (2005) - F-Statistic	7.24			6.88		
Range of critical values for Stock-Yogo test (2005)	5.53-16.38			5.53-16.38		
Number of clusters	94	94	94	94	94	94
Number of observations	11,754	11,754	11,754	11,754	11,754	11,754

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Notes: Dependent variable is a dummy indicating eligibility for a matriculation diploma. Individual controls include: gender, parents' average years of schooling, number of siblings, ethnicity, a dummy variable indicating immigrant status, dummies indicating the mother's and father's continent of birth, and indicators for missing values of number of siblings, parents' average years of schooling, mother's continent of birth and father's continent of birth. Class and school controls include: percent females in the class and disadvantaged index of the school for the year 2008-2009. Standard errors clustered at the enrollment level are in parentheses.

	Specification	
	(1)	(2)
p	Second order polynomial function of enrollment	Piecewise linear trend
0	0.003	0.005
0.1	0	-0.003
0.2	-0.004	-0.011
0.3	-0.008	-0.018
0.4	-0.011	-0.026
0.5	-0.015	-0.034
0.6	-0.019	-0.042
0.7	-0.023	-0.050
0.8	-0.026	-0.057
0.9	-0.030	-0.065
1	-0.034	-0.073

Table 8. Corrected class size effects for different values of p

Table 9. Regressions of percent disadvantaged on the Maimonides' rule and the smooth functions of enrollment		
	(1)	(2)
Dependent variable	**Percent disadvantaged**	**Percent disadvantaged**
Constant	27.629*** (2.098)	25.851*** (2.156)
Expected average class size (Maimonides' rule)	0.103 (0.080)	0.101 (0.077)
Enrollment		-0.271*** (0.029)
Enrollment squared/100		0.080*** (0.013)
Piecewise linear trend	-0.308*** (0.026)	
Number of clusters	143	151
Number of observations	2,009	2,059

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Notes: Standard errors clustered at the enrollment level are in parentheses.

Table 10. Behavior of percent disadvantaged at the cut-off points	
	OLS
Dependent variable	**Percent disadvantaged**
Constant	21.190*** (3.619)
1{enroll≥41}	-1.617 (2.220)
1{enroll≥81}	-1.557 (1.331)
1{enroll≥121}	-1.797 (1.764)
1{enroll≥161}	1.083 (4.482)
1{enroll≥201}	-6.425 (7.045)
enrollment	0.008 (0.125)
(enroll-41)*1{enroll≥41}	-0.202 (0.133)
(enroll-81)*1{enroll≥81}	0.155** (0.063)
(enroll-121)*1{enroll≥121}	0.045 (0.079)
(enroll-161)*1{enroll≥161}	0.071 (0.222)
(enroll-201)*1{enroll≥201}	0.041 (0.305)
Number of clusters	151
Number of observations	2,059

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Notes: Standard errors clustered at the enrollment level are in parentheses.

Table 11. Discontinuities in percent disadvantaged at each cut-off point		
Sample	**Explanatory Variables**	
36≤enroll≤45	(enroll-40)	-2.096*** (0.313)
	(enroll-40)∗1{enroll≥41}	2.454 (1.425)
	1{enroll≥41}	-3.119 (2.835)
76≤enroll≤85	(enroll-80)	0.177 (0.489)
	(enroll-80)∗1{enroll≥81}	-1.156 (1.032)
	1{enroll≥81}	1.907 (3.928)
116≤enroll≤125	(enroll-120)	1.092 (0.770)
	(enroll-120)∗1{enroll≥121}	-1.759 (1.746)
	1{enroll≥121}	-5.013 (6.145)

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Notes: Standard errors clustered at the enrollment level are in parentheses.