Class Size and Sorting in Market Equilibrium:
Theory and Evidence

Miguel Urquiola[†]
Eric Verhoogen[‡]

Sept. 2006

Abstract

This paper examines how schools choose class size and how households sort in response to those choices. Focusing on the highly liberalized Chilean education market, we develop a model in which schools are heterogeneous in an underlying productivity parameter, class size is a component of school quality, a class-size cap applies to some schools, and households are heterogeneous in income and hence willingness to pay for quality. The model offers an explanation for two distinct empirical patterns: (i) There is an inverted-U relation between class size and household income in equilibrium, which will tend to bias cross-sectional estimates of the effect of class size on student performance. (ii) Some schools at the class size cap adjust prices and/or enrollments to avoid adding another classroom, which produces stacking at enrollments that are multiples of the class size cap. This results in discontinuities in the relationship between enrollment and students' income at those points, violating the assumptions underlying regression-discontinuity (RD) research designs. An implication is that RD approaches should not be applied in settings in which parents have substantial school choice and schools are free to set prices and influence their enrollments.

1 Introduction

There has been a long and heated debate on whether class-size reductions improve educational performance. Hanushek (1995, 2003) reviews an extensive literature and concludes that class size has no systematic effect on student achievement in either developed or developing countries. Krueger (2003), Kremer (1995) and others have countered that this conclusion is based largely on cross-sectional evidence and subject to multiple potential sources of bias, including the endogenous sorting of students into classes of different sizes, and have called for further analyses using experimental and quasi-experimental designs. In the latter category, an influential approach has been the regression-discontinuity (RD) design of Angrist and Lavy (1999), which exploits the discontinuous relationship between enrollment and class size that results from class-size caps.[1]

Despite a general awareness of the possible endogeneity of class size, relatively little attention has been paid to how schools choose class size or to how households sort in response to those choices. In this paper, we develop a model of class-size choices by heterogeneous schools and of school choices by heterogeneous households, show that its central predictions are borne out in data on Chilean schools, and argue that these findings have important implications for attempts to estimate the effect of class size on student outcomes. Chile's educational market is well-suited to such an investigation in part because private schools account for approximately half of the market, and a majority of them are operated on a for-profit basis. This makes it straightforward to specify schools' objective functions—an otherwise difficult task in many public-sector contexts.

In the model, schools are assumed to be monopolistically competitive, to be heterogeneous in an underlying productivity parameter, and to offer quality-differentiated "products," where class size is a component of school quality. Households are assumed to be heterogeneous in income and hence in willingness to pay for quality. Schools face three constraints which correspond to real restrictions faced by Chilean schools: (1) a class size cap at 45 students, which applies to private schools accepting government subsidies; (2) an integer constraint on the number of classrooms, which applies to all schools; and (3) the restriction that enrollment (a choice variable of schools) cannot exceed demand, which also applies to all schools.

---

[1]The RD approach has also been used to study the effects of class size by Browning and Heinesen (2003) in Denmark, Dobbelsteen, Levin, and Oosterbeek (2002) in Holland, Hoxby (2000) in the U.S., McEwan and Urquiola (2005) in Chile, and Urquiola (2006) in Bolivia. For a formal treatment of identification issues in the RD context, see Hahn, Todd, and der Klaauw (2001).

The model delivers two main empirical predictions, both of which find support in the data. First, there is an inverted-U relation between class size and household income in cross-section. The model predicts that higher-income households sort into higher-productivity, higher-quality schools, as one might expect. The inverted U arises from the interaction of two effects: higher productivity both enables schools to better fill their existing classrooms and leads schools to add classrooms and reduce class size to appeal to higher income households. The former tends to dominate at lower levels of productivity, and the latter at higher levels. The inverted-U relation between class size and income will tend to confound attempts to estimate the effect of class size on student outcomes in cross-sectional regressions.

Second, in the presence of the class-size cap and the integer constraint on the number of classrooms, schools at the cap adjust price and/or enrollment to avoid having to add an additional classroom. This results in stacking at enrollment levels that are multiples of 45. Because higher-income households sort into higher-productivity schools, the stacking implies discontinuous changes in average family income and hence in other correlates of income, such as mothers' schooling, at these multiples. The resulting discontinuities violate the assumptions underlying the RD designs that have been used to estimate the effect of class size. Our results thus provide a concrete illustration of how endogenous sorting around discontinuities may invalidate RD designs (Lee, 2005; McCrary, 2005). We view these results as a cautionary note that such designs should not be applied in contexts where schools are able to set prices and influence their enrollments, and parents have substantial school choice.[2] As we discuss below, we have no reason to believe that this conclusion generalizes to previous studies, which we interpret as focusing on situations in which students are required to attend local schools, and in which schools cannot control their enrollments but rather react mechanically to them.

In addition to the papers cited above, our work is relevant to several existing literatures. First, it is related to theoretical models of school choice (Manski, 1992; Epple and Romano, 1998, 2002; Epple, Figlio, and Romano, 2002). In these frameworks, schools are essentially passive "clubs" whose main attribute is the average ability and income of their students. In our model, in contrast, schools actively choose the level of educational quality to supply. This comes at a cost, as we must abstract from peer effects in order to maintain tractability. In the long run it

---

[2]As we discuss below, private schools in Chile can turn away students for a wide variety of reasons, and parents in turn can use any public or private voucher school that is willing to accept their children.

would clearly be desirable to combine both peer effects and the elements of quality differentiation we emphasize.

Second, in seeking to understand the mechanisms behind the determination of class size, we view our work as complementary to Lazear (2001), which focuses on how schools allocate students with heterogeneous levels of self-discipline into classes of different sizes. We abstract from sorting within schools and instead focus on sorting between schools with different average class sizes.

Third, our results are related to studies of school choice and stratification. One strand of this literature analyzes how greater choice due to greater school district availability affects sorting outcomes (Bayer, McMillan, and Rueben, 2004; Clotfelter, 1999; Rothstein, forthcoming; Urquiola, 2005), while another considers the effects of the introduction of vouchers (Nechyba, 2003; Hsieh and Urquiola, 2006). In contrast, we focus on how a regulatory constraint on class size affects sorting outcomes in a market that is already largely liberalized.

Fourth, in its focus on how households of different incomes sort into schools of different qualities, our approach has elements in common with hedonic models of matching between heterogeneous consumers and heterogeneous producers (Rosen, 1974; Ekeland, Heckman, and Nesheim, 2004), assignment models of matching between heterogeneous workers and jobs requiring heterogeneous skills (Tinbergen, 1956; Sattinger, 1993; Teulings, 1995),[3] and the discrete-choice model that forms the basis of demand-system estimation in Berry, Levinsohn, and Pakes (1995). Our work differs from these literatures both in that we focus on the role of institutional constraints in the matching process, and in that we look for evidence of simple reduced-form patterns predicted by our model, rather than attempting to estimate underlying preference or technology parameters.

Finally, this paper is related to work on quality choice by firms (Gabszewicz and Thisse, 1979; Shaked and Sutton, 1982; Anderson and de Palma, 2001), and in particular to Verhoogen (2006), which models quality choice by Mexican firms facing heterogeneous consumers in the domestic and export markets. The application of a model of firm quality choice to the education sector appears to be novel. The main advantages of this paper over the existing quality-choice literature are that class size is arguably a better measure of product quality than has been previously available, and that we allow for—and have data on—consumer heterogeneity at the household, rather than market, level.

---

[3]Nesheim (2002) integrates the peer-effects mechanism discussed above into a hedonic model along the lines of Ekeland, Heckman, and Nesheim (2004).

The remainder of the paper is organized as follows. Section 2 provides institutional background, and section 3 sets out the model. Section 4 describes the data. Section 5 discusses testable implications and presents the results. Section 6 concludes.

2   Chile's School System

We focus on Chile's primary (K-8) school sector, which comprises three types of schools:

1. *Public* or *municipal* schools are run by roughly 300 municipalities which receive a per-student "voucher" payment from the central government. These schools cannot turn away students unless oversubscribed, and are limited to a maximum class size of 45.[4] In most municipalities, they are the suppliers of last resort.

2. *Private subsidized* or *voucher* schools are independent, and since 1981 have received exactly the same per student subsidy as municipal schools.[5] They are also constrained to a maximum class size of 45, but unlike public schools, have wide latitude regarding student selection.

3. *Private unsubsidized* schools are also independent, but receive no explicit subsidies.

We focus on primary schools because class size, a central variable in our analysis, is more clearly defined at the primary than at the secondary level.

Private schools (both voucher and unsubsidized) account for about 40 percent of all schools, and voucher schools alone account for about 34 percent. In urban areas, these shares are 58 and 47 percent, respectively. Private schools can be explicitly for-profit, and using their tax status to classify them, Elacqua (2005) calculates that about 70 percent of them are indeed operated as such. Further, even non-profit schools can legally distribute dividends to principals or board members. A handful of private schools are run by privately or publicly held corporations that control chains of schools, but the modal one is owned and managed by a single principal/entrepreneur.

Public primary schools are not allowed to charge "add on" tuition supplemental to the voucher subsidy.[6] While initially voucher private schools were subject to the same constraint, this restric-

---

[4]In very few instances schools are temporarily authorized to have classes of 46 or 47, but they receive no payments for the students above 45.

[5]The payment varies somewhat by location, but within an area voucher and municipal schools receive equal payments. For further details on the creation of the voucher system, see Hsieh and Urquiola (2006).

[6]Public secondary schools can charge add-ons, but in practice very few do

tion was eased beginning with the 1994 school year. Since then, they have been able to charge tuition as high as four times the voucher payment. The resources these institutions raise through tuition are equal to about 20 percent of their State funding, although their distribution is highly unequal.

A final relevant fact is that as elsewhere, primary schools in Chile are not large; 95 percent of urban ones have fewer than 135 students in the 4th grade.[7] As Figure 1 illustrates, they therefore run relatively few classes per grade. In 2002, for instance, 53 percent of urban private schools had only one 4th grade class, while 86 and 95 percent had two or fewer or three or fewer, respectively. Public schools run a slightly higher average number of classes, but 91 percent of them still operate three or fewer 4th grades. Below, we use these facts to motivate an integer constraint on the number of classrooms per grade.

3　The Model

This section develops a model of quality differentiation and sorting in the Chilean school market. We model parents' demand for education in a standard discrete-choice framework with quality differentiation (McFadden, 1974; Anderson, de Palma, and Thisse, 1992; Berry, Levinsohn, and Pakes, 1995). We solve the optimization problems of private, profit-maximizing schools in two cases corresponding to the different constraints facing unsubsidized and voucher schools.

To simplify the model, we take the set of schools in each segment of the private education market—unsubsidized and voucher—as given. This is a strong assumption, but our view is that including a detailed analysis of entry and possible switching between segments would add more tedious complication than real insight. Under the assumption that each school thinks of itself as small relative to the market as a whole, the extent of entry would not affect the optimizing decisions of particular schools, and our two main implications would continue to hold. For these reasons, we focus on the optimizing decisions of schools conditional on being in the market.

It is worth emphasizing that our two main implications do not hold for all possible parameter values in our model. Rather, we show that there exists a set of parameter values for which the implications do hold. In Section 4, we examine whether there is empirical support for the

---

[7]For reasons discussed below, we focus on urban schools and primarily on 4th grade observations. The results for the full sample and for other grades, however, are quite similar.

5

implications.

3.1 Basic Set-up

Schools are assumed to be heterogeneous in a productivity parameter λ, which one can think of as the ability of their principal/entrepreneur or their reputation. In each market segment, there is a continuum of schools with density $f_m(\lambda)$ over the interval $[\underline{\lambda}_m, \overline{\lambda}_m)$, where $m = u$ for "unsubsidized" or $m = v$ for "voucher." The λ parameter is a fixed characteristic, and identifies each school uniquely within each segment. To simplify notation, we do not include a subscript on λ indicating the market segment; this should be clear from the context.

There is a continuum of households of mass M, heterogeneous in income. Each is assumed to have one child and to enroll the child in a private school.[8] We assume that households have the following indirect utility function:

$$U(p, q \,;\, \theta) = \theta q - p + \varepsilon \tag{1}$$

where q is school quality, p is tuition, and ε is a random term capturing the utility of a particular household-school match. This specification follows from a direct utility function in which households differ only in income and in which θ, a monotonically increasing function of household income, can be interpreted as households' willingness to pay for quality.[9] We assume that θ has a distribution $g(\theta)$ with positive support over $(\underline{\theta}, \overline{\theta})$ where $\underline{\theta}, \overline{\theta} > 0$, reflecting the underlying distribution of income among households. We assume the random-utility term ε is i.i.d. across households with a double-exponential distribution with c.d.f. $F(\varepsilon) = \exp\left[-\exp\left(-\frac{\varepsilon}{\mu} + \gamma\right)\right]$,[10] where μ is a positive constant that captures the degree of differentiation between schools.[11]

---

[8]Including the public school sector in the choice set would add additional terms to the demand function below, but would not alter the main conclusions.

[9]Suppose:

$$\widetilde{U}(z, q) = u(z) + q + \widetilde{\varepsilon}$$

where z is a non-differentiated numeraire good, q is the quality of education, $\widetilde{\varepsilon}$ is a mean-zero random term, and the sub-utility function $u(\cdot)$ has $u'(\cdot) > 0$ and $u''(\cdot) < 0$. If households are on their budget constraint, then indirect utility is:

$$\widetilde{U}(p, q \,;\, y) = u(y - p) + q + \widetilde{\varepsilon}$$

where p is tuition. Taking a first-order approximation of $u(\cdot)$ around y, and setting $\theta \equiv 1/u'(y)$, $U \equiv \frac{\widetilde{U}}{u'(y)} - \frac{u(y)}{u'(y)}$, and $\varepsilon \equiv \frac{\widetilde{\varepsilon}}{u'(y)}$, we have (1). Note that the $\frac{u(y)}{u'(y)}$ term is constant across schools and does not affect the household's choice probabilities.

[10]We assume $\gamma = 0.5772$ (Euler's constant), ensuring that the expectation of ε is zero.

[11]As $\mu \to 0$, the distribution of household-school-specific utility terms collapses to a point, and the model

A standard derivation yields the probability that a household chooses school λ in a given segment, conditional on having willingness to pay for quality θ:[12]

$$s(\lambda \mid \theta) = \frac{1}{\Omega(\theta)} \exp\left(\frac{\theta q - p}{\mu}\right) \tag{2}$$

where

$$\Omega(\theta) \equiv \int_{\underline{\lambda}_v}^{\overline{\lambda}_v} \exp\left(\frac{\theta q - p}{\mu}\right) f_v(\lambda)\, \mathrm{d}\lambda + \int_{\underline{\lambda}_u}^{\overline{\lambda}_u} \exp\left(\frac{\theta q - p}{\mu}\right) f_u(\lambda)\, \mathrm{d}\lambda$$

We assume schools cannot discriminate among households, and price and quality are equal for all households in a given school. Equation (2) represents the expected demand of a household with willingness to pay θ for school λ. As is common in monopolistic-competition models, we treat individual schools as small relative to the market as a whole, and assume that they ignore their effect on the aggregate $\Omega(\theta)$.

The expected market share of school λ, integrating over all households, is:

$$s(\lambda) = \int_{\underline{\theta}}^{\overline{\theta}} s(\lambda|\theta) g(\theta) \mathrm{d}\theta \tag{3}$$

Demand is then:

$$d(\lambda) = M s(\lambda) \tag{4}$$

Demand for school λ is declining in price and increasing in quality, and higher-θ households are more sensitive to quality for a given price. Note that this specification combines horizontal differentiation, in the sense that if all schools' tuitions are equal each will face positive demand with positive probability, with vertical differentiation, in the sense that if tuitions are equal, higher-quality schools will face higher demand. Throughout we will assume schools are risk-neutral, and ignore the fact that the expression for $d(\lambda)$ represents an expectation.

Each school is constrained to offer just one "product," and is assumed to produce quality with a technology

$$q = \lambda \ln\left(\frac{T}{x/n}\right) \tag{5}$$

where x is enrollment, n is the number of classrooms, the denominator is class size, and T is a

approaches perfect competition.

[12] See for example Anderson, de Palma, and Thisse (1992, theorem 2.2, p. 39).

constant that represents the technological maximum of class size. The term in parentheses is by assumption always greater than or equal to one. This specification captures the idea that the larger is class size the less teacher attention is available for each individual student.[13] Note that a given reduction in class size raises quality more at higher-λ schools. This complementarity will be crucial in what follows.

Combining (2), (3), (4) and (5), we have the following expression for demand:

$$d(\lambda) = \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{\Omega(\theta)} \left(\frac{nT}{x} \right)^{\frac{\lambda\theta}{\mu}} e^{-\frac{p}{\mu}} M g(\theta) \mathrm{d}\theta \tag{6}$$

In order to guarantee an interior solution for the school's optimization problem we must impose a lower bound on the degree of differentiation between schools. The condition:

$$\mu > \overline{\lambda}_m \overline{\theta} \quad \text{for} \quad m = u, v \tag{7}$$

ensures that the exponent on the $\left(\frac{nT}{x} \right)$ term in (6) is less than one, and will be sufficient. Intuitively, the lower bound on the degree of differentiation between schools limits the extent to which demand for a school increases with a given class-size reduction.

It will be convenient to define the average willingness to pay of households that send their children to school λ:

$$\Theta(\lambda) \equiv E(\theta|\lambda) = \int_{\underline{\theta}}^{\overline{\theta}} \theta \left[\frac{s(\lambda|\ \theta)g(\theta)}{s(\lambda)} \right] \mathrm{d}\theta \tag{8}$$

where by Bayes' rule the term in brackets represents the probability density of θ conditional on households sending their children to school λ.[14]

To derive schools' profits, let p be the tuition schools charge, and let τ be the per-student subsidy they receive from the government (which is zero for private unsubsidized schools). We assume that τ is greater than the average variable cost per student if schools have class sizes

---

[13] An interesting extension might be to include an endogenous term in the numerator representing teacher quality, an additional choice variable for schools. We leave this task for future work, in part because we do not have data on teacher salaries or other teacher characteristics.

[14] Note that the assumption that schools cannot price discriminate implies that the price term can be brought outside the integral in (6), and Θ can be written as:

$$\Theta(\lambda) = \frac{\int_{\underline{\theta}}^{\overline{\theta}} \frac{\theta}{\Omega(\theta)} \left(\frac{nT}{x} \right)^{\frac{\theta\lambda}{\mu}} g(\theta) \mathrm{d}\theta}{\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{\Omega(\theta)} \left(\frac{nT}{x} \right)^{\frac{\theta\lambda}{\mu}} g(\theta) \mathrm{d}\theta} \tag{9}$$

of 45: $\tau > c + \frac{F_c}{45}$. We also assume that τ is less than the tuition that schools would charge in the absence of the subsidy; this will eliminate the possibility that the optimal price is negative. Suppose there is a fixed cost F_s of running a school, a fixed cost F_c of operating a classroom, and a constant variable cost c for each student. Profit is then:

$$\pi(p, n, x; \lambda) = (p + \tau - c)\, x - nF_c - F_s \tag{10}$$

The presence of F_s generates increasing returns to scale at the school level. There is no cost of differentiation. As a consequence, every school differentiates its "product" and has a monopoly over the product it offers.

3.2 Schools' Optimization Problem

The problem facing schools is to maximize profit over the choice of the number of classrooms, tuition, and enrollment:

$$\max_{p,x,n} \ \pi(p, x, n; \lambda) \tag{11}$$

This optimization is subject to three constraints, not all of which apply at all times:

1. Enrollment cannot exceed demand:

$$x \leq d(\lambda) \tag{12}$$

   where $d(\lambda)$ is given by (6). This constraint applies in all cases.[15]

2. The number of classrooms must be a positive integer

$$n \in \mathbb{N} \tag{13}$$

   where $\mathbb{N}$ is the set of natural numbers $\{1, 2, 3, ...\}$. As discussed above, this restriction is relevant to Chilean primary schools, the vast majority of which run three or fewer classes per grade. Given their generally small size, it would probably also apply to primary schools in most countries.

---

[15]This constraint ends up binding in every case we present here, and we could treat it as an equality constraint or substitute $d(\lambda)$ for x in (10) and (11). But there exist realistic cases in which it does not bind—i.e. if there is also a tuition constraint, as there was for voucher schools pre-1994. For conceptual clarity, we leave the constraint as an inequality.

3. The class-size cap:

$$\frac{x}{n} \le 45 \tag{14}$$

which applies only to voucher schools in Chile. Class size caps are certainly not universal, but are also relevant in jurisdictions including other countries and states in the U.S.

For expository purposes, we consider first the case of private unsubsidized schools, which are not subject to the class-size cap, and then the case of voucher schools. The integer restriction complicates the solution to the optimization problem, since we cannot simply solve a set of first-order conditions. A common approach is to first relax this constraint, and then compare solutions with and without the relaxation, and this is how we proceed below. In the main text, we report key results; derivations of those results appear in Appendix A, with section numbers corresponding to the cases. To reduce clutter, we do not write explicitly the dependence of the endogenous variables—price, enrollment, number of classrooms, demand, and average willingness to pay—on λ, but this dependence should be understood.

Case 1: Private Unsubsidized Schools

Case 1.1: Divisible Classrooms

In this case, schools are subject only to the constraint that enrollment not exceed demand (12). The state subsidy, τ, is zero, but we leave it in the equations for purposes of comparison with the cases that follow. The unique solution to the first-order conditions for a given value of λ is:

$$p^* = \mu + c - \tau + \lambda\Theta \tag{15a}$$

$$x^* = d \tag{15b}$$

$$n^* = \frac{x^*\lambda\Theta}{F_c} \tag{15c}$$

where d is given by (6) and Θ by (8). In order for the second-order conditions to be satisfied and for this solution to represent a maximum, it must be the case that:

$$\Psi \equiv \Theta - \frac{\lambda\sigma_{\theta|\lambda}^2}{\mu} > 0 \tag{16}$$

where $\sigma^2_{\theta|\lambda}$ is the variance of θ among households with children attending school λ.[16] Assumption (7) guarantees that this condition holds. (See Appendix A.1.1.)

Unfortunately, there is no explicit analytical solution to (15a)-(15c). But using the implicit function theorem, we can nonetheless sign the relationship between the various endogenous variables and the underlying productivity parameter, λ. In particular:

$$\frac{\partial p}{\partial \lambda} > 0 \tag{17a}$$

$$\frac{\partial x}{\partial \lambda} > 0 \tag{17b}$$

$$\frac{\partial n}{\partial \lambda} > 0 \tag{17c}$$

$$\frac{\partial \Theta}{\partial \lambda} > 0 \tag{17d}$$

$$\frac{\partial}{\partial \lambda} \left(\frac{x}{n} \right) < 0 \tag{17e}$$

Higher-λ schools charge higher tuition, have larger enrollments, operate more classrooms, have smaller class sizes, and attract students whose families are on average wealthier and have higher willingness to pay for quality. All of these relationships are monotonic in λ. The monotonically declining relationship between class size and λ is illustrated in Figure 2.

It will be convenient to define $\gamma_k \equiv n^{*-1}(k)$, where $n(\cdot)$ is the function defined by (15c), and the fact that $\frac{\partial n}{\partial \lambda} > 0$ everywhere ensures that it is invertible. Then γ_k is the value of λ at which the optimal number of classrooms is k in this divisible-classrooms case.

Case 1.2: Indivisible Classrooms

Now add the restriction that the number of classrooms must be an integer (13), maintaining the demand constraint (12). Our strategy for dealing with the integer constraint is to first solve the optimization problem for a given number of classrooms, and to then characterize the sets of schools that choose each integer number of classrooms.

To begin, suppose that n is fixed and think of it as a parameter. The solution to the opti-

---

[16]That is, define:

$$\sigma^2_{\theta|\lambda} \equiv \int_{\underline{\theta}}^{\overline{\theta}} \theta^2 \left[\frac{s(\lambda|\,\theta)g(\theta)}{s(\lambda)} \right] \mathrm{d}\theta - \Theta^2$$

11

mization problem of school λ is:

$$p^* = \mu + c - \tau + \lambda\Theta \tag{18a}$$

$$x^* = d \tag{18b}$$

where d is given by (6) and Θ by (8).

Price and average willingness to pay (i.e. average household income) are again unambiguously increasing in λ:

$$\frac{\partial p}{\partial \lambda} > 0 \tag{19a}$$

$$\frac{\partial \Theta}{\partial \lambda} > 0 \tag{19b}$$

where we use partial derivatives to indicate that n is being held constant.

There is a subtlety in the relationship between enrollment and λ. On one hand, there is a direct effect of a higher λ on demand: for a given class size, households prefer higher-λ schools. On the other hand, there is an indirect effect: as is evident in (6), at higher values of λ a given increase in enrollment has a larger negative effect on demand. It is theoretically possible in this case that the latter effect dominates, making it optimal for higher-λ schools to raise prices such that enrollment, conditional on a given number of classrooms, is decreasing in λ. In that case, our testable implications (discussed in the introduction and in more detail below) do not hold. We focus instead on the case where enrollment is increasing in λ. A necessary and sufficient condition for this, which we assume hereafter, is:[17]

$$\ln\left(\frac{nT}{x}\right) > \frac{\Theta}{\Psi} \tag{20}$$

---

[17]If we replace the production function for quality (5) by a general function $q(x, n, \lambda)$, then the condition is:

$$-\frac{\partial^2 q}{\partial x \partial \lambda} < \left(\frac{\sigma_{\theta|\lambda}^2}{\mu\Theta}\frac{\partial q}{\partial x} + \frac{1}{x}\right)\frac{\partial q}{\partial \lambda}$$

which makes it clear that the indirect effect of higher λ described above (represented by $\frac{\partial^2 q}{\partial x \partial \lambda}$) must be small in magnitude relative to the direct effect (represented by $\frac{\partial q}{\partial \lambda}$).

where Ψ is defined as in (16).[18] Under this assumption, we have

$$\frac{\partial x}{\partial \lambda} > 0 \tag{21}$$

Since n is fixed, (21) implies that $\frac{\partial}{\partial \lambda}\left(\frac{x}{n}\right) > 0$; for a given number of classrooms, class size is *increasing* in λ. Conditional on n, higher-λ schools are better able to fill their classrooms.

Now consider the issue of which integer number of classrooms schools choose. Let

$$\pi^*(n, \lambda) = \pi(p^*(n, \lambda), x^*(n, \lambda), n; \lambda) \tag{22}$$

be school λ's optimal profit when the number of classrooms is fixed at n, where $p^*(n, \lambda)$ and $x^*(n, \lambda)$ are given by (18a) and (18b). Define Λ_k to be the set of all schools for which a given integer k is the optimal number of classrooms:

$$\Lambda_k = \{\lambda : \pi^*(k, \lambda) \geq \pi^*(j, \lambda) \ \forall \ j \neq k, \ j \in \mathbb{N}\} \tag{23}$$

Note that γ_k, the unique value of λ at which the optimal number of classrooms is k in the divisible-classrooms case, is an element of Λ_k.[19] The following lemma characterizes the sets Λ_k.

Lemma 1. *In Case 1.2, under assumptions (1), (5), (7), (10), and (20), there exist unique positive integers $\underline{k}$ and $\overline{k}$, and a unique set of critical values $\rho_{\underline{k}-1}, \rho_{\underline{k}}, ..., \rho_{\overline{k}-1}, \rho_{\overline{k}}$ such that:*

$$\Lambda_k \quad = \quad \{\lambda : \rho_{k-1} \leq \lambda < \rho_k\} \ for \ k = \underline{k}, \underline{k}+1, ..., \overline{k} - 1, \overline{k}$$

where $\underline{\lambda}_u = \rho_{\underline{k}-1} < \rho_{\underline{k}} < ... < \rho_{\overline{k}} = \overline{\lambda}_u$.

For the proof, see Appendix A.1.2. The lemma indicates that the set of private unsubsidized schools can be partitioned into a set of intervals, $[\rho_{\underline{k}-1}, \rho_{\underline{k}})$, $[\rho_{\underline{k}}, \rho_{\underline{k}+1})$ etc., where the optimal integer number of classrooms is $\underline{k}$ in the first interval, $\underline{k}+1$ in the next, and so on. Within each of the intervals, the results (18a)-(19b) and (21) hold.

---

[18]Since $\Psi \leq \Theta$, a sufficient condition is simply that $T \geq e * \left(\frac{x}{n}\right)$.

[19]By the definition of γ_k:

$$\pi^*(k, \gamma_k) > \pi^*(j, \gamma_k) \ \forall \ j \neq k, \ \text{where} \ j \in \mathbb{R}$$

Since this is true $\forall \ k \in \mathbb{R}$, it must be true $\forall \ j \in \mathbb{N}$. Hence $\gamma_k \in \Lambda_k$.

The appendix shows that at the critical values $\rho_{\underline{k}}, \rho_{\underline{k}+1}, ..., \rho_{\overline{k}-1}$, there are positive discontinuities in tuition, enrollment and average willingness to pay, and negative discontinuities in class size. The facts that tuition, enrollment and average willingness to pay are increasing in λ between the critical values and that they jump up at the critical values imply that all are monotonically increasing in λ for all λ.

Figure 3 plots class size vs. λ in the case where $\underline{k} = 1$ and $\overline{k} = 6$. The fact that $\gamma_k \in \Lambda_k$ means that each upward-sloping portion of the graph intersects the graph for the continuous-n case, indicated by the dashed line. The result is a saw-tooth pattern around the downward-sloping curve (represented by a dashed line) from Case 1.1.

Case 2: Voucher Schools

Case 2.1: Divisible Classrooms

As stated above, private voucher schools are subject to a policy-induced constraint: the class size cap at 45. In this case, with divisible classrooms, there is a single critical value of λ, call it α, below which the class-size cap binds and above which it does not. If $\lambda \leq \alpha$ and the cap binds:

$$p^* = c - \tau + \mu + \frac{F_c}{45} \tag{24a}$$

$$x^* = d \tag{24b}$$

$$n^* = \frac{x^*}{45} \tag{24c}$$

where d is given by (6) and Θ by (8). The slopes with respect to λ in this sub-case are:

$$\frac{\partial p}{\partial \lambda} = 0 \tag{25a}$$

$$\frac{\partial x}{\partial \lambda} > 0 \tag{25b}$$

$$\frac{\partial n}{\partial \lambda} > 0 \tag{25c}$$

$$\frac{\partial \Theta}{\partial \lambda} > 0 \tag{25d}$$

$$\frac{\partial}{\partial \lambda}\left(\frac{x}{n}\right) = 0 \tag{25e}$$

Although class size and (perhaps surprisingly) price are constant, enrollment, the number of classrooms, and average household income are increasing in the productivity parameter. The last

14

fact is true because λ raises quality conditional on class size.

If $\lambda > \alpha$ and the class-size cap does not bind, then we are back in the no-class-size-cap case (Case 1.1) but with $\tau > 0$. The solution is given by (15a)-(15c) and the results (17a)-(17e) again hold. The critical value α is defined implicitly by the equation

$$\alpha\Theta(\alpha) = \frac{F_c}{45} \tag{26}$$

This is the value of λ at which $\frac{x}{n} = 45$ in Case 1.1. Note that there is no guarantee that $\alpha \in (\underline{\lambda}_v, \overline{\lambda}_v)$, i.e. that the critical value falls in the relevant range of λ. At the critical value, the optimal choices p^*, x^*, and n^* are equal in the two sub-cases (cap binding, cap non binding). Hence p^*, x^* and n^* are continuous, p^* is weakly monotonically increasing and x^* and n^* are strictly monotonically increasing over the entire range of λ.

Again it will be convenient to define the value of λ at which the optimal number of classrooms is k: $\delta_k \equiv n^{*-1}(k)$ where $n^*(\cdot)$ is defined by (24c) in the binding cap sub-case and by (15c) in the non-binding-cap sub-case, with invertibility following from (25c) and (17c). Also, note that in both sub-cases $p^* > 0$ as long as the subsidy, τ, is less than what a school would charge in the absence of the subsidy, as we have assumed.

Figure 4 illustrates the relationship between class size and λ in the case where $\underline{\lambda}_v < \alpha < \overline{\lambda}_v$. To the left of α, class size is flat at 45; to the right, it coincides with the case of private unsubsidized schools illustrated in Figure 2.

Case 2.2: Indivisible Classrooms

Now consider the case of voucher schools with indivisible classrooms. We proceed as in Case 1.2 above, first solving the optimization problem for a given value of n, and then characterizing the sets of schools that choose each n.

Fix n. For a given n there is a single critical value of λ, call it β_n, *above* which the class-size cap binds and *below* which it does not. If $\lambda \leq \beta_n$ and the class-size cap does not bind, then the solution to the school's optimization problem is the same as for private unsubsidized schools in Case 1.2, given by (18a)-(18b), with slopes vs. λ given by (19a), (19b), and (21). As in that case, price and average willingness to pay are unambiguously increasing in λ, and under assumption (20) enrollment and class size are increasing in λ as well.

15

If $\lambda > \beta_n$ and the class-size cap binds, then we have two endogenous variables and two binding constraints. The constraints pin down the optimal values of p and x:

$$x^* = 45n \tag{27a}$$

$$p^* = \mu \ln \Sigma - \mu \ln(45n) \tag{27b}$$

where

$$\Sigma \equiv \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{\Omega(\theta)} \left(\frac{T}{45} \right)^{\frac{\theta \lambda}{\mu}} g(\theta) \mathrm{d}\theta \tag{28}$$

The slopes with respect to λ are:

$$\frac{\partial p}{\partial \lambda} > 0 \tag{29a}$$

$$\frac{\partial x}{\partial \lambda} = \frac{\partial}{\partial \lambda} \left(\frac{x}{n} \right) = 0 \tag{29b}$$

$$\frac{\partial \Theta}{\partial \lambda} > 0 \tag{29c}$$

The critical value β_n is defined implicitly by the equation:

$$\beta_n \, \Theta(\beta_n) = \mu \ln \Sigma - \mu \ln(45n) - c + \tau - \mu \tag{30}$$

where Θ is given by (8). The critical value of λ is the point at which class size reaches 45 in Case 1.2. At this value, the optimal choices of p and x are the same in the two sub-cases (class-size cap binding and not binding). Hence for a given n, p^*, x^*, and Θ are continuous, p^* and Θ are strictly monotonically increasing in λ, and x^* and $\frac{x^*}{n}$ are weakly monotonically increasing in λ.

Define Λ_k as the set of schools for which k is the optimal integer number of classrooms, as in (23). We again have a lemma to characterize these sets:

Lemma 2. *In Case 2.2, under assumptions (1), (5), (7), (10), and (20), there exist unique integers $\underline{k}$ and $\overline{k}$ and a unique set of critical values $\nu_{\underline{k}}, \nu_{\underline{k}+1}, ... \nu_{\overline{k}-1}, \nu_{\overline{k}}$ such that:*

$$\Lambda_k = \{ \lambda : \nu_{k-1} \leq \lambda < \nu_k \} \text{ for } k = \underline{k}, \underline{k}+1, ..., \overline{k}$$

where $\underline{\lambda}_v = \nu_{\underline{k}-1} < \nu_{\underline{k}} < ... < \nu_{\overline{k}-1} < \nu_{\overline{k}} = \overline{\lambda}_v$.

The proof is in Appendix A.2.2. Within each of the subsets Λ_k the above results for fixed n hold.

Note that there is no guarantee that the value of λ at which the class-size cap starts to bind for a given integer k, β_k, is to the left of the value of λ at which it becomes optimal to add an additional classroom, ν_k. In the empirical part of the paper, we present evidence consistent with the hypothesis that $\beta_k < \nu_k$ for low values of k.

The appendix shows that at the critical values $\nu_{\underline{k}}, \nu_{\underline{k}+1}, ..., \nu_{\overline{k}-1}$, enrollment is increasing, average willingness to pay is non-decreasing, and class size is non-increasing. Hence enrollment is strictly monotonically increasing and average willingness to pay is weakly monotonically increasing in λ for all λ.[20]

Figure 5 plots class size against λ for the case where $\beta_k < \nu_k$ for $k = 1$ and $k = 2$ but not thereafter. Again, the curves for each integer k intersect the curve from the divisible-classrooms case (dashed line) at the values of λ at which the optimal n for the divisible-classrooms case (Case 2.1) is an integer.

Figure 5 exhibits an approximately inverted-U relationship between class size and λ. It results from the interaction of two effects: (1) conditional on a value of n, class size is increasing in λ, since greater values of λ make schools better able to fill their classrooms; and (2) across values of n, class size is declining in λ, since greater λ leads schools to increase the number of classrooms, reducing class size. While λ is not observable, the model predicts that average willingness to pay, Θ, and hence average household income are (weakly) monotonically increasing in λ. We thus have our first testable implication:

Testable Implication 1 There is an approximately inverted-U relationship between class size and average household income in equilibrium.

Figure 5 also illustrates that schools between β_1 and ν_1 have enrollment 45 and schools between β_2 and ν_2 have enrollment 90. Intuitively, in these regions schools raise tuition rather than incurring the fixed cost of starting a new classroom.[21] Although we do not model the possibility explicitly, one can easily imagine that in the presence of stochasticity in demand and menu costs of

---

[20]The direction of the change in price at each critical value is ambiguous in this case. The slope $\frac{\partial p}{\partial \lambda}$ is greater when the class-size cap binds than when it does not bind, since higher λ leads schools raise prices to keep class-size pegged at 45. Consequently, price may be greater to the left of ν_k than to the right.

[21]The appendix shows that for a given number of classrooms, tuition (p) is more steeply sloped in λ in the region where the class-size cap binds than in the region where the cap does not bind.

changing tuition, schools might turn away potential students for the same reason.[22] Since average household income is monotonically increasing in λ, the stacking implies discontinuous changes in average household income with respect to enrollment at enrollments of 45, 90, and so on.[23] More generally, we have our second testable implication:

Testable Implication 2 Schools may stack at enrollments that are multiples of 45, implying discontinuous changes in average household income with respect to enrollment at those points.

4   Data

To examine our model's implications, we draw on two sources of information. The first is administrative information on schools' grade-specific enrollments and the number of classrooms they operate, from which we calculate their average class sizes. The second is data from the SIMCE testing system[24] which tracks schools' math and language performance. SIMCE data are available at the school level since 1988. Since 1997, they also exist at the individual level and include information on students' household income, parental schooling, and other characteristics (this information is collected via a questionnaire sent home for parental response).

Depending on the year, the SIMCE tests 4th, 8th, or 10th graders. We focus on the 4th grade because class size, one of the key variables in our model, is best-defined in early primary grades. We also focus on the 2002 cross section because it is the most recent 4th grade testing round for which we have complete data.[25] We note, however, that the general conclusions we obtain emerge in other cross-sections we have analyzed (for instance, the 1999 4th grade and the 2004 8th grade waves).[26]

Table 1 presents descriptive statistics for each type of school in 2002. In cross-section, students in private unsubsidized schools tend to be of higher socioeconomic status than students

---

[22]In Chile, private schools have wide latitude regarding student selection, and can turn away students for reasons ranging from the desire to maintain a given class size, to the desire to maintain religious uniformity.

[23]Technically speaking, the model predicts a discontinuity in Θ even in the absence of stacking. Average willingness to pay, Θ, is a smooth, monotonically increasing function of class size, $\frac{x}{n}$. (See (59) and (60) in Appendix A.2.2.) Class size jumps discontinuously at the critical values ν_k. Hence Θ must also jump at ν_k. Our view, however, is that the discontinuity due to stacking is the more empirically important one.

[24]SIMCE, which stands for *Sistema de Medición de la Calidad de la Educación* (Educational Quality Measurement System), is Chile's standardized testing program.

[25]The 2003 and 2004 testing rounds were for the 10th and 8th grades, respectively.

[26]The 8th grade is still part of primary school in Chile.

in voucher schools. Their household income and mothers' and fathers' schooling are higher, and unsurprisingly, test scores follow the same pattern. In addition, unsubsidized schools have lower average class sizes. All of these facts are consistent with the hypothesis that unsubsidized schools tend to have higher λ than voucher schools. Note, however, that these schools are typically smaller in terms of enrollment than voucher schools. On this dimension, our model is not an accurate stylization.[27] Figure 6 presents densities of log income and mothers' schooling by type of school, showing that unsubsidized schools tend to attract richer households, voucher schools middle-income households, and public schools poorer households.

5 Results

We now take the testable implications to the data. We review each implication, discuss how it relates to the existing literature, and present the empirical results. We focus on urban schools because we want to consider settings where enrollment and class size are determined by schools' and households' choices, and not constrained by the size of the market, as could happen in rural areas. The conclusions of our empirical analyses, however, turn out to not be much affected by this selection.

5.1 Class Size and Income in Cross-Section: The Inverted U

As discussed in Section 3, the first testable prediction is an inverted-U relationship between class size and average household income. The upward-sloping portion in such a pattern reflects the fact that low-λ schools may have trouble filling their existing classroom(s) to achieve the desired class size. The downward-sloping portion reflects the fact that among the higher-λ schools used by higher-income households, quality considerations dominate: these schools find it profitable to restrict class size and charge higher tuition. These mechanisms are consistent with anecdotal evidence from Chile, where there is a widespread perception that many lower-quality voucher schools are small "mom and pop" operations that struggle to fill their classrooms. In contrast, voucher schools run by larger firms have sufficient demand to operate multiple classrooms, and are generally perceived to be of higher quality.

---

[27]In a model with peer effects, schools might have an incentive to keep enrollment low to avoid diluting the quality of the student pool. In the long run, it would be important to incorporate such concerns into a framework such as ours.

Figure 7 plots class size against log average household income among all private urban schools (Panel A), voucher schools only (Panel B), and unsubsidized schools only (Panel C). In all three panels, the thicker line plots fitted values of a locally weighted regression of class size on log income, and the thinner lines plot the fitted values, along with 95 percent confidence intervals, of a regression of class size on a fifth-order polynomial in log income. In Panels A and B, a clear overall inverted-U pattern is observed. The pattern is driven by the voucher schools, which make up more than 75 percent of the private school market. There is no inverted U among the unsubsidized schools (panel C); this is consistent with the idea that unsubsidized schools are high-λ institutions for which filling classrooms is not the primary challenge. Panels D, E and F present similar evidence using mothers' schooling rather than income on the x-axis. The overall patterns also describe an inverted U, and illustrate that among all urban private schools, average class size rises with mothers' schooling up to about the point where the average mother is a high school graduate, and declines thereafter.

A possible concern with these figures is that the inverted-U patterns reflect the composition of schools across regions, rather than cross-sectional patterns within markets. To examine this possibility, Table 2 reports simple regressions of class size on polynomials in log income (Panel A) and mother's schooling (Panel B). To facilitate the interpretation of the coefficients, we use second-order polynomials rather than the fifth-order polynomials used in Figure 7. The inverted-U pattern is not sensitive to the order of the polynomial, and figures that control for regional composition are available from the authors. Columns 1, 4 and 7 report results without region dummies for all private schools, voucher schools and unsubsidized private schools respectively; Columns 2, 5 and 8 include dummies for each of Chile's 13 regions; and Columns 3, 6 and 9 include dummies for 318 communes or municipalities. Among all private schools or voucher schools, the quadratic term is uniformly negative and significant, and not much affected by the regional controls. The inverted-U patterns seem to hold even within much more narrowly defined urban markets.[28]

Another potential concern is that average household income is a poor proxy for school productivity, λ, and that the inverted U is generated by a mechanism unrelated to the school-quality choice that we model. To investigate this, Figure 8 plots a locally weighted regression of tuition

---

[28]We have replicated these results using using 8th rather than 4th grade class size as the dependent variable, with similar results.

against average household income, as well as a fitted fifth-order polynomial and 95 percent confidence interval, as in Figure 7. Consistent with the predictions of the model, we find a strong positive correlation between tuition and income. While ours is obviously not the only model that would predict such a correlation, we take this result as reassuring that our model is consistent with first-order patterns in the data and that average household income is a correlate of school quality.

The inverted-U finding is relevant to the literature on the effect of class size on student achievement. In this literature, it is common to see cross-sectional estimates that are of the "wrong" sign or essentially equal to zero. Since student achievement tends to be strongly correlated with household income and since imprecision in measurement makes it is difficult to control completely for differences in income, the inverted-U pattern suggests that cross-sectional regressions are likely to understate the effect of class size reductions among lower-income voucher schools, and to overstate it among higher-income ones.[29]

Additionally, previous work has revealed positive correlations between class size and enrollment and between enrollment and household socioeconomic status among public schools in Israel (Angrist and Lavy, 1999) and Bolivia (Urquiola, 2006),[30] but to our knowledge our paper is the first to provide a theoretical rationale or empirical evidence for a non-linear relationship between class size and household income. We conjecture that the inverted-U pattern is likely to arise among private primary schools in other countries.[31]

5.2 Stacking at Multiples of Class-Size Cap

Our second testable implication is related to regression discontinuity (RD) designs that exploit the discontinuous relation between enrollment and class size induced by class-size caps.[32] Figure 9 shows that the Chilean setting appears to be a promising one for an RD-based evaluation of the effect of class size. The solid line plots the relation between class size and enrollment that

---

[29]Consistent with Figure 7 (panel A), for instance, we find that a cross-sectional bi-variate regression of test scores on class size among all urban private schools results is an insignificant point estimate. If the sample is restricted to schools with mean mothers' schooling below 12 years of age, however, the coefficient is positive and significant. If it is restricted to schools with a mean above 12, it is negative and significant at the 10 percent level.

[30]Mizala and Romaguera (2002) present evidence of a positive correlation between enrollment and household socioeconomic status in Chile.

[31]Note that the inverted-U pattern can arise even in the absence of a class size cap, as Figure 3 illustrates.

[32]For overviews and history of the RD design, see Angrist and Krueger (1999), van der Klaauw (2002), and Shadish, Cook, and Campbell (2002).

would be observed if schools mechanically expanded class size with enrollment until reaching the cap (45, in Chile), i.e., if class size were determined by:

$$\left(\frac{x}{n}\right)^p = \frac{x}{int\left(\frac{x-1}{45}\right) + 1} \tag{31}$$

where the superscript p indicates this is the predicted level. This results in a "saw-tooth" pattern in which class size increases one for one with enrollment until, at 46, a new class is added and average class size falls to 23, with other discontinuities observed at 90, 135, etc. Using data for voucher schools for 2002, the circles plot enrollment-cell means of class size, and the dotted line plots a smoothed value of these, showing that the rule predicts actual class size quite closely. Aggregated to the enrollment-cell level as in this figure, a regression of actual on predicted class size would produce an R^2 greater than 0.9—a better "first stage" than that in any RD-based class-size study we are aware of.[33]

The idea behind RD designs is that discontinuities like those in Figure 9 can be used to identify the causal effect of class size even if enrollment is systematically related to factors that affect students' outcomes. The key assumption is that enrollment is smoothly related to student characteristics and other factors that affect achievement at multiples of the class-size cap. If this is the case, students in schools with enrollments of 45 arguably provide an adequate control group for those in schools with enrollments of 46, for example, and differences in their performance can be attributed to the very different class sizes they experience.

Table 3 reports the results of a standard RD analysis using school-level data from 2002. Column (1) presents a regression of class size on a piecewise linear spline for enrollment, as in van der Klaauw (2002). The first four dummy variables indicate whether schools' enrollments are above the first four cutoffs, and their corresponding coefficients thus provide direct estimates of the average decline in class size that takes place in the vicinity of those breaks.[34] Consistent with the visual evidence in Figure 9, the first one suggests that class size drops by about 17 students at the first threshold. The declines at the first three of the four cutoffs are statistically significant,

---

[33]For visual clarity, Figure 9 excludes schools that declare 4th grade enrollments above 180 students (less than two percent of all schools), thus focusing on only the first three discontinuities in the enrollment/class size relation.

[34]For the sake of space, Table 2 and all subsequent ones exclude schools that declare 4th grade enrollments above 225 (less than one percent of all schools), thus focusing on only the first four discontinuities in the enrollment/class size relation.

and become progressively smaller.[35] In this specification, all standard errors are clustered by enrollment levels, as Lee and Card (2004) suggest is appropriate in RD settings in which the assignment variable (here enrollment) is discrete.

There is prima-facie evidence that the standard RD strategy would generate significant results. Figure 10 presents "raw" enrollment-cell means of math and language test scores, along with the fitted values of a locally weighted regression calculated within each enrollment segment. Particularly around the first cutoff, which accounts for the greatest density of schools (see Figure 1), the discrete reduction in class size is mirrored by an associated increase in average test scores. This observation is also borne out by the regression results. Columns 2-3 of Table 3 present reduced-form regressions of average math and language scores on the piecewise linear spline in enrollment. We see positive and significant increases in scores at the first cutoff, and generally positive (although not significant) increases at subsequent ones. Columns 4-5 report instrumental-variables (IV) specifications, where dummy variables for the first four cutoffs are used as instruments for class size. In both cases, class size appears to have a negative and significant effect on test scores.[36]

Focusing more narrowly around the discontinuities, Columns 1-3 of Table 4 select bands of 5 students (panels A and C for math and language, respectively) and 3 students (panels B and D) around the first three breaks.[37] The IV specifications in these columns regress schools' average test scores on class size, where the latter is instrumented by an indicator for whether schools' enrollment is above the respective cutoff. As van der Klaauw (2002) indicates, these are equivalent to Wald estimates of the effect of class size around each discontinuity.[38] Column 4 (panels A and B) produces similar estimates pooling all three local samples.[39] In these pooled samples, the point estimates of the effect of class size on test scores are uniformly negative, although they are not statistically significant.[40]

---

[35]Although we omit the results, adding controls for individuals' characteristics has essentially no effect on the key coefficients.

[36]The conclusions from Table 3 are similar if the control function includes quadratic and not just linear terms for enrollment. (For more detail on using splines of higher-order polynomials, see van der Klaauw (2002).)

[37]Separate results around the fourth cutoff are omitted for the sake of space; they account for less than 1 percent of all school observations. The erratic results in Column 2 are due to outliers close to the 90-student cutoff; when we replicate the results using 1999 data, the point estimates and standard errors around the second cutoff are in line with those around the other cutoffs.

[38]In other words, the point estimates could be replicated by dividing the difference in average test scores between the schools above and below the cutoff within each band, by the difference in their respective average class sizes.

[39]In this case dummies for whether enrollments are above the three cutoffs, $1\{x > 45\}$, $1\{x > 90\}$, and $1\{x > 135\}$, as well as three sample-specific intercepts serve as instruments; see van der Klaauw (2002).

[40]Note that clustering by enrollment level, as suggested by Lee and Card (2004), lowers significance levels.

23

One might be tempted to interpret these results as estimates of the causal effect of class size on achievement. If the second testable implication of our model is correct, however, then the smoothness conditions required for valid RD-based inference are likely to be violated.[41] First, note that in the case of voucher schools illustrated in Figure 5, there is a non-negligible mass of schools offering one or two classrooms for which the class-size cap binds. In the figure, all schools with productivity parameters between β_1 and ν_1 have enrollment 45, and all those between β_2 and ν_2 have enrollment 90. We describe this as stacking or "bunching up". Panel A of Figure 11 presents a histogram of 4th grade enrollments among urban voucher schools, and the evidence of such stacking is clear: more than 5 times as many schools report enrollments of 45 as report enrollments of 46. The same happens at higher cutoffs as well: more than 7 times as many schools have 90 4th graders as have 91, for instance.[42] Panel B shows that there is no evidence of stacking among private unsubsidized schools, which are not subject to the class-size cap; it appears that the stacking among voucher schools is not due to technological factors unrelated to the cap. Together, Panels A and B of Figure 11 provide a clear illustration of what McCrary (2005) terms manipulation of the running variable—enrollment in this case.

Second, note that the model predicts that higher-income households on average sort into higher-λ schools. If so, then the stacking will generate discontinuities in the relationship between enrollment and student characteristics close to the cutoff points, violating the smoothness assumptions underlying the RD approach. Again, Figure 5 illustrates the intuition: because of the stacking, the average value of λ among schools at the cap is strictly less than the average value just above the cap; since average household income is weakly monotonically increasing in λ (see Case 2.2 above), the stacking generates discontinuous changes in household income at the class-size cutoffs. It is worth emphasizing that stacking alone may not violate the RD assumptions in our context;[43] the violation of the RD assumptions arises from the interaction of the stacking and the endogenous sorting of households.

Panel A of Figure 12 plots the fitted values from locally weighted regressions of log average household income (calculated within enrollment cells) on enrollment. The size of each circle is

---

[41]For this reason, we do not discuss in detail the magnitudes of the effects in tables 3 and 4.

[42]Similar stacking occurs if 1st or 8th grade data are used.

[43]If student performance depended only on class size and not directly on λ, and there were no sorting of students, then students on one side of the class-size cutoff would still serve as a valid control group for students on the other side.

proportional to the number of student observations in each enrollment cell. As expected, the circles at 45, 90, and 135 are relatively large—a reflection of stacking at these points. There is clear evidence that student income changes discontinuously around the first cutoff. Schools with enrollments of 46 students have student incomes about 20 percent higher than schools with 45 students. Panel C shows, not surprisingly, that the former also have higher average mothers' schooling—more than half a year. While jumps at the subsequent cutoff points are less evident, the clear discontinuities at the first one are sufficient to cast doubt on the RD approach in this context, especially since much of the density in the distribution of voucher schools is concentrated around the first cutoff. (Refer to Figure 11). For further detail, panels B and D present the corresponding (for income and mothers' schooling, respectively) "raw" enrollment-cell means, along with the fitted values of a locally weighted regression calculated within each enrollment segment.

Columns 1-3 of Table 5 present regressions of household characteristics on the piecewise linear spline in enrollment. The results are consistent with the visual evidence from Figure 12. In particular, they confirm that income, mothers' schooling, and fathers' schooling display substantial and statistically significant jumps at the first enrollment cutoff; the coefficients for subsequent cutoffs are generally positive but not significant. Columns 4-5 show that the IV results from Columns 4-5 of Table 3 are sensitive to the inclusion of socioeconomic controls. The coefficient on class size for the math-score specification drops in magnitude from -0.7 and significant (Table 3, Column 4) to -0.1 and insignificant (Table 5, Column 4) with the inclusion of the controls. For the language-score specification, the drop is from -0.6 to -0.1. The coefficients on mothers' schooling and income are strongly significant in the test-score regressions; fathers' schooling is significant at the 10 percent level in the math-score specification.[44] This is strong evidence that the exclusion restriction required for the IV estimates in Table 3 is invalid: the cutoff dummies used as instruments are correlated with household characteristics that are omitted from the Table 3 specification, and those characteristics are in turn correlated with the test-score outcomes. For completeness, Table 6 replicates the within-band estimates from Table 4, but including controls, with similar conclusions.

In short, these results provide a concrete illustration of Lee's (2005) observation that "economic

---

[44]These results are qualitatively similar if we use the predicted class size (31) as an instrument for class size in place of the piecewise linear spline.

behavior can corrupt the RD design." It is worth emphasizing, however, that our results apply to settings in which for-profit schools can set prices and directly influence their enrollments, and in which households enjoy substantial freedom to sort between schools; we have no reason to believe that they extend to public-school contexts typically studied. For instance, Angrist and Lavy (1999) point out that in the Israeli public school context they analyze, pupils are required to attend their neighborhood schools, and schools in turn must accept applicants.[45] Further, migration and immigration may render it difficult for schools to predict enrollments, and private participation is limited to orthodox schools.[46]

6 Conclusion

The model developed in this paper offers an explanation for two distinct empirical patterns observed in the Chilean data. First, there is an inverted-U cross-sectional relationship between class size and household income, which is likely to bias non-experimental estimates of the effect of class size. Second, schools' enrollments tend to stack at multiples of the class-size cap, which, in conjunction with the sorting of households into schools of different quality, generates discontinuities in student characteristics at these points. These in turn violate the assumptions required for regression-discontinuity analyses of class size in the private-school context we analyze. The fact that a relatively parsimonious model can can account for these distinct phenomena suggests that it is a useful way to organize our thinking about class size and sorting in liberalized education markets. Our findings recommend caution in interpreting cross-sectional and RD estimates of the effect of class size in such settings, and underline the value of randomized experiments to estimate class-size effects in contexts where schools are free to set prices and/or turn away students, and households are free to sort between schools.[47]

This paper has also sought to show that a regulatory constraint on quality (the class size cap), as well as lumpiness in the provision of the service being regulated (the integer constraint),

---

[45]Similarly, in some exercises Urquiola (2006) considers Bolivian schools in rural towns in which school choice is likely to be very limited.

[46]The observation that Israeli institutions prevent strategic behavior of the kind we emphasize in this paper is consistent with the finding of Angrist and Lavy (1999) that controlling for secular enrollment effects, adding controls for the proportion of students with low socioeconomic backgrounds does not affect their key estimates.

[47]For a broader argument in favor of randomization in estimating the effects of school inputs in developing countries, see Duflo and Kremer (2004). Banerjee, Cole, Duflo, and Linden (2004) present randomized evaluations of two programs to increase teacher attention per student in India.

can have important and perhaps unexpected consequences for the matching process between heterogeneous consumers and heterogeneous producers.[48] Such consequences should be taken into account in designing regulations and policy interventions in education markets, as well as in attempting to use those institutional features as sources of exogenous variation in empirical investigations.

---

[48]In this sense, the paper seeks to contribute to the broader literature on price and quality regulation; see Sappington (2005) and Armstrong and Sappington (2006) for reviews.

A Theory Appendix

A.1 Case 1: Private Unsubsidized Schools

A.1.1 Case 1.1: Divisible Classrooms

Form the Lagrangian from (11), where the only constraint in effect is (12):

$$\mathcal{L} = (p - c + \tau)\, x - nF_c - F_s - \phi\,(x - d) \tag{32}$$

The first-order conditions for an optimum are:

$$\frac{\partial \mathcal{L}}{\partial p} = x + \phi\left(-\frac{d}{\mu}\right) = 0 \tag{33a}$$

$$\frac{\partial \mathcal{L}}{\partial n} = -F_c + \phi \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial s(\lambda|\theta)}{\partial n} Mg(\theta)\, \mathrm{d}\theta \tag{33b}$$

$$= -F_c + \phi\left(\frac{\lambda \Theta d}{\mu n}\right) = 0$$

$$\frac{\partial \mathcal{L}}{\partial x} = p - c + \tau - \phi\left(1 - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial s(\lambda|\theta)}{\partial x} g(\theta)\, \mathrm{d}\theta\right) \tag{33c}$$

$$= p - c + \tau - \phi\left(1 + \frac{\lambda \Theta s}{\mu x}\right) = 0$$

$$\frac{\partial \mathcal{L}}{\partial \phi} \geq 0,\; \phi \geq 0,\; \text{and}\; \phi\frac{\partial \mathcal{L}}{\partial \phi} = 0 \tag{33d}$$

The interchanging of the partial derivatives and the integrals is justified by a standard property of integrals (see e.g. Bartle (1976, Theorem 31.7, p. 245)) and the continuity of $s(\lambda|\theta)$ and its partial derivatives.

If $\phi = 0$ and $\frac{\partial \mathcal{L}}{\partial \phi} > 0$, then (33b) implies $F_c = 0$, which is false. If there is a solution it must be the case that $\frac{\partial \mathcal{L}}{\partial \phi} = 0$. If so, then $x = s$ and (33a) implies $\phi = \mu$. The solution given by (15a)-(15c) follows.

To check the second-order conditions, form the bordered Hessian:

$$H \equiv \begin{pmatrix} 0 & \frac{\partial h}{\partial p} & \frac{\partial h}{\partial n} & \frac{\partial h}{\partial x} \\ \frac{\partial h}{\partial p} & \frac{\partial^2 \mathcal{L}}{\partial p^2} & \frac{\partial^2 \mathcal{L}}{\partial n \partial p} & \frac{\partial^2 \mathcal{L}}{\partial x \partial p} \\ \frac{\partial h}{\partial n} & \frac{\partial^2 \mathcal{L}}{\partial p \partial n} & \frac{\partial^2 \mathcal{L}}{\partial n^2} & \frac{\partial^2 \mathcal{L}}{\partial x \partial n} \\ \frac{\partial h}{\partial x} & \frac{\partial^2 \mathcal{L}}{\partial p \partial x} & \frac{\partial^2 \mathcal{L}}{\partial n \partial x} & \frac{\partial^2 \mathcal{L}}{\partial x^2} \end{pmatrix}$$

where $h(p, n, x; \lambda) \equiv x - s \leq 0$ is the (binding) inequality constraint. It is then straightforward (if tedious) to show that the last two leading principal minors alternate in sign with the last one negative (and hence the Hessian of $\mathcal{L}$ is negative definite on the constraint set) if and only if (16) holds. Rewriting (16),

$$\int_{\underline{\theta}}^{\overline{\theta}} \theta \left[\frac{s(\lambda|\theta)g(\theta)}{s(\lambda)}\right]\, \mathrm{d}\theta > \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{\lambda\theta}{\mu}\right)\theta \left[\frac{s(\lambda|\theta)g(\theta)}{s(\lambda)}\right]\, \mathrm{d}\theta - \frac{\lambda\Theta^2}{\mu}$$

Under assumption (7), $\frac{\lambda\theta}{\mu} < 1$ for all values of λ and θ. Hence the left-hand-side term is greater than the first term on the right-hand side. Since the second term on the right-hand side is negative, it follows that the solution to the first-order conditions given is a local constrained maximum (Simon and Blume, 1994, theorem 19.8, p. 466). Moreover, the negative-definiteness of $\mathcal{L}$ on the constraint set holds for all p, n and all $x > 0$, hence the local maximum is a unique global maximum of the constrained optimization problem.

Rewrite (15a)-(15c), together with (8), noting that $\phi = \mu$ and $x = d$:

$$G_1 \;\; = \;\; p - (\mu + c - \tau + \lambda\Theta) = 0 \tag{34a}$$

$$G_2 \;\; = \;\; x - \int_{\underline{\theta}}^{\overline{\theta}} s(\lambda|\theta) M g(\theta) \, \mathrm{d}\theta = 0 \tag{34b}$$

$$G_3 \;\; \equiv \;\; \Theta - \frac{\int_{\underline{\theta}}^{\overline{\theta}} \theta s(\lambda|\theta) g(\theta) \, \mathrm{d}\theta}{\int_{\underline{\theta}}^{\overline{\theta}} s(\lambda|\theta) g(\theta) \, \mathrm{d}\theta} = 0 \tag{34c}$$

$$G_4 \;\; = \;\; -F_c + \frac{\lambda\Theta x}{n} = 0 \tag{34d}$$

where $s(\lambda|\theta)$ is given by (2). It is convenient to define $z = \frac{x}{n}$ (class size) and analyze (34a)-(34d) as a set of four equations with four endogenous variables, p, x, Θ, and z, and one exogenous variable, λ. Let J be the Jacobian:

$$J \equiv \begin{pmatrix} \frac{\partial G_1}{\partial p} & \frac{\partial G_1}{\partial x} & \frac{\partial G_1}{\partial z} & \frac{\partial G_1}{\partial \Theta} \\ \frac{\partial G_2}{\partial p} & \frac{\partial G_2}{\partial x} & \frac{\partial G_2}{\partial z} & \frac{\partial G_2}{\partial \Theta} \\ \frac{\partial G_3}{\partial p} & \frac{\partial G_3}{\partial x} & \frac{\partial G_3}{\partial z} & \frac{\partial G_3}{\partial \Theta} \\ \frac{\partial G_4}{\partial p} & \frac{\partial G_4}{\partial x} & \frac{\partial G_4}{\partial z} & \frac{\partial G_4}{\partial \Theta} \end{pmatrix}$$

By the implicit function theorem (e.g. Simon and Blume (1994, theorem 15.7, p. 355)):

$$\begin{pmatrix} \frac{\partial p}{\partial \lambda} \\ \frac{\partial x}{\partial \lambda} \\ \frac{\partial z}{\partial \lambda} \\ \frac{\partial \Theta}{\partial \lambda} \end{pmatrix} = -J^{-1} \begin{pmatrix} \frac{\partial G_1}{\partial \lambda} \\ \frac{\partial G_2}{\partial \lambda} \\ \frac{\partial G_3}{\partial \lambda} \\ \frac{\partial G_4}{\partial \lambda} \end{pmatrix} \tag{35}$$

It is straightforward to show that:

$$\det J = -\frac{\Psi}{\Theta} < 0 \tag{36}$$

since $\Psi > 0$ (refer to (16)). Simplifying (35), we have:

$$\begin{pmatrix} \frac{\partial p}{\partial \lambda} \\ \frac{\partial x}{\partial \lambda} \\ \frac{\partial z}{\partial \lambda} \\ \frac{\partial \Theta}{\partial \lambda} \end{pmatrix} = \begin{pmatrix} \frac{\Theta}{\Psi} \left[\Theta + \frac{\lambda}{\mu} \ln \left(\frac{T}{z} \right) \sigma_{\theta|\lambda}^2 \right] \\ \frac{x\Theta}{\mu} \ln \left(\frac{T}{z} \right) \\ -\frac{z}{\lambda\Psi} \left[\frac{\lambda}{\mu} \ln \left(\frac{T}{z} \right) \sigma_{\theta|\lambda}^2 + \Theta \right] \\ \frac{\Theta \sigma_{\theta|\lambda}^2}{\mu\Psi} \left[\ln \left(\frac{T}{z} \right) + 1 \right] \end{pmatrix} \tag{37}$$

at the optimum.

Note that

$$\sigma_{\theta|\lambda}^2 = \int_{\underline{\theta}}^{\overline{\theta}} (\theta - \overline{\theta})^2 \left[\frac{s(\lambda|\,\theta)g(\theta)}{s(\lambda)} \right] \, \mathrm{d}\theta$$

and that because the double-exponential distribution yields a non-zero probability that any given household will choose any given school, the term in brackets is non-zero for all θ. Hence as long as $\theta \neq \Theta$ for some θ, which we assumed when we assumed θ has positive support over $(0, \overline{\theta})$, we have:

$$\sigma_{\theta|\lambda}^2 > 0 \tag{38}$$

The results (17a), (17b), (17d) and (17e) follow from (16), (37) and (38). Finally, we have:

$$\frac{\partial n}{\partial \lambda} = \frac{\partial}{\partial \lambda}\left(\frac{x}{z}\right) = \frac{1}{z}\frac{\partial x}{\partial \lambda} - \frac{x}{z^2}\frac{\partial z}{\partial \lambda} > 0$$

where the inequality follows from (17b) and (17e). This gives (17c).

A.1.2 Case 1.2: Indivisible Classrooms

The Lagrangian for this case is the same as (32), but where n is now interpreted as a parameter. The first-order conditions are given by (33a), (33c), and (33d) above.

If $\phi = 0$, then we have $x = 0$ and the second-order conditions for a maximum are not satisfied. If there is a solution it must be the case that $\frac{\partial \mathcal{L}}{\partial \phi} = 0$. If so, then (18a)-(18b) follow. As in Case 1.1, $\phi = \mu$.

The second-order conditions can be verified by evaluating the determinant of the bordered Hessian (3x3 in this case). The determinant is positive, and the second-order conditions are satisfied, if condition (16) is satisfied. This condition is guaranteed by (7). We again have that the Hessian of the profit function is negative definite on the constraint set, and hence that the solution given by (18a)-(18b) is a global maximum.

Rewrite (18a)-(18b), together with (8):

$$G_1 \;=\; p(n,\lambda) - [\mu + c - \tau + \lambda\Theta(n,\lambda)] = 0 \tag{39a}$$

$$G_2 \;\equiv\; x - \int_{\underline{\theta}}^{\overline{\theta}} s(\lambda|\theta)Mg(\theta)\,\mathrm{d}\theta = 0 \tag{39b}$$

$$G_3 \;\equiv\; \Theta - \frac{\int_{\underline{\theta}}^{\overline{\theta}} \theta s(\lambda|\theta)g(\theta)\,\mathrm{d}\theta}{\int_{\underline{\theta}}^{\overline{\theta}} s(\lambda|\theta)g(\theta)\,\mathrm{d}\theta} = 0 \tag{39c}$$

Note that these are the same as (34a)-(34c). The number of classrooms is now treated as a parameter, and we no longer have (34d). Inverting the Jacobian (3x3 in this case) and using the implicit function theorem (as in (35)) applied to (39a)-(39c), we have:

$$\begin{pmatrix} \frac{\partial p}{\partial \lambda} \\ \frac{\partial x}{\partial \lambda} \\ \frac{\partial \Theta}{\partial \lambda} \end{pmatrix} \;=\; \frac{1}{1 + \frac{\lambda\Psi}{\mu}} \begin{pmatrix} \Theta + \frac{\lambda\Theta^2}{\mu} + \frac{\lambda}{\mu}\ln\left(\frac{nT}{x}\right)\sigma_{\theta|\lambda}^2 \\ \frac{x}{\mu}\left[\Psi\ln\left(\frac{nT}{x}\right) - \Theta\right] \\ \frac{\sigma_{\theta|\lambda}^2}{\mu}\left[\frac{\lambda\Theta}{\mu} + \ln\left(\frac{nT}{x}\right)\right] \end{pmatrix} \tag{40}$$

The results (19a),(19b) and (21) follow from (16), (20), (38), and (40).

30

Proof of Lemma 1

i. By the envelope theorem, we have:
$$\frac{\partial \pi^*}{\partial n} = \frac{\partial \mathcal{L}}{\partial n}$$

where $\mathcal{L}$ is given by (32) (where n is now interpreted as a parameter) and $\frac{\partial \pi^*}{\partial n}$ allows p and x to vary (holding λ constant) but $\frac{\partial \mathcal{L}}{\partial n}$ holds p, x and λ constant. Then by (33b):

$$
\begin{aligned}
\frac{\partial^2 \pi^*}{\partial n^2} = \frac{\partial}{\partial n}\left(\frac{\partial \mathcal{L}}{\partial n}\right) &= \frac{\partial}{\partial n}\left(-F_c + \frac{\lambda \Theta x}{n}\right) \\
&= -\frac{\lambda \Theta x}{n^2} + \frac{\lambda \Theta}{n}\frac{\partial x}{\partial n} + \frac{\lambda x}{n}\frac{\partial \Theta}{\partial n}
\end{aligned}
\tag{41}
$$

Applying the implicit function theorem (as in (35)) to (39a)-(39c),

$$
\begin{pmatrix}
\frac{\partial p}{\partial n} \\
\frac{\partial x}{\partial n} \\
\frac{\partial \Theta}{\partial n}
\end{pmatrix} = \frac{\lambda}{\mu + \lambda \Psi}
\begin{pmatrix}
\frac{\lambda \sigma_{\theta|\lambda}^2}{n} \\
\frac{x \Psi}{n} \\
\frac{\sigma_{\theta|\lambda}^2}{n}
\end{pmatrix}
\tag{42}
$$

Plugging into (41),
$$\frac{\partial^2 \pi^*}{\partial n^2} = -\frac{\lambda x \Psi}{n^2 \left(1 + \frac{\lambda \Psi}{\mu}\right)} < 0$$

Hence $\pi^*(n, \lambda)$ is globally concave in n for all λ.

ii. Since $n^*(\lambda)$ is monotonically increasing in λ in Case 1.1 (refer to (15c) and (17c)), $\lambda \in (\gamma_k, \gamma_{k+1})$ implies $n^*(\lambda) \in (k, k+1)$. From the global concavity of $\pi^*(n, \lambda)$, it follows that $\pi^*(n, \lambda)$ increases to the left of n^* and decreases to the right of n^* for a given λ. Hence for a given λ within the interval (γ_k, γ_{k+1}) either k or $k+1$ must be the optimal integer number of classrooms.

iii. For all $\lambda \in (\gamma_k, \gamma_{k+1})$, define

$$\Pi(\lambda) \equiv \pi^*(k+1, \lambda) - \pi^*(k, \lambda)$$

Since k is the unique optimal choice of number of classrooms at γ_k in the divisible-classrooms case, we know:
$$\Pi(\gamma_k) = \pi^*(k+1, \gamma_k) - \pi^*(k, \gamma_k) < 0$$

Similarly,
$$\Pi(\gamma_{k+1}) = \pi^*(k+1, \gamma_{k+1}) - \pi^*(k, \gamma_{k+1}) > 0$$

Note that:
$$\frac{\partial^2 \pi^*}{\partial n \partial \lambda} = \frac{\partial}{\partial \lambda}\left(\frac{\partial \pi^*}{\partial n}\right) = \frac{\partial}{\partial \lambda}\left(\frac{\partial \mathcal{L}}{\partial n}\right) = \frac{\partial}{\partial \lambda}\left(-F_c + \frac{\lambda x \Theta}{n}\right) > 0$$

where the second equation follows from the envelope theorem (as above), the third equation

follows from (33b), and the inequality follows from (19b) and (21). Hence:

$$\frac{d\Pi}{d\lambda} = \frac{\partial \pi^*(k+1, \lambda)}{\partial \lambda} - \frac{\partial \pi^*(k, \lambda)}{\partial \lambda} > 0$$

for $\lambda \in (\gamma_k, \gamma_{k+1})$. Since $\Pi(\lambda)$ is differentiable, monotonically increasing, negative at γ_k and positive at γ_{k+1}, we know that there is exactly one value of λ, call it ρ_k, at which $\Pi(\rho_k, k) = 0$. In the interval $[\gamma_k, \rho_k)$, k is the optimal integer number of classrooms; in the interval $[\rho_k, \gamma_{k+1})$, $k + 1$ is the optimal choice.

iv. It remains to consider the regions at the extremes of the support of λ. Without loss of generality, let $\underline{j}$ be the largest integer such that $\gamma_{\underline{j}} \leq \underline{\lambda}_u$,[49] and let $\overline{j}$ be the smallest integer such that $\overline{\lambda}_u \leq \gamma_{\overline{j}}$. Within each interval, $[\gamma_{\underline{j}}, \gamma_{\underline{j}+1})$, $[\gamma_{\underline{j}+1}, \gamma_{\underline{j}+2})$, ..., $[\gamma_{\overline{j}-1}, \gamma_{\overline{j}})$, the result from part iii above holds. Truncate the interval $(\gamma_{\underline{j}}, \gamma_{\overline{j}})$ at $\underline{\lambda}_u$ below and $\overline{\lambda}_u$ above. If $\rho_{\underline{j}} \leq \underline{\lambda}_u$, then let $\underline{k} = \underline{j} + 1$; else if $\underline{\lambda}_u < \rho_{\underline{j}}$ then let $\underline{k} = \underline{j}$. If $\rho_{\overline{j}-1} < \overline{\lambda}_u$, then let $\overline{k} = \overline{j}$; else if $\overline{\lambda}_u \leq \rho_{\overline{j}-1}$, then let $\overline{k} = \overline{j} - 1$. Let $\rho_{\underline{k}-1} = \underline{\lambda}_u$ and $\rho_{\overline{k}} = \overline{\lambda}_u$. Then

$$\underline{\lambda}_u = \rho_{\underline{k}-1} < \rho_{\underline{k}} < ... < \rho_{\overline{k}} = \overline{\lambda}_u$$

form a partition of the set of unsubsidized schools, with the optimal integer number of classrooms equal to $\underline{k}, \underline{k} + 1, ..., \overline{k}$ between consecutive values, and the lemma is proved.

Discontinuities at Critical Values

Consider a given ρ_k from Lemma 1, where $\underline{k} < k < \overline{k}$. Note that $\lim_{\lambda \to \rho_k^-} p^*$, $\lim_{\lambda \to \rho_k^-} x^*$ and $\lim_{\lambda \to \rho_k^-} \Theta$ (the limits as λ approaches ρ_k from the left) are given by (18a), (18b), and (8) (combined with (18a) and (18b)) with $n = k$. $\lim_{\lambda \to \rho_k^+} p^*$, $\lim_{\lambda \to \rho_k^+} x^*$ and $\lim_{\lambda \to \rho_k^+} \Theta$ are given by the same expressions with $n = k + 1$. The differences in the left and right limits then have the same signs as the partial derivatives of the variables with respect to n. By equation (42), we have immediately that $\frac{\partial p}{\partial n} > 0$, $\frac{\partial x}{\partial n} > 0$, and $\frac{\partial \Theta}{\partial n} > 0$. Hence:

$$\lim_{\lambda \to \rho_k^-} p^* < \lim_{\lambda \to \rho_k^+} p^* \tag{43a}$$

$$\lim_{\lambda \to \rho_k^-} x^* < \lim_{\lambda \to \rho_k^+} x^* \tag{43b}$$

$$\lim_{\lambda \to \rho_k^-} \Theta < \lim_{\lambda \to \rho_k^+} \Theta \tag{43c}$$

Moreover,

$$\begin{aligned}
\frac{\partial}{\partial n}\left(\frac{x}{n}\right) &= \frac{1}{n}\frac{\partial x}{\partial n} - \frac{x}{n^2} \\
&= \frac{x}{n^2}\left(\frac{\lambda \Psi}{\mu + \lambda \Psi} - 1\right) < 0
\end{aligned}$$

---

[49]If $\underline{\lambda}_v < \gamma_1$ then let $\underline{j} = 0$ and $\gamma_0 = 0$.

and hence:

$$\lim_{\lambda \to \rho_k^-} \frac{x^*}{n} > \lim_{\lambda \to \rho_k^+} \frac{x^*}{n} \tag{44}$$

A.2   Case 2: Voucher Schools, Post-1994

A.2.1   Case 2.1: Divisible Classrooms

The Lagrangian in this case is:

$$\mathcal{L} = (p - c + \tau)\, x - nF_c - F_s - \phi_1 (x - d) - \phi_2 \left(\frac{x}{n} - 45 \right) \tag{45}$$

The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial p} = x + \phi_1 \left(-\frac{d}{\mu} \right) = 0 \tag{46a}$$

$$\frac{\partial \mathcal{L}}{\partial n} = -F_c + \phi_1 \left(\frac{\lambda \Theta d}{\mu n} \right) + \phi_2 \left(\frac{x}{n^2} \right) = 0 \tag{46b}$$

$$\frac{\partial \mathcal{L}}{\partial x} = p - c + \tau - \phi_1 \left(1 + \frac{\lambda \Theta d}{\mu x} \right) - \phi_2 \left(\frac{1}{n} \right) = 0 \tag{46c}$$

$$\frac{\partial \mathcal{L}}{\partial \phi_1} \geq 0, \ \phi_1 \geq 0, \ \text{and } \phi_1 \frac{\partial \mathcal{L}}{\partial \phi_1} = 0 \tag{46d}$$

$$\frac{\partial \mathcal{L}}{\partial \phi_2} \geq 0, \ \phi_2 \geq 0, \ \text{and } \phi_2 \frac{\partial \mathcal{L}}{\partial \phi_2} = 0 \tag{46e}$$

Suppose $\phi_1 = 0$ and $\frac{\partial \mathcal{L}}{\partial \phi_1} > 0$, i.e. the demand constraint is not binding. Then (46a) implies $x = 0$, and (46b) in turn implies $F_c = 0$, which is false. Hence if there is a solution it must be that $\frac{\partial \mathcal{L}}{\partial \phi_1} = 0$ and the demand constraint is binding: $x = s$. Hence $\phi_1 = \mu$ by (46a).

We then have two sub-cases:

1. Class size cap binding: $\phi_2 \geq 0$ and $\frac{\partial \mathcal{L}}{\partial \phi_2} = 0 \Rightarrow \frac{x}{n} = 45$. Then by (46b), $\frac{\phi_2}{n} = \frac{F}{45} - \lambda \Theta$ and $\phi_2 \geq 0 \Rightarrow \lambda \Theta \leq \frac{F}{45}$. Algebra yields (24a)-(24c). To verify the second-order conditions, the bordered Hessian is:

$$H \equiv \begin{pmatrix} 0 & 0 & \frac{\partial h_1}{\partial p} & \frac{\partial h_1}{\partial n} & \frac{\partial h_1}{\partial x} \\ 0 & 0 & \frac{\partial h_2}{\partial p} & \frac{\partial h_2}{\partial n} & \frac{\partial h_2}{\partial x} \\ \frac{\partial h_1}{\partial p} & \frac{\partial h_2}{\partial p} & \frac{\partial^2 \mathcal{L}}{\partial p^2} & \frac{\partial^2 \mathcal{L}}{\partial n \partial p} & \frac{\partial^2 \mathcal{L}}{\partial x \partial p} \\ \frac{\partial h_1}{\partial n} & \frac{\partial h_2}{\partial n} & \frac{\partial^2 \mathcal{L}}{\partial p \partial n} & \frac{\partial^2 \mathcal{L}}{\partial n^2} & \frac{\partial^2 \mathcal{L}}{\partial x \partial n} \\ \frac{\partial h_1}{\partial x} & \frac{\partial h_2}{\partial x} & \frac{\partial^2 \mathcal{L}}{\partial p \partial x} & \frac{\partial^2 \mathcal{L}}{\partial n \partial x} & \frac{\partial^2 \mathcal{L}}{\partial x^2} \end{pmatrix} \tag{47}$$

where $h_1(p, n, x; \lambda) = x - d$ and $h_2(p, n, x; \lambda) = \frac{x}{n} - 45$ and $\mathcal{L}$ is given in (45). It is easily verified that $\det H = -\frac{x^3}{\mu n^4} < 0$ at the optimum, and hence that the second-order conditions for a maximum are satisfied.

To analyze the slopes with respect to λ in this sub-case, rewrite (24a)-(24c) with (8) as:

$$G_1 \quad = \quad p - \left(c - \tau + \mu + \frac{F_c}{45} \right) = 0 \tag{48a}$$

$$G_2 \quad = \quad x - \int_{\underline{\theta}}^{\overline{\theta}} s(\lambda|\theta) M g(\theta) \, \mathrm{d}\theta = 0 \tag{48b}$$

$$G_3 \quad \equiv \quad \Theta - \frac{\int_{\underline{\theta}}^{\overline{\theta}} \theta s(\lambda|\theta) g(\theta) \, \mathrm{d}\theta}{\int_{\underline{\theta}}^{\overline{\theta}} s(\lambda|\theta) g(\theta) \, \mathrm{d}\theta} = 0 \tag{48c}$$

$$G_4 \quad = \quad n - \frac{x}{45} = 0 \tag{48d}$$

where $s(\lambda|\theta)$ is given by (2).

Applying the implicit function theorem (as in (35)) to (48a)-(48d), we have:

$$\begin{pmatrix} \frac{\partial p}{\partial \lambda} \\ \frac{\partial x}{\partial \lambda} \\ \frac{\partial n}{\partial \lambda} \\ \frac{\partial \Theta}{\partial \lambda} \end{pmatrix} = \frac{1}{\mu} \ln \left(\frac{T}{45} \right) \begin{pmatrix} 0 \\ x\Theta \\ \frac{x\Theta}{45} \\ \sigma_{\theta|\lambda}^2 \end{pmatrix} \tag{49}$$

which in turn implies (25a)-(25e).

2. Class size cap non-binding: $\phi_2 = 0$ and $\frac{\partial \mathcal{L}}{\partial \phi_2} \geq 0$. By (46b), $\frac{x}{n} = \frac{F_c}{\lambda \Theta}$ and $\frac{\partial \mathcal{L}}{\partial \phi_2} \geq 0$ implies $\lambda \Theta \geq \frac{F_c}{45}$. When this condition is satisfied, the expressions for p, x, n, and Θ, the second-order conditions and the slopes with respect to λ are the same as in Case 1.1.

The fact that $\frac{\partial \Theta}{\partial \lambda} > 0$ in both sub-cases guarantees that there is at most one critical value of λ, call it α, at which $\alpha \Theta(\alpha) = \frac{F_c}{45}$.

A.2.2   Case 2.2: Indivisible Classrooms

The Lagrangian in this case is:

$$\mathcal{L} = (p - c + \tau) x - n F_c - F_s - \phi_1 (x - d) - \phi_2 \left(\frac{x}{n} - 45 \right) \tag{50}$$

The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial p} \quad = \quad x - \phi_1 \left(\frac{d}{\mu} \right) = 0 \tag{51a}$$

$$\frac{\partial \mathcal{L}}{\partial x} \quad = \quad p - c + \tau - \phi_1 \left(1 + \frac{\lambda \Theta d}{\mu x} \right) - \phi_2 \left(\frac{1}{n} \right) = 0 \tag{51b}$$

$$\frac{\partial \mathcal{L}}{\partial \phi_1} \quad \geq \quad 0, \ \phi_1 \geq 0, \ \text{and} \ \phi_1 \frac{\partial \mathcal{L}}{\partial \phi_1} = 0 \tag{51c}$$

$$\frac{\partial \mathcal{L}}{\partial \phi_2} \quad \geq \quad 0, \ \phi_2 \geq 0, \ \text{and} \ \phi_2 \frac{\partial \mathcal{L}}{\partial \phi_2} = 0 \tag{51d}$$

There are four sub-cases to be considered:

1. Demand constraint not binding, class-size cap not binding: $\phi_1 = 0$, $\frac{\partial \mathcal{L}}{\partial \phi_1} \geq 0$, $\phi_2 = 0$ and $\frac{\partial \mathcal{L}}{\partial \phi_2} \geq 0$. By (51a), $x = 0$ and by (51b), $p = c - \tau$. It is straightforward to verify that the second-order conditions are not satisfied in this sub-case.

2. Demand constraint not binding, class-size cap binding: $\phi_1 = 0$, $\frac{\partial \mathcal{L}}{\partial \phi_1} \geq 0$, $\phi_2 \geq 0$ and $\frac{\partial \mathcal{L}}{\partial \phi_2} = 0$. By (51a), we have $x = 0$, but then the class-size constraint ($\frac{x}{n} = 45$) is violated, hence there is no solution in this case.

3. Demand constraint binding, class-size cap non-binding: $\phi_1 \geq 0$, $\frac{\partial \mathcal{L}}{\partial \phi_1} = 0$, $\phi_2 = 0$, $\frac{\partial \mathcal{L}}{\partial \phi_2} \geq 0$. The solution in this case is the same as in Case 1.2, given by (18a)-(18b), with slopes vs. λ given by (40), and the second-order conditions satisfied as discussed in Appendix A.1.2 above. The condition $\frac{\partial \mathcal{L}}{\partial \phi_2} \geq 0$ requires:

$$45n \leq \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{\Omega(\theta)} \left(\frac{T}{45} \right)^{\frac{\theta\lambda}{\mu}} \exp \left(-\frac{c - \tau + \mu + \lambda\Theta}{\mu} \right) M g(\theta) \, \mathrm{d}\theta$$

Taking the exponential term out of the integral and solving for $\lambda\Theta$, we have:

$$\lambda\Theta \leq \mu \ln \Sigma - \mu \ln(45n) - c + \tau - \mu$$

Setting this inequality to an equality yields (30), which implicitly defines β_n, the value of λ at which the class-size cap begins to bind.

4. Demand constraint binding, class-size cap binding: $\phi_1 \geq 0$, $\frac{\partial \mathcal{L}}{\partial \phi_1} = 0$, $\phi_2 \geq 0$ and $\frac{\partial \mathcal{L}}{\partial \phi_2} = 0$. The two constraints, $x = d$ and $x = 45n$, pin down the values of p and x, and the results (27a)-(29b) follow immediately. As in the previous sub-case, $\phi_1 = \mu > 0$.

   The fact that $\frac{\partial x}{\partial \lambda} = \frac{\partial}{\partial \lambda} \left(\frac{x}{n} \right) = 0$ follows immediately. It is straightforward to show that

$$\begin{aligned} \frac{\partial p}{\partial \lambda} &= \Theta \ln \left(\frac{T}{45} \right) > 0 \qquad\qquad (52) \\ \frac{\partial \Theta}{\partial \lambda} &= \frac{1}{\mu} \sigma_{\theta|\lambda}^2 \ln \left(\frac{T}{45} \right) > 0 \end{aligned}$$

It remains to establish the set of schools over which the condition $\phi_2 \geq 0$ is satisfied. The first order conditions imply:

$$\phi_2 = n \left\{ \mu \ln \Sigma - \mu \ln(45n) - c + \tau - \mu - \lambda\Theta \right\} \qquad\qquad (53)$$

By the definition of β_n above, if $\lambda = \beta_n$ then $\phi_2 = 0$. Partially differentiating (53):

$$\begin{aligned} \frac{\partial \phi_2}{\partial \lambda} &= n \left\{ \frac{\mu}{\Sigma} \frac{\partial \Sigma}{\partial \lambda} - \lambda \frac{\partial \Theta}{\partial \lambda} - \Theta \right\} \\ &= n \left\{ \Psi \ln \left(\frac{T}{45} \right) - \Theta \right\} > 0 \end{aligned}$$

by assumption (20). Thus $\phi_2 \geq 0$ for $\lambda \geq \beta_n$.

Finally, we note that at the critical value β_n, it follows from (40), (52), and (20) that:

$$\lim_{\lambda \to \beta_n^-} \frac{\partial p}{\partial \lambda} < \lim_{\lambda \to \beta_n^+} \frac{\partial p}{\partial \lambda}$$

That is, the slope of p with respect to λ is steeper to the right of the critical value, in the region where the class-size cap binds.

Proof of Lemma 2

Our strategy is similar to that of the proof of Lemma 1 in Appendix A.1.2 above, with additional steps to deal with the presence of the class-size cap.

i. Treating β_n as a function of n and implicitly differentiating (30) with respect to n, we have:

$$\frac{\partial \beta_n}{\partial n} = \frac{\mu}{n \left(\Psi \ln \left(\frac{T}{45} \right) - \Theta \right)} > 0$$

for all n, where the inequality follows from (20). Hence β_n is monotonically increasing in n and invertible. For a given λ, the class-size cap binds for $n < \beta_n^{-1}(\lambda)$ and does not bind for $n \geq \beta_n^{-1}(\lambda)$.

ii. For all λ, define:

$$\pi^{**}(n, \lambda) = \pi(p^*, x^*, n, \lambda)$$

where x^* and p^* are given by (27a) and (27b); this is optimal profit when the inequality constraint on class size is replaced by an equality constraint. Recall that $\pi^*(n, \lambda)$ (defined in (22)) is optimal profit when there is no class-size cap. For each λ, there is a single value of n, namely $\beta_n^{-1}(\lambda)$, at which the two profit expressions are equal:

$$\pi^*(\beta_n^{-1}(\lambda), \lambda) = \pi^{**}(\beta_n^{-1}(\lambda), \lambda)$$

While $\pi^*(n, \lambda)$ is a solution to the optimizing problem holding n constant, $\pi^{**}(n, \lambda)$ is a solution to the same problem holding n constant *and* holding x constant (at $45n$). Hence the curves are tangent at the point at which they coincide, $\beta_n^{-1}(\lambda)$;[50] for a proof of this standard result, see Dixit (1976, Ch. 3). The concavity of $\pi^*(n, \lambda)$ was established in the proof of Lemma 1. Combining (10), (27b) and (27a), and differentiating twice with respect to n, we have that $\frac{\partial^2 \pi^{**}}{\partial n^2} = -\frac{45\mu}{n} < 0$ and hence that $\pi^{**}(n, \lambda)$ is concave as well. Now define:

$$\widetilde{\pi}(n, \lambda) = \begin{cases} \pi^*(n, \lambda) & \text{if} \quad n \geq \beta_n^{-1}(\lambda) \\ \pi^{**}(n, \lambda) & \text{if} \quad n < \beta_n^{-1}(\lambda) \end{cases} \tag{54}$$

This function represents optimal profit, taking into account whether or not the class-size cap is binding. For $n < \beta_n^{-1}(\lambda)$, $\frac{\partial \widetilde{\pi}}{\partial n}$ is decreasing in n by the concavity of $\pi^{**}(n, \lambda)$. For $n \geq \beta_n^{-1}(\lambda)$, $\frac{\partial \widetilde{\pi}}{\partial n}$ is decreasing in n by the concavity of $\pi^*(n, \lambda)$. At $n = \beta_n^{-1}(\lambda)$ the two curves are tangent and $\frac{\partial \pi^*}{\partial n} = \frac{\partial \pi^{**}}{\partial n}$. It follows that $\frac{\partial \widetilde{\pi}}{\partial n}$ is decreasing in n for all n and all λ. Hence $\widetilde{\pi}(n, \lambda)$ is globally concave in n for all λ.

---

[50]This argument relies on the fact that both functions are continuously differentiable.

iii. Since n^* is monotonically increasing in λ in Case 2.1, $\lambda \in (\delta_k, \delta_{k+1})$ implies $n^*(\lambda) \in (k, k+1)$. From the concavity of $\widetilde{\pi}(n, \lambda)$ it follows that $\widetilde{\pi}(n, \lambda)$ increases to the left of n^* and decreases to the right of n^*. Hence within the interval (δ_k, δ_{k+1}) either k or $k+1$ must be the optimal integer number of classrooms.

iv. For $\lambda \in (\delta_k, \delta_{k+1})$, define:
$$\widetilde{\Pi}(\lambda) \equiv \widetilde{\pi}(k+1, \lambda) - \widetilde{\pi}(k, \lambda)$$

Since k is the unique optimal choice of number of classrooms at δ_k in the divisible-classrooms case (Case 2.1),
$$\widetilde{\Pi}(\delta_k) = \widetilde{\pi}(k+1, \delta_k) - \widetilde{\pi}(k, \delta_k) < 0$$

Similarly,
$$\widetilde{\Pi}(\delta_{k+1}) = \widetilde{\pi}(k+1, \delta_{k+1}) - \widetilde{\pi}(k, \delta_{k+1}) > 0$$

From part i above, we have $\beta_k < \beta_{k+1}$. Using (54), the definition of $\widetilde{\Pi}(\lambda)$ can be restated:
$$\widetilde{\Pi}(\lambda) = \begin{cases} \pi^*(k+1, \lambda) - \pi^*(k, \lambda) & \text{if} \quad \lambda \leq \beta_k \\ \pi^*(k+1, \lambda) - \pi^{**}(k, \lambda) & \text{if} \quad \beta_k < \lambda \leq \beta_{k+1} \\ \pi^{**}(k+1, \lambda) - \pi^{**}(k, \lambda) & \text{if} \quad \beta_{k+1} < \lambda \end{cases}$$

Consider each interval in turn:

a. $\lambda \leq \beta_k$. The class-size cap binds neither for $n = k$ nor for $n = k + 1$. From the proof of Lemma 1, $\frac{\partial^2 \pi^*}{\partial n \partial \lambda} > 0$, hence $\frac{\partial \pi^*(k+1, \lambda)}{\partial \lambda} > \frac{\partial \pi^*(k, \lambda)}{\partial \lambda}$, hence $\frac{d\widetilde{\Pi}}{d\lambda} > 0$.

b. $\beta_k < \lambda \leq \beta_{k+1}$. The class-size cap binds for $n = k$ but not for $n = k + 1$. Note that
$$\frac{\partial \pi^*}{\partial \lambda} = \frac{\partial \mathcal{L}}{\partial \lambda} = x\Theta \ln\left(\frac{nT}{x}\right) \tag{55}$$

where $\mathcal{L}$ is given by (50), and the first equality follows by the envelope theorem. Similarly,
$$\frac{\partial \pi^{**}}{\partial \lambda} = \frac{\partial \mathcal{L}}{\partial \lambda} = 45n\Theta \ln\left(\frac{T}{45}\right) \tag{56}$$

At $n = k$, the optimal enrollment if there were no class-size cap would be greater than or equal to $45n$; otherwise the cap would not be binding. Hence, comparing (55) and (56),
$$\frac{\partial \pi^*(k, \lambda)}{\partial \lambda} \geq \frac{\partial \pi^{**}(k, \lambda)}{\partial \lambda}$$

Using part iv.a,
$$\frac{\partial \pi^*(k+1, \lambda)}{\partial \lambda} > \frac{\partial \pi^*(k, \lambda)}{\partial \lambda} \geq \frac{\partial \pi^{**}(k, \lambda)}{\partial \lambda}$$

Hence $\frac{d\widetilde{\Pi}}{d\lambda} > 0$

c. $\beta_{k+1} < \lambda$. The class-size cap binds for both $n = k$ and $n = k + 1$. Partially differentiating (56) and using (25d),
$$\frac{\partial}{\partial n}\left(\frac{\partial \pi^{**}}{\partial \lambda}\right) = 45\ln\left(\frac{T}{45}\right)\left\{\Theta + n\frac{\partial \Theta}{\partial n}\right\} > 0$$

37

Hence $\frac{\partial \pi^{**}(k+1,\lambda)}{\partial \lambda} > \frac{\partial \pi^{**}(k,\lambda)}{\partial \lambda}$ and $\frac{d\widetilde{\Pi}}{d\lambda} > 0$.

Thus $\widetilde{\Pi}(\lambda)$ is differentiable and monotonically increasing in λ for all λ. Together with the fact that it is negative at δ_k and positive at δ_{k+1}, this implies that there is exactly one, call it ν_k, at which $\Pi(\nu_k) = 0$. For $\lambda \in [\delta_k, \nu_k)$, k is the optimal integer number of classrooms; for $\lambda \in (\nu_k, \delta_{k+1})$, $k+1$ is optimal.

v. It remains to consider the regions at the extremes of the support of λ. Without loss of generality, let $\underline{j}$ be the largest integer such that $\delta_{\underline{j}} \leq \underline{\lambda}_v$,[51] and let $\overline{j}$ be the smallest integer such that $\overline{\lambda}_v \leq \delta_{\overline{j}}$. Within each interval, $[\delta_{\underline{j}}, \delta_{\underline{j}+1})$, $[\delta_{\underline{j}+1}, \delta_{\underline{j}+2})$, ..., $[\delta_{\overline{j}-1}, \delta_{\overline{j}})$, the result from part v holds. Truncate the interval $(\delta_{\underline{j}}, \delta_{\overline{j}})$ at $\underline{\lambda}_v$ below and $\overline{\lambda}_v$ above. If $\nu_{\underline{j}} \leq \underline{\lambda}_v$, then let $\underline{k} = \underline{j} + 1$; else if $\underline{\lambda}_v < \nu_{\underline{j}}$ then let $\underline{k} = \underline{j}$. If $\nu_{\overline{j}-1} < \overline{\lambda}_v$, then let $\overline{k} = \overline{j}$; else if $\overline{\lambda}_v \leq \nu_{\overline{j}-1}$, then let $\overline{k} = \overline{j} - 1$. Let $\nu_{\underline{k}-1} = \underline{\lambda}_v$ and $\nu_{\overline{k}} = \overline{\lambda}_v$. Then

$$\nu_{\underline{k}-1} < \nu_{\underline{k}} < ... < \nu_{\overline{k}}$$

form a partition of the set of voucher schools, with the optimal integer number of classrooms equal to $\underline{k}, \underline{k}+1, ..., \overline{k}$ between consecutive values, and the lemma is proved.

By the intermediate value theorem, there must be at least one $\lambda \in (\delta_k, \delta_{k+1})$ where $\widetilde{\Pi}(\lambda) = 0$.

Discontinuities at Critical Values

Consider a given ν_k from Lemma 2, where $\underline{k} < k < \overline{k}$. There are two cases to consider:

1. If $\beta_k \geq \nu_k$ and the class-size cap is not binding to the left of ν_k, then (43a)-(44) from Appendix A.1.2 apply to the particular critical value ν_k.

2. If $\beta_k < \nu_k$ and the class-size cap is binding to the left of ν_k, then consider x, $\frac{x}{n}$, and Θ in turn:

   (a) Suppose there were no class-size cap. Then for $\lambda \in (\beta_k, \nu_k)$, we would have $\frac{\partial x}{\partial \lambda} > 0$ and (43b) from Appendix A.1.2 would apply. In the presence of the class-size cap, for $\lambda \in (\beta_k \geq \nu_k)$ we have $\frac{\partial x}{\partial \lambda} = 0$. Hence

$$\lim_{\lambda \to \rho_k^-} x^* \bigg|_{cap} < \lim_{\lambda \to \rho_k^-} x^* \bigg|_{no\ cap}$$

   If to the right of ν_k the class size cap is not binding, then by (43b),

$$\lim_{\lambda \to \rho_k^-} x^* < \lim_{\lambda \to \rho_k^+} x^*$$

   Else if to the right of ν_k the class size cap is binding, then

$$\lim_{\lambda \to \rho_k^-} x^* = 45k < 45(k+1) = \lim_{\lambda \to \rho_k^+} x^*$$

---

[51] If $\underline{\lambda}_v < \delta_1$ then let $\underline{j} = 0$ and $\delta_0 = 0$.

(b) Let $z^* = \frac{x^*}{n}$. If to the right of ν_k the class-size cap is not binding, then:

$$\lim_{\lambda \to \rho_k^-} z^* = 45 > \lim_{\lambda \to \rho_k^+} z^* \tag{57}$$

Else if to the right of ν_k the class-size cap is binding

$$\lim_{\lambda \to \rho_k^-} z^* = \lim_{\lambda \to \rho_k^+} z^* = 45 \tag{58}$$

(c) Average willingness to pay can be written:

$$\Theta(z) = \frac{\int_{\underline{\theta}}^{\overline{\theta}} \theta \frac{1}{\Omega(\theta)} \left(\frac{T}{z}\right)^{\frac{\theta \lambda}{\mu}} g(\theta) \mathrm{d}\theta}{\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{\Omega(\theta)} \left(\frac{T}{z}\right)^{\frac{\theta \lambda}{\mu}} g(\theta) \mathrm{d}\theta} \tag{59}$$

where $z = \frac{x}{n}$. Differentiating,

$$\frac{\partial \Theta}{\partial z} = -\frac{\lambda}{\mu z} \sigma_{\theta|\lambda}^2 > 0 \tag{60}$$

If to the right of ν_k the class-size cap is not binding, then by (57) and (60):

$$\lim_{\lambda \to \rho_k^-} \Theta < \lim_{\lambda \to \rho_k^+} \Theta$$

Else if to the right of ν_k the class-size cap is binding, then by (58) and (60):

$$\lim_{\lambda \to \rho_k^-} \Theta = \lim_{\lambda \to \rho_k^+} \Theta$$

References

ANDERSON, S., AND A. DE PALMA (2001): "Product Diversity in Asymmetric Oligopoly: Is the Quality of Consumer Goods Too Low?," *Journal of Industrial Economics*, 49(2), 113–135.

ANDERSON, S., A. DE PALMA, AND J.-F. THISSE (1992): *Discrete-Choice Theory of Product Differentiation*. MIT Press, Cambridge, MA.

ANGRIST, J., AND A. KRUEGER (1999): "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card. Elsevier Science.

ANGRIST, J. D., AND V. LAVY (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114(2), 533–575.

ARMSTRONG, M., AND D. E. M. SAPPINGTON (2006): "Regulation, Competition and Liberalization," *Journal of Economic Literature*, 44(2), 325–366.

BANERJEE, A., S. COLE, E. DUFLO, AND L. LINDEN (2004): "Remedying Education: Evidence from Two Randomized Experiments in India," Unpub. paper, MIT.

BARTLE, R. G. (1976): *The Elements of Real Analysis*. John Wiley & Sons, New York, 2nd edition edn.

BAYER, P., R. MCMILLAN, AND K. RUEBEN (2004): "An Equilibrium Model of Sorting in an Urban Housing Market," NBER Working Paper # 10865.

BERRY, S. T., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

BROWNING, M., AND E. HEINESEN (2003): "Class Size, Teacher Hours and Educational Attainment," Centre for Applied Microeconometrics Working Paper 2003-15, Institute of Economics, University of Copenhagen.

CLOTFELTER, C. T. (1999): "Public School Segregation in Metropolitan Areas," *Land Economics*, 75(4), 487–504.

DIXIT, A. K. (1976): *Optimization in Economic Theory*. Oxford University Press, Oxford, UK.

DOBBELSTEEN, S., J. LEVIN, AND H. OOSTERBEEK (2002): "The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition," *Oxford Bulletin of Economics and Statistics*, 64(1), 17–38.

DUFLO, E., AND M. KREMER (2004): "Use of Randomization in the Evaluation of Development Effectiveness," in *Evaluating Development Effectiveness: World Bank Series on Evaluation and Development, Volume 7*, ed. by O. Feinstein, G. K. Ingram, and G. K. Pitman, pp. 205–232. Transaction Publishers, New Brunswick, NJ.

EKELAND, I., J. J. HECKMAN, AND L. NESHEIM (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(1), S60–109.
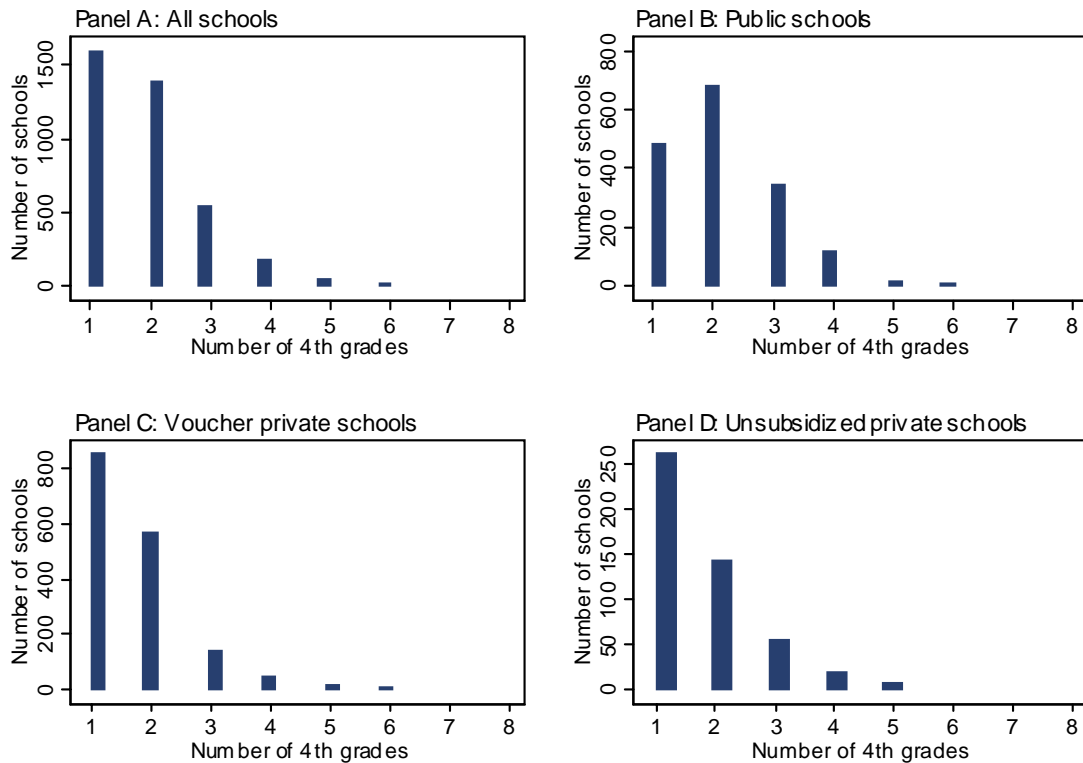
ELACQUA, G. (2005): "How Do For-Profit Schools Respond to Voucher Funding? Evidence from Chile," Occasional Paper Series National Center for Study of Privatization in Education, Teachers College, Columbia University.

EPPLE, D., D. FIGLIO, AND R. ROMANO (2002): "Competition Between Private and Public Schools: Testing Stratification and Pricing Predictions," Unpub. paper, University of Florida.

EPPLE, D., AND R. ROMANO (1998): "Competition Between Private and Public Schools, Vouchers, and Peer Group Effects," *American Economic Review*, 88(1), 33–62.

————— (2002): "Education Vouchers and Cream Skimming," NBER Working Paper #9354.

GABSZEWICZ, J. J., AND J.-F. THISSE (1979): "Price Competition, Quality, and Income Disparities," *Journal of Economic Theory*, 20, 340–359.

HAHN, J., P. TODD, AND W. V. DER KLAAUW (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69(1), 201–209.

HANUSHEK, E. A. (1995): "Interpreting Recent Research on Schooling in Developing Countries," *World Bank Research Observer*, 10, 227–246.

————— (2003): "The Failure of Input-Based Schooling Policies," *Economic Journal*, 113(485), F64–F98.

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.

HSIEH, C.-T., AND M. URQUIOLA (2006): "The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile's School Voucher Program," *Journal of Public Economics*, 90, 1477–1503.

KREMER, M. R. (1995): "Research on Schooling: What We Know and What We Don't: A Comment," *World Bank Research Observer*, 10(2), 247–254.

KRUEGER, A. B. (2003): "Economic Considerations and Class Size," *Economic Journal*, 113(485), F34–63.

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

LEE, D. S. (2005): "Randomized Experiments from Non-Random Selection in U.S. House Elections," Unpub. paper, University of California at Berkeley.

LEE, D. S., AND D. CARD (2004): "Regression Discontinuity Inference with Specification Error," Working Paper No. 74, Center for Labor Economics, University of California at Berkeley.

MANSKI, C. (1992): "Educational Choice (Vouchers) and Social Mobility," *Economics of Education Review*, 11(4), 351–369.

McCRARY, J. (2005): "Manipulation of the Running Variable in the Regression Discontinuity Design," Unpub. paper, University of Michigan.

41

McEwan, P., and M. Urquiola (2005): "Precise Sorting Around Cutoffs in the Regression-Discontinuity Design: Evidence from Class Size Reduction," Unpub. paper, Columbia University.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. Academic Press, New York.

Mizala, A., and P. Romaguera (2002): "Equity and Educational Performance," *Economía*, 2(2), 219–262.

Nechyba, T. J. (2003): "Centralization, Fiscal Federalism, and Private School Attendance," *International Economic Review*, 44(1), 179–204.

Nesheim, L. (2002): "Equilibrium Sorting of Heterogeneous Consumers Across Locations: Theory and Empirical Implications," CEMMAP Working Paper CWP08/02, University College London.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82(1), 34–55.

Rothstein, J. (forthcoming): "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Effects of Inter-District Competition," *American Economic Review*.

Sappington, D. E. M. (2005): "Regulating Service Quality: A Survey," *Journal of Regulatory Economics*, 27(2), 123–154.

Sattinger, M. (1993): "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31(2), 831–880.

Shadish, W. R., T. D. Cook, and D. T. Campbell (2002): *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston.

Shaked, A., and J. Sutton (1982): "Relaxing Price Competition through Product Differentiation," *Review of Economic Studies*, 49, 3–13.

Simon, C. P., and L. Blume (1994): *Mathematics for Economists*. W. W. Norton & Company, New York.

Teulings, C. N. (1995): "The Wage Distribution in a Model of the Assignment of Skills to Jobs," *Journal of Political Economy*, 103(2), 280–315.

Tinbergen, J. (1956): "On the Theory of Income Distribution," *Weltwirtschaftlches Archiv*, 77(2), 155–173.

Urquiola, M. (2005): "Does School Choice Lead to Sorting? Evidence from Tiebout Variation," *American Economic Review*, 95(4), 1310–1326.

——— (2006): "Identifying Class Size Effects in Developing Countries: Evidence from Rural Schools in Bolivia," *Review of Economics and Statistics*, 88(1).

van der Klaauw, W. (2002): "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach," *International Economic Review*, 43(4), 1249–1287.

Verhoogen, E. A. (2006): "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector," Unpub. paper, Columbia University.

Figure 1: Histograms of the number of 4th grades in urban schools, 2002

Panel A: All schools

Panel B: Public schools

Panel C: Voucher private schools

Panel D: Unsubsidized private schools

Note: Based on 2002 administrative data for schools with positive 4th grade enrollments. The figures cover only schools Chile's Ministry of education classifies as urban. For voucher schools, panel C excludes about 0.2 percent of schools which report having more than eight 4th grade classes.

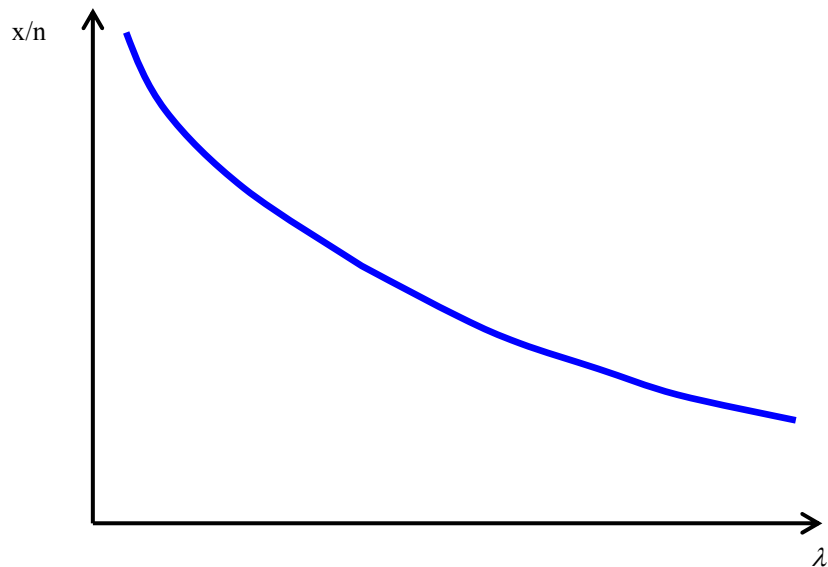Figure 2: Case 1.1—Private unsubsidized schools with divisible classrooms

Figure 3: Case 1.2——Private unsubsidized schools with indivisible classrooms
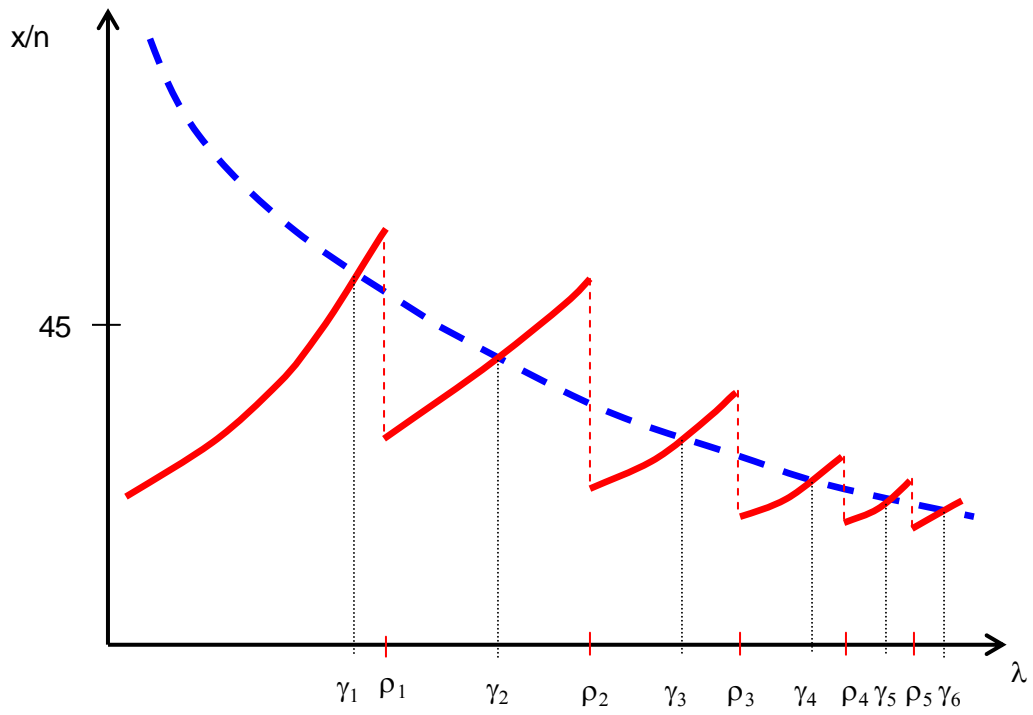

Figure 4: Case 2.1—Voucher schools with divisible classrooms
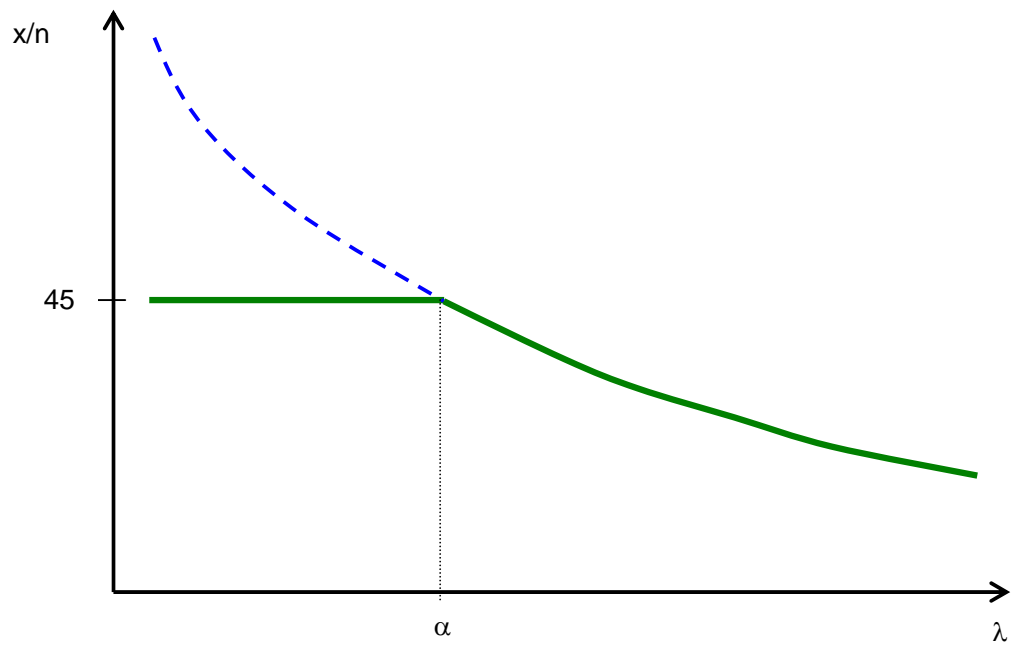
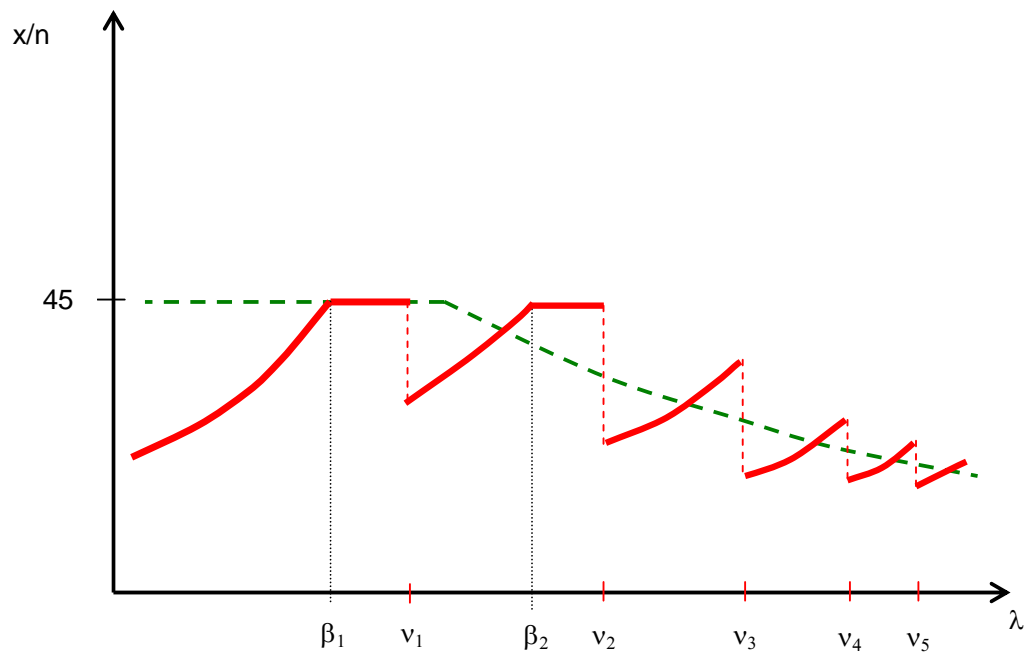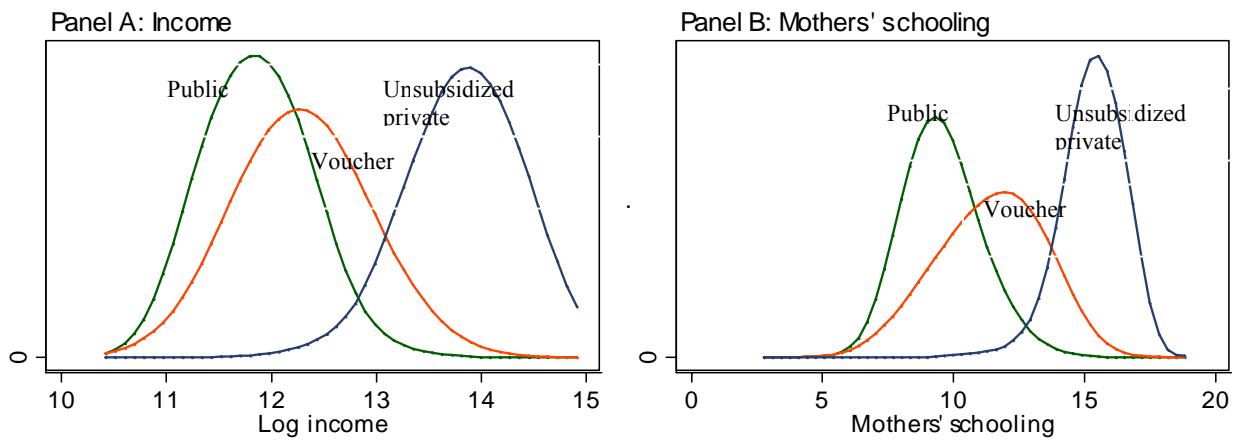Figure 5: Case 2.2—Voucher schools with indivisible classrooms

Figure 6: Densities of log income and mothers' schooling by type of urban school, 2002

Notes: The figures plot kernel densities of log average household income and average mothers' schooling. The data are from 2002 individual level SIMCE information aggregated to the school level. Both panels refer to urban schools only.

Figure 7: Class size and income/mothers' schooling among urban private schools, 2002
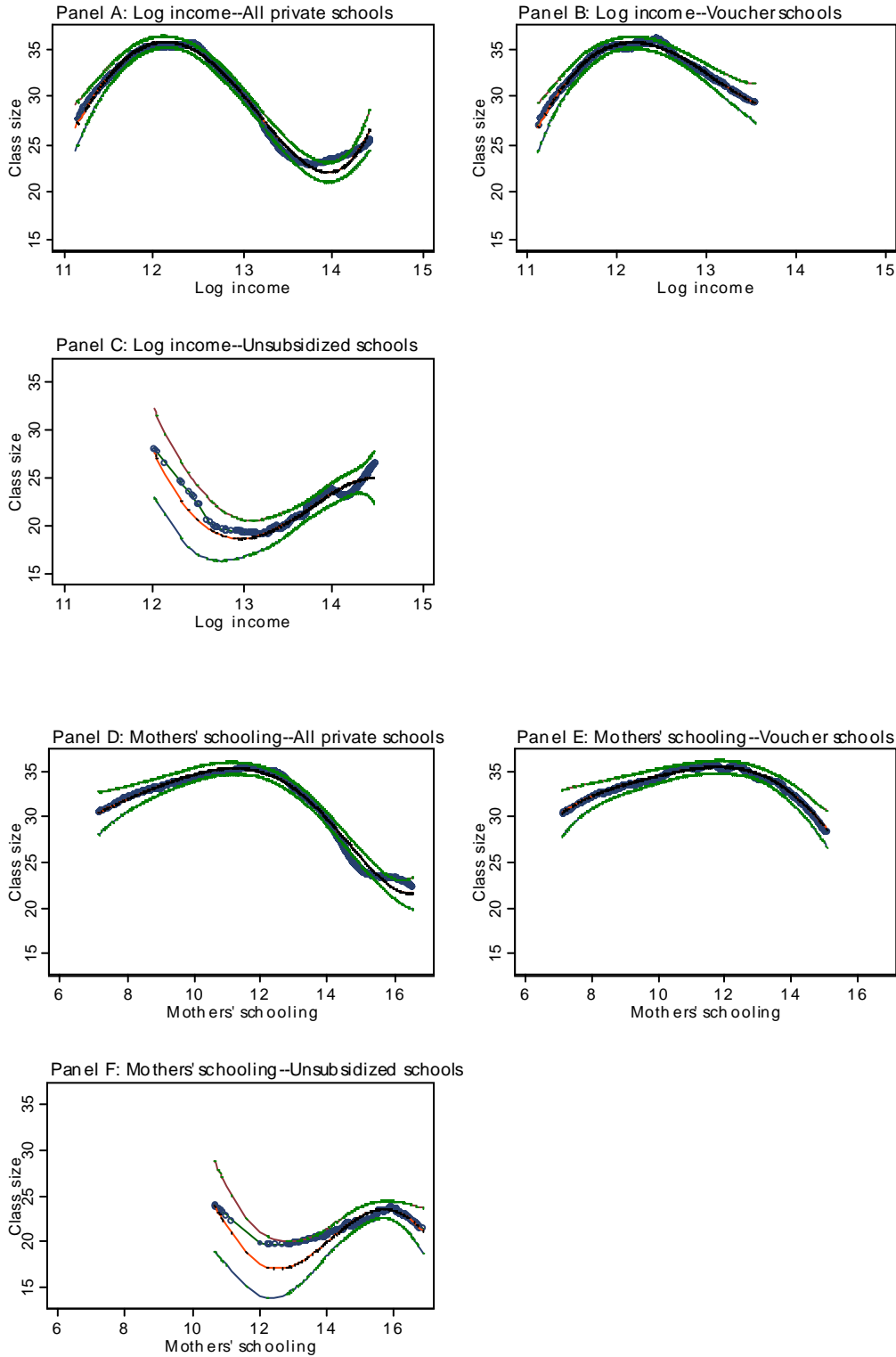
Note: Income and mothers' schooling come from 2002 individual-level SIMCE data aggregated to the school level. Class size is from 2002 administrative information. In each panel, the thicker lines plot fitted values of locally weighted regressions of class size on log income (panels A-C) and mothers' schooling (panels D-F), using a bandwidth of 0.2. The thinner lines plot fitted values, along with the 5th and 95th percentile confidence interval, of a regression of class size on a 5th order polynomial of log income or mothers' schooling. Within each set of schools, the figures omit observations below and above the 1st and 99th percentile of income or mothers' schooling.

Figure 8: Average tuition and log income among urban voucher schools, 2002
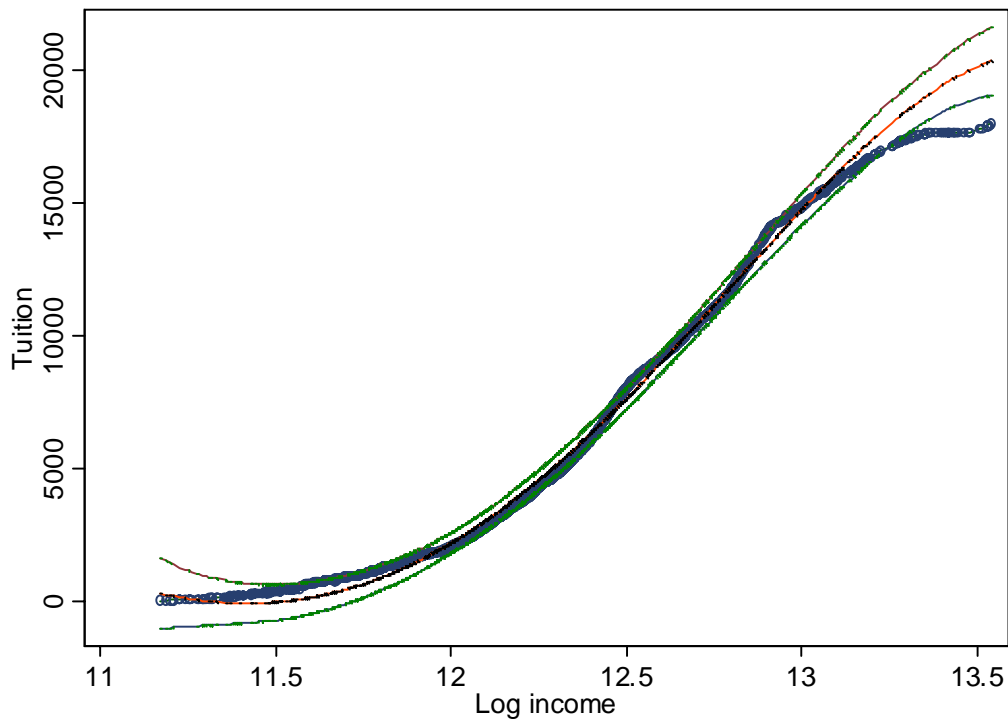
Note: Tuition information comes from school-level administrative data for 2002, and is in monthly Chilean pesos. Income is from 2002 individual-level SIMCE test data, aggregated to the school level. The thicker line plots fitted values of locally weighted regressions of tuition on log income using a bandwidth of 0.2. The thinner lines plot fitted values, along with the 5th and 95th percentile confidence interval, of a regression of monthly tuition on a 5th order polynomial of log income. The figure omits observations below and above the 1st and 99th percentile of log income.

Figure 9: 4th grade enrollment and class size in urban private voucher schools, 2002
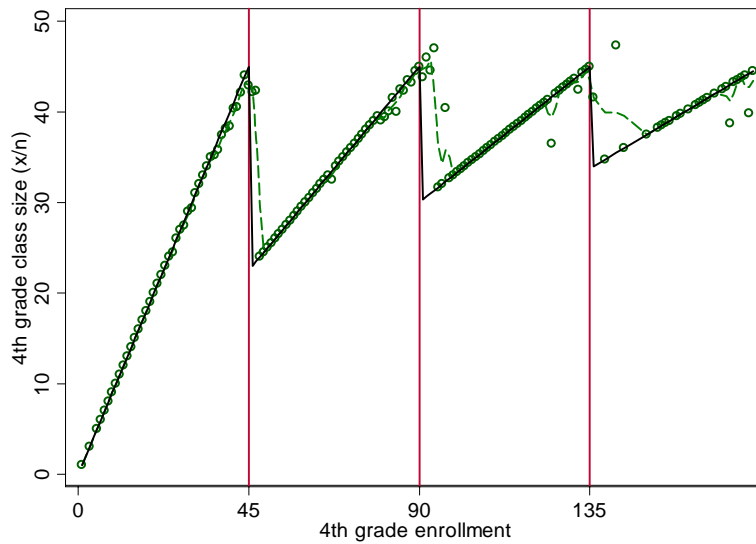
Note: Based on administrative data for 2002. The solid line describes the relationship between enrollment and class size that would exist if the class size rule (equation 30 in the text) were applied mechanically. The circles plot the enrollment cell means of 4th grade class size, and the dotted line plots fitted values from a locally weighted regression (using a bandwidth of 0.05) of class size on enrollment. Only data for schools with 4th grade enrollments below 180 are plotted; this excludes less than two percent of all schools.

Figure 10: Math scores and enrollment in urban private voucher schools, 2002
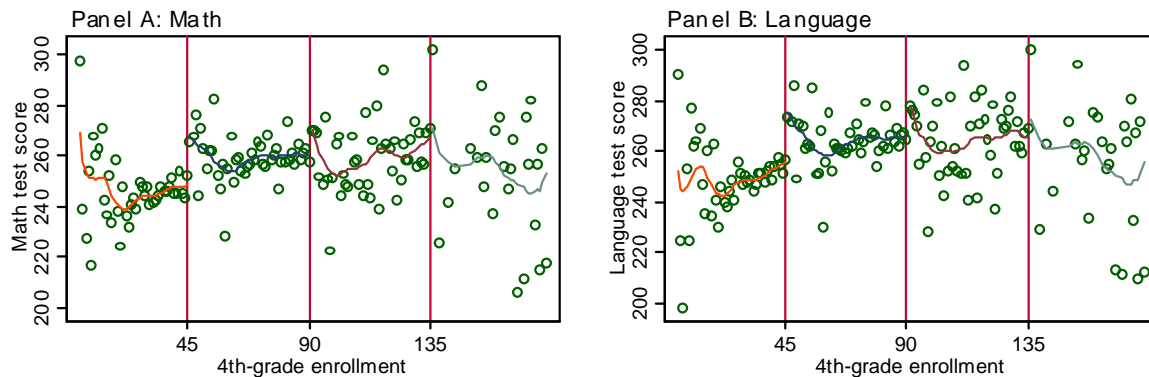
Note: Test scores come from 2002 individual-level SIMCE information aggregated to the school level, and enrollment is from administrative information for the same year. The figures plot "raw" enrollment-cell means of test scores, along with the fitted values of a locally weighted regression calculated within each enrollment segment

Figure 11: Histograms of 4th grade enrollment in urban private schools, 2002

Panel A: Voucher private

Panel B: Unsubsidized private
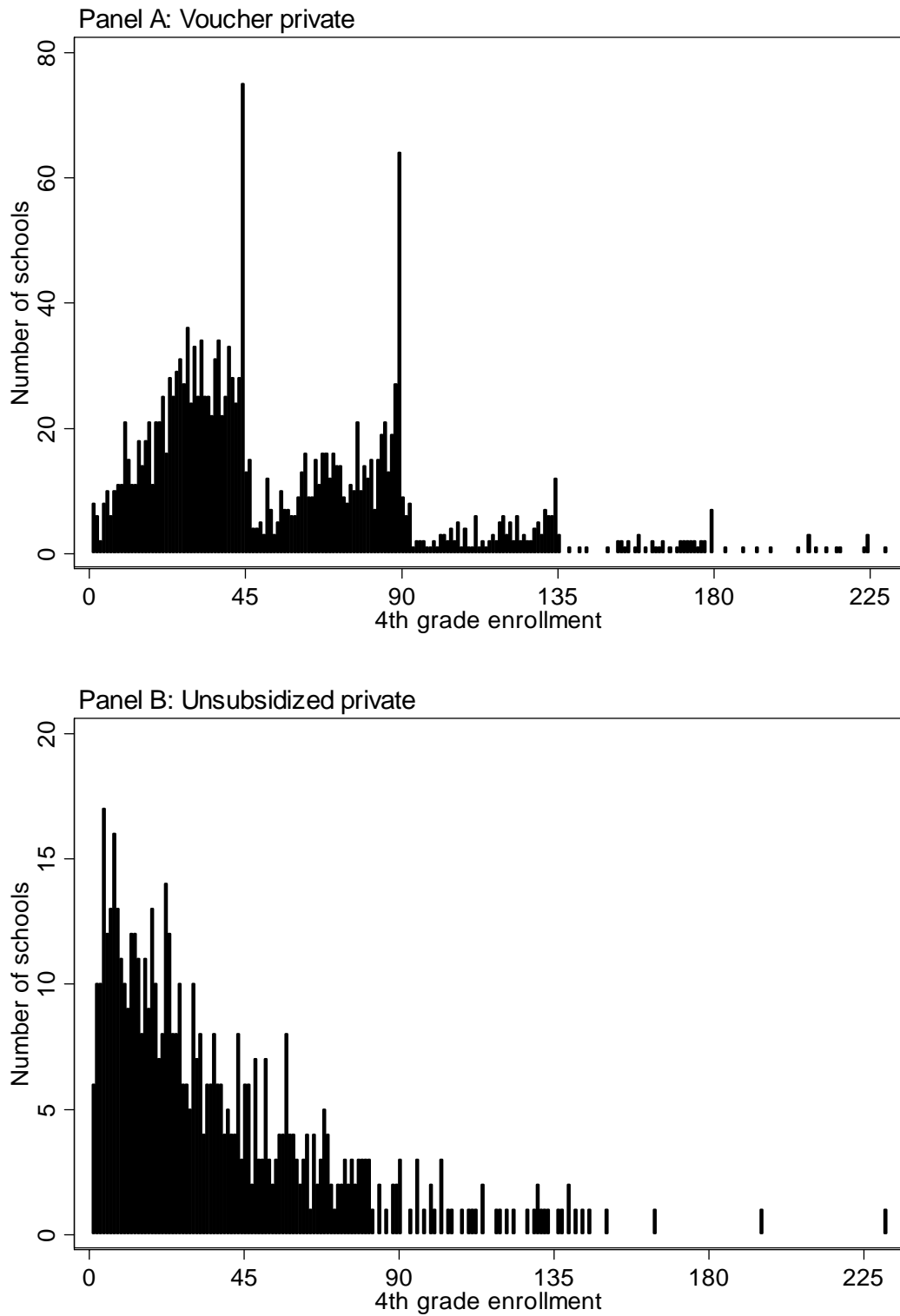
Note: Based on administrative data for 2002. For visual clarity, only schools with 4th grade enrollments below 225 are displayed. This excludes less than one percent of all schools.

Figure 12: Student characteristics, test scores, and enrollment in urban private voucher schools, 2002
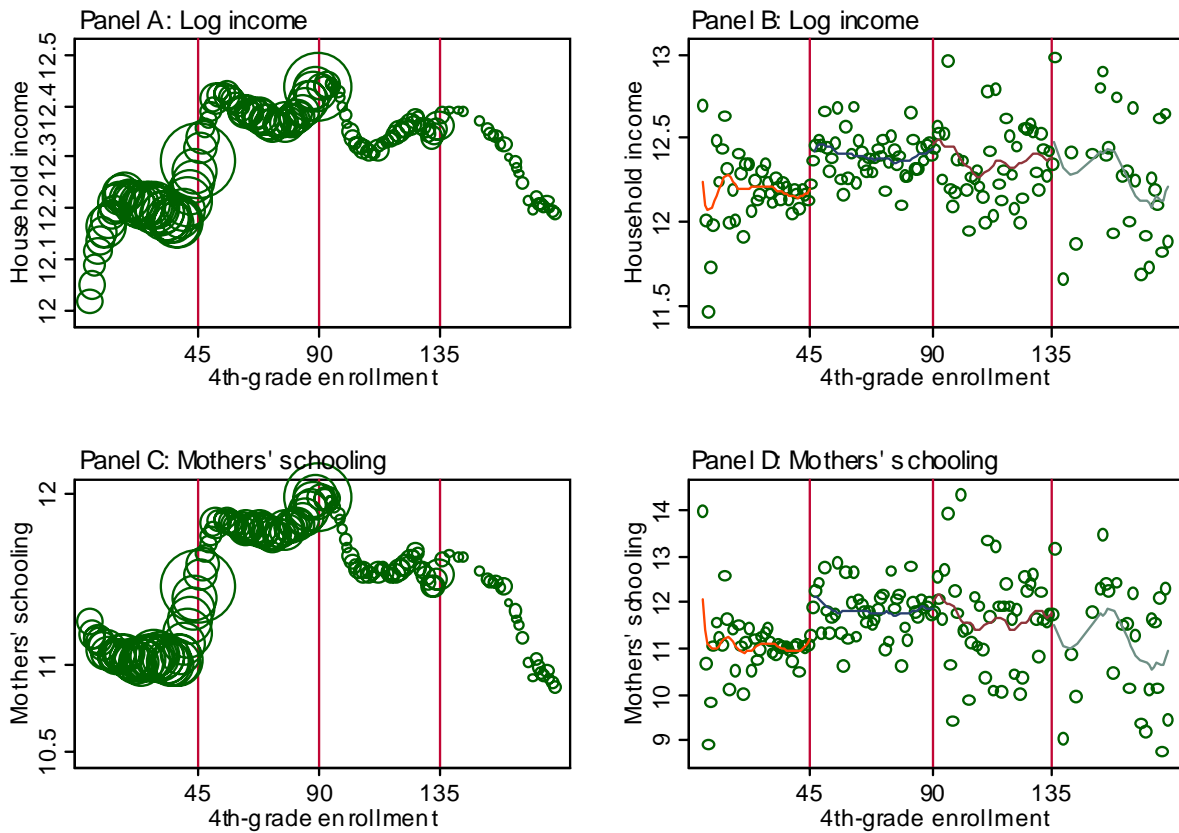
Note: Income and mothers' schooling come from 2002 individual SIMCE information aggregated to the school level. Enrollment is from administrative data for the same year. Panels A and C present the fitted values of a locally-weighted regression of average log income and mothers' schooling on enrollment, where the size of each circle is proportional to the number of student observations in each enrollment cell. Panels B and D present the corresponding "raw" enrollment-cell means, along with the fitted values of a locally weighted regression calculated within each enrollment segment. Only data for schools with 4th grade enrollments below 180 are plotted; this excludes less than two percent of all schools.

Table 1: Descriptive statistics for urban schools, 2002

Sample/variable			Quantile				
	Mean	S.D.	10th	25th	50th	75th	90th
Panel A: Full sample							
Income	311.2	341.7	102.5	130	180	303.7	781.8
Mothers' schooling	11.1	2.4	8.3	9.3	10.7	12.7	14.9
Fathers' schooling	11.3	2.5	8.5	9.4	10.8	12.9	15.2
Math score	249.3	30	212.6	227.7	246.9	269.3	291.9
Language score	253.1	30.5	215.2	231.3	251.5	275.1	296
4th grade class size	32.9	9.2	20	27	34	40.3	44.3
No. of 4th grade clases	1.88	1.09	1	1	2	2	3
4th grade enrollment	65	45.1	21	33	56	86	122
Panel B: Public schools							
Income	152.8	69.1	94.4	113.6	138.6	172.1	217
Mothers' schooling	9.6	1.3	8.1	8.6	9.4	10.4	11.3
Fathers' schooling	9.7	1.4	8.1	8.8	9.6	10.6	11.5
Math score	235.1	21.4	208.7	220.6	233.8	248.8	262.2
Language score	237.9	21.6	211.2	223.9	237.5	251.8	265.8
4th grade class size	34.7	7.2	25	30	35.5	40.3	44
No. of 4th grade clases	2.1	1	1	1	2	3	3
4th grade enrollment	75.4	42	29	41.5	70	98	130
Panel C: Voucher private schools							
Income	250.8	142.6	115.7	155.8	213.1	307.1	428.2
Mothers' schooling	11.5	1.8	8.9	10.2	11.6	12.8	13.8
Fathers' schooling	11.5	1.9	9	10.2	11.6	12.9	13.9
Math score	252.2	27.5	215.1	234.4	254.3	271.3	287.5
Language score	256.9	28.4	218.4	239.1	259.7	277.4	290.7
4th grade class size	34.2	9.3	21	28	36	42	45
No. of 4th grade clases	1.7	1.07	1	1	1	2	3
4th grade enrollment	61.4	48.1	21	31	45	82.5	112
Panel D: Unsubsidized private schools							
Income	1050.2	419.3	506.3	770.9	1003.6	1350	1673.5
Mothers' schooling	15.1	1.2	13.8	14.7	15.4	15.9	16.3
Fathers' schooling	15.6	1.3	14	15	15.9	16.5	16.9
Math score	288.1	25.3	254.3	276	292.9	304.7	314.1
Language score	291.8	23.7	261.8	283	297.5	307	315.1
4th grade class size	22.4	8.6	10	16.8	23	28.5	34
No. of 4th grade clases	1.7	1	1	1	1	2	3
4th grade enrollment	41.7	33.2	10	17	31.5	57.5	85

Note: Data on income, parental schooling, and test scores are from 2002 individual SIMCE test information aggregated to the school level. Class size, the number of classes operated, and enrollment come from administrative data for the same year. The table covers only urban schools. Panel A describes all 3,776 schools in the sample, panel B covers 1,652 public schools, panel C refers to 1,636 voucher private schools, and panel D is based on 488 private unsubsidized institutions.

Table 2: Class size and income and mothers' schooling among urban private schools, 2002

	All private			Voucher private			Unsubsidized private		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A-dep. var: 4th grade class size									
Log income	60.0***	59.8***	78.6***	120.3***	125.5***	135.3***	-106.9***	-90.2***	-115.8***
	(7.7)	(7.8)	(9.0)	(15.2)	(15.6)	(16.7)	(26.0)	(26.2)	(34.3)
Log income^2	-2.5***	-2.5***	-3.3***	-4.9***	-5.1***	-5.5***	4.1***	3.4***	4.5***
	(0.3)	(0.3)	(0.4)	(0.6)	(0.6)	(0.7)	(1.0)	(1.0)	(1.3)
13 region dummies	No	Yes	No	No	Yes	No	No	Yes	No
318 commune dummies	No	No	Yes	No	No	Yes	No	No	Yes
R^2	0.161	0.188	0.276	0.038	0.073	0.209	0.052	0.112	0.265
N	2,124	2,124	2,124	1,636	1,636	1,636	488	488	488
Panel B-dep. var: 4th grade class size									
Mothers' schooling	9.8***	10.2***	11.9***	8.5***	8.7***	9.8***	-7.6**	-6.3*	-6.0
	(0.8)	(0.9)	(1.0)	(1.2)	(1.3)	(1.4)	(3.4)	(3.4)	(4.2)
Mothers' schooling^2	-0.5***	-0.5***	-0.5***	-0.4***	-0.4***	-0.4***	0.3**	0.2**	0.2
	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
13 region dummies	No	Yes	No	No	Yes	No	No	Yes	No
318 commune dummies	No	No	Yes	No	No	Yes	No	No	Yes
R^2	0.171	0.191	0.278	0.029	0.061	0.200	0.013	0.086	0.224
N	2,124	2,124	2,124	1,636	1,636	1,636	488	488	488

Note: Income and mothers' schooling are from 2002 individual-level SIMCE data aggregated to the school level. Class size comes from administrative data for the same year. *** indicates statistical significance at the 1% level; ** at 5%, and * at 10%.

Table 3: 1$^{\text{st}}$ stage, reduced form, and base IV specifications; urban private voucher schools, 2002

	1$^{\text{st}}$ stage	Reduced form		IV	
	Class size	Math score	Language score	Math score	Language score
	(1)	(2)	(3)	(4)	(5)
Class size				-0.7***	-0.6**
				(0.3)	(0.3)
1{$x \geq 46$}	-16.5***	11.8***	9.9***		
	(2.7)	(3.2)	(3.3)		
1{$x \geq 91$}	-4.9**	0.0	1.6		
	(2.3)	(4.0)	(4.0)		
1{$x \geq 136$}	-4.3**	11.5	10.9		
	(2.0)	(13.6)	(12.9)		
1{$x \geq 181$}	-3.4	11.2	11.5		
	(3.0)	(10.6)	(13.9)		
x	0.95***	0.1	0.2*	0.8***	0.8***
	(0.01)	(0.1)	(0.1)	(0.2)	(0.3)
(x-46)*1{$x \geq 46$}	-0.6***	-0.1	-0.2	-0.6**	-0.6**
	(0.1)	(0.2)	(0.2)	(0.3)	(0.3)
(x-91)*1{$x \geq 91$}	-0.3**	0.0	-0.1	-0.2*	-0.3**
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
(x-136)*1{$x \geq 136$}	0.0	-0.6	-0.4	-0.2	-0.1
	(0.1)	(0.4)	(0.4)	(0.2)	(0.2)
(x-181)*1{$x \geq 181$}	-0.1	0.2	0.2	0.1	0.1
	(0.1)	(0.5)	(0.6)	(0.3)	(0.4)
N	1,623	1,623	1,623	1,623	1,623
R^2	0.844	0.069	0.072		

Note: Test scores are from 2002 SIMCE individual-level data, aggregated to the school level. Class size and enrollment come from administrative information for the same year. *** indicates statistical significance at the 1% level; ** at 5%, and * at 10%. All regressions are clustered by enrollment levels, as Lee and Card (2004) suggest is appropriate in RD settings in which the assignment variable is discrete. The table focuses only on effects around the first fours cutoffs, excluding the less than 1 percent of schools that report 4$^{\text{th}}$ grade enrollments in excess of 225 students.

Table 4: Within-enrollment band regressions; urban private voucher schools, 2002

	Cutoff			Pooled cutoffs
	1st (45 students)	2nd (90 students)	3rd (135 students)	
	(1)	(2)	(3)	(4)
Panel A: 5 student interval				
Dep. var.: Math score				
Class size	-3.3	-32.6	-4.4	-3.4
	(2.5)	(185.3)	(7.3)	(2.4)
N	249	186	41	476
Panel B: 3 student interval				
Dep. var.: Math score				
Class size	-6.9	-55.2	-13.9**	-7.4
	(6.7)	(197.9)	(4.4)	(6.1)
N	185	145	33	363
Panel C: 5 student interval				
Dep. var.: Language score				
Class size	-3.1	-38.1	-4.0	-3.2
	(2.4)	(216.0)	(6.7)	(2.3)
N	249	186	41	476
Panel D: 3 student interval				
Dep. var.: Language score				
Class size	-7.0	59.7	-13.0**	-7.3
	(6.6)	(213.3)	(3.7)	(5.9)
N	185	145	33	363

Notes: Test scores are from 2002 SIMCE individual-level data, aggregated to the school level. Class size and enrollment come from administrative information for the same year. Columns present regressions within 5 (panels A and C) and 3 (panels B and D) student enrollment bands around the first three cutoffs. Separate results around the fourth cutoff are omitted for the sake of space; they account for less than 1 percent of all school observations. The IV specifications in these columns regress schools' average scores on class size, where the latter is instrumented by using an indicator for whether schools' enrollment is above the respective cutoff. As van der Klaauw (2002) indicates, these are equivalent to Wald estimates of the effect of class size around each discontinuity. Column 4 produces similar estimates pooling all three local samples. In this case, the three cutoffs ($1\{x>45\}$, $1\{x>90\}$, and $1\{x>135\}$) and three sample-specific intercepts serve as instruments; see van der Klaauw (2002). All regressions are clustered around enrollment levels, see Lee and Card (2004).

Table 5: Behavior of selected variables around enrollment cutoffs and IV specifications; urban private voucher schools, 2002

	Mothers' schooling (1)	Fathers' schooling (2)	Household income (3)	IV	
				Math (4)	Language (5)
Class size				-0.1	0.1
				(0.1)	(0.1)
$1\{x \geq 46\}$	0.93***	0.94***	66.6***		
	(0.2)	(0.2)	(14.1)		
$1\{x \geq 91\}$	0.03	0.03	17.6		
	(0.2)	(0.2)	(17.3)		
$1\{x \geq 136\}$	0.66	0.86	143.7*		
	(0.7)	(0.8)	(79.4)		
$1\{x \geq 181\}$	0.66	0.71	53.1		
	(1.1)	(1.1)	(77.7)		
x	-0.02*	-0.02*	-2.4***	0.4***	0.4***
	(0.0)	(0.0)	(0.8)	(0.1)	(0.1)
$(x-46)*1\{x \geq 46\}$	0.02*	0.01	2.3***	-0.4***	-0.4***
	(0.0)	(0.0)	(0.8)	(0.1)	(0.1)
$(x-91)*1\{x \geq 91\}$	-0.01	0	-0.7	0.1	0
	(0.0)	(0.0)	(0.6)	(0.1)	(0.1)
$(x-136)*1\{x \geq 136\}$	-0.02	-0.03	-3.5	-0.2**	-0.1
	(0.0)	(0.0)	(2.3)	(0.1)	(0.1)
$(x-181)*1\{x \geq 181\}$	0.01	0.02	4	0.1	0.1
	(0.0)	(0.0)	(3.4)	(0.1)	(0.2)
Mothers' schooling				8.5***	9.5***
				(0.9)	(1.0)
Fathers' schooling				1.6*	1.1
				(0.9)	(0.9)
Household income				13.4**	16.6***
				(5.4)	(5.5)
N	1,623	1,623	1,623	1,623	1,623
R^2	0.034	0.032	0.029		

Notes: Test scores are from 2002 SIMCE individual-level data, aggregated to the school level. Class size and enrollment come from administrative information for the same year. *** indicates statistical significance at 1%; ** at 5%, and * at 10%. All regressions are clustered by enrollment levels. The table focuses only on effects around the first fours cutoffs, excluding the less than 1 percent of schools that report 4th grade enrollments in excess of 225 students.

Table 6: Within-enrollment band regressions; urban private voucher schools, 2002

	Cutoff			Pooled cutoffs
	1st (45 students)	2nd (90 students)	3rd (135 students)	
	(1)	(2)	(3)	(4)
Panel A: 5 student interval; Dep. var: Math score				
Class size	-1.1	-16.1	-11.7	-1.1
	(1.2)	(83.1)	(25.8)	(1.0)
Mothers' schooling	7.5**	13.2	11.5	8.7**
	(2.5)	(24.2)	(12.1)	(1.9)
Fathers' schooling	2.8**	-4.9	19.2	1.4
	(1.4)	(25.3)	(43.9)	(1.6)
Household income	-2.9	7.7	-254.2	9.8
	(32.5)	(155.4)	(506.2)	(14.7)
N	249	186	41	476
Panel B: 3 student interval; Dep. var: Math score				
Class size	-3.3	39.6	25.7	-2.8
	(3.8)	(154.9)	(26.3)	(2.6)
Mothers' schooling	8.8	-9.6	12	10.9***
	(5.4)	(85.9)	(18.3)	(3.6)
Fathers' schooling	-0.3	10.1	-26.9**	-1.6
	(3.2)	(58.9)	(13.5)	(2.7)
Household income	8.6	76.6	469.2	11.5
	(52.4)	(120.3)	(413.2)	(25.5)
N	185	145	33	363
Panel C: 5 student interval; Dep. var--Language score				
Class size	-0.8	-20.2	-9.8	-0.8
	(1.0)	(103.7)	(21.1)	(0.8)
Mothers' schooling	7.7**	15.7	8.6	9.1***
	(2.1)	(29.8)	(10.7)	(1.6)
Fathers' schooling	2.6**	-5.7	18.5	1.6
	(1.1)	(31.2)	(35.7)	(1.2)
Household income	9.1	62.4	-212	11.2
	(25.0)	(194.8)	(413.3)	(12.6)
N	249	186	41	476
Panel D: 3 student interval; Dep. var.--Language score				
Class size	-3	42.9	21.1	-2.4
	(3.0)	(167.3)	(21.2)	(2.0)
Mothers' schooling	9.4**	-10.6	6.8	11.1***
	(4.5)	(91.7)	(15.4)	(2.9)
Fathers' schooling	-0.3	12.9	-17.7*	-0.7
	(2.6)	(62.8)	(6.8)	(2.0)
Household income	13	61.2	386.6	9.1
	(46.2)	(131.5)	(326.6)	(21.1)
N	185	145	33	363

Notes: Test scores are from 2002 SIMCE individual-level data, aggregated to the school level. Class size and enrollment come from administrative information for the same year. Columns present regressions within 5 (panels A and C) and 3 (panels B and D) student enrollment bands around the first three cutoffs. Separate results around the fourth cutoff are omitted; they account for less than 1 percent of observations. These specifications regress schools' average math scores on class size, where the latter is instrumented using an indicator for whether schools' enrollment is above the respective cutoff. As van der Klaauw (2002) indicates, these are equivalent to Wald estimates of the effect of class size around each discontinuity. Column 4 produces similar estimates pooling all three local samples. In this case, the three cutoffs ($1\{x>45\}$, $1\{x>90\}$, and $1\{x>135\}$) and three sample-specific intercepts serve as instruments; see van der Klaauw (2002). All regressions are clustered around enrollment levels to adjust for the fact that the assignment variable (enrollment) is discrete.