

## Class-Size Caps, Sorting, and the Regression-Discontinuity Design

By MIGUEL URQUIOLA AND ERIC VERHOOGEN\*

*This paper examines how schools' choices of class size and households' choices of schools affect regression-discontinuity-based estimates of the effect of class size on student outcomes. We build a model in which schools are subject to a class-size cap and an integer constraint on the number of classrooms, and higher-income households sort into higher-quality schools. The key prediction, borne out in data from Chile's liberalized education market, is that schools at the class-size cap adjust prices (or enrollments) to avoid adding an additional classroom, which generates discontinuities in the relationship between enrollment and household characteristics, violating the assumptions underlying regression-discontinuity research designs. (JEL D12, I21, I28, O15)*

There has been a long and heated debate on the effect of class size on student performance. Eric A. 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. Alan B. Krueger (2003), Michael R. 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 Joshua D. Angrist and Victor Lavy (1999), which exploits the discontinuous relationship between enrollment and class size that results from class-size caps.<sup>1</sup>

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 two 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

\* Urquiola: Columbia University, 420 W. 118th St., Room 1022, New York NY 10027, and NBER (e-mail: miguel.urquiola@columbia.edu.); Verhoogen: Columbia University, 420 W. 118th St., Room 1022, New York, NY 10027, and BREAD, CEPR, IZA, and NBER (e-mail: eric.verhoogen@columbia.edu.) We thank Kensuke Teshima for excellent research assistance. For useful comments we thank (without implicating) Josh Angrist, Jere Behrman, David Card, Ken Chay, Pierre-André Chiappori, Gregory Elacqua, Helios Herrera, Kate Ho, Larry Katz, David Lee, Richard Romano, Bernard Salanié, and many seminar participants. We especially thank Patrick McEwan, who was involved in the early stages of the project. This paper was previously circulated under the title "Class Size and Sorting in Market Equilibrium: Theory and Evidence."

<sup>1</sup> The RD approach has also been used to study the effects of class size by Caroline M. Hoxby (2000) in the United States, Simone Dobbela, Jesse D. Levin, and Hessel Oosterbeek (2002) in Holland, Martin Browning and Eskil Heinesen (2003) in Denmark, Pascal Bressoux, Francis Kramarz, and Corinne Prost (2005) and Thomas Piketty and Mathieu Valdenaire (2006) in France, M. Niaz Asadullah (2005) in Bangladesh, Ludger Wössmann (2005) in ten European countries, Maciej Jakubowski and Pawel Sakowski (2006) in Poland, and Urquiola (2006) in Bolivia.

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, corresponding to real restrictions faced by private schools that accept vouchers in Chile: a class size cap at 45 students; an integer constraint on the number of classrooms; and the restriction that enrollment (a choice variable of schools) cannot exceed demand.

The model delivers two main empirical predictions, both of which find support in the data. First, there is an inverted-U relationship 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 enables schools to fill up their existing classrooms, and it also leads them 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 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 (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 (David S. Lee 2008; Justin McCrary 2008). We view these results as a cautionary note regarding the application of such designs in contexts where schools are able to set prices and influence their enrollments, and where parents have substantial school choice. As we discuss below, we have no reason to believe that this conclusion generalizes to the public school settings typically studied, 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, this paper is related to a growing body of theoretical and empirical work on sorting in education markets, including Charles F. Manski (1992), Dennis N. Epple and Richard E. Romano (1998), Charles T. Clotfelter (1999), Epple, David N. Figlio, and Romano (2004), Epple and Romano (2008), Lars Nesheim (2002), Elizabeth Caucutt (2002), Thomas J. Nechyba (2003), Patrick J. Bayer, Robert McMillan, and Kim Rueben (2004), Joseph G. Altonji, Ching-I Huang, and Christopher R. Taber (2005), Urquiola (2005), Damon Clark (2005), Epple, Romano, and Holger Sieg (2006), Chang-Tai Hsieh and Urquiola (2006), Jesse M. Rothstein (2006), and Maria Marta Ferreyra (2007). Much of this work is focused on the impact of greater school choice—either through greater school district availability or through vouchers—on sorting outcomes. The distinctive aspect of this paper is our focus on the role of institutional constraints—the class-size cap and the integer constraint—in a market that is already largely liberalized.

One caveat is that this paper does not consider the role of peer effects, which play a central role in much of the previous theoretical work on sorting in educational markets (Epple and Romano 1998, 2008; Epple et al. 2004). In many of these models, schools are essentially passive “clubs” whose main attribute is the average ability and income of their students. In our model,

as in Epple et al. (2006), schools actively choose the level of educational quality to supply. It is difficult to integrate both the peer-effects and the quality-choice elements in an analytically tractable model. Epple et al. (2006) maintain both elements, but must rely on numerical methods to compute equilibria. Our approach is to abstract from peer effects in order to arrive at analytical results. In the long run it would clearly be desirable to develop an analytically tractable model that combines both elements.

This paper is also related to work on quality choice by firms (Michael Mussa and Sherwin Rosen 1978; J. Jaskold Gabszewicz and Jacques-François Thisse 1979; Avner Shaked and John Sutton 1982; Simon P. Anderson and André de Palma 2001). The structure of the theoretical model is similar to that of Verhoogen (2008), which models quality choice by heterogeneous Mexican firms facing heterogeneous consumers in the domestic and export markets.

Finally, in seeking to understand the mechanisms behind the determination of class size, we view our work as complementary to that of Edward P. 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.

The remainder of the paper is organized as follows. Section I provides institutional background and Section II sets out the model. Section III describes the data. Section IV discusses testable implications and presents the results. Section V concludes.

## I. Chile's School System

There are three main types of schools in Chile:

- (i) *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 demand exceeds capacity, and are limited to a maximum class size of 45.<sup>2</sup> In most municipalities, they are the suppliers of last resort.
- (ii) *Private subsidized* or *voucher* schools are independent, and since 1981 have received exactly the same per-student subsidy as municipal schools.<sup>3</sup> They are also constrained to a maximum class size of 45, but, unlike public schools, have wide latitude regarding student selection.
- (iii) *Private unsubsidized* schools are independent, do not accept vouchers, receive no other explicit subsidies, and are not bound by the class-size cap.

Parents can use the per-student voucher in any public or private voucher school that is willing to accept their children. In 2003, private schools (both voucher and unsubsidized) accounted for about 45 percent of all schools, and voucher schools alone accounted for about 36 percent. In urban areas, these shares were 62 and 48 percent, respectively. Private schools can be explicitly for-profit, and using their tax status to classify them, Gregory Elacqua (2005) calculates that about 70 percent of them are indeed operated as such. Further, even nonprofit schools can legally

<sup>2</sup> In some instances schools are temporarily authorized to have classes of 46 or 47, but they receive no payments for the students above 45.

<sup>3</sup> 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).

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.<sup>4</sup> While initially voucher private schools were subject to the same constraint, this restriction was eased beginning in 1994. At present, they can charge tuition as high as approximately 1.7 times the voucher payment. In practice, this constraint appears not to be important for most voucher schools; in 2006, for instance, fewer than 4 percent of them had per-student revenues within 25 percent of the tuition cap.<sup>5</sup> The resources these institutions raise through tuition are equal to about 20 percent of their state funding.

Rather than attempt to analyze the entire Chilean educational sector, we narrow our focus in four important ways. First, we restrict attention to primary (K–8) schools because class size, a central variable in our analysis, is more clearly defined at the primary than at the secondary level. Second, we focus on private schools since, as mentioned above, we can plausibly assume that they are profit-maximizing. Third, we focus on urban areas 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.<sup>6</sup> Fourth, we focus on voucher schools, the private schools subject to the class-size cap, and not on unsubsidized schools. We do so in part because we are primarily interested in how the class-size cap affects sorting outcomes, and in part because the unsubsidized schools serve a very distinct, elite population,<sup>7</sup> and appear to be governed by considerations that would be difficult to incorporate tractably into our theoretical framework.<sup>8</sup>

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 fourth grade.<sup>9</sup> As Figure 1 illustrates, they run relatively few classes per grade. In 2002, for instance, 53 percent of urban private schools had only one fourth 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 fourth grades. While in theory schools could combine students from more than one grade into a single classroom, in practice very few do, especially in the urban areas we focus on. In 2002, for instance, only 4.3 percent of urban voucher schools reported they combined more than one grade into a class.<sup>10</sup> These facts motivate the integer constraint in our model.

<sup>4</sup> Public secondary schools can charge add-ons, but in practice very few do.

<sup>5</sup> The administrative data on which this figure is based contain information on average revenue and not posted prices. The former can be lower than posted tuition if some students receive discounts.

<sup>6</sup> The qualitative conclusions of our empirical analyses turn out not to be much affected by this restriction.

<sup>7</sup> The summary statistics in Table 1, discussed in more detail below, indicate that the students attending private unsubsidized schools are from markedly richer households than those in voucher or public schools; for instance, the average household income at the tenth percentile of the income distribution in unsubsidized schools is greater than the average household income at the ninetieth percentile of the income distribution in voucher schools.

<sup>8</sup> Indeed, in the context of our model, it is a puzzle that many unsubsidized schools refuse to accept vouchers. Essentially all these schools have class sizes below 45, and while we do not have reliable data on their tuition, it appears that many of them also charge average fees well below the maximum allowed for voucher schools. Anecdotally, it appears that an important reason these schools do not accept vouchers is that exclusivity is part of their appeal. This appears to be related to peer effects and considerations of social status that are difficult to model tractably.

<sup>9</sup> As discussed below, we focus primarily on fourth grade observations because our testing data are at that grade level. The results for other grades are, however, quite similar.

<sup>10</sup> The administrative data do not allow us to discern how often this happened specifically at the fourth grade level, the one we focus on below. They simply report that the school did this for some combination of grades.

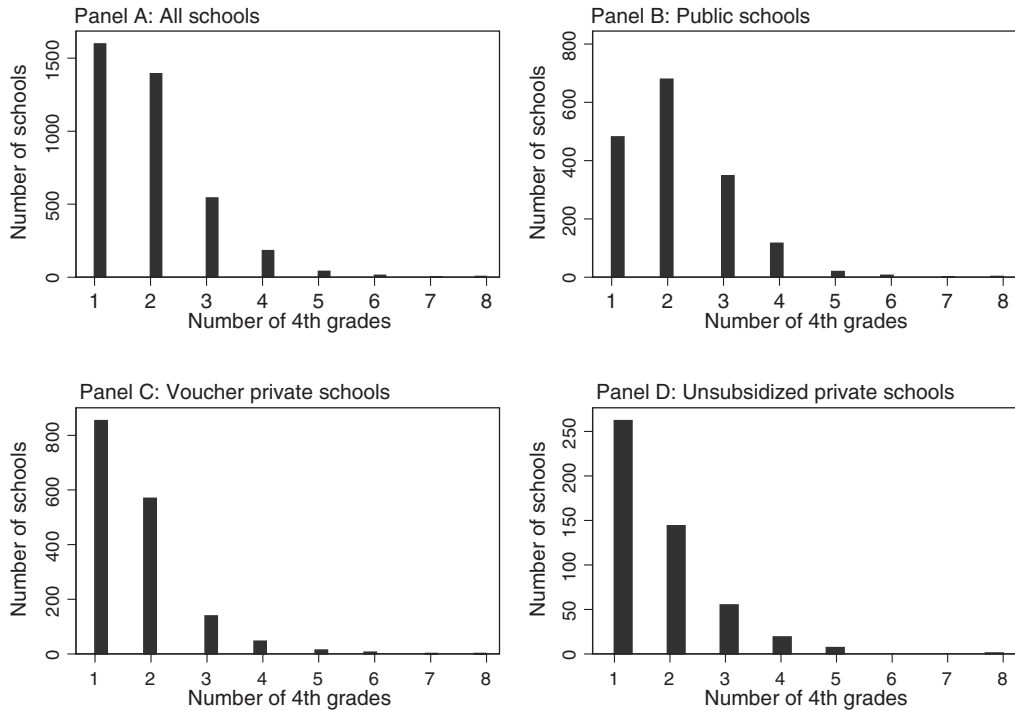


FIGURE 1. HISTOGRAMS OF THE NUMBER OF FOURTH GRADES IN URBAN SCHOOLS, 2002

*Notes:* Based on 2002 administrative data for urban schools with positive fourth grade enrollments. The figures cover only schools classified as urban by Chile's Ministry of Education. For voucher schools, panel C excludes about 0.2 percent of schools that report having more than eight fourth grade classes.

## II. 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 (Daniel L. McFadden 1973; Anderson et al. 1992). We solve the optimization problem of profit-maximizing voucher schools under realistic constraints. To simplify the model, we take the set of voucher schools as given and abstract from entry decisions. This is a strong assumption, but our view is that including a detailed analysis of entry 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. It is worth emphasizing that these two 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, and in Section IV we examine whether there is empirical support for them.

### A. Demand

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 school. The parameter  $\lambda$ , discussed in more detail below, indexes schools. Let  $x(\lambda)$ ,  $n(\lambda)$ , and  $p(\lambda)$  represent the enrollment, number of classrooms,

and tuition of school  $\lambda$ . We assume that school quality is observed by households and depends on enrollment, the number of classrooms, and  $\lambda$  in a manner to be made clear below. Households are assumed to have the following indirect utility function:

$$(1) \quad U(p(\lambda), q(x(\lambda), n(\lambda); \lambda); \theta) = \theta q(x(\lambda), n(\lambda); \lambda) - p(\lambda) + \varepsilon,$$

where  $q(\cdot)$  is school quality and  $\varepsilon$  is a random term capturing the utility of a particular household-school match.<sup>11</sup> This specification follows from a direct utility function in which households have identical utility functions and differ only in income.<sup>12</sup> The parameter  $\theta$  represents households' willingness to pay for quality, and is a monotonically increasing function of household income. We assume that  $\theta$  has a distribution  $g(\theta)$  with positive support over  $(\underline{\theta}, \bar{\theta})$ , where  $0 < \underline{\theta} < \bar{\theta}$ ; this distribution reflects the underlying distribution of income among households. We assume the random-utility term  $\varepsilon$  is i.i.d. across households with a double-exponential distribution with c.d.f.  $F(\varepsilon) = \exp[-\exp(-\varepsilon/\mu + \chi)]$ ,<sup>13</sup> where  $\mu$  is a positive constant that captures the degree of differentiation between schools.<sup>14</sup>

A standard derivation yields the probability that a household chooses school  $\lambda$ , conditional on having willingness to pay for quality  $\theta$  and on the qualities and prices of all schools:<sup>15</sup>

$$(2) \quad s(\lambda | \theta, q(\cdot), p(\cdot)) = \frac{1}{\Omega(\theta)} \exp\left(\frac{\theta q(x(\lambda), n(\lambda); \lambda) - p(\lambda)}{\mu}\right),$$

where

$$\Omega(\theta) \equiv \int_{\tilde{\lambda} \in \Lambda} \exp\left(\frac{\theta q(x(\tilde{\lambda}), n(\tilde{\lambda}); \tilde{\lambda}) - p(\tilde{\lambda})}{\mu}\right) f(\tilde{\lambda}) d\tilde{\lambda},$$

and  $\Lambda$  is the set of all schools in the market. We assume that schools cannot discriminate among households, and hence that price and quality are equal for all households in a given school. As is common in monopolistic-competition models, we will 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  $\lambda$ , integrating over all households, is

$$(3) \quad s(q(x(\lambda), n(\lambda); \lambda), p(\lambda)) = \int_{\underline{\theta}}^{\bar{\theta}} s(\lambda | \theta, q(x(\lambda), n(\lambda); \lambda), p(\lambda)) g(\theta) d\theta.$$

<sup>11</sup> As will become clear below, school quality will depend on class size,  $x/n$ , rather than on  $x$  and  $n$  separately. But because  $x$  and  $n$  will be separate choice variables of each school, it is convenient to treat them separately in this expression.

<sup>12</sup> Suppose  $\tilde{U}(z, q) = u(z) + q + \tilde{\varepsilon}$ , where  $z$  is a nondifferentiated numeraire good,  $q$  is the quality of education,  $\tilde{\varepsilon}$  is a mean-zero random term, and the subutility function  $u(\cdot)$  has  $u'(\cdot) > 0$  and  $u''(\cdot) < 0$ . If households are on their budget constraint, then indirect utility is  $\tilde{U}(p, q; y) = u(y - p) + q + \tilde{\varepsilon}$ . Taking a first-order approximation of  $u(\cdot)$  around  $y$ , and setting  $\theta \equiv 1/u'(y)$ ,  $U \equiv (\tilde{U}/u'(y)) - (u(y)/u'(y))$ , and  $\varepsilon \equiv \tilde{\varepsilon}/u'(y)$ , we have (1). Note that the  $u(y)/u'(y)$  term is constant across schools and does not affect the household's choice probabilities.

<sup>13</sup> We assume  $\chi = 0.5772$  (Euler's constant) to ensure that the expectation of  $\varepsilon$  is zero.

<sup>14</sup> As  $\mu \rightarrow 0$ , the distribution of household-school-specific utility terms collapses to a point, and the model approaches perfect competition.

<sup>15</sup> See, for example, Anderson et al. (1992, 39, theorem 2.2).

Expected demand for school  $\lambda$  is then

$$(4) \quad d\left(q(x(\lambda), n(\lambda); \lambda); p(\lambda)\right) = Ms\left(q(x(\lambda), n(\lambda); \lambda), p(\lambda)\right).$$

The key implications of this demand specification are that demand for school  $\lambda$  is declining in price and increasing in quality, and that higher- $\theta$  households are more sensitive to quality for a given price. Note that the 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  $d(\cdot)$  represents an expectation.

It will be convenient to define the expected willingness to pay of households that send their children to school  $\lambda$ :

$$(5) \quad \Theta\left(q(x(\lambda), n(\lambda); \lambda); p(\lambda)\right) \equiv E\left(\theta \mid q(x(\lambda), n(\lambda); \lambda), p(\lambda)\right) \\ = \int_{\underline{\theta}}^{\bar{\theta}} \theta \left[ \frac{s\left(\lambda \mid \theta, q(x(\lambda), n(\lambda), \lambda), p(\lambda)\right) g(\theta)}{s\left(q(x(\lambda), n(\lambda), \lambda), p(\lambda)\right)} \right] d\theta,$$

where by Bayes's rule the term in brackets represents the probability density of  $\theta$  conditional on households sending their children to school  $\lambda$ .

### B. Production

We now think of  $\lambda$  as an exogenously fixed productivity parameter in which schools are heterogeneous. It can be interpreted as the ability of the school principal/entrepreneur.<sup>16</sup> We assume that there is a continuum of schools with continuous density  $f(\lambda)$  over the interval  $[\underline{\lambda}, \bar{\lambda}]$ . Each school is uniquely identified by its value of  $\lambda$ , which justifies our use of  $\lambda$  as an index above.

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

$$(6) \quad q(x, n; \lambda) = \lambda \ln\left(\frac{T}{x/n}\right),$$

where  $x$  is enrollment,  $n$  is the number of classrooms, the denominator is class size, and  $T$  is a 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 the class size, the less teacher attention is available for each individual student.<sup>17</sup> Note that a given reduction in class size raises quality more at higher- $\lambda$  schools. This complementarity will be crucial in what follows.

<sup>16</sup> In a more complex, dynamic model, one might think of  $\lambda$  as reputation. The important point is that it affects households' perceptions of quality and is unaffected by schools' decisions in the short run.

<sup>17</sup> 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.

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

$$(7) \quad \mu > \bar{\lambda} \bar{\theta}$$

will be sufficient. Intuitively, this will limit the extent to which demand for a school increases with a given class-size reduction.

We suppose that 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. Recall that  $p$  is tuition, and let  $\tau$  be the per-student subsidy that schools receive from the government. Profit is then

$$(8) \quad \pi(p, n, x; \lambda) = (p + \tau - c)x - nF_c - F_s.$$

There is assumed to be no cost of differentiation; hence every school differentiates its "product" and has a monopoly over the product it offers.

### C. Schools' Optimization Problem

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

$$(9) \quad \max_{p, x, n} \pi(p, x, n; \lambda).$$

This optimization is subject to three constraints:

- (i) The number of classrooms must be a positive integer,

$$(10) \quad n \in \mathbb{N},$$

where  $\mathbb{N}$  is the set of natural numbers  $\{1, 2, 3, \dots\}$ .

- (ii) Class size cannot exceed the class-size cap:

$$(11) \quad \frac{x}{n} \leq 45.$$

- (iii) Enrollment cannot exceed demand:

$$(12) \quad x \leq d(q(x, n; \lambda), p),$$

where  $q(\cdot)$  is given by (6) and  $d(\cdot)$  by (4).

These three constraints correspond to realistic constraints facing Chilean schools. The first and third clearly generalize to other countries. The second constraint, the class-size cap, also exists in many (but not all) settings.<sup>18</sup>

<sup>18</sup> The third constraint ends up binding in every case we present here, and we could treat it as an equality constraint or substitute  $d(\cdot)$  for  $x$  in (8) and (9). But there exist realistic cases in which it would not bind—e.g., if there were a tuition constraint. For conceptual clarity, we leave the constraint as an inequality.



D. Characterization of Equilibrium

The integer restriction complicates the solution of schools' optimization problems, since we cannot simply solve a set of first-order conditions. A common approach to such problems is to first relax this constraint, then compare solutions with and without the relaxation. This is how we proceed below. In the main text, we report key results; derivations of those results appear in the mathematical appendices (Appendix A appears at the end of this article, and Appendix B is available online at <http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.1.179>).

Case 1: Divisible Classrooms

If the integer constraint (10) is relaxed, it turns out that in equilibrium there is a critical value of the entrepreneurial-ability parameter, call it  $\alpha$ , to the right of which the class-size cap does not bind and to the left of which it does bind. Consider each of these subcases in turn.

Subcase 1.1: Class-Size Cap Nonbinding

If  $\lambda > \alpha$  and the cap does not bind, then in equilibrium, schools' optimal choices are defined implicitly by the following:

$$(13a) \quad p^*(\lambda) = \mu + c - \tau + \lambda \Theta(q(x^*(\lambda), n^*(\lambda); \lambda); p^*(\lambda)),$$

$$(13b) \quad x^*(\lambda) = d(q(x^*(\lambda), n^*(\lambda); \lambda); p^*(\lambda)),$$

$$(13c) \quad n^*(\lambda) = \frac{1}{F_c} \lambda x^*(\lambda) \Theta(q(x^*(\lambda), n^*(\lambda); \lambda); p^*(\lambda)),$$

where  $\Theta(\cdot)$  is given by (5),  $d(\cdot)$  is given by (4), and the asterisks indicate equilibrium values. In order for the second-order conditions for a maximum to be satisfied, it must be the case that

$$(14) \quad \Psi \equiv \Theta(q(x^*(\lambda); n^*(\lambda); \lambda), p^*(\lambda)) - \frac{\lambda \sigma_{\theta|\lambda}^2}{\mu} > 0,$$

where  $\sigma_{\theta|\lambda}^2$  is the variance of  $\theta$  among households with children attending school  $\lambda$  in equilibrium.<sup>19</sup> Assumption (7) guarantees that this condition holds. (See the discussion of this subcase in Appendix A.) Unfortunately, there is no explicit analytical solution to (13a)–(13c). Nonetheless, using the implicit function theorem, we can sign the relationship between the various endogenous variables and the underlying productivity parameter,  $\lambda$ . In particular:

$$(16a) \quad \frac{dp^*}{d\lambda} > 0,$$

$$(16b) \quad \frac{dx^*}{d\lambda} > 0,$$

$$(16c) \quad \frac{dn^*}{d\lambda} > 0,$$

<sup>19</sup> That is, define:

$$(15) \quad \sigma_{\theta|\lambda}^2 \equiv \int_{\underline{\theta}}^{\bar{\theta}} (\theta - \Theta(q(x^*(\lambda), n^*(\lambda); \lambda), p^*(\lambda)))^2 \left[ \frac{s(\lambda | \theta, q(x^*(\lambda), n^*(\lambda); \lambda), p^*(\lambda)) g(\theta)}{s(q(x^*(\lambda), n^*(\lambda); \lambda), p^*(\lambda))} \right] d\theta.$$

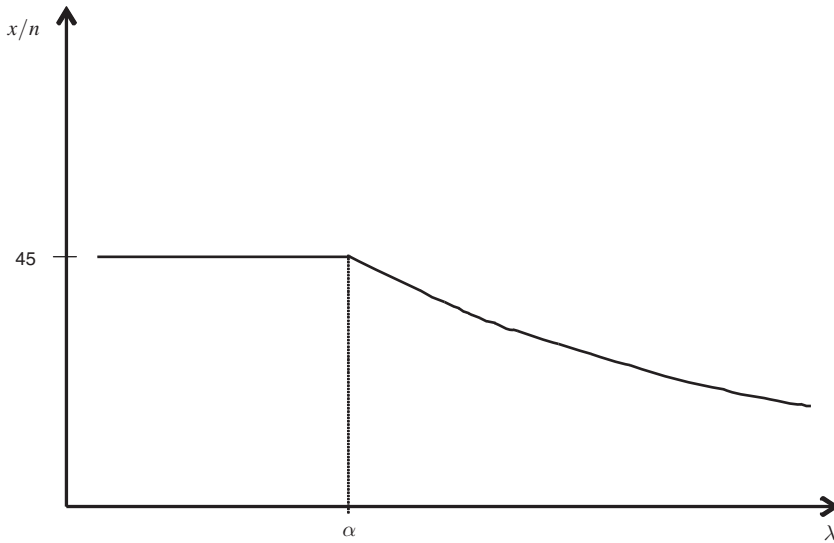


FIGURE 2. CASE 1— DIVISIBLE CLASSROOMS

$$(16d) \quad \frac{d\Theta^*}{d\lambda} > 0,$$

$$(16e) \quad \frac{d}{d\lambda} \left( \frac{x^*}{n^*} \right) < 0,$$

where  $\Theta^*$  is shorthand for the equilibrium value of  $\Theta(\cdot)$  for a given  $\lambda$ . In equilibrium, higher- $\lambda$  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  $\lambda$ .

In Figure 2, which plots class size versus  $\lambda$  in the divisible-classrooms case, this subcase corresponds to the declining portion of the curve, to the right of the critical value  $\alpha$ . Intuitively, the fact that class size is declining in  $\lambda$  is a consequence of the fact that  $\lambda$  and class-size reductions are complementary in the quality production function (6).

#### *Subcase 1.2: Class-Size Cap Binding*

If  $\lambda \leq \alpha$  and the class-size cap binds, then in equilibrium schools' optimal choices are defined implicitly by:

$$(17a) \quad p^*(\lambda) = c - \tau + \mu + \frac{F_c}{45},$$

$$(17b) \quad x^*(\lambda) = d(q(x^*(\lambda), n^*(\lambda)); \lambda, p^*(\lambda)),$$

$$(17c) \quad n^*(\lambda) = \frac{x^*(\lambda)}{45}.$$

The slopes with respect to  $\lambda$  in this subcase are

$$(18a) \quad \frac{dp^*}{d\lambda} = 0,$$

$$(18b) \quad \frac{dx^*}{d\lambda} > 0,$$

$$(18c) \quad \frac{dn^*}{d\lambda} > 0,$$

$$(18d) \quad \frac{d\Theta^*}{d\lambda} > 0,$$

$$(18e) \quad \frac{d}{d\lambda} \left( \frac{x^*}{n^*} \right) = 0.$$

Although class size and price are constant, enrollment, the number of classrooms, and average household income are increasing in the productivity parameter. Average household income is increasing in  $\lambda$  because  $\lambda$  raises quality conditional on class size. It is notable that profits are also increasing in  $\lambda$ , even though price is constant, because enrollment is increasing in  $\lambda$  and schools charge a mark-up over costs. In Figure 2, this subcase corresponds to the portion of the curve to the left of  $\alpha$  where class size is flat at 45.

The critical value  $\alpha$  is defined implicitly by the equation

$$(19) \quad \alpha\Theta(q(x^*(\alpha), n^*(\alpha); \alpha), p^*(\alpha)) = \frac{F_c}{45}.$$

This is the value of  $\lambda$  at which  $x/n = 45$  in the first subcase above. Note that there is no guarantee that  $\alpha \in (\underline{\lambda}, \bar{\lambda})$ , i.e., that the class-size cap will be binding on any schools in the market. At the critical value, the optimal choices  $p^*$ ,  $x^*$ , and  $n^*$  are equal in the two subcases (cap binding, cap nonbinding). Hence, over the entire range of  $\lambda$  we have that  $p^*$ ,  $x^*$ ,  $n^*$ , and  $\Theta^*$  are continuous;  $p^*$  is weakly monotonically increasing;  $x^*$ ,  $n^*$ , and  $\Theta^*$  are strictly monotonically increasing; and  $x^*/n^*$  is weakly monotonically decreasing.

### Case 2: Indivisible Classrooms

Now add the restriction that the number of classrooms must be an integer (10). Our strategy for dealing with the integer constraint is first to characterize the optimal choices of schools for a given number of classrooms, and then to characterize the sets of schools that choose each integer number of classrooms.

#### Fixed Number of Classrooms

To begin, suppose that  $n$  is fixed and think of it as a parameter. It will turn out that for a given  $n$  there is a single critical value of  $\lambda$ , call it  $\beta(n)$ , below which the class-size cap does not bind and above which it binds. Again, consider the two subcases in turn.

##### Subcase 2.1: Class-Size Cap Nonbinding

If  $\lambda \leq \beta(n)$  and the class-size cap does not bind, then schools' optimal choices are implicitly defined by

$$(20a) \quad p^*(n, \lambda) = \mu + c - \tau + \lambda\Theta(q(x^*(n, \lambda); n, \lambda), p^*(n, \lambda)),$$

$$(20b) \quad x^*(n, \lambda) = d(q(x^*(n, \lambda); n, \lambda), p^*(n, \lambda)).$$

Price and average willingness to pay (i.e., average household income) are unambiguously increasing in  $\lambda$ :

$$(21a) \quad \frac{\partial p^*}{\partial \lambda} > 0,$$

$$(21b) \quad \frac{\partial \Theta^*}{\partial \lambda} > 0.$$

There is a subtlety in the relationship between enrollment and  $\lambda$ . On one hand, there is a direct effect of a higher  $\lambda$  on demand: for a given class size, households prefer higher- $\lambda$  schools. On the other hand, there is an indirect effect: we see in (6) that at higher values of  $\lambda$  a given increase in enrollment has a larger negative effect on quality and hence on demand. It is theoretically possible in this case that the latter effect dominates, making it optimal for higher- $\lambda$  schools to raise prices such that enrollment, conditional on a given number of classrooms, is decreasing in  $\lambda$ . 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  $\lambda$ . A necessary and sufficient condition for this, which we assume hereafter, is<sup>20</sup>

$$(22) \quad \ln\left(\frac{nT}{x^*(n, \lambda)}\right) > \frac{\Theta(q(x^*(n, \lambda); n, \lambda), p^*(n, \lambda))}{\Psi},$$

where  $\Psi$  is defined as in (14). Under this assumption, we have

$$(23) \quad \frac{\partial x^*}{\partial \lambda} > 0.$$

Since  $n$  is fixed, (23) implies that  $(\partial/\partial\lambda)(x^*/n) > 0$ ; for a given number of classrooms, class size is *increasing* in  $\lambda$ . That is, conditional on  $n$ , higher- $\lambda$  schools are better able to fill their classrooms.

#### *Subcase 2.2: Class-Size Cap Binding*

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 values of  $p$  and  $x$ :

$$(24a) \quad p^*(n, \lambda) = \mu \ln \Sigma - \mu \ln(45n),$$

$$(24b) \quad x^*(n, \lambda) = 45n,$$

<sup>20</sup> If we replace the production function for quality (6) 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  $\lambda$  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

$$(25) \quad \Sigma \equiv \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{\Omega(\theta)} \left( \frac{T}{45} \right)^{\frac{\theta\lambda}{\mu}} g(\theta) d\theta.$$

The slopes with respect to  $\lambda$  are

$$(26a) \quad \frac{\partial p^*}{\partial \lambda} > 0,$$

$$(26b) \quad \frac{\partial x^*}{\partial \lambda} = \frac{\partial}{\partial \lambda} \left( \frac{x^*}{n} \right) = 0,$$

$$(26c) \quad \frac{\partial \Theta^*}{\partial \lambda} > 0.$$

The critical value  $\beta(n)$  for a given  $n$  is defined implicitly by the equation

$$(27) \quad \beta(n) \Theta \left( q(x^*(\beta(n)); n, \beta(n)), p^*(\beta(n)) \right) = \mu \ln \Sigma - \mu \ln(45n) - c + \tau - \mu.$$

The critical value of  $\lambda$  is the point at which class size reaches 45 in subcase 2.1. At this value, the optimal choices  $p^*$  and  $x^*$  are the same in the two subcases. Hence, for a given  $n$ ,  $p^*$ ,  $x^*$ , and  $\Theta^*$  are continuous at  $\beta(n)$ ,  $p^*$  and  $\Theta^*$  are strictly monotonically increasing in  $\lambda$ , and  $x^*$  and  $x^*/n$  are weakly monotonically increasing in  $\lambda$ .

### Optimal Choice of Number of Classrooms

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

$$(28) \quad \tilde{\pi}(n, \lambda) \equiv \pi(p^*(n, \lambda), x^*(n, \lambda); n, \lambda)$$

be school  $\lambda$ 's optimal profit when the number of classrooms is fixed at  $n$ , where  $p^*(n, \lambda)$  and  $x^*(n, \lambda)$  are given by (20a)–(20b) for  $\lambda \leq \beta(n)$  and by (24a)–(24b) for  $\lambda > \beta(n)$ . Define  $\Lambda_k$  to be the set of all schools for which a given integer  $k$  is the optimal number of classrooms:

$$(29) \quad \Lambda_k = \{ \lambda : \tilde{\pi}(k, \lambda) \geq \tilde{\pi}(j, \lambda) \forall j \neq k, j, k \in \mathbb{N} \}.$$

The following lemma characterizes the sets  $\Lambda_k$ :

**LEMMA 1:** *There exist unique positive integers  $\underline{k}$  and  $\bar{k}$  and a unique set of critical values  $\nu_{\underline{k}}, \nu_{\underline{k}+1}, \dots, \nu_{\bar{k}-1}, \nu_{\bar{k}}$  such that:*

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

where  $\underline{\lambda} = \nu_{\underline{k}-1} < \nu_{\underline{k}} < \dots < \nu_{\bar{k}-1} < \nu_{\bar{k}} = \bar{\lambda}$ .

The proof is in online Appendix B. The lemma indicates that the set of schools can be partitioned into a set of intervals,  $[\nu_{\underline{k}-1}, \nu_{\underline{k}})$ ,  $[\nu_{\underline{k}}, \nu_{\underline{k}+1})$ ,  $[\nu_{\underline{k}+1}, \nu_{\underline{k}+2})$ , etc., where the optimal integer number of

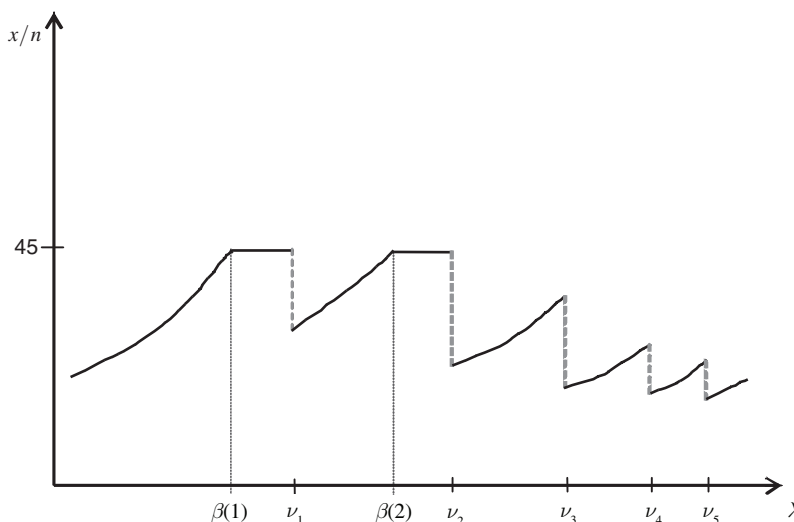


FIGURE 3. CASE 2 — INDIVISIBLE CLASSROOMS

classrooms is  $\underline{k}$  in the first interval,  $\underline{k} + 1$  in the next,  $\underline{k} + 2$  in the next, and so on. Within each of the subsets  $\Lambda_k$ , the results above for fixed  $n$  hold.

The discussion of discontinuities in Appendix B shows that at the critical values  $\nu_k, \nu_{k+1}, \dots, \nu_{\underline{k}-1}$ , enrollment is strictly increasing, class size is weakly decreasing, and average willingness to pay is strictly increasing in  $\lambda$ .<sup>21</sup> The fact that average willingness to pay is strictly increasing at the critical values, and that it is strictly increasing between the critical values (refer to (21b) and (26c)), means that in equilibrium average willingness to pay  $\Theta^*$  is monotonically increasing in  $\lambda$  for all  $\lambda$ .

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

Figure 3 plots class size against  $\lambda$  for the case where  $\beta(k) < \nu_k$  for  $k = 1$  and  $k = 2$  but not thereafter. The curve roughly resembles the curve for the divisible-classrooms case (Figure 2), but with an overlaid saw-tooth pattern generated by the integer constraint. The figure exhibits an approximately inverted-U relationship between class size and  $\lambda$ . It results from the interaction of two effects: conditional on a value of  $n$ , class size is increasing in  $\lambda$ , since greater values of  $\lambda$  make schools better able to fill their classrooms; and across values of  $n$ , class size is declining in  $\lambda$ , since greater  $\lambda$  leads schools to increase the number of classrooms, reducing class size.

The inverted-U relationship between class size and  $\lambda$  is not itself testable, because  $\lambda$  is unobserved. But since  $\Theta^*$  is strictly increasing in  $\lambda$  for all  $\lambda$ , we would also expect an inverted-U relationship in equilibrium between class size and  $\Theta^*$ . This gives us our first testable implication:

*Testable Implication 1: In equilibrium, there is an approximately inverted-U relationship between class size and average household income.*

<sup>21</sup> The direction of the change in price at each critical value is ambiguous in this case. The slope  $\partial p / \partial \lambda$  is greater when the class-size cap binds than when it does not bind, since higher  $\lambda$  leads schools to raise prices to keep class-size pegged at 45. Consequently, price may be greater to the left of  $\nu_k$  than to the right.

Figure 3 also illustrates that schools between  $\beta(1)$  and  $\nu_1$  have enrollment of 45 and schools between  $\beta(2)$  and  $\nu_2$  have enrollment of 90. Intuitively, in these regions schools raise tuition rather than incur the fixed cost of starting a new classroom.<sup>22</sup> Although we do not model the possibility explicitly, one can easily imagine that in the presence of stochasticity in demand and menu costs of changing tuition, schools might turn away potential students for the same reason.<sup>23</sup> Since average household income is monotonically increasing in  $\lambda$ , the stacking implies discontinuous changes in average household income with respect to enrollment at enrollments of 45, 90, and so on.<sup>24</sup> More generally, we have our second testable implication:

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

It is worth considering briefly what our model predicts for private unsubsidized schools, the elite private schools that do not accept vouchers and are not subject to the class-size cap. In the context of the model, the results for unsubsidized schools would be similar to those for voucher schools in the subcases where the class-size cap is not binding. That is, if we were to set the value of the voucher to zero and suppose that  $\beta(k) > \nu_k$  for all  $k$ , then the results in subcases 1.1 and 2.1 would carry over to unsubsidized schools. In this case, the figure analogous to Figure 3 would not have the flat regions between  $\beta(1)$  and  $\nu_1$  and between  $\beta(2)$  and  $\nu_2$ .

### III. Data

To examine our model's implications, we draw on two sources of information. The first is school-level administrative information on grade-specific enrollments and the number of classrooms from the Chilean Ministry of Education; we use these data to calculate average class sizes in each grade. The second source of information is the SIMCE testing system,<sup>25</sup> which tracks schools' math and language performance. SIMCE data are available at the school level from 1988 on. Since 1997, they also exist at the individual level and include information on students' household income, parental schooling, and other characteristics from a parental questionnaire sent home with students.

Depending on the year, the SIMCE tests fourth, eighth, or tenth graders. We focus on the fourth grade because, as indicated above, class size is best defined in early primary grades. We focus on the 2002 cross section because it is the most recent fourth grade testing round for which we have data. We note, however, that the general conclusions we obtain emerge in other cross sections we have analyzed (for instance, the 1999 fourth grade and the 2004 eighth grade waves).

Table 1 presents descriptive statistics for public schools, private voucher schools, and public unsubsidized schools. As noted in Section I, the table indicates that private unsubsidized schools serve a very different demographic group than either the voucher or the public schools. Household

<sup>22</sup> The discussion of Case 2 in Appendix A shows that for a given number of classrooms, tuition ( $p$ ) is more steeply sloped in  $\lambda$  in the region where the class-size cap binds than in the region where the cap does not bind.

<sup>23</sup> 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.

<sup>24</sup> Technically speaking, as long as the class-size cap is not binding to the right of the critical value,  $\nu_k$ , the model predicts a discontinuity in  $\Theta^*$  even in the absence of stacking. Average willingness to pay in equilibrium is a smooth, monotonically increasing function of class size. (See the discussions leading up to (A25c) and (A30) in online Appendix B.) Class size jumps discontinuously at the critical values  $\nu_k$ . Hence  $\Theta^*$  must also jump at  $\nu_k$ . Our view, however, is that the discontinuity due to stacking is the more empirically important one.

<sup>25</sup> SIMCE, which stands for *Sistema de Medición de la Calidad de la Educación* (Educational Quality Measurement System), is Chile's standardized testing program.

TABLE 1—DESCRIPTIVE STATISTICS FOR URBAN SCHOOLS, 2002

Sample/variable	Quantile						
	Mean	Standard deviation	10th	25th	50th	75th	90th
<i>Panel A: Full sample</i>							
Income	311.2	341.7	102.5	130.0	180.0	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	30.0	212.6	227.7	246.9	269.3	291.9
Language score	253.1	30.5	215.2	231.3	251.5	275.1	296.0
4th grade class size	32.9	9.2	20.0	27.0	34.0	40.3	44.3
Number of 4th grade classes	1.9	1.0	1.0	1.0	2.0	2.0	3.0
4th grade enrollment	65.0	45.1	21.0	33.0	56.0	86.0	122.0
<i>Panel B: Public schools</i>							
Income	152.8	69.1	94.4	113.6	138.6	172.1	217.0
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	25.0	30.0	35.5	40.3	44.0
Number of 4th grade classes	2.1	1.0	1.0	1.0	2.0	3.0	3.0
4th grade enrollment	75.4	41.9	29.0	41.5	70.0	98.0	130.0
<i>Panel C: Voucher private schools</i>							
Income	250.8	142.6	115.7	155.8	213.1	307.1	428.3
Mothers' schooling	11.5	1.8	8.9	10.2	11.6	12.8	13.8
Fathers' schooling	11.5	1.9	9.0	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	21.0	28.0	36.0	42.0	45.0
Number of 4th grade classes	1.7	1.1	1.0	1.0	1.0	2.0	3.0
4th grade enrollment	61.4	48.1	21.0	31.0	45.0	82.5	112.0
<i>Panel D: Unsubsidized private schools</i>							
Income	1,050.2	419.3	506.3	770.9	1,003.6	1,350.0	1,673.5
Mothers' schooling	15.1	1.2	13.8	14.7	15.4	15.9	16.3
Fathers' schooling	15.6	1.3	14.0	15.0	15.9	16.5	16.9
Math score	288.1	25.3	254.3	276.0	292.9	304.7	314.1
Language score	291.8	23.7	261.8	283.0	297.5	307.0	315.1
4th grade class size	22.4	8.6	10.0	16.8	23.0	28.5	34.0
Number of 4th grade classes	1.7	1.0	1.0	1.0	1.0	2.0	3.0
4th grade enrollment	41.7	33.2	10.0	17.0	31.5	57.5	85.0

*Notes:* Data on monthly income, parental schooling, and test scores come from 2002 individual-level SIMCE data aggregated at the school level. Income is in thousands of Chilean pesos (the average exchange rate for 2002 was 704 pesos per US dollar). 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 refers to 488 private unsubsidized institutions.

income and parents' schooling are markedly higher in the unsubsidized schools. Students' test scores, unsurprisingly, are highly correlated with these measures of socioeconomic status.

#### IV. Results

We now take our two testable implications to the data. We review each implication, discuss how it relates to the existing literature, and present the empirical results.



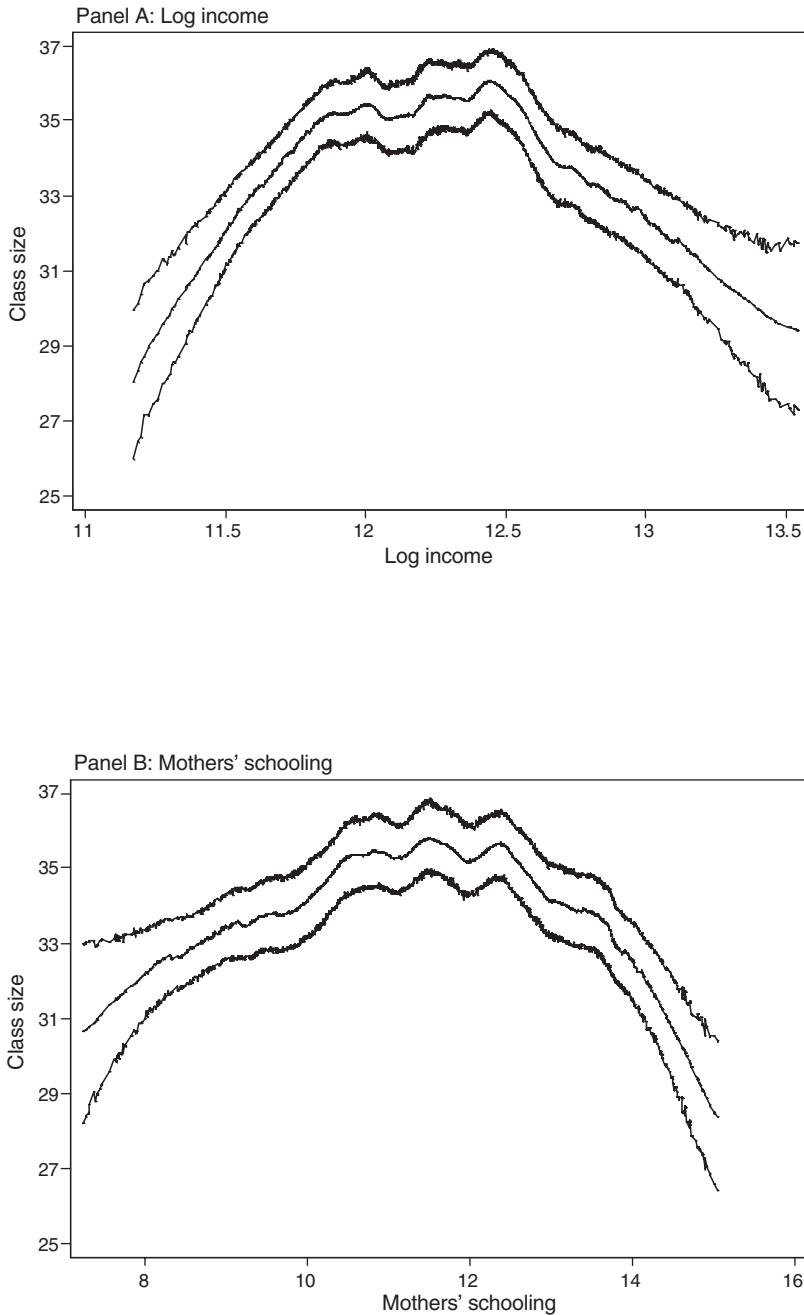


FIGURE 4. CLASS SIZE AND INCOME AND MOTHERS' SCHOOLING AMONG URBAN PRIVATE VOUCHER SCHOOLS, 2002

*Notes:* Income and mothers' schooling figures come from 2002 individual-level SIMCE data aggregated at the school level. Class size is based on 2002 administrative information. In each panel, the lines plot fitted values, along with point-wise confidence intervals, of locally weighted regressions (using Stata's *lowess* command and a bandwidth of 0.2) of class size on log income (panel A) and mothers' schooling (panel B). The figures omit observations below the first and above the ninety-ninth percentile of income or mothers' schooling. Standard errors are based on a bootstrap procedure with 1,000 replications. We note that the results are very similar if derived using regressions of class size on a fifth-order polynomial of the independent variables.

### A. Class Size and Income in Cross Section: The Inverted U

The first testable prediction is an inverted-U relationship between class size and average household income. The upward-sloping portion reflects the fact that low- $\lambda$  schools may have trouble filling their existing classrooms to achieve the desired class size. The downward-sloping portion reflects the fact that higher- $\lambda$  schools find it profitable to add classrooms and reduce class size to appeal to richer households. 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.

Panel A in Figure 4 plots class size against log average household income among urban voucher schools. The central line plots fitted values of a locally weighted regression of class size on log income, and the outer lines plot point-wise confidence intervals generated by a bootstrap procedure. A clear inverted-U pattern is evident. Panel B produces a similar conclusion, using mothers' schooling rather than income on the x-axis. 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 pattern may reflect the aggregation 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) among urban voucher schools. To facilitate interpretation, we use second-order polynomials. Column 1 reports results without region dummies, and columns 2 and 3 include dummies for 13 regions and 318 municipalities, respectively. The quadratic term is uniformly negative and significant, and not much affected by the regional controls. That is, the inverted-U pattern holds even within narrowly defined urban markets.

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 achievement tends to be strongly correlated with household income, and since income is often unobserved (or is observed with error), 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.<sup>26</sup>

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),<sup>27</sup> but to our knowledge our paper is the first to provide either a theoretical rationale or empirical evidence for a nonlinear 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.

### B. Stacking at Multiples of Class-Size Cap

Our second testable implication is related to regression-discontinuity (RD) designs that exploit the discontinuous relationship between enrollment and class size induced by class-size caps.<sup>28</sup>

<sup>26</sup> Consistent with Figure 4, for instance, we find that a cross-sectional bivariate regression of test scores on class size among all urban voucher schools results in 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.

<sup>27</sup> Alexandra Mizala and Pilar Romaguera (2002) present evidence of a positive correlation between enrollment and household socioeconomic status in Chile.

<sup>28</sup> For overviews and history of the RD design, see Angrist and Krueger (1999), Wilbert van der Klaauw (2002), and William R. Shadish, Thomas D. Cook, and Donald T. Campbell (2002).

TABLE 2—CLASS SIZE AND INCOME AND MOTHERS' SCHOOLING AMONG URBAN PRIVATE VOUCHER SCHOOLS, 2002

	Dependent variable: Fourth grade class size		
	(1)	(2)	(3)
<i>Panel A</i>			
Log income	120.3*** (15.2)	125.5*** (15.6)	135.3*** (16.7)
Log income <sup>2</sup>	-4.9*** (0.6)	-5.1*** (0.6)	-5.5*** (0.7)
13 region dummies	No	Yes	No
318 commune dummies	No	No	Yes
<i>R</i> <sup>2</sup>	0.038	0.073	0.209
<i>N</i>	1,636	1,636	1,636
<i>Panel B</i>			
Mothers' schooling	8.5*** (1.2)	8.7*** (1.3)	9.8*** (1.4)
Mothers' schooling <sup>2</sup>	-0.4*** (0.1)	-0.4*** (0.1)	-0.4*** (0.1)
13 region dummies	No	Yes	No
318 commune dummies	No	No	Yes
<i>R</i> <sup>2</sup>	0.029	0.061	0.201
<i>N</i>	1,636	1,636	1,636

Notes: Income and mothers' schooling figures are drawn from 2002 individual-level SIMCE data aggregated at the school level. Class size comes from administrative data for the same year.

\*\*\* Significant at the 1 percent level.

Figure 5 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 would be observed if schools mechanically expanded class size with enrollment until reaching the class-size cap, i.e., if class size were determined by

$$(30) \quad \left(\frac{x}{n}\right)^p = \frac{x}{\text{int}\left(\frac{x-1}{45}\right) + 1},$$

where the superscript  $p$  indicates the predicted level and the  $\text{int}(\cdot)$  function takes the greatest integer less than the given argument. 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 urban voucher schools for 2002, the circles plot enrollment-cell means of class size. Aggregated to the enrollment-cell level as in this figure, a regression of actual on predicted class size produces an  $R^2$  greater than 0.9—a clear "first stage."<sup>29</sup>

The idea behind RD designs, originally proposed by Donald L. Thistlewaite and Campbell (1960), is that discontinuities like those in Figure 5 can be used to identify the causal effect of class size even if enrollment is systematically related to factors that affect students' outcomes. Intuitively, if enrollment is smoothly related to student characteristics and other factors that affect achievement at multiples of the class-size cap, then, for example, students in schools with enrollments of 45 provide an adequate control group for those in schools with enrollments of 46.

<sup>29</sup> For visual clarity, Figure 5 excludes schools that declare fourth grade enrollments above 180 students (less than 2 percent of all schools), thus focusing on only the first three discontinuities in the enrollment/class size relation.

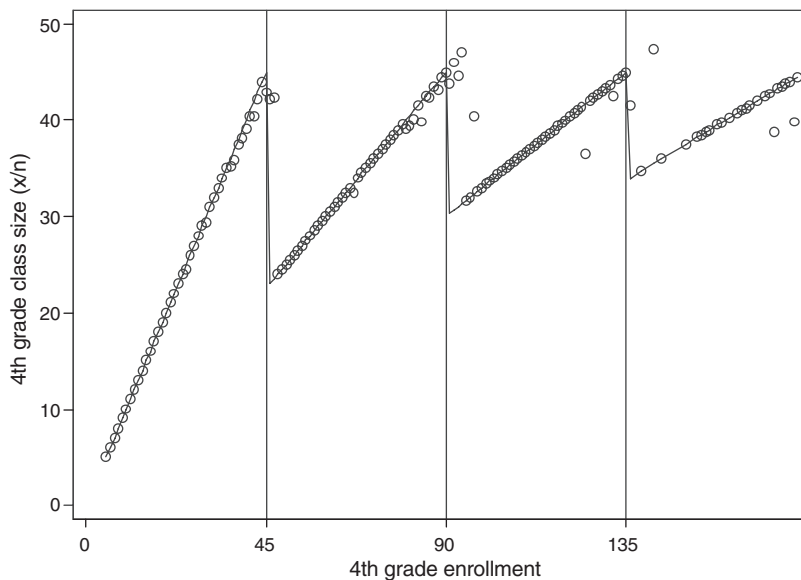


FIGURE 5. FOURTH GRADE ENROLLMENT AND CLASS SIZE IN URBAN PRIVATE VOUCHER SCHOOLS, 2002

*Notes:* 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 actual enrollment cell means of fourth grade class size. Only data for schools with fourth grade enrollments below 180 are plotted; this excludes less than 2 percent of all schools.

In such a case, differences in students' performance can be attributed to the very different class sizes they experience.

More formally, consider a standard RD model, assuming one class-size cutoff and a homogeneous effect of class size on test scores (Van der Klaauw 2002):

$$(31a) \quad TS_i = \gamma E(CS_i | X_i) + a(X_i) + u_i,$$

$$(31b) \quad E(CS_i | X_i) = \eta 1(X_i \geq \tilde{X}_0) + b(X_i),$$

where  $i$  indexes schools,  $TS_i$  is the average fourth-grade test score in school  $i$ ,  $CS_i$  is average fourth-grade class size,  $X_i$  is fourth-grade enrollment,  $\tilde{X}_0$  is the value of the class-size cutoff (i.e., 45),  $a(\cdot)$  and  $b(\cdot)$  are flexible functions of enrollment, and  $E(u_i | X_i) = 0$ . In the present setting this model corresponds to a “fuzzy” (as opposed to “sharp”) RD design, since, as Figure 5 indicates, enrollment affects, but does not perfectly explain, class size.

In the context of this model, if  $a(\cdot)$  and  $b(\cdot)$  are continuous at  $\tilde{X}_0$  and the mean of class size conditional on enrollment,  $E(CS_i | X_i)$ , is discontinuous at  $\tilde{X}_0$ , then the class-size effect,  $\gamma$ , is non-parametrically identified at the cutoff (Jinyong Hahn, Petra E. Todd, and Van der Klaauw 2001). Intuitively, in a small enough neighborhood around the cutoff,  $a(\cdot)$  and  $b(\cdot)$  are constant and any discontinuity in test scores can be attributed to the discontinuity in the conditional mean of class size. In practice, one rarely has enough data in neighborhoods around the cutoff to estimate  $\gamma$  precisely. As Lee and David Card (2008) point out, if  $X_i$  is discrete, as in our case, one cannot estimate the class-size effect nonparametrically, even with an infinite amount of data, and one must therefore choose parametric specifications for  $a(\cdot)$  and  $b(\cdot)$ . If these are specified correctly,

TABLE 3—FIRST-STAGE, REDUCED-FORM, AND BASE IV SPECIFICATIONS: URBAN PRIVATE VOUCHER SCHOOLS, 2002

	First stage	Reduced form		IV	
	Class size (1)	Math score (2)	Language score (3)	Math score (4)	Language score (5)
Class size				-0.7*** (0.3)	-0.6** (0.3)
$1\{x \geq 46\}$	-16.5*** (2.7)	11.8*** (3.2)	9.9*** (3.3)		
$1\{x \geq 91\}$	-4.9** (2.3)	0.0 (4.0)	1.6 (4.0)		
$1\{x \geq 136\}$	-4.3** (2.0)	11.5 (13.6)	10.9 (12.9)		
$1\{x \geq 181\}$	-3.4 (3.0)	11.2 (10.6)	11.5 (13.9)		
$x$	1.0*** (0.01)	0.1 (0.1)	0.2* (0.1)	0.8*** (0.2)	0.8*** (0.3)
$(x - 46) \times 1\{x \geq 46\}$	-0.6*** (0.1)	-0.1 (0.2)	-0.2 (0.2)	-0.6** (0.3)	-0.6** (0.3)
$(x - 91) \times 1\{x \geq 91\}$	-0.3** (0.1)	0.0 (0.1)	-0.1 (0.1)	-0.2* (0.1)	-0.3** (0.1)
$(x - 136) \times 1\{x \geq 136\}$	0.0 (0.1)	-0.6 (0.4)	-0.4 (0.4)	-0.2 (0.2)	-0.1 (0.2)
$(x - 181) \times 1\{x \geq 181\}$	-0.1 (0.1)	0.2 (0.5)	0.2 (0.6)	0.1 (0.3)	0.1 (0.4)
$R^2$	0.844	0.069	0.072		
$N$	1,623	1,623	1,623	1,623	1,623

Notes: Test scores are based on 2002 SIMCE individual-level data, aggregated at the school level. Class size and enrollment come from administrative information for the same year. All regressions are clustered by enrollment levels; see Lee and Card (2008). The table focuses only on effects around the first four cutoffs, excluding the less than 1 percent of schools that report fourth grade enrollments in excess of 225 students.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

then they capture all dependence of class size and test scores on enrollment away from the cutoff, and the instrumental-variable (IV) procedure corresponding to (31a)–(31b) will consistently estimate  $\gamma$ , effectively using only the discontinuity in  $E(CS_i | X_i)$ . In our baseline specifications, we use four class-size cutoffs, rather than just one as in (31b), and piecewise linear splines (with “kinks” at the values of the cutoffs) for  $a(\cdot)$  and  $b(\cdot)$ .<sup>30</sup>

Table 3 reports the results of a standard RD analysis using school-level data from 2002. Column 1 presents the first-stage regression of class size on indicators for whether enrollment is above the first four cutoffs, along with the piecewise linear spline for enrollment. The coefficients on the cutoff indicators are estimates of the average decline in class size at those breaks.<sup>31</sup> Consistent

<sup>30</sup> We return below to the possibility of using higher-order polynomials. We note that the piecewise splines are appropriate to the extent that equation (30) suggests that if the class size rule were strictly applied, the slope of the relationship between class size and enrollment would be different in each segment.

<sup>31</sup> For the sake of space, Table 3 and all subsequent tables exclude the small number of schools that declare fourth grade enrollments above 225 (less than 1 percent of all schools), thus focusing on only the first four discontinuities in the enrollment/class size relation.

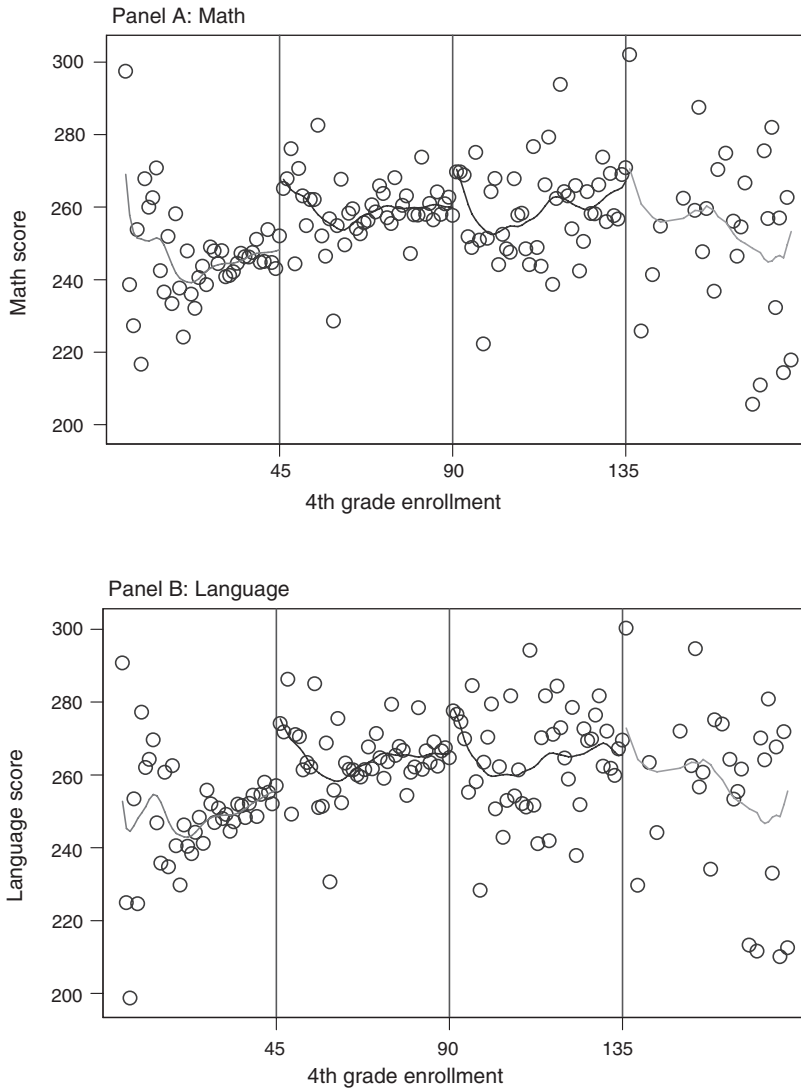


FIGURE 6. TEST SCORES AND ENROLLMENT IN URBAN PRIVATE VOUCHER SCHOOLS, 2002

*Notes:* Test scores come from 2002 individual-level SIMCE data aggregated at the school level, and enrollment is drawn from administrative data 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.

with the visual evidence in Figure 5, 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, but become progressively smaller.<sup>32</sup> In this specification, all standard errors are clustered by enrollment levels, as Lee and Card (2008) suggest is appropriate when the assignment variable (here enrollment) is discrete.

<sup>32</sup> Although we omit the results, adding controls for individuals’ characteristics has essentially no effect on the key coefficients.

TABLE 4—WITHIN-ENROLLMENT BAND IV REGRESSIONS: URBAN PRIVATE VOUCHER SCHOOLS, 2002

	Cutoff			Pooled cutoffs (4)
	1st (45 students) (1)	2nd (90 students) (2)	3rd (135 students) (3)	
<i>Panel A: Five-student interval</i>				
Dep. var.: Math score				
Class size	-3.3 (2.5)	-32.6 (185.3)	-4.4 (7.3)	-3.4 (2.4)
<i>N</i>	249	186	41	476
<i>Panel B: Three-student interval</i>				
Dep. var.: Math score				
Class size	-6.9 (6.7)	55.2 (197.9)	-13.9** (4.4)	-7.4 (6.1)
<i>N</i>	185	145	33	363
<i>Panel C: Five-student interval</i>				
Dep. var.: Language score				
Class size	-3.1 (2.4)	-38.1 (216.0)	-4.0 (6.7)	-3.2 (2.3)
<i>N</i>	249	186	41	476
<i>Panel D: Three-student interval</i>				
Dep. var.: Language score				
Class size	-7.0 (6.5)	59.7 (213.3)	-13.0** (3.7)	-7.3 (5.9)
<i>N</i>	185	145	33	363

*Notes:* Test scores are from 2002 SIMCE individual-level data, aggregated at the school level. Class size and enrollment come from administrative information for the same year. Columns present regressions within five (panels A and C) and three (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 observations. Columns 1–3 are IV specifications using an indicator for whether schools' enrollment is above the respective cutoff as an instrument. As Van der Klaauw (2002) indicates, these are equivalent to simple Wald estimates of the effect of class size around each discontinuity. Column 4 produces similar estimates pooling all three local samples and using the three cutoffs ( $1\{x > 45\}$ ,  $1\{x > 90\}$ , and  $1\{x > 135\}$ ) as instruments. All regressions are clustered around enrollment levels (see Lee and Card 2008).

\*\* Significant at the 5 percent level.

There is *prima facie* evidence that the standard RD strategy would generate significant results. Figure 6 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, the discrete reduction in class size is accompanied by an 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, showing positive and significant increases in scores at the first cutoff, and positive (although not significant) increases at subsequent ones. Columns 4–5 report IV specifications, where dummy variables for the first four cutoffs are used as instruments for class size. In both columns, *class size* appears to have a negative and significant effect on test scores.

Focusing more narrowly around the discontinuities as in Van der Klaauw (2002), columns 1–3 of Table 4 present IV results for bands of five students (panels A and C for math and language, respectively) and three students (panels B and D) around the first three breaks, omitting the

piecewise spline in enrollment.<sup>33</sup> These are equivalent to simple Wald estimates of the effect of class size within the bands around each cutoff.<sup>34</sup> Column 4 produces similar estimates pooling all three local samples.<sup>35</sup> In these pooled samples, the point estimates of the effect of class size on test scores are uniformly negative, although not statistically significant.<sup>36</sup>

One might be tempted to interpret Tables 3 and 4 as producing consistent 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. Recall that the theoretical model predicts a nonnegligible mass of schools stacking at enrollments of 45; in Figure 3, for instance, all schools with productivity parameters between  $\beta(1)$  and  $\nu_1$  have an enrollment of 45, and all those between  $\beta(2)$  and  $\nu_2$  have an enrollment of 90. Panel A of Figure 7 presents a histogram of fourth grade enrollment among urban voucher schools, and the evidence of such stacking is clear: more than five times as many schools report enrollments of 45 as report enrollments of 46. The same happens at higher cutoffs: more than seven times as many schools have 90 fourth graders as have 91, for instance.<sup>37</sup> Panel B of Figure 7 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. In short, Figure 7 provides a clear illustration of what McCrary (2008) terms manipulation of the running variable—enrollment, in this case.

Recall also that the model predicts that higher-income households on average sort into higher- $\lambda$  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 3 illustrates the intuition: because of the stacking, the average value of  $\lambda$  among schools at the cap is strictly less than the average value just above the cap; since average household income is strictly monotonically increasing in  $\lambda$  (see the discussion in Section IID above), the stacking generates discontinuous changes in household income at the class-size cutoffs.<sup>38</sup>

Panel A of Figure 8 presents enrollment cell means of household income, along with the fitted values of a locally weighted regression calculated within each enrollment segment. There is visual evidence that income changes discontinuously around the first cutoff—schools just to the right contain students who come from wealthier households. Panel B shows, not surprisingly, that they also have students with higher average mothers' schooling—as much as a full year. While jumps at the subsequent cutoff points are less evident, the clear discontinuities at the first one (which provides the bulk of observations, as shown in Figure 7) are sufficient to cast doubt on the RD approach in this context.

Columns 1–3 of Table 5 present regressions of household characteristics on the piecewise linear spline in enrollment and indicator variables for the class-size cutoffs. The results are consistent with the visual evidence from Figure 8. In particular, they confirm that *mothers' schooling*,

<sup>33</sup> The somewhat 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.

<sup>34</sup> 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.

<sup>35</sup> 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).

<sup>36</sup> Note that clustering by enrollment level, as suggested by Lee and Card (2008), lowers significance levels.

<sup>37</sup> Similar stacking occurs if first or eighth grade data are used.

<sup>38</sup> It is worth emphasizing that stacking alone may not violate the RD assumptions in our context. If student performance depended only on class size and not directly on  $\lambda$ , and there were no sorting, then the students on one side of the class-size cutoff would still serve as a valid control group for those on the other side. The violation of the RD assumptions arises from the interaction of the stacking and the endogenous sorting of households.



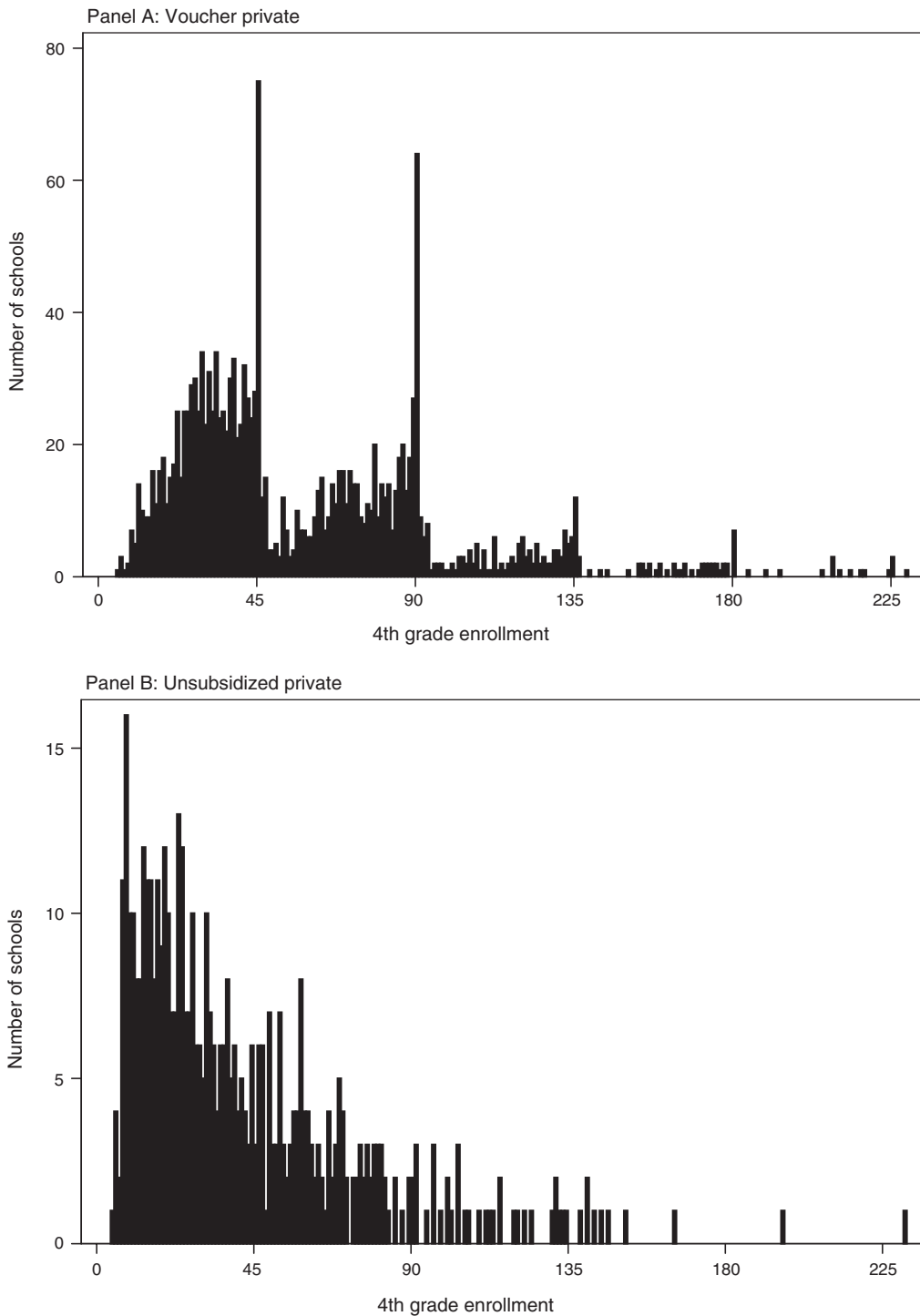


FIGURE 7. HISTOGRAMS OF FOURTH GRADE ENROLLMENT IN URBAN PRIVATE SCHOOLS, 2002

Notes: Enrollment is drawn from administrative data for 2002. For visual clarity, only schools with fourth grade enrollment below 225 are displayed. This excludes less than 1 percent of all schools.

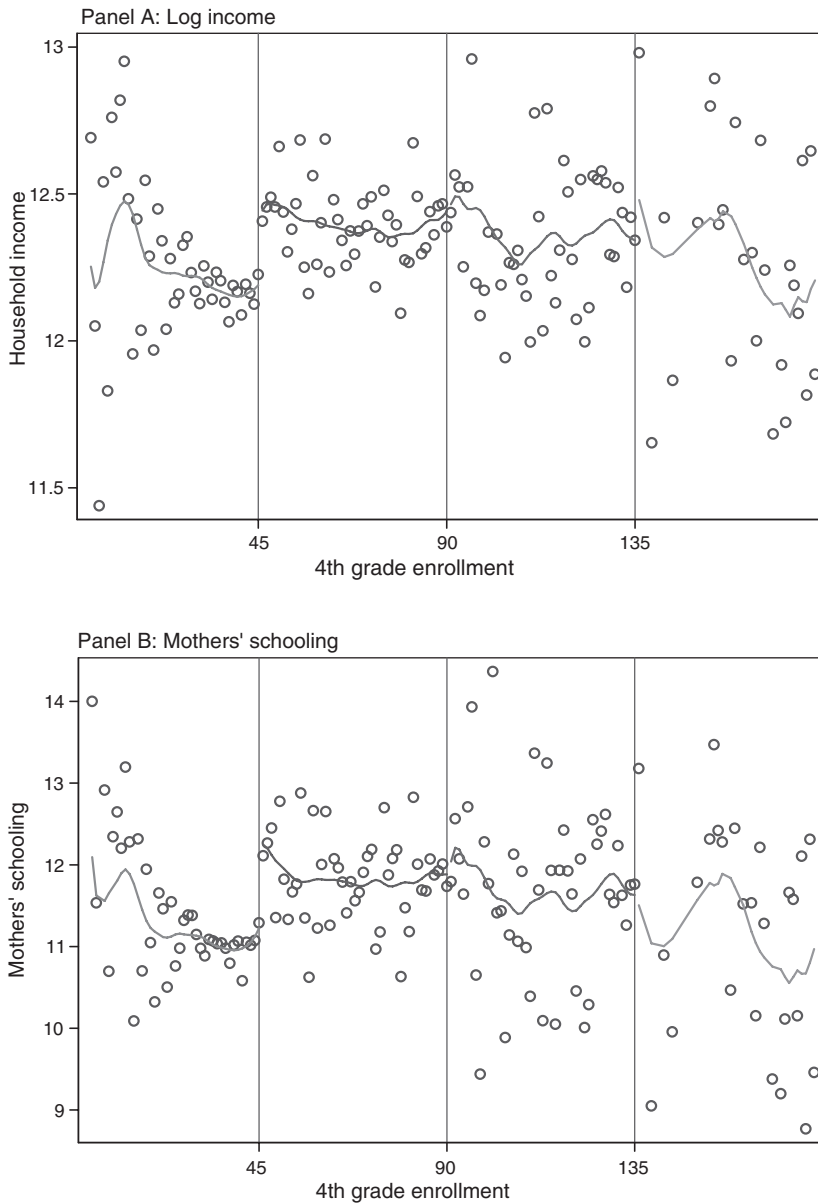


FIGURE 8. STUDENT CHARACTERISTICS AND ENROLLMENT IN URBAN PRIVATE VOUCHER SCHOOLS, 2002

*Notes:* Income and mothers' schooling come from 2002 individual-level SIMCE data aggregated at the school level. Enrollment is drawn from administrative data for the same year. The figure presents "raw" enrollment-cell means, along with the fitted values of a locally weighted regression calculated within each enrollment segment. Only data for schools with fourth grade enrollment below 180 are plotted; this excludes less than 2 percent of all schools.

*fathers' schooling*, and *income* display substantial and statistically significant jumps at the first enrollment cutoff; the coefficients for subsequent cutoffs are positive but not significant. This is further evidence that the continuity assumption on the function  $a(X_i)$  in (31a)—which is presumed to capture the effect of all variables that vary with enrollment except class size—is violated.

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*** (0.2)	0.94*** (0.2)	66.6*** (14.1)		
$1\{x \geq 91\}$	0.03 (0.2)	0.03 (0.2)	17.6 (17.3)		
$1\{x \geq 136\}$	0.66 (0.7)	0.86 (0.8)	143.7* (79.4)		
$1\{x \geq 181\}$	0.66 (1.1)	0.71 (1.1)	53.1 (77.7)		
$x$	-0.02* (0.0)	-0.02* (0.0)	-2.4*** (0.8)	0.4*** (0.1)	0.4*** (0.1)
$(x - 46) \times 1\{x \geq 46\}$	0.02* (0.0)	0.01 (0.0)	2.3*** (0.8)	-0.4*** (0.1)	-0.4*** (0.1)
$(x - 91) \times 1\{x \geq 91\}$	-0.01 (0.0)	0.00 (0.0)	-0.7 (0.6)	0.1 (0.1)	0 (0.1)
$(x - 136) \times 1\{x \geq 136\}$	-0.02 (0.0)	-0.03 (0.0)	-3.5 (2.3)	-0.2** (0.1)	-0.1 (0.1)
$(x - 181) \times 1\{x \geq 181\}$	0.01 (0.0)	0.02 (0.0)	4.0 (3.4)	0.0 (0.2)	0.1 (0.2)
Mothers' schooling				8.5*** (0.9)	9.6*** (1.0)
Fathers' schooling				1.6* (0.9)	1.1 (0.9)
Household income				13.5** (5.4)	16.6*** (5.5)
$R^2$	0.034	0.032	0.029		
$N$	1,623	1,623	1,623	1,623	1,623

Notes: Test scores and socioeconomic status measures are from 2002 SIMCE individual-level data, aggregated at the school level. Class size and enrollment come from administrative information for the same year. All regressions are clustered by enrollment levels. The table focuses only on effects around the first four cutoffs, excluding the less than 1 percent of schools that report fourth grade enrollments in excess of 225.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

For another indication that something is amiss in the RD design, consider the sensitivity of the estimates of the class-size effect to the inclusion of socioeconomic controls. If the RD approach were valid in this context, then including socioeconomic controls in (31a)–(31b) would have little effect on the estimate of  $\gamma$ . If the conditional means of the controls with respect to enrollment were continuous, then the component of each control that varied with enrollment would be captured by  $a(X_i)$  and  $b(X_i)$ , and the component that was orthogonal to enrollment would also be orthogonal to the instrument. In fact, columns 4–5 of Table 5 show that the IV estimates 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 change 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.<sup>39</sup> This is clear 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.

Finally, we explore to extent to which the observed discontinuities in the treatment and outcome variables, and in other covariates, are robust to more flexible specifications of  $a(X_i)$  and  $b(X_i)$  in (31a)–(31b). While the literature has not produced a consensus on the correct specification of these functions,<sup>40</sup> two practical guidelines for credible RD estimation seem uncontroversial. First, there should not be evidence of discontinuities in relevant covariates at the cutoffs using the same specification of  $a(X_i)$  used in the outcome equation (31a), as, for instance, we found in columns 1–3 of Table 5. Second, it is worrisome if the results, either for the estimates of  $\gamma$  or for the absence of discontinuities in other covariates, are not robust to the choice of functional form. To investigate this, Table 6 presents reduced-form regressions analogous to those in columns 1–3 of Tables 3 and 5, using higher-order specifications of the piecewise control function, as in Van der Klaauw (2002).<sup>41</sup> Note that the first stage (panel A) is not robust to the addition of more flexible controls—the estimated declines in class size become small or even positive at some cutoffs. This is somewhat surprising given the apparent strength of the first stage observed in Figure 5.<sup>42</sup> But, more important for our story, the discontinuous jumps in test scores and socioeconomic status (panels B–F) are robust and, if anything, grow stronger as one moves to the right in the table. In short, Table 6 reaffirms that sorting across the cutoffs appears too strong to allow for reliable implementation of an RD approach.

Lee and Card (2008) propose an additional practical procedure to guide the choice of polynomial for  $a(X_i)$  in the outcome equation in a context such as ours in which the running variable is discrete. The test is based on a goodness-of-fit statistic for a “restricted” low-order polynomial specification relative to an “unrestricted” model with a dummy for each value of the discrete running variable. This is not a definitive test—Lee and Card note that rejection of a given polynomial does not necessarily imply that the corresponding estimate of  $\gamma$  is inconsistent—but our confidence in a chosen polynomial specification increases if it cannot be rejected by this Lee-Card test. For the outcome equation (31a) in our case, the test does not reject *any* of the specifications in columns 1–4 of Table 6. This is consistent with our finding in panels B–F of Table 6 that the observed discontinuities in the outcome variables and household characteristics are robust to the choice of functional form of  $a(\cdot)$  and  $b(\cdot)$ .

To summarize, our results provide a concrete illustration of Lee’s (2008) observation that “economic behavior can corrupt the RD design.” It is worth emphasizing 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.<sup>43</sup> Further,

<sup>39</sup> These results are qualitatively similar if we use the predicted class size in equation (30) as an instrument for class size in place of the piecewise linear spline. Additionally, although we omit the exercise for the sake of space, we note that the simple within-band IV results in Table 4 are similarly sensitive to the addition of controls for socioeconomic status (results available from the authors).

<sup>40</sup> In their very useful review of practical issues in implementing RD designs, Guido W. Imbens and Thomas Lemieux (2008) do not make a specific recommendation.

<sup>41</sup> For the sake of space, we report only the coefficients on the first two cutoff indicators.

<sup>42</sup> We have verified that other years and grades of the administrative enrollment data produce similar results.

<sup>43</sup> Similarly, in some exercises, Urquiola (2006) considers Bolivian schools in rural towns in which school choice is likely to be very limited.

TABLE 6—REDUCED-FORM REGRESSIONS WITH PIECEWISE CONTROL FUNCTIONS OF HIGHER ORDER

	Specification of the piecewise control function			
	1st order (1)	2nd order (2)	3rd order (3)	4th order (4)
<i>Panel A—Dep. var.: Class size</i>				
1{x ≥ 46}	−16.5*** (2.7)	−10.0** (4.1)	−4.6 (4.0)	1.5 (3.1)
1{x ≥ 91}	−4.9** (2.3)	−2.4 (1.6)	2.7 (1.5)	3.3 (2.1)
<i>Panel B—Dep. var.: Math score</i>				
1{x ≥ 46}	11.8*** (3.2)	11.5*** (3.8)	20.2*** (3.7)	22.3*** (4.7)
1{x ≥ 91}	−0.0 (4.0)	6.6** (3.3)	13.6*** (2.4)	19.8*** (4.0)
<i>Panel C—Dep. var.: Language score</i>				
1{x ≥ 46}	9.9*** (3.3)	10.9*** (3.9)	19.3*** (3.1)	24.2*** (4.0)
1{x ≥ 91}	1.6 (4.0)	7.4** (3.2)	17.4*** (2.3)	18.6*** (3.4)
<i>Panel D—Dep. var.: Mothers' schooling</i>				
1{x ≥ 46}	0.93*** (0.20)	0.91*** (0.22)	1.09*** (0.21)	1.11*** (0.24)
1{x ≥ 91}	0.03 (0.22)	0.29 (0.26)	0.50 (0.31)	0.40 (0.40)
<i>Panel E—Dep. var.: Fathers' schooling</i>				
1{x ≥ 46}	0.94*** (0.20)	0.93*** (0.24)	1.02*** (0.24)	1.05*** (0.28)
1{x ≥ 91}	0.03 (0.22)	0.25 (0.26)	0.56* (0.32)	0.47 (0.41)
<i>Panel F—Dep. var.: Household income</i>				
1{x ≥ 46}	66.6*** (14.1)	76.1*** (14.5)	65.4*** (14.4)	68.4*** (19.0)
1{x ≥ 91}	17.6 (17.3)	29.6 (20.1)	38.7 (25.4)	22.3 (31.1)
<i>N</i>	1,623	1,623	1,623	1,623

*Notes:* Test scores and socioeconomic status measures are from 2002 SIMCE individual-level data, aggregated at the school level. Class size and enrollment come from administrative information for the same year. All regressions are clustered by enrollment levels. This table analyzes the robustness of reduced-form results to the inclusion of higher-order piecewise control functions (Van der Klaauw 2002). The results in column 1 (panels A–C) replicate those in Table 3 (columns 1–3). The results in column 1 (panels D–F) replicate those in Table 5 (columns 1–3). The remaining columns then explore how these results change as higher-order piecewise control functions are added. For the sake of space, these panels report only the coefficients on the first two cutoffs, which concentrate the majority of observations.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

migration and immigration may render it difficult for schools to predict enrollments, and private participation is limited to orthodox schools.<sup>44</sup>

<sup>44</sup> 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. For an alternative

## V. 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 nonexperimental 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 household characteristics at these points. These in turn violate the assumptions required for regression-discontinuity analyses of class size. The fact that a single, relatively parsimonious model can account for these two 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.<sup>45</sup>

Our results also recommend caution more broadly in the application and interpretation of RD designs. It has become increasingly common to use discontinuities in the application of regulations to estimate causal effects of those regulations. But in contexts where economic agents interact freely in markets, such discontinuities are likely to generate discontinuities along more than one dimension, undermining the RD approach. This general observation has been made before, by Lee (2008) and others. This paper has sought to provide a concrete illustration of such behavior, in a setting with heterogeneity and sorting on two sides of an important market.

## APPENDIX A

To avoid clutter, we do not write explicitly the dependence of  $x$ ,  $n$ , and  $p$  on  $\lambda$ , of  $q(\cdot)$  on  $x$ ,  $n$ , and  $\lambda$ , or of  $d(\cdot)$  and  $\Theta(\cdot)$  on  $p$  and  $q$ , but this dependence should be understood. As above, asterisks indicate equilibrium values. The proof of Lemma 1 and the results for discontinuities at critical values for the case of indivisible classrooms (Case 2) appear in Appendix B (online).

### Case 1: Divisible Classrooms

To solve the school's optimization problem in this case, form the Lagrangian from (9), where the constraints are (12) and (11):

$$(A1) \quad \mathcal{L}(p, x, n; \lambda) = (p - c + \tau) x - nF_c - F_s - \phi_1 (x - d) - \phi_2 \left( \frac{x}{n} - 45 \right).$$

The first-order conditions are

$$(A2a) \quad \frac{\partial \mathcal{L}}{\partial p} = x + \phi_1 \left( -\frac{d}{\mu} \right) = 0,$$

$$\frac{\partial \mathcal{L}}{\partial n} = -F_c + \phi_1 \int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial s(\lambda | \theta, q, p)}{\partial n} Mg(\theta) d\theta + \phi_2 \left( \frac{x}{n^2} \right)$$

approach to checking for discontinuities in covariates at discontinuities, see Patrick J. McEwan and Joseph S. Shapiro (2008).

<sup>45</sup> Abhijit V. Banerjee et al. (2007) present randomized evaluations of two programs to increase teacher attention per student in India.

$$(A2b) \quad = -F_c + \phi_1 \left( \frac{\lambda \Theta d}{\mu n} \right) + \phi_2 \left( \frac{x}{n^2} \right) = 0,$$

$$(A2c) \quad \frac{\partial \mathcal{L}}{\partial x} = p - c + \tau - \phi_1 \left( 1 - \int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial s(\lambda | \theta, q, p)}{\partial x} g(\theta) d\theta \right) - \phi_2 \left( \frac{1}{n} \right) \\ = p - c + \tau - \phi_1 \left( 1 + \frac{\lambda \Theta d}{\mu x} \right) - \phi_2 \left( \frac{1}{n} \right) = 0,$$

$$(A2d) \quad \frac{\partial \mathcal{L}}{\partial \phi_1} \geq 0, \quad \phi_1 \geq 0, \quad \text{and} \quad \phi_1 \frac{\partial \mathcal{L}}{\partial \phi_1} = 0,$$

$$(A2e) \quad \frac{\partial \mathcal{L}}{\partial \phi_2} \geq 0, \quad \phi_2 \geq 0, \quad \text{and} \quad \phi_2 \frac{\partial \mathcal{L}}{\partial \phi_2} = 0,$$

where  $s(\lambda | \theta, q, p)$  is given by (2). The interchanging of the partial derivatives and the integrals is justified by a standard property of integrals (see, e.g., Robert G. Bartle 1976, 245, theorem 31.7) and the continuity of  $s(\lambda | \theta, q, p)$  and its partial derivatives.

Suppose  $\phi_1 = 0$  and  $\partial \mathcal{L} / \partial \phi_1 > 0$ , i.e., the demand constraint is not binding. Then (A2a) implies  $x = 0$ , and (A2b) in turn implies  $F_c = 0$ , which is false. Hence, if there is a solution, it must be that  $\partial \mathcal{L} / \partial \phi_1 = 0$  and the demand constraint is binding:  $x = d$ . It follows from (A2a) that  $\phi_1 = \mu$ . We then have two subcases.

*Subcase 1.1: Class-Size Cap Nonbinding*

In this subcase, the class size cap is nonbinding:  $\phi_2 = 0$  and  $\partial \mathcal{L} / \partial \phi_2 \geq 0$ . By (A2b),  $x/n = F_c / \lambda \Theta$  and  $\partial \mathcal{L} / \partial \phi_2 \geq 0$  imply  $\lambda \Theta \geq F_c / 45$ . Simple algebra then yields (13a)–(13c). It is straightforward (if tedious) to show that the last two leading principal minors of the corresponding bordered Hessian alternate in sign with the last one negative (and hence that the Hessian of  $\mathcal{L}$  is negative definite on the constraint set) if and only if (14) holds.<sup>46</sup> Rewriting (14),

$$\int_{\underline{\theta}}^{\bar{\theta}} \theta \left[ \frac{s(\lambda | \theta, q^*, p^*) g(\theta)}{s(q^*, p^*)} \right] d\theta > \int_{\underline{\theta}}^{\bar{\theta}} \left( \frac{\lambda \theta}{\mu} \right) \theta \left[ \frac{s(\lambda | \theta, q^*, p^*) g(\theta)}{s(q^*, p^*)} \right] d\theta - \frac{\lambda \Theta^*}{\mu}.$$

Under assumption (7),  $\lambda \theta / \mu < 1$  for all values of  $\lambda$  and  $\theta$ . 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. 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 (13a)–(13c), together with (5), noting that  $\phi = \mu$  and  $x = d$ :

$$(A3a) \quad G_1 \equiv p^* - (\mu + c - \tau + \lambda \Theta^*) = 0,$$

<sup>46</sup> The bordered Hessians for this subcase and for subcase 1.2 are written out explicitly in the working paper (Urquiola and Verhoogen 2007).

$$(A3b) \quad G_2 \equiv x^* - \int_{\underline{\theta}}^{\bar{\theta}} s(\lambda | \theta, q^*, p^*) M g(\theta) \, d\theta = 0,$$

$$(A3c) \quad G_3 \equiv \Theta^* - \int_{\underline{\theta}}^{\bar{\theta}} \theta \left[ \frac{s(\lambda | \theta, q^*, p^*) g(\theta)}{s(q^*, p^*)} \right] \, d\theta = 0,$$

$$(A3d) \quad G_4 \equiv -F_c + \frac{\lambda \Theta^* x^*}{n^*} = 0.$$

It is convenient to define  $z^* = x^*/n^*$  (class size) and analyze (A3a)–(A3d) as a set of four equations with four endogenous variables,  $p^*$ ,  $x^*$ ,  $\Theta^*$ , and  $z^*$ , and one exogenous variable,  $\lambda$ . By the implicit function theorem:

$$(A4) \quad \begin{pmatrix} \frac{dp^*}{d\lambda} \\ \frac{dx^*}{d\lambda} \\ \frac{dz^*}{d\lambda} \\ \frac{d\Theta^*}{d\lambda} \end{pmatrix} = - \begin{pmatrix} \frac{\partial G_1}{\partial p^*} & \frac{\partial G_1}{\partial x^*} & \frac{\partial G_1}{\partial z^*} & \frac{\partial G_1}{\partial \Theta^*} \\ \frac{\partial G_2}{\partial p^*} & \frac{\partial G_2}{\partial x^*} & \frac{\partial G_2}{\partial z^*} & \frac{\partial G_2}{\partial \Theta^*} \\ \frac{\partial G_3}{\partial p^*} & \frac{\partial G_3}{\partial x^*} & \frac{\partial G_3}{\partial z^*} & \frac{\partial G_3}{\partial \Theta^*} \\ \frac{\partial G_4}{\partial p^*} & \frac{\partial G_4}{\partial x^*} & \frac{\partial G_4}{\partial z^*} & \frac{\partial G_4}{\partial \Theta^*} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial G_1}{\partial \lambda} \\ \frac{\partial G_2}{\partial \lambda} \\ \frac{\partial G_3}{\partial \lambda} \\ \frac{\partial G_4}{\partial \lambda} \end{pmatrix}$$

where the matrix in the first set of parentheses on the right-hand side is the Jacobian, call it  $\mathbf{J}$ . It is straightforward to show that:  $\det \mathbf{J} = -\Psi/\Theta^* < 0$ , since  $\Psi > 0$  (refer to (14)). Simplifying (A4), we have

$$(A5) \quad \begin{pmatrix} \frac{dp^*}{d\lambda} \\ \frac{dx^*}{d\lambda} \\ \frac{dz^*}{d\lambda} \\ \frac{d\Theta^*}{d\lambda} \end{pmatrix} = \begin{pmatrix} \frac{\Theta^*}{\Psi} \left[ \Theta^* + \frac{\lambda}{\mu} \ln \left( \frac{T}{z^*} \right) \sigma_{\theta|\lambda}^2 \right] \\ \frac{x^* \Theta^*}{\mu} \ln \left( \frac{T}{z^*} \right) \\ - \frac{z^*}{\lambda \Psi} \left[ \frac{\lambda}{\mu} \ln \left( \frac{T}{z^*} \right) \sigma_{\theta|\lambda}^2 + \Theta^* \right] \\ \frac{\Theta^* \sigma_{\theta|\lambda}^2}{\mu \Psi} \left[ \ln \left( \frac{T}{z^*} \right) + 1 \right] \end{pmatrix}$$

at the optimum.

In the definition of  $\sigma_{\theta|\lambda}^2$  in (15), the fact that the double-exponential distribution yields a non-zero probability that any given household will choose any given school implies that the term in brackets is nonzero for all  $\theta$ . Hence, as long as  $\theta \neq \Theta^*$  for some  $\theta$ , which follows from the assumption that  $\theta$  has positive support over  $(\underline{\theta}, \bar{\theta})$ , we have

$$(A6) \quad \sigma_{\theta|\lambda}^2 > 0.$$



The results (16a), (16b), (16d), and (16e) follow from (14), (A5), and (A6). Finally, we have

$$\frac{dn^*}{d\lambda} = \frac{d}{d\lambda} \left( \frac{x^*}{z^*} \right) = \frac{1}{z^*} \frac{dx^*}{d\lambda} - \frac{x^*}{z^{*2}} \frac{dz^*}{d\lambda} > 0,$$

where the inequality follows from (16b) and (16e). This gives (16c).

*Subcase 1.2: Class-Size Cap Binding*

In this subcase, the class-size cap is binding:  $\phi_2 \geq 0$  and  $\partial\mathcal{L}/\partial\phi_2 = 0$  (that is,  $x/n = 45$ ). By (A2b) and the fact that  $x = d$  and  $\phi_1 = \mu$ , we have  $\phi_2/n = F/45 - \lambda\Theta$ . The fact that  $\phi_2 \geq 0$  implies  $\lambda\Theta \leq F/45$ . Algebra yields (17a)–(17c). It is straightforward to verify that the determinant of the corresponding bordered Hessian is negative at the optimum, and hence that the second-order conditions for a maximum are satisfied. To analyze the slopes with respect to  $\lambda$  in this subcase, rewrite (17a)–(17c) with (5) as

$$(A7a) \quad G_1 \equiv p^* - \left( c - \tau + \mu + \frac{F_c}{45} \right) = 0,$$

$$(A7b) \quad G_2 \equiv x^* - \int_{\underline{\theta}}^{\bar{\theta}} s(\lambda | \theta, q^*, p^*) M g(\theta) \, d\theta = 0,$$

$$(A7c) \quad G_3 \equiv \Theta^* - \int_{\underline{\theta}}^{\bar{\theta}} \theta \left[ \frac{s(\lambda | \theta, q^*, p^*) g(\theta)}{s(q^*, p^*)} \right] \, d\theta = 0,$$

$$(A7d) \quad G_4 \equiv n^* - \frac{x^*}{45} = 0.$$

Applying the implicit function theorem (as in (A4)) to (A7a)–(A7d), we have

$$(A8) \quad \begin{pmatrix} \frac{dp^*}{d\lambda} \\ \frac{dx^*}{d\lambda} \\ \frac{dn^*}{d\lambda} \\ \frac{d\Theta^*}{d\lambda} \end{pmatrix} = \frac{1}{\mu} \ln \left( \frac{T}{45} \right) \begin{pmatrix} 0 \\ x^* \Theta^* \\ \frac{x^* \Theta^*}{45} \\ \sigma_{\theta|\lambda}^2 \end{pmatrix}$$

which in turn implies (18a)–(18e).

The fact that  $d\Theta^*/d\lambda > 0$  in both subcases implies that  $\lambda\Theta^*$  is monotonically increasing in  $\lambda$  and guarantees that there is at most one critical value of  $\lambda$ , call it  $\alpha$ , at which (19) holds. The results for the first subcase apply for  $\lambda \geq \alpha$  and the results for the second subcase apply for  $\lambda < \alpha$ .

*Case 2: Indivisible Classrooms*

The Lagrangian in this case is the same as (A1), but  $n$  is now interpreted as a parameter

$$(A9) \quad \mathcal{L}(p, x; n, \lambda) = (p - c + \tau) x - nF_c - F_s - \phi_1 (x - d) - \phi_2 \left( \frac{x}{n} - 45 \right).$$

The first-order conditions are

$$(A10a) \quad \frac{\partial \mathcal{L}}{\partial p} = x - \phi_1 \left( \frac{d}{\mu} \right) = 0,$$

$$(A10b) \quad \frac{\partial \mathcal{L}}{\partial x} = p - c + \tau - \phi_1 \left( 1 + \frac{\lambda \Theta d}{\mu x} \right) - \phi_2 \left( \frac{1}{n} \right) = 0,$$

$$(A10c) \quad \frac{\partial \mathcal{L}}{\partial \phi_1} \geq 0, \quad \phi_1 \geq 0, \quad \text{and} \quad \phi_1 \frac{\partial \mathcal{L}}{\partial \phi_1} = 0,$$

$$(A10d) \quad \frac{\partial \mathcal{L}}{\partial \phi_2} \geq 0, \quad \phi_2 \geq 0, \quad \text{and} \quad \phi_2 \frac{\partial \mathcal{L}}{\partial \phi_2} = 0.$$

We can quickly rule out the possibility that the demand constraint is not binding:

1. If neither the demand constraint nor the class-size cap is binding ( $\phi_1 = 0$ ,  $\partial \mathcal{L} / \partial \phi_1 \geq 0$ ,  $\phi_2 = 0$ , and  $\partial \mathcal{L} / \partial \phi_2 \geq 0$ ), then by (A10a),  $x = 0$  and by (A10b),  $p = c - \tau$ . It is straightforward to verify that the second-order conditions are not satisfied in this subcase.
2. If the demand constraint is not binding but the class-size cap is binding ( $\phi_1 = 0$ ,  $\partial \mathcal{L} / \partial \phi_1 \geq 0$ ,  $\phi_2 \geq 0$ , and  $\partial \mathcal{L} / \partial \phi_2 = 0$ ), then by (A10a), we have  $x = 0$ . This violates the class-size constraint ( $x/n = 45$ ); hence there is no solution in this case.

There are then two subcases to be considered.

#### *Subcase 2.1: Class-Size Cap Nonbinding*

In this subcase, the demand constraint is binding and the class-size cap is nonbinding:  $\phi_1 \geq 0$ ,  $\partial \mathcal{L} / \partial \phi_1 = 0$ ,  $\phi_2 = 0$ ,  $\partial \mathcal{L} / \partial \phi_2 \geq 0$ . Equations (20a)–(20b) follow immediately from (A10a)–(A10b).

The second-order conditions can be verified by evaluating the determinant of the bordered Hessian ( $3 \times 3$  in this case). The determinant is positive, and the second-order conditions are satisfied, if condition (14) 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 (20a)–(20b) is a global maximum.

Rewrite (20a)–(20b), together with (5)

$$(A11a) \quad G_1 \equiv p^* - [\mu + c - \tau + \lambda \Theta^*] = 0,$$

$$(A11b) \quad G_2 \equiv x^* - \int_{\underline{\theta}}^{\bar{\theta}} s(\lambda | \theta, q^*, p^*) M g(\theta) d\theta = 0,$$

$$(A11c) \quad G_3 \equiv \Theta^* - \int_{\underline{\theta}}^{\bar{\theta}} \theta \left[ \frac{s(\lambda | \theta, q^*, p^*) g(\theta)}{s(q^*, p^*)} \right] d\theta = 0.$$

Note that these are the same as (A3a)–(A3c). (The number of classrooms is now treated as a parameter, and we no longer have (A3d).) Inverting the Jacobian ( $3 \times 3$  in this case) and using the implicit function theorem (as in (A4)) applied to (A11a)–(A11c), we have

$$(A12) \quad \begin{pmatrix} \frac{\partial p^*}{\partial \lambda} \\ \frac{\partial x^*}{\partial \lambda} \\ \frac{\partial \Theta^*}{\partial \lambda} \end{pmatrix} = \frac{1}{1 + \frac{\lambda \Psi}{\mu}} \begin{pmatrix} \Theta^* + \frac{\lambda \Theta^{*2}}{\mu} + \frac{\lambda}{\mu} \ln\left(\frac{nT}{x^*}\right) \sigma_{\theta|\lambda}^2 \\ \frac{x^*}{\mu} \left[ \Psi \ln\left(\frac{nT}{x^*}\right) - \Theta^* \right] \\ \frac{\sigma_{\theta|\lambda}^2}{\mu} \left[ \frac{\lambda \Theta^*}{\mu} + \ln\left(\frac{nT}{x^*}\right) \right] \end{pmatrix}$$

The results (21a), (21b), and (23) follow from (14), (22), (A6), and (A12). The condition  $\partial \mathcal{L} / \partial \phi_2 \geq 0$  requires

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

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

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

where  $\Sigma$  is defined as in (25). Setting this inequality to an equality yields (27), which implicitly defines  $\beta(n)$ , the value of  $\lambda$  at which the class-size cap begins to bind.

*Subcase 2.2: Class-Size Cap Binding*

In this subcase, both the demand constraint and class-size cap are binding:  $\phi_1 \geq 0$ ,  $\partial \mathcal{L} / \partial \phi_1 = 0$ ,  $\phi_2 \geq 0$ , and  $\partial \mathcal{L} / \partial \phi_2 = 0$ . The two constraints,  $x = d$  and  $x = 45n$ , pin down the values of  $p$  and  $x$ , and the results (24a)–(24b) follow immediately.

The fact that  $\partial x^* / \partial \lambda = (\partial / \partial \lambda) (x/n) = 0$  follows immediately. It is straightforward to show that

$$(A13a) \quad \frac{\partial p^*}{\partial \lambda} = \Theta^* \ln\left(\frac{T}{45}\right) > 0,$$

$$(A13b) \quad \frac{\partial \Theta^*}{\partial \lambda} = \frac{1}{\mu} \sigma_{\theta|\lambda}^2 \ln\left(\frac{T}{45}\right) > 0.$$

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

$$(A14) \quad \phi_2 = n \{ \mu \ln \Sigma - \mu \ln(45n) - c + \tau - \mu - \lambda \Theta^* \}.$$

By the definition of  $\beta(n)$  above, if  $\lambda = \beta(n)$  then  $\phi_2 = 0$ . Partially differentiating (A14):

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

by assumption (22). Thus,  $\phi_2 \geq 0$  for  $\lambda \geq \beta(n)$ . The first subcase applies for  $\lambda \leq \beta(n)$  and the second applies for  $\lambda > \beta(n)$ .

Finally, it follows from (A12), (A13a), and (22) that:  $\lim_{\lambda \rightarrow \beta(n)^-} \partial p^* / \partial \lambda < \lim_{\lambda \rightarrow \beta(n)^+} \partial p^* / \partial \lambda$ . That is, the slope of  $p^*$  with respect to  $\lambda$  is steeper to the right of the critical value, in the region where the class-size cap binds.

## REFERENCES

- Altonji, Joseph G., Ching-I Huang, and Christopher R. Taber. 2005. "Estimating the Cream Skimming Effect of Private School Vouchers on Public School Students." Unpublished.
- Anderson, Simon P., and André de Palma. 2001. "Product Diversity in Asymmetric Oligopoly: Is the Quality of Consumer Goods Too Low?" *Journal of Industrial Economics*, 49(2): 113–35.
- Anderson, Simon P., André de Palma, and Jacques-François Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.
- Angrist, Joshua D., and Alan B. Krueger. 1999. "Empirical Strategies in Labor Economics." In *Handbook of Labor Economics*. Vol. 3, ed. Orley C. Ashenfelter and David Card, 1277–1366. Amsterdam: Elsevier.
- ▶ Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, 114(2): 533–75.
- ▶ Asadullah, M. Niaz. 2005. "The Effect of Class Size on Student Achievement: Evidence from Bangladesh." *Applied Economics Letters*, 12(4): 217–21.
- ▶ Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics*, 122(3): 1235–64.
- Bartle, Robert G. 1976. *The Elements of Real Analysis*. 2nd ed. New York: John Wiley & Sons.
- Bayer, Patrick J., Robert McMillan, and Kim Reuben. 2004. "An Equilibrium Model of Sorting in an Urban Housing Market." National Bureau of Economic Research Working Paper 10865.
- Bressoux, Pascal, Francis Kramarz, and Corinne Prost. 2005. "Teachers' Training, Class Size and Students' Outcomes: Evidence from Third Grade Classes in France." Unpublished.
- Browning, Martin, and Eskil Heinesen. 2003. "Class Size, Teacher Hours and Educational Attainment." Centre for Applied Microeconometrics Working Paper 2003–15.
- ▶ Caucutt, Elizabeth M. 2002. "Educational Vouchers When There Are Peer Group Effects: Size Matters." *International Economic Review*, 43(1): 195–222.
- Clark, Damon. 2005. "Politics, Markets and Schools: Quasi-Experimental Evidence on the Impact of Autonomy and Competition from a Truly Revolutionary UK Reform." Unpublished.
- ▶ Clotfelter, Charles T. 1999. "Public School Segregation in Metropolitan Areas." *Land Economics*, 75(4): 487–504.
- Dixit, Avinash K. 1976. *Optimization in Economic Theory*. Oxford, UK: Oxford University Press.
- ▶ Dobbela, Simone, Jesse D. Levin, and Hessel 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.
- Elacqua, Gregory. 2006. "Enrollment Practices in Response to Vouchers: Evidence from Chile." National Center for Study of Privatization in Education, Teachers College, Columbia University, Occasional Paper 125.
- ▶ Epple, Dennis N., David N. Figlio, and Richard E. Romano. 2004. "Competition Between Private and Public Schools: Testing Stratification and Pricing Predictions." *Journal of Public Economics*, 88(7–8): 1215–45.
- Epple, Dennis N., and Richard E. Romano. 1998. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review*, 88(1): 33–62.
- Epple, Dennis N., and Richard E. Romano. 2008. "Educational Vouchers and Cream Skimming." *International Economic Review*, 49(4): 1395–1435.
- ▶ Epple, Dennis N., Richard E. Romano, and Holger Sieg. 2006. "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education." *Econometrica*, 74(4): 885–928.
- ▶ Ferreyra, Maria Marta. 2007. "Estimating the Effects of Private School Vouchers in Multidistrict Economies." *American Economic Review*, 97(3): 789–817.
- Gabszewicz, J. Jaskold, and Jacques-François Thisse. 1979. "Price Competition, Quality and Income Disparities." *Journal of Economic Theory*, 2(3): 340–59.
- ▶ Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69(1): 201–09.
- ▶ Hanushek, Eric A. 1995. "Interpreting Recent Research on Schooling in Developing Countries." *World Bank Research Observer*, 10(2): 227–46.

- ▶ **Hanushek, Eric A.** 2003. "The Failure of Input-Based Schooling Policies." *Economic Journal*, 113(485): F64–98.
- ▶ **Hoxby, Caroline M.** 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics*, 115(4): 1239–85.
- ▶ **Hsieh, Chang-Tai, and Miguel Urquiola.** 2006. "The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile's Voucher Program." *Journal of Public Economics*, 90(8–9): 1477–1503.
- ▶ **Imbens, Guido, and Thomas Lemieux.** 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics*, 142(2): 615–35.
- ▶ **Jakubowski, Maciej, and Pawel Sakowski.** 2006. "Quasi-Experimental Estimates of Class Size Effect in Primary Schools in Poland." *International Journal of Educational Research*, 45(3): 202–15.
- ▶ **Kremer, Michael R.** 1995. "Research on Schooling: What We Know and What We Don't: A Comment." *World Bank Research Observer*, 10(2): 247–54.
- ▶ **Krueger, Alan B.** 2003. "Economic Considerations and Class Size." *Economic Journal*, 113(485): F34–63.
- ▶ **Lazear, Edward P.** 2001. "Educational Production." *Quarterly Journal of Economics*, 116(3): 777–803.
- ▶ **Lee, David S.** 2008. "Randomized Experiments from Non-random Selection in US House Elections." *Journal of Econometrics*, 142(2): 675–97.
- ▶ **Lee, David S., and David Card.** 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics*, 142(2): 655–74.
- ▶ **Manski, Charles.** 1992. "Educational Choice (Vouchers) and Social Mobility." *Economics of Education Review*, 11(4): 351–69.
- ▶ **McCrary, Justin.** 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142(2): 698–714.
- ▶ **McEwan, Patrick J., and Joseph S. Shapiro.** 2008. "The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates." *Journal of Human Resources*, 43(1): 1–29.
- ▶ **McFadden, Daniel.** 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, ed. Paul Zarembka, 105–142. New York: Academic Press.
- ▶ **Mizala, Alexandra, and Pilar Romaguera.** 2002. "Equity and Educational Performance." *Economía: Journal of the Latin American and Caribbean Economic Association*, 2(2): 219–62.
- ▶ **Mussa, Michael, and Sherwin Rosen.** 1978. "Monopoly and Product Quality." *Journal of Economic Theory*, 18(2): 301–17.
- ▶ **Nechyba, Thomas J.** 2003. "Centralization, Fiscal Federalism, and Private School Attendance." *International Economic Review*, 44(1): 179–204.
- ▶ **Nesheim, Lars.** 2002. "Equilibrium Sorting of Heterogeneous Consumers Across Locations: Theory and Empirical Implications." Centre for Microdata Methods and Practice Working Paper CWP08/02.
- ▶ **Piketty, Thomas, and Mathieu Valdenaire.** 2006. "L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français" [The Impact of Class Size on Student Success in French Primary and Secondary Schools]. Les dossiers évaluations et statistiques, Ministère de l'éducation nationale, No. 173.
- ▶ **Rothstein, Jesse M.** 2006. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *American Economic Review*, 96(4): 1333–50.
- ▶ **Shadish, W. R., T. D. Cook, and D. T. Campbell.** 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- ▶ **Shaked, Avner, and John Sutton.** 1982. "Relaxing Price Competition through Product Differentiation." *Review of Economic Studies*, 49(1): 3–13.
- ▶ **Thistlewaite, D., and D. T. Campbell.** 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology*, 51(6): 309–317.
- ▶ **Urquiola, Miguel.** 2005. "Does School Choice Lead to Sorting? Evidence from Tiebout Variation." *American Economic Review*, 95(4): 1310–26.
- ▶ **Urquiola, Miguel.** 2006. "Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia." *Review of Economics and Statistics*, 88(1): 171–77.
- ▶ **Urquiola, Miguel, and Eric Verhoogen.** 2007. "Class Size and Sorting in Market Equilibrium: Theory and Evidence." National Bureau of Economic Research Working Paper 13303.
- ▶ **van der Klaauw, Wilbert.** 2002. "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach." *International Economic Review*, 43(4): 1249–87.
- ▶ **Verhoogen, Eric A.** 2008. "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector." *Quarterly Journal of Economics*, 123(2): 489–530.
- ▶ **Wössmann, Ludger.** 2005. "Educational Production in Europe." *Economic Policy*, 43: 445–93.