# Teacher Labor Markets, School Vouchers and Student Cognitive Achievement: Evidence from Chile

Michela Tincani[*]

University College London

October 19, 2012

## Abstract

This paper develops and structurally estimates an equilibrium model of the Chilean school system and uses the model to assess the effect of teacher wage and accreditation policies on student achievement. In the model, potential teachers choose between teaching in a public school, teaching in a private school, working in the non-teaching sector and not working; parents choose whether to enroll their children in public or private schools; and private schools set tuition and teacher wages. The equilibrium of the model determines the distribution of student cognitive achievement. The estimation is based on a unique dataset that combines three rich data sources from Chile: student test score data, household survey data and teacher survey data. Chile is an ideal environment for studying school choice, because it has thirty years of experience with a school voucher system and a large private school sector. I use a two-step estimation approach that addresses the issue of potentially multiple equilibria. The estimated model fits the data well. Policy experiments show that an increase in the wages of public school teachers accompanied by minimum teacher competency requirements would increase average student test scores without a large increase in government costs. A larger increase in test scores could be achieved by making the wages of public school teachers more directly tied to their skills than in the current system. Under the existing voucher plan, both policies could be partially financed by the increased voucher revenues that result from an increased demand for public education.

# 1 Introduction

Improving student achievement is a central goal of education policies in many countries. One promising way to increase achievement is to improve teacher quality, as empirical evidence suggests that teacher quality is an important determinant of student achievement.[1] Teacher quality can be improved by increasing the effectiveness of current teachers or by attracting more highly skilled individuals into the teaching profession. This paper focuses on the second margin. Its goal is to evaluate the effect of alternative teacher wage and accreditation policies on student achievement. Moreover, the paper explores the implications of school choice programs for the evaluation and design of teacher policies. As school choice in many countries expands (e.g., charter schools in the U.S.), understanding its implications for other education policy instruments becomes increasingly important. To this end, I develop and structurally estimate a new model that incorporates teacher labor markets into a school choice framework. The model is estimated using data from Chile, which provides an ideal environment to study the role of school choice because it has thirty years of experience with a universal school voucher system.

Up until recently, studies of teacher quality have pursued two mostly separate lines of research. The first investigates how the supply of teacher characteristics, such as teacher certifications and teaching experience, reacts to changes in wages and non-pecuniary job characteristics.[2] A second line of research is concerned with quantifying teacher quality by evaluating the effect of teachers on student achievement.[3] Teachers are found to be important for student achievement, however, it has been noted that unobserved characteristics often have a larger impact on achievement than most observed ones.[4]

As Hanushek and Rivkin (2004) noted, the separate treatment of teacher labor supply from the evaluation of teacher quality poses a challenge to evaluating the impact on achievement of policies targeted at attracting skilled teachers. Because teacher quality does not correlate well with most observed characteristics, studies of the supply of observed teacher characteristics are not very informative on the supply of teacher quality.[5] Moreover, in a context with a large-scale school choice program, like the one considered in this paper, correctly estimating the effect of a policy on student achievement requires also taking into account the reactions of parents and non-public schools.

When parents can choose which school their child attends, they may change their choice of school as a reaction to policies that affect the composition of teachers across schools. This reaction must be taken into account because student characteristics, such as parental investments and innate ability to learn, are important determinants of achievement. Moreover, if some schools are not covered by centralized teacher contract agreements, such as charter or voucher schools, then those schools might also react to policies

---

[1] See, for example, Rivkin, Hanushek and Kain (2005).

[2] See, for example, Murnane (1981), Dolton and van der Klaauw (1999), Stinebrickner (2001), Hanushek, Kain and Rivkin (2004), Corcoran, Evans and Schwab (2004), Boyd, Lankford, Loeb and Wyckoff (2005), and Buckley, Schneider and Shang (2005).

[3] See, for example, Rivkin, Hanushek and Kain (2005). A third line of research is concerned with ways to improve teacher quality by providing incentive pay schemes that encourage teacher effort. See, for example, Barlevy and Neal (2011). These studies are especially related to the literature on policies to improve the quality of current teachers.

[4] Teaching experience is the only observed characteristic that is found to have an impact on achievement across multiple studies. See Hanushek and Rivkin (2004) for a literature review.

[5] This observation, and the availability of new data, has spurred a new and active area of research that links teacher labor supply to student achievement. An early example in this research area is Hanushek, Kain, O'Brien and Rivkin (2005). The authors evaluate individual teachers' quality using a teacher value added approach and then analyze whether higher paying districts attract the teachers who are estimated to be most effective. They conclude that higher paying districts do not attract the best teachers.

targeted at public sector teachers by changing their teacher contracts. Ignoring these equilibrium responses is likely to give misleading predictions for policy evaluation.

The objective of this paper is to develop and estimate an equilibrium model that can inform policy makers about the likely impact and cost of various policies that affect teacher quality. The model predicts the policy reaction of the supply of teacher quality, which is allowed to depend on both observable and unobservable characteristics. It also encompasses the equilibrium responses of parents and of non-public schools.

Chile is an ideal setting for studying teacher labor market policies in the presence of school choice. A private subsidized sector provides schooling to 47.5% of students and employs 39.1% of teachers.[6] Public sector teacher contracts are centrally negotiated between the government and the national teacher association, the *Colegio de Profesores*. The contracts of teachers in the private sector are not determined by centralized bargaining agreements: each school is allowed to independently determine them. Every child in the country receives a voucher, which can be used toward entire coverage of tuition fees in public (or municipal) schools, or toward partial or entire coverage of tuition fees in private subsidized schools.[7] The law mandates private subsidized schools to grant fellowships mostly based on socio-economic status whenever their tuition, which is subject to a legal cap, exceeds the voucher amount.[8]

The design of the Chilean system is similar to that of some charter school systems in the U.S., which are present in 41 states. They are often financed through capitation grants that follow the students in their school of choice and therefore provide enrollment-based subsidies similar to those of a voucher system. Moreover, 32 of the states with charter schools do not require charter school teachers to be covered by district bargaining agreements.[9] We can, therefore, gain valuable lessons for U.S. education policy by studying Chile.

The model developed in this paper incorporates key features of the Chilean education system. It assumes that each market includes both a private voucher school and a municipal school. Potential teachers, parents and the voucher school are the model's decision makers. Potential teachers care about wages and non-pecuniary job characteristics. They receive a wage offer from public and private schools and from the non-teaching sector. Potential teachers with differing teaching skills make labor supply decisions by either accepting one of the wage offers or not working.[10] All parents receive a voucher. If the voucher is not enough to cover tuition at the private school, eligible students receive a fellowship for private education. Parents care about consumption and the cognitive achievement of their child, which is produced as a function of home inputs, the student's innate ability and teacher skills. They choose a school for their child. Teacher skills supplied to the two schools are determined endogenously by the occupational choices of potential teachers. Both students and potential teachers differ in terms of characteristics that are not observed by the researcher and that affect their abilities and preferences. The model allows for unobserved abilities because the literature identifies them as important determinants of achievement. Finally, the private voucher school chooses teacher wages and tuition so as to maximize its profits subject to a legal cap on tuition.[11] The municipal school is

---

[6]Sources: 2006 Teacher Census and 2006 SIMCE dataset.

[7]In Chile there is a third type of school, private unsubsidized schools, where the voucher cannot be used and from which my work abstracts. These schools account for about 6% of enrollment. The parents of students in these schools earn about six times as much as parents of students in municipal and voucher schools, and they pay tuition fees that are about six times as high as tuition fees in private subsidized schools.

[8]Fellowships are financed by the schools and by the government.

[9]See the 2011 Report on Charter School Laws prepared by the Center for Education Reform.

[10]The modeling of labor supply decisions builds on a rich literature in labor economics that analyzes self-selection into occupations. Roy (1951) is the seminal paper.

[11]Using data from the Ministry of Education, Elacqua (2006) calculates that in 2003 73.60% of all Chilean voucher schools reported being for-profit. In Urquiola and Verhoogen (2009), which presents a model of the Chilean marker for education,

the residual claimant of the teachers' supply and the demand for enrollment not absorbed by private schools.

The model is sequential, with two stages. In the first stage, the private school sets tuition and the price of one unit of teaching skills. In the second stage, parents and potential teachers sort across schools and occupations. I show that an equilibrium exists, that the equilibrium of the second stage is unique, and that the only source of potential multiplicity comes from the first stage. The model determines as joint equilibrium outcomes the school choices of parents, the labor supply decisions of potential teachers and the tuition level and price of teaching skills in the private school. The distribution of student achievement is implied by the equilibrium of the model. I use the estimated model to simulate this distribution under alternative teacher wage and accreditation policies. For cost-benefit analysis purposes, I also simulate government spending under the alternative policies.

The empirical implementation of the model is demanding of the data in that it requires observations on all three decision makers: schools, parents and potential teachers. I constructed a unique dataset that can serve as a basis for estimation by combining three rich Chilean data sources. I used the 2006 *Encuesta de Caracterización Socioeconómica Nacional* (CASEN), a nationally representative dataset, to identify a sample of $3,520$ individuals holding a college degree, who serve as the pool of potential entrants into the teaching profession. To augment the sample of teachers, I obtained a sample of $3,195$ teachers from the 2009 *Encuesta Longitudinal Docente* (ELD), a teacher survey with information on both private and public sector teachers.[12] For all individuals in the two samples I obtained the same set of characteristics, along with occupational choices and teaching and non-teaching wages. Finally, I used a sample of $100,000$ students from the restricted version of the 2006 *Sistema de Medición de Calidad de la Educación* (SIMCE) dataset. This dataset contains rich information on all $4^{th}$ and $10^{th}$ graders in the country, including administrative test scores in math and Spanish, which I used to build a measure of cognitive achievement.

The dataset is unique because it has observations on multiple markets, making it representative of the entire population of students and potential teachers in the country;[13] it contains individual-level information on both public and private school teachers and students (private school information is typically unavailable); and it contains a unique market identifier that is homogeneous across data sources, which allowed me to match teachers, students and schools at the market level.

I use a two-step estimation approach that addresses the issue of potentially multiple solutions to the voucher school's profit maximization problem in the first stage of the model.[14] I show that by exploiting some key model features, in particular the fact that for every solution to the profit maximization there is only one equilibrium sorting of parents and potential teachers, I can separately estimate the parameters related to potential teachers and parents, which represent almost all parameters in the model, without assuming an equilibrium selection rule. These parameters are estimated in the first step by the Method of Simulated Moments (McFadden (1989)). The second step uses simulation and kernel methods to estimate the private school's parameters by Nonparametric Simulated Maximum Likelihood (Laroque and Salanié (1989), Fermanian and Salanié (2004)). Additional assumptions are needed in this step of the estimation to

---

schools are also assumed to be profit maximizers.

[12]In estimation I account for the fact that this is a choice-based sample. The ELD survey administered in 2009 contains information on 2006, which is the year used in the analysis. This year was chosen because it was the only year for which a rich dataset with all the necessary information on both potential teachers and students could be constructed.

[13]Only two rural regions were excluded for sample size reasons: Aisén in the South and Tarapacá in the North.

[14]This method draws on an original idea by Moro (2003). Similar methods have been used by, for example, Fang (2006) and Fu (2011).

account for the possibility of multiple solutions to the school's maximization problem.

Parameter estimates reveal that private schools are better able to attract skilled teachers than public schools. Results of counterfactual policy experiments show that increasing the wages of public sector teachers by an identical amount for individuals of all skills would improve student cognitive achievement; however, such a policy would be less effective at raising achievement than a more radical policy that substitutes the current municipal sector wage formulae with wage schedules that are more reflective of a teacher's skills. Moreover, I find that, under both policies, the cost of increasing the public sector teacher wage bill would be partly offset by the increased voucher revenues that result from an increased demand for public education. The voucher plan, therefore, plays an important role in the determination of the policy costs. The next two paragraphs describe examples of these two alternative policies.

In the first policy experiment, a flat increase in municipal teacher wages of CLP 90,000 (approx. $180) a month, coupled with a minimum competency cutoff, results, in equilibrium, in a 9.86% increase in average municipal teacher wages.[15] The policy attracts both non-teachers and voucher school teachers into the municipal school sector, where mean teaching skills improve. Mean teaching skills improve in the voucher school sector too, because the individuals who leave the voucher school for the municipal school and who are, on average, better teachers than the incumbent municipal sector teachers are also the worst voucher school teachers. Among the non-teachers, more skilled individuals are less likely to move to the municipal school sector than less skilled individuals, indicating that the policy is not able to attract the most highly skilled individuals into the teaching profession. As a result of the change in the composition of teachers across schools, the demand for public education increases (the public education enrollment share increases by three percentage points). This determines an inflow of voucher revenues to the municipal sector that partially finances the wage bill increase. Overall, government spending must increase by only 3%. Mean test scores increase by 8% of a standard deviation.

Results from the second policy experiment show that a pay-per-skill municipal wage schedule, i.e., a (linear) function of skills, coupled with minimum competency requirements, is able to attract more skilled individuals into the municipal schools than a flat increase in municipal wages.[16] The policy attracts both non-teachers and voucher school teachers into the municipal sector. However, twice as many non-teachers are attracted than under a flat wage increase, and among non-teachers, more skilled individuals are more likely to move to the municipal school sector than less skilled individuals. The mean skills of teachers in the municipal school sector increase considerably both because of the inflow of highly skilled non-teachers, and because of the inflow of highly skilled voucher school teachers. In spite of the fact that voucher schools increase their wage offers to retain teachers, they lose their best teachers to the municipal school. As a result of the new teacher composition across schools, demand for public education increases considerably, resulting in a twelve-percentage-point increase in enrollment share. This determines a flow of voucher revenues to the municipal sector that partially finances the required increase in the municipal wage bill. On average, municipal school wages increase by 67%, and they are financed through a 28% increase in government spending. Because this

---

[15]The cutoff is set equal to the pre-policy bottom quartile in the distribution of municipal teachers' skills. Minimum competency requirements are introduced as a way for the municipal sector to select only the best teachers among those willing to accept their wage offers. The cutoffs are set in a way that keeps class size in the municipal sector equal to the pre-policy class size.

[16]The minimum competency cutoff can be set at a higher level than in the previous policy experiment, in this example it is equal to the median in the distribution of skills, because pay-per-skill schemes are better able to attract teachers. With governmental wage schedules, high minimum competency requirements would induce an insufficient supply of teachers.

policy attracts better teachers into the teaching profession than a flat wage increase, test scores increase by more: mean test scores increase by 18% of a standard deviation, a substantial increase.

This paper is related to three strands in the literature: the already mentioned literature on the supply of teacher quality, the literature on school choice, and the literature on education policy in developing countries. The school choice literature is extensive. A large body of work examines whether provision of choice improves student achievement. School choice proponents argue that choice encourages the entry into the market for education of schools that are effective at raising student achievement. Competition forces would then encourage public schools to improve quality, ultimately benefiting all students (Friedman (1955)). Studies include comparisons of the effectiveness of public and voucher schools (e.g., Hoxby(2003), Greene (1996), Rouse (2998)) and of public and charter schools (e.g., Hoxby and Rockoff (2004), Hanushek et al (2007), Angrist et al. (2010)), examinations of whether school competition increases school quality (e.g., Hoxby (2003), Sass (2006), Chakrabarti (2008), Mehta (2011)), and studies of patterns of sorting of students across schools (e.g., Schneider and Buckley (2002), Hoxby (2003), Urquiola (2005)). Within the school choice literature there has been a growing interest in equilibrium models of school systems, examples include models of parental sorting across schools and locations (e.g., Nachyba (2000), Ferreyra (2007)) and of parental sorting and optimal school tuition policies in the presence of peer effects (e.g., Epple and Romano (1998, 2008), Epple, Romano and Sieg (2006)). The aim of this paper is not to evaluate school choice programs, rather its goal is to understand the implications that school choice programs have for the evaluation and the design of teacher policies. The literature so far has documented the existence of a connection between teacher labor markets and parental school choice: Hoxby (2002) finds that when schools face competition, teacher characteristics that are valued more by parents are also rewarded more on the labor market, and Hanushek and Rivkin (2003) find that better teachers are hired in districts with more competition. No previous study, however, has explored how the intuitive, and documented, connection between teacher labor markets and parental school choice works, and what its policy implications are. This is the first paper to explore this connection, and it is one of the first papers to provide a link between the literature on school choice and the literature on teacher labor markets.

Finally, this work is related to studies of education policy in developing countries. Behrman, Parker and Todd (2011) provide an overview of a large body of work that estimates the impact of education incentive programs in developing countries, such as Mexico's cash transfer program conditional on school attendance, performance incentive programs (e.g., in Kenya and Mexico), school food programs (e.g., in the Philippines and Bangladesh) and school voucher programs (e.g., in Chile and Colombia).[17] A number of studies have focused on Chile because of the availability of rich data and because Chile is unusual in having thirty years of experience with a universal voucher program. A body of work analyzes how the Chilean voucher program has affected student test scores and stratification (e.g., Mizala and Romaguera (2000), Contreras (2001), McEwan (2001), Sapelli and Vial (2002, 2003), Hsieh and Urquiola (2003, 2006), McEwan, Urquiola and Vegas (2008)) and labor market outcomes (e.g., Bravo, Mukhopadhyay, and Todd (2010)) and the evidence is mixed. For example, Bravo, Mukhopadhyay, and Todd (2010) find positive effects of the voucher reform on high-school graduation, whereas Hsieh and Urquiola (2006) find no positive effect on student test scores. Sapelli and Vial (2002) estimate the average treatment effect of attending the voucher school on achievement

---

[17]See also the Handbook of Development Economics chapter by Behrman (2010) for a comprehensive treatment of educational investment in developing countries.

(ATE), as well as the treatment effect on the treated (TT). Using the estimated model to simulate these treatment effects, I find similar results: a small and negative ATE, but a large and positive TT. Finally, Urquiola and Verhoogen (2009) is the only other paper that develops a model of parents' and schools' (but not teachers') optimal behaviors in Chile. Their model however differs substantially from mine, as described in the model section.

To summarize, this paper finds that wage increases in the public school sector, accompanied by minimum competency requirements, would increase mean test scores at a modest cost in terms of government spending on education. However, a more effective way to increase student achievement would be the introduction of new wages for public school teachers that are more reflective of a teacher's talent than the current government formulae. Offering higher wages to more talented individuals attracts talented teachers into the public school sector more effectively than offering higher wages to all. Moreover, under both policies the improvement in the quality of public sector teachers determines an increase in the demand for public education. This demand increase results in higher voucher revenues in the municipal school sector, which can be used to partially cover the policy costs. The universal voucher plan, therefore, plays an important role in the design of policies targeted at teachers.

The rest of the paper is organized as follows. Section 2 describes the institutional features included in the model, which is presented in section 3. Section 4 describes the data, section 5 illustrates the estimation approach, and section 6 shows how the estimated model fits the data. Finally, section 7 presents the results and section 8 concludes. The technical appendices follow.

## 2   Institutional Features Incorporated into the Model

Each Chilean family receives a voucher that can be spent toward coverage of tuition fees in a municipal school or in a private subsidized school. The voucher does not vary by family characteristics and it can be used in both primary (grades one to eight) and secondary (grades nine to twelve) schools. In December 2005 its value was CLP $27,391.903$, approximately \$50.[18] It cannot be used in private unsubsidized schools, which account for around 6% of enrollment and which I do not include in the model.[19] Tuition fees in municipal schools can never exceed the value of the voucher. Private voucher schools are allowed to charge a fee that exceeds the value of the voucher, up to a legal cap. The cap is equal to four *USE*, *Unidad de Subvención Educacional*. The value of one *USE* is annually adjusted for inflation and in the year covered by my sample (2006) its value was CLP $13,504.692$ per month, approximately \$25. Depending primarily on their socio-economic status, some children are eligible for a fellowship for private education, *beca*, that covers partially or totally the tuition fees in excess of the value of the voucher. According to my calculations using the SIMCE dataset, in 2006 around 60% of all Chilean children enrolled in private subsidized schools received a fellowship. Prior to the start of the academic year the school owner must announce the criteria according to which fellowships are awarded. The government sets guidelines for these criteria. In the model

---

[18]This is an average across education levels (primary vs. secondary) and types of education (e.g., scientific-humanistic vs. vocational), among which the model does not distinguish.

[19]Parents in these school are substantially different from those in the public and private subsidized sector. They pay tuition fees that are, on average, 6.31 times those at private subsidized schools and their income is 4.3 times that of families sending children to subsidized schools and 6.7 times that of parents who send their children to municipal schools. Teachers earn, on average 17%, more than teachers in the public and subsidized sectors. This suggests that parents and teachers in unsubsidized schools are unlikely to change their behavior as a reaction to the policies that I analyze.

the fellowship assignment criteria are treated as an exogenous formula of student characteristics that parents know when making their school choice. The value of the voucher and the cap on private school tuition are contained in the the law *Decreto con Fuerza de Ley Nº 2, De Education, De 20.08.98* and in the law on shared financing, *Financiamiento Compartido, Ley Nº 19.532*. The guidelines for fellowship assignment can be found in articles 24 and 27 of the *Ley de Subvenciones, Decreto con Fuerza de Ley Nº 2, 20.08.98*.

Individuals who intend to become teachers must obtain a teaching certification. The teachers' statute, *Estatuto Docente, Ley Nº 10.070*, stipulates four ways to become certified to teach: i) getting an undergraduate degree in education (6-11 semesters), ii) getting a degree in another area (10-12 semesters) and subsequently getting a graduate degree in education (2-4 semesters), iii) having some teaching experience and subsequently getting a short university degree (2-5 semesters), or iv) having some teaching experience and subsequently getting a special government authorization. According to the 2006 teacher census (*Idoneidad Docentes*), 95% of all teachers get certified through channels one or two and only 5% through channels three or four. The model allows only channels i) and ii) by requiring a college degree for entering the teaching profession. All the teachers in my sample hold a college degree. In the data, I do not observe whether a potential teacher has a teaching certification. The model assumes that teaching wage offers are conditional on the individual obtaining a teaching certification before starting employment. An individual who holds a college degree in an area different from education must obtain a graduate degree in education prior to accepting a wage offer in the teaching sector. To account for the fact that such an individual must incur a cost to accept a teaching wage offer, the model assumes that potential teachers have a direct preference (or cost) for teaching that varies by characteristics that are not observed by the researcher. One such characteristic is whether the individual holds a teaching certification.

Teachers' wages in the municipal sector are determined by rigid formualae that are regularly negotiated between the government and the National Teachers' Association, *Colegio de Profesores*. Wages are subject to seniority increments and other adjustments, such as compensation for working in difficult conditions. Teachers in private schools are subject to the Private Labor Code. Their wages can be freely set by the school.

To compute the revenues of private voucher schools the model uses the formulae contained in article 25 of the law *Decreto con Fuerza de Ley Nº 2, De Educación, de 20.08.98*. Schools must use these formulae to compute the amount of revenues they are required to devote to partially fund the fellowship system, and to compute adjustments to the per-pupil subsidy. Fellowship financing increases progressively in the average payments made by parents. The amount of the per-pupil subsidy decreases progressively as tuition payments increase. Details on the formulae are provided in Appendix A.

# 3 Model

## 3.1 Decision Makers, Timing and Information Structure

In a market there are a municipal school, a private voucher school, a mass $P$ of parents and a mass $W$ of potential teachers. The parents, the potential teachers and the voucher school are decision makers in a sequential problem. Before the problem is solved, all decision makers draw shocks to their technology and/or preferences. Parents and potential teachers also draw a type, a discrete random variable that also affects

their preferences and technology. Types are used in estimation to capture characteristics not observed by the econometrician: they index subsets of the populations of parents and potential teachers that differ in terms of characteristics not observed by the researcher. Some structural parameters are allowed to vary by type (Heckman and Singer (1984)). The distributions of shocks and types are common knowledge, and decision makers know the realization of their own shock and type. Moreover, the voucher school observes the realizations of the shocks and types of all potential teachers. The municipal school is the residual claimant of the demand for enrollment and supply of teachers not absorbed by the voucher school.[20]

The timing of the problem after the realization of shocks and types is as follows:

*Stage 1: Private School Sets Tuition and Wages.* The private voucher school announces tuition $p$ and the skill price $r$ of one unit of teaching skills.

*Stage 2: Labor Supply of Potential Teachers and School Choice of Parents.* Potential teachers make a labor supply decision. Parents choose a school for their child.[21]

I solve the problem by backward induction.

Although the model refers to a unique market, across-market variation is needed to estimate a subset of the parameters in the model. I estimate the model over a sample of $D$ markets, and I assume that each market is a closed market, i.e., there is no across-market mobility of potential teachers and parents. The sequential problem described above is independently solved in each market. In describing the model I indicate explicitly which parameters are assumed to vary across markets.

## 3.2    Primitives of Stage Two: Potential Teachers

The problem of potential teachers is a Roy model (Roy (1951)) with two types of skills, teaching and non-teaching, extended to incorporate a non-market option and utility (as opposed to earnings) maximization. After drawing their shock and type, potential teachers receive a positive wage offer from the non-teaching sector $(NT)$, from the municipal school $(M)$ and from the private voucher school $(V)$ and they decide whether to accept one of those offers or to not work $(H)$. Choices are mutually exclusive. Let $L_i \in \{M, V, NT, H\}$ denote the labor supply decision of individual $i$. Given the wage offers $w_i^M, w_i^V, w_i^{NT}$, the choice-specific utilities of potential teacher $i$ are:

$$
\begin{aligned}
u_i^M &= ln(w_i^M) + \mu_i^M \\
u_i^V &= ln(w_i^V) + \mu_i^V \\
u_i^{NT} &= ln(w_i^{NT}) \\
u_i^H &= \mu_i^H + \epsilon_i^H
\end{aligned}
$$

---

[20]Public schools are not subject to significant incentives to compete. For example, individual schools do not have control over teacher wages and recruitment. There is evidence that few public schools have been forced to close since the introduction of the voucher plan, in spite of the growth of the private sector.

[21]Because of data limitations, I do not jointly model school and labor supply choices. This is likely to have a negligible effect on my policy experiments: the only policy response I am not accounting for is how the choice of a school for a parent who is also a potential teacher is affected by policies that affect her labor supply decisions. Because college graduates, the potential teachers in the model, represent only 10% of the Chilean population, and only a fraction of them has a school-aged child, ignoring the school choice for their child has a likely negligible effect on the policy outcomes.

where $\mu_i^M, \mu_i^V$ and $\mu_i^H$ are individual $i$'s non-pecuniary utilities associated with each choice. The non-pecuniary term for the non-teaching sector has been normalized to zero because it is not separately identified.[22] A potential teacher chooses labor supply $L_i$ so as to maximize his utility.

### 3.2.1 Wages

The wage offers depend on a vector of individual characteristics $Z_i$, on the realization of the potential teacher's type $l_i \in \{1, ..., \bar{L}\}$, which is drawn from a probability mass function with proportions $\psi_1, ..., \psi_{\bar{L}}$, and from the realization of a vector of sector-specific shocks $\epsilon_i = [\epsilon_i^M \quad \epsilon_i^V \quad \epsilon_i^{NT}]'$, which is *iid* across individuals and distributed as $N(0, \Sigma)$. I assume that the wage shocks are independent; therefore, $\Sigma$ is a diagonal matrix with diagonal elements $\sigma_M^2, \sigma_V^2, \sigma_{NT}^2$. Any correlation between wage offers derives from the type distribution. I do not make any *a priori* assumption on the magnitude or sign of this correlation. The characteristics that enter $Z_i$ are age, gender, whether the individual holds a professional certification (*perfeccionamiento*), and whether the individual holds a graduate university degree (master's or Ph.D.).[23]

The wage offer functions in each sector are:

- Municipal school: $w_i^M = exp\left(\alpha_{0M}(l_i) + \alpha_M' Z_i + \epsilon_i^M\right)$. The wage offer is an exogenous function that represents the rigid formulae negotiated by the government and the national teachers' association. The formulae include adjustments to the wage for characteristics not observed in my sample; for example, whether the teacher is working in difficult conditions, or if he has received teaching prizes. To capture these adjustments I let the intercept of the log-wage vary by type $l_i$, and I let the wage offer depend on a shock $\epsilon_i^M$. Individual municipal schools in Chile do not have the power to negotiate over these wage adjustments, which are set *a priori* by law. Although the wage offer is not a function of an individual's teaching skills, some of the wage adjustments, like teaching prizes, might reflect the true teacher's skills.

  Finally, estimation is performed over multiple markets. The intercept of the log-wage, $\alpha_{0M}(l_i)$, is assumed to vary by market. A model with an intercept that does not vary by market is unable to fit the data. To limit the number of parameters to estimate, I assume that the intercept of type 1 varies by market, whereas the intercepts of the other types are obtained as an affine linear function of the intercept of type 1. This function is constant across markets. Letting $d$ index a market: $\alpha_{0M}^d(l_i) = \alpha_{0M}^d(1) + \Delta(l_i)$ if $l_i \neq 1$. Letting the number of markets be $D$, this reduces the number of parameters to estimate from $D \times \bar{L}$ (when the intercepts of all types vary by market) to $D + \bar{L} - 1$.[24]

- Voucher school: $w_i^V = rs_i$. The wage offer is equal to the product of the price of one unit of teaching skills, $r$, and the number of teaching skills possessed by individual $i$, $s_i$. The price of skills $r$ is chosen by the private school in the first stage of the model.[25] The teaching skills that individual $i$ possesses

---

[22]Because no earnings are observed if the home sector is chosen, the constant in the utility from one of the four sectors cannot be separately identified and must be set to zero. See Heckman and Honoré (1990), Theorem 6.

[23]Although it would be desirable to include teaching and non-teaching work experience, these variables are not available for the sample of potential teachers. I have ascertained that age and gender correlate substantially with teaching and non-teaching work experience by using a dataset that is not suitable for the estimation of the model in this paper because it lacks geographical information, the *Encuesta de Protección Social*.

[24]This parameter reduction restricts the correlation between municipal wages and teaching skills to be constant across markets.

[25]In estimation, the price of skills $r$ varies by market.

are equal to:

$$s_i = exp\left(\alpha_{0V}(l_i) + \alpha_V' Z_i + \epsilon_i^V\right) \tag{1}$$

where $\epsilon_i^V$ is a production shock. The log-skill intercept $\alpha_{0V}(l_i)$ can be interpreted as the teaching skill endowment of individual $i$. This is the amount of teaching skills that is not affected by the $Z_i$ characteristics: it does not vary, for example, by an individual's degrees or age.

- Non-teaching sector: $w_i^{NT} = r^{NT} s_i^{NT} = exp\left(\alpha_{0NT}(l_i) + \alpha_{NT}' Z_i + \epsilon_i^{NT}\right)$. The wage offer depends on the amount of non-teaching skills possessed by the individual, $s_i^{NT}$, and on the price of those skills, $r^{NT}$. I assume that $r^{NT}$ is exogenous and invariant to the policies considered in this paper. Under this assumption, quantifying the response to policy of teaching skill supply does not require separately identifying non-teaching skills from their price. Therefore, I let the log-wage intercept $\alpha_{0NT}(l_i)$ in the above formula absorb the log-price of skills and the log-skill intercept. The shock $\epsilon_i^{NT}$ is a shock to the production of non-teaching skills. In estimation, the log-wage intercept $\alpha_{0NT}(l_i)$ varies by market.[26] The covariance between teaching and non-teaching skills is entirely driven by the distribution of types.

### 3.2.2 Non-Pecuniary Utility

Individuals enjoy the following choice-specific non-pecuniary utilities:

- Municipal school: $\mu_i^M = \mu_{0M}(l_i) + \mu_{0Teach} D_i^f$ where $D_i^f$ is a gender dummy equal to one if $i$ is female.[27]

- Voucher school: $\mu_i^V = \mu_{0V}(l_i) + \mu_{0Teach} D_i^f$.

- Home: $\mu_i^H = \mu_{0H}(l_i) + \mu_H' D_i$ where $D_i$ is a vector of individual characteristics containing a gender dummy, the number of children, the number of children younger than two and between two and six years of age, age and age squared. The non-pecuniary utility is also subject to a preference shock $\epsilon_i^H \sim N(0, \sigma_H^2)$ independent of the wage shocks. The variance of the non-pecuniary utility from home is identified by virtue of the assumption that the non-pecuniary utilities from the other options are degenerate random variables. See Heckman and Sedlacek (1986) for a formal discussion of the role of this assumption in identification.

## 3.3 Primitives of Stage Two: Parents

Parental decision making follows a unitary model. Each couple has one school-aged child and, after drawing their shock and type, must choose between the private voucher and the municipal school. Parents receive a voucher $v$ entirely covering tuition in the municipal school, and they are eligible for a fellowship $0 \leq f(X_h, p) \leq p - v$, which depends on their characteristics $X_h$ and on the tuition charged by the voucher school, $p$, and which entirely or partially covers the difference between tuition and the voucher. Some parents are not eligible for a fellowship, for them $f(\cdot) = 0$. School choices are mutually exclusive. Let $E_h \in \{M, V\}$ denote the school choice of parents $h$.

---

[26] I do not adopt the same parameter reduction technique used in the municipal sector: this parameter varies both by type $l_i$ and market $d$. Therefore, both the within market expectation and the variance of this parameter with respect to the type distribution vary by market. As a consequence the correlation between teaching and non-teaching skills is not restricted to be constant across markets.

[27] The term $\mu_{0Teach}$ proved important in fitting occupational choices and accepted wages.

Parents care about consumption, $c_h$, and the cognitive achievement of their child, $a_h$. They also have a direct preference for the type of school. This term varies by type $k_h \in \{1, ..., \bar{K}\}$ to capture unobservable characteristics of the household that affect the utility from a choice of school, such as the relative distance from a school.[28]

I assume that parents are credit constrained: if private tuition net of all financial aid is greater than or equal to household income, $y_h$, the utility from choosing the voucher school is negative infinity.[29] Let $a_h^M$ and $a_h^V$ denote cognitive achievement of child $h$ in the municipal and the voucher school and let $\eta_h^M(k_h)$ denote a direct preference for school $M$. The choice-specific utilities of parents $h$ are:

$$
\begin{aligned}
u_h^M &= \tau(k_h)ln(c_h) + a_h^M + \eta_h^M(k_h) + \nu_{h\eta} \\
u_h^V &= \tau(k_h)ln(c_h) + a_h^V \quad \text{if } y_h - p + v + f(X_h, p) > 0 \\
u_h^V &= -\infty \quad \text{if } y_h - p + v + f(X_h, p) \leq 0
\end{aligned}
$$

where $\nu_{h\eta} \sim N(0, \sigma_\eta^2)$ is a shock to the preference for the municipal school, and where the direct preference for the voucher school has been normalized to zero because it is not separately identified. The parameter $\tau(k_h)$ is a weight that parents place on the utility from consumption and that measures the rate at which they are willing to trade off utility from consumption for child achievement. The household's type $k_h$ is drawn from a probability mass function with proportions $\pi_1, ..., \pi_{\bar{K}}$. Consumption $c_h$ is equal to household income if the municipal school is chosen, and to household income minus tuition payments if the voucher school is chosen:

$$
c_h = \begin{cases} y_h & \text{if M chosen} \\ y_h - (p - v - f(X_h, p))) & \text{if V chosen.} \end{cases}
$$

In the empirical implementation household income is defined as the income of all members of a household, which could include members of the extended family. Parents make the school choice $E_h$ that maximizes their utility.

### 3.3.1 Cognitive Achievement

Cognitive achievement is produced as a function of home inputs, the child's innate ability and mean teacher skills. Home inputs are an exogenous, policy-invariant function of parental characteristics and type. The characteristics used are household income per capita, household income per capita squared and average parental education.[30] The child's innate ability is captured by a student's type, which is the same as a household's type. Some of the parameters of the cognitive achievement production function vary by type.

Given a vector of parental characteristics $X_h$ and a vector of average teaching skills supplied to each school, $\bar{s} = [\bar{s}^M \quad \bar{s}^V]'$, the school-specific cognitive achievement of the child of couple $h$ is:

---

[28]This term also captures the possibility that the type of school affects future outcomes of the child independently of its effect on the child's achievement. For a rationalization of this preference term, which is outside the scope of this paper, see, for example, MacLeod and Urquiola (2009), where schools are characterized by a reputation that affects the job market outcomes of their graduates. (Fully incorporating MacLeod and Urquiola's reputation term into the present framework would require letting the preference for a school depend on the characteristics of the children at that school, which would introduce peer effects and therefore estimation challenges.)

[29]In the sample, I never observe parents choosing the voucher school when their income is smaller than tuition.

[30]Households are often composed of extended families. The older members are typically not income earners, and their presence might reduce the income available for educational expenses.

$$a_h^M = \beta_{0M}(k_h) + \beta_{1M}(k_h)\bar{s}^M + \beta_{2M}(k_h)^{'} X_h + \nu_h^M$$

$$a_h^V = \beta_{0V}(k_h) + \beta_{1V}(k_h)\bar{s}^V + \beta_{2V}(k_h)^{'} X_h + \nu_h^V$$

where the shocks $\nu_h^M$ and $\nu_h^V$ are distributed as independent mean-zero normal random variables with variances, respectively, $\sigma_{\nu M}^2$ and $\sigma_{\nu V}^2$. These shocks are independent of the preference shock $\nu_{h\eta}$. For ease of exposition, I grouped all exogenous characteristics into vector $X_h$ and expressed its coefficient vectors as a function of type: $\beta_{2M}(k_h), \beta_{2V}(k_h)$. The coefficient on income per capita squared, however, does not vary by type; all the other coefficients do.

Three aspects of the production function proved to be determinant in fitting school choices and the distribution of cognitive achievement. The first aspect is the dependence of the technology parameters on the type of school. This suggests that the two schools differ in some important components. There are at least three rationalizations. First, the two schools might have different productivity, so that the same inputs have different effects on achievement. Second, there might be unobserved school inputs that are different across schools and that interact with the observed inputs. Third, the parental demand for home inputs may vary depending on the type of school the parents choose. The last two rationalizations are admissible interpretations within the framework of this model under the assumptions that the unobserved school inputs and the demands for home inputs are policy invariant. The second aspect that was determinant in fitting the data is the dependence of the intercept and of some of the coefficients on the type. This means that unobserved heterogeneity across students is important in explaining observed outcomes. I incrementally injected unobserved heterogeneity into the model until it was able to fit the data. Finally, the intercept in both production functions is subject to geographical variation. In order to be able to separately identify the constant term from the effect of teacher skills, I restrict the constant to be the same in two markets: the Santiago metropolitan region and the Valparaíso region. These two markets, which are both in the central part of the country, are the two largest metropolitan regions. They share similar features in terms of population density and socioeconomic characteristics. The constants in all other markets are allowed to vary by market. To limit the number of parameters, I assume that the intercept of type 1 varies by market (except for the aforementioned restriction), whereas the intercepts for the other types are obtained as an affine linear function of the intercept of type 1, and this function is constant across markets. Letting $d$ denote a market: $\beta_{0M}^d(k) = \beta_{0M}^d(1) + \Delta_M(k)$ and $\beta_{0V}^d(k) = \beta_{0V}^d(1) + \Delta_V(k)$. The number of intercepts to estimate is reduced from $(D-1)*K$ per school type to $D - 1 + K - 1$ per school type. The parameter reduction amounts to imposing additional structure on the way that unobserved heterogeneity affects cognitive achievement, and it allows me to fit the data as well as a less parsimonious model with one intercept per type and per market.

Because this specification of cognitive achievement is linear in teachers' skills and because cognitive achievement enters linearly in the utility function, parents can be interpreted as expected utility maximizers where achievement depends on the skills of the teacher assigned to their child, and the expectation is taken with respect to the distribution of teacher skills in each school and students are randomly assigned to teachers. In the remainder of the paper, however, I adopt the interpretation that mean teaching skills enter the cognitive achievement production function and that households maximize a deterministic utility.

The model features interactions between the unobserved types and both teacher skills and parental

characteristics, in the form of type-specific coefficients. Including an interaction term between parental char-acteristics and teacher skills does not improve the model fit over the current model with only an interaction between type and teacher skills. To obtain a fit as good as the one I obtain with this model, the coefficients on the interactions between parental characteristics and teacher skills must be close to zero. Therefore, the current specification, which does not include these interactions, is empirically supported.

The specification of the cognitive achievement production function generates sorting based on both the observable characteristics $X_h$ and the unobservable characteristics captured by the type. Notice, however, that other aspects of the model of parental sorting also generate these sorting patterns. Sorting based on parental income, for example, also occurs because the utility from consumption is concave.[31] Sorting based on the unobserved type also derives from the fact that types enter the direct utility for the municipal school and the weight $\tau_h$ on the utility from consumption.

### 3.3.2 Preference for the Type of School

Parents have a direct preference for the municipal school, $\eta_h^M$, which is independent of how the school affects achievement. This term is assumed to depend on the parents' type $k_h$, on the school level (primary or secondary) and on whether the family lives in a rural area:[32]

$$\eta_h^M = \eta(k_h) + \eta_1 primaria_h + \eta_2 rural_h$$

where $primaria_h$ is equal to one if the student is in primary school.

### 3.3.3 Fellowship Formula

There is an exogenous, nationwide fellowship formula:

$$f(X_h, p) = b_0 + b_1(p - v) + b_2 primaria_h + b_3 nfam_h + b_4 rural_h + b_5 y_h$$

where $nfam_h$ is the size of the extended family residing in the same house. In estimation it is assumed that the fellowship is observed with a measurement error.

## 3.4 Solution of Stage Two

The order in which households and potential teachers make their decisions is irrelevant. This is because the decisions of potential teachers do not depend on the parental school choices, and because parents, whose decisions depend on the choices of potential teachers, are able to derive the mean skills supplied to the two schools. This is because parents know the wage offer functions, the distribution of exogenous potential teachers' characteristics, the distribution of their shocks, their utility functions and the fact that they are utility maximizers. Because there is a continuum of potential teachers, parents face no uncertainty in terms

---

[31] Two families with identical $a_h^M, a_h^V$ and with identical $\tau_h$ and preference shock realizations might make different school choices if their incomes are different because, for the wealthier family, private education costs less in terms of consumption utiles. The model, therefore, generates a correlation between parental income and school choice that is independent of the differential in the mean teaching skills in the two schools.

[32] The dependence on school level and rural status also captures the presence of primary and secondary private schools in rural and urban areas. Private schools are more scarce in rural areas, indicating that the traveling cost could be higher if the few private schools are not located near the household's residence.

of the mean skills optimally supplied by potential teachers to each sector. A model in which parents and potential teachers move simultaneously, one in which parents move first and one in which potential teachers move first are equivalent.

### 3.4.1 Labor Supply of Potential Teachers

Potential teachers move after the private school has chosen $r$. Their choices determine the amounts of teaching and non-teaching skills supplied to each sector, as well as the mass of individuals in each occupation. In this section I derive the mean teaching skills supplied to each school, which enter the parental problem, and the total skills and mass of teachers supplied to the private voucher school, which enter the voucher school's problem in the first stage.

Potential teacher $i$ chooses to work in the voucher school if and only if:[33]

$$\left(u_i^V \geq u_i^M\right) \wedge \left(u_i^V \geq u_i^{NT}\right) \wedge \left(u_i^V \geq u_i^H\right). \tag{2}$$

The proportion of the population of potential teachers working in the voucher school is the proportion for whom these inequalities are true. Let $Q_i = [Z_i \quad D_i]'$ be the variables that determine wage offers and preference for home, and assume that $Q_i$ has a well-defined joint density $f^q(q)$. The proportion of individuals choosing the voucher school is:

$$\Pr_W^V = \int_q \Pr(u_i^V \geq u_i^M|q)\Pr(u_i^V \geq u_i^{NT}|u_i^V \geq u_i^M, q)\Pr(u_i^V \geq u_i^H|u_i^V \geq u_i^M, u_i^V \geq u_i^H, q)f^q(q)dq \tag{3}$$

where $\Pr(u_i^V \geq u_i^J|q)$ is the proportion of individuals with characteristics $Q_i = q$ who prefer the voucher school to sector, $J \in \{M, NT, H\}$.[34] The mass of individuals choosing the voucher school, $SV$, is given by this proportion multiplied by the total mass potential teachers in the market, $W$:

$$SV = W\Pr_W^V. \tag{4}$$

Refer to Appendix B for a derivation of 3.

To compute the mean teaching skills supplied to the voucher school I derive the density of teaching skills conditional on the voucher school being chosen, which in general is different from the population density of teaching skills. Recall that the teaching skills of an individual with type draw $l_i$ and shock draw $\epsilon_i^V$ are:

$$s_i = s(\epsilon_i^V, l_i) = exp\left(\alpha_{0V}(l_i) + \alpha_V' Z_i + \epsilon_i^V\right)$$

with $\epsilon_i^V \sim N(0, \sigma_V^2)$. That is, conditional on type, skills are log-normally distributed. Given $Z_i = z$, the density of teaching skills depends both on the density of the shock $\epsilon_i^V$ and on the type probability $\psi_{l_i}$:[35]

---

[33]Because all shocks are continuously distributed in the population and are non-degenerate random variables, the population proportion of individuals who are indifferent between sectors is negligible and $\geq$ below can be replaced by $>$.

[34]For ease of exposition I replaced the multiple integral signs with respect to the elements of vector $Q_i$ by a unique integral sign indexed by $q$. Integration is over the support of $f^q(q)$.

[35]If $ln(x) \sim N(0, \sigma^2)$, $x$ has density $\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(lnx-\mu)^2}{2\sigma^2}}$ with $x \geq 0$.

$$f^s(s_i|z) = \frac{\psi_{l_i}}{s_i \sigma_V \sqrt{2\pi}} exp\left\{-\frac{(lns_i - \alpha_{0V}(l) - \alpha_V' z)^2}{2\sigma_V^2}\right\}.$$

The population density is obtained by integrating over the distribution of $z$, $f^z(z)$:

$$f^s(s_i) = \int \frac{\psi_{l_i}}{s_i \sigma_V \sqrt{2\pi}} exp\left\{-\frac{(lns_i - \alpha_{0V}(l) - \alpha_V' z)^2}{2\sigma_V^2}\right\} f^z(z)dz.$$

To derive the density of teaching skills in the voucher school, define $A(q, \epsilon_i^V, l_i)$ as the subset of $\mathbb{R}^3$ that is such that if $\epsilon_i^{-V} = [\epsilon_i^M \quad \epsilon_i^{NT} \quad \epsilon_i^H]' \in A(q, \epsilon_i^V, l_i)$, an individual with characteristics $q$, shock realization $\epsilon_i^V$ and type realization $l_i$ chooses the voucher school. This subset is defined by the inequalities in (2). The density of teaching skills in sector $V$ may be written as:

$$g^V(s_i|\text{sector V chosen}) = \frac{1}{\Pr_W^V} \psi_{l_i} \int_{\epsilon_i^{-V} \in A} f^s(s_i) f^{-V}(\epsilon_i^{-V}) d\epsilon_i^{-V}$$

where I let $\int_{\epsilon_i^{-V} \in A}$ denote multiple integration with respect to $\epsilon_i^M, \epsilon_i^{NT}, \epsilon_i^H$ over the area $\epsilon_i^{-V} \in A(q, \epsilon_i^V, l_i)$, and where the joint density of the shocks in sectors $M, NT$ and $H$ is:

$$f^{-V}(\epsilon_i^{-V}) = \frac{1}{\sigma_M \sigma_{NT} \sigma_H} \phi\left(\frac{\epsilon_i^M}{\sigma_M}\right) \phi\left(\frac{\epsilon_i^{NT}}{\sigma_{NT}}\right) \phi\left(\frac{\epsilon_i^H}{\sigma_H}\right).$$

The density of teaching skills in the municipal school, $g^M(s_i|\text{sector M chosen})$, can be derived by a similar argument.[36]

The mean skills supplied to each sector are obtained using the conditional densities $g^M, g^V$:

$$
\begin{aligned}
\bar{s}^M &= \sum_{l_i} \psi_{l_i} \int s(\epsilon_i^V, l_i) g^M(s(\epsilon_i^V, l_i)|\text{sector M chosen}) d\epsilon_i^V \\
\bar{s}^V &= \sum_{l_i} \psi_{l_i} \int s(\epsilon_i^V, l_i) g^V(s(\epsilon_i^V, l_i)|\text{sector V chosen}) d\epsilon_i^V.
\end{aligned}
\tag{5}
$$

### 3.4.2 School Choice of Parents

Parents move after the private school has chosen tuition $p$ and skill price $r$. Parents derive the mean skills optimally supplied by potential teachers to each sector, given in (5). Parents' choice of school determines the mass of children enrolled in each school for each pair $(p, r)$, which I refer to as demand for enrollment. In this section I derive the demand for enrollment of the private voucher school, which enters the voucher school's problem.

Parents $h$ choose to enroll their child in the private voucher school if and only if:[37]

$$u_h^V \geq u_h^M. \tag{6}$$

---

[36] First define the proportion of potential teachers choosing the municipal school, $Pr_W^M$. Then define the area $B(q, \epsilon_i^V, l_i)$ that is such that if $[\epsilon_i^M \quad \epsilon_i^{NT} \quad \epsilon_i^H]' \in B(q, \epsilon_i^V, l_i)$, an individual with characteristics $q$, shock realization $\epsilon_i^V$ and type realization $l_i$ chooses the municipal school.

[37] Because all shocks are continuously distributed in the population and are non-degenerate random variables, the population proportion of parents who are indifferent between schools is negligible and $\geq$ below can be replaced by $>$.

The proportion of the population of parents enrolling their child in the voucher school is the proportion for whom this inequality is true. Recall that $X_h$ includes household's income, which affects consumption, and other variables that determine home input demands and fellowship amount. Assume that $X_h$ has a well-defined joint density $f^x(x)$. The proportion of parents choosing the voucher school is:

$$Pr_P^V = \int_x Pr\left(u_h^V \geq u_h^M | x\right) f^x(x) dx$$

where $Pr(u_h^V \geq u_h^M | x)$ is the proportion of parents with characteristics $X_h = x$ who choose the voucher school.[38] The mass of parents choosing the voucher school, $DV$, is given by this proportion multiplied by the total mass of parents in the market:

$$DV = PPr_P^V. \tag{7}$$

Next, I derive $Pr(u_h^V \geq u_h^M | x)$. Denote by $b(k_h, \nu_h^V, \nu_{h\eta}; x)$ the cutoff value that is such that if $\nu_h^M < b(k_h, \nu_h^V, \nu_{h\eta}; x)$, the voucher school is preferred to the municipal school by parents with characteristics $X_h = x$.[39] This cutoff depends on the realization of the cognitive achievement production shock in the voucher school $(\nu_h^V)$, of the preference shock for the municipal school $\nu_{h\eta}$ and on the realization of the parents' type $k_h$. The proportion of parents with characteristics $X_h = x$ who prefer the voucher school to the municipal school is:

$$Pr\left(u_h^V \geq u_h^M | x\right) = \sum_{k_h=1}^{\bar{K}} \pi_{k_h} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{b(k_h, \nu_h^V, \nu_{h\eta}; x)} f\left(\nu_h^V, \nu_h^M, \nu_{h\eta}\right) d\nu_h^M d\nu_h^V d\nu_{h\eta}$$

where under the assumption of normality and independence of the shocks, the joint density $f\left(\nu_h^V, \nu_h^M, \nu_{h\eta}\right)$ is equal to $\frac{1}{\sigma_{\nu V}\sigma_{\nu M}\sigma_\eta} \phi\left(\frac{\nu_h^V}{\sigma}\right) \phi\left(\frac{\nu_h^M}{\sigma_{\nu M}}\right) \phi\left(\frac{\nu_{h\eta}}{\sigma_\eta}\right)$.

### 3.4.3 Equilibrium of Stage Two

An equilibrium in the second stage is attained when, given $p$ and $r$, all parents and all potential teachers choose the option that maximizes their utility. Let $E_h^*(p, r) \in C_P = \{M, V\}$ denote the optimal school choice of parents $h$, and let $L_i^*(r) \in C_W = \{M, V, NT, H\}$ denote the optimal labor supply of potential teacher $i$.[40]

*Existence and Uniqueness.* The equilibrium of stage two exists and is unique. Existence is easily established: given $(p, r)$, each potential teacher and each household have at least one most-preferred choice. This is true by construction: utilities of potential teachers and of households are well-defined for every admissible value of $r$ and $p$.[41] Because teaching skills and preferences of potential teachers are continuously distributed in the population and are non-degenerate random variables, the population proportion of potential teachers who are indifferent between sectors has measure zero. As a consequence, all potential teachers have a most-preferred sector and one can define the optimal action profiles of potential teachers as a function $c_W \in C_W^{[0,W]}$. Similarly, because parental preferences and shocks to the production of student achievement are continuously distributed in the population and are non-degenerate random variables, indifference of par-

---

[38]For ease of exposition I replaced the multiple integral signs with respect to the elements of vector $X_h$ by a unique integral sign indexed by $x$. Integration is over the support of $f^x(x)$.

[39]The expression for this cutoff can be found in Appendix B.

[40]The school choice of parents depends on $r$ because parents use knowledge of $r$ to derive the mean teaching skills optimally supplied to each school.

[41]Skill price $r$ must be positive.

ents between schools occurs on a set of measure zero. All parents have a most preferred school and one can define the optimal action profiles of parents as a function $c_P \in C_P^{[0,P]}$. This establishes uniqueness of the equilibrium: for every pair $(p, r)$ there is one and only one profile of parents' and potential teachers' optimal choices.[42]

For the purposes of solving the voucher school's problem in the first stage of the model, I define four objects that enter the voucher school's payoff and that derive from the equilibrium of the second stage. These are the demand for enrollment in the private voucher school, $DV(p.r)$, the supply of teachers to the private voucher school, $SV(r)$, the total skills supplied to the voucher school, $TSV(r)$, and the mean tuition paid by the households who choose the voucher school, $EPV(p, r)$.[43] All four objects enter the profit function of the school. The least intuitive is the mean tuition, which enters the profit function only because of the governmental formulae for the per-pupil subsidy received by the school and for the per-pupil funding of financial aid that is due by the school. These formulae can be found in Appendix A. In particular, the subsidy given to the school is reduced in a progressive manner as average tuition payments made by parents increase. The fraction of financial aid in the form of fellowships that has to be financed by the school is an increasing function of average tuition payments. I explicitly incorporate the official governmental formulae in the computation of the voucher school's revenues.

*Definition of Optimal Sorting.* The solution to the households' and potential teachers' problems in the second stage of the model yields a vector valued optimal sorting function $OS(p, r) = [DV(p, r), SV(r), TSV(r), EPV(p, r)]'$ of tuition and wage rates. In particular, equations (7), (4), (5) and equation (13) in Appendix A yield the enrollment, the supply of teachers, the mean teacher skills and the mean tuition payments in the voucher school as a function of $(p, r)$.

## 3.5 Private Voucher School: Stage One

In the first stage, the private voucher school chooses tuition and teachers' wages to maximize its profits subject to the function $OS(p, r)$, i.e., subject to its demand for enrollment, supply of teachers, supply of teaching skills and mean tuition payments, all of which are the outcome of the optimal sorting of parents and potential teachers in stage two.[44] Moreover, there exists a legal cap on the tuition that private subsidized schools can charge. Profit maximization is an appropriate assumption in the Chilean context because, as Elacqua (2006) documents using 2003 data from the Ministry of Education, 73.60% of all Chilean voucher schools report being for-profit.[45] Let $g^r(EPV(p, r))$ denote the governmental formula for the per-pupil

---

[42]For the purposes of estimation, the presence of a zero-measure mass of individuals who are indifferent between alternatives is irrelevant. This is why I specify the action profiles as functions.

[43]Different households receive different levels of financial aid, and therefore, they are responsible for different tuition payments. $EPV(p, r)$ is the average of these payments.

[44]Each voucher school offers both primary and secondary education. Although it chooses a unique tuition for the two levels, parents of students in primary and secondary school receive different fellowships. The model therefore accounts for differences in tuition payments across education levels.

[45]McEwan, Urquiola and Vegas (2008), in describing the Chilean educational institutions, identify Chilean voucher schools as mostly profit maximizing. Urquiola and Verhoogen (2009) build a model of school competition in Chile and assume that schools are profit maximizing. Their paper is the only other paper including a model of the market for education in Chile. Our models share a few features, but they are substantially different. Both papers model the behaviors of parents and schools. They take private schools as profit maximizers, they assume that the effect of schools on achievement is known by the decision makers and they provide a random utility model of parental preferences, which depend on the school's effect on student achievement and on consumption. The choice set of schools, the constraints they are subject to and the structure of the market are different in the two models. A major departure of my model, due to its different goal, is that it includes the labor supply decision of potential teachers and private schools' wage setting.

revenues before any fellowship expenses, and $g^f(EPV(p,r))$ for the per-pupil expenses devoted to finance the fellowships. The formulae can be found in Appendix A. Formally, the problem of the private subsidized school is:

$$\max_{(p,r)} \Pi = \underbrace{g^r(EPV(p,r))DV(p,r)}_{\text{GrossRevenues}} - \underbrace{g^f(EPV(p,r))DV(p,r)}_{\text{FinancialAid}} - \underbrace{((c_1+\epsilon_{cost})DV(p,r) + c_2 DV(p,r)^2)}_{\text{Variable\quad Cost}}$$

$$- \underbrace{rTSV(r)}_{\text{Tot\quad Wages}} - \underbrace{c_3 \frac{DV(p,r)/SV(r)}{45}}_{\text{NClasses\quad per\quad Teacher}} - \underbrace{OC}_{\text{Fixed\quad Operating\quad Costs}}$$

$$\text{s.t.}[\quad DV(p,r) \quad SV(r) \quad TSV(r) \quad EPV(p,r)]' = OS(p,v)$$

$$p \le \bar{p} \quad \text{with multiplier} \quad \lambda.$$

The total wage payments are given by the teaching skills supplied to the school, $TSV$, multiplied by the rental rate. The variable cost is subject to a shock $\epsilon_{cost}$ which is distributed according to a truncated log-normal distribution with parameters 0 and $\sigma^2_{cost}$, where the truncation is such that profits are non-negative.[46] The truncation is needed to guarantee that there is a private school in the market.[47] The term $c_3 \frac{DV/SV}{45}$ is a cost proportional to the minimum average number of classes taught by the same teacher. The legal cap on class size in Chile is 45, so if there are $x$ students per teacher, those students must be split into at least $x/45$ classes.[48] Finally, I assume that if $DV = 0$ or $SV = 0$, the private voucher school does not operate.

### 3.5.1 Existence of a Solution to the Private School's Problem and Discussion of Multiplicity

**Existence**

The existence of a solution requires continuity of $\Pi$ and compactness of the set in which $(p,r)$ lie. I assume that $(p,r) \in [0, \bar{p}] \times [\delta, \bar{r}]$ for $\delta > 0$ arbitrarily small, which is compact. Over this domain, $\Pi$ is continuous because it is a continuous function of continuous functions. A solution exists by the Extreme Value Theorem.[49] The skill price $r$ is restricted to be positive because at $r = 0$ the profit function is not defined. This is because the problem of potential teachers is not well-defined for non-positive wage offers. When $r > 0$

---

[46]If $\nu \sim N(0, \sigma^2_{cost})$, $\epsilon_{cost} = exp(\nu)$. I choose a log-normal distribution to restrict the $\epsilon_{cost}$ to be positive. The parameter $c_1$ is restricted to be greater or equal to zero. As a consequence, $c_1 + \epsilon_{cost}$ is always positive.

[47]Although I do not model the entry decision, the assumption that the error is drawn from a truncated distribution is equivalent to assuming that there is a continuum of potential entrants, each of which draws a cost shock and enters only if the profits are non-negative. The assumption of a continuum guarantees that with probability one there is at least one entrant. Assume that the actual entrant is chosen at random among those willing to enter. Finally, in estimation the truncation assumes that the fixed operating costs are zero because these costs are not identified.

[48]This term avoids the outcome that a school optimally sets $p$ and $r$ so that very few but highly skilled individuals teach a large number of students. An alternative way to avoid this outcome is to have either cognitive achievement or teachers' utility, or both, depend on class size. This modeling option, however, introduces peer effects that pose challenges to the estimation of the models' parameters.

[49]The legal cap $\bar{p}$ must be below the income of the wealthiest family to guarantee that enrollment in the voucher school is not zero, and hence that profits are continuous. The legal cap is in fact smaller than the highest income.

the proportion of potential teachers choosing the voucher school is positive because the area of the shock space over which the voucher school is chosen has a positive measure, which implies that a strictly positive fraction of the $W$ mass of potential teachers chooses the voucher school.

**Discussion of Multiplicity**

Uniqueness of a solution depends on the curvature properties of the profit function at the true parameter value, and it is not guaranteed. I address the possibility of multiple solutions in estimation. Refer to section 5.1 on the two-step estimation approach for a more detailed discussion.

### 3.5.2 Finding an Optimum

The function $\Pi$ and its first and second derivative functions do not admit a closed form; hence, $\Pi$'s critical points and curvature properties can't be derived analytically.[50] I approximate profits with a function with a known closed form and I maximize the approximated profit function. I evaluate numerically the profit function at a large number of points $(p^r, r^r)$ and approximate it using an interpolating regression. To evaluate the true profit function I derive numerically the vector $OS$ at each evaluation point $(p^r, v^r)$ by solving the second stage of the model. I then plug $OS$ into the profits. I use a cubic interpolating polynomial:

$$\hat{\Pi} = \hat{a}_1 + \hat{a}_2 p + \hat{a}_3 p^2 + \hat{a}_4 r + \hat{a}_5 r^2 + \hat{a}_6 pr + \hat{a}_7 p^3 + \hat{a}_8 r^3 + \hat{a}_9 p^2 r + \hat{a}_{10} pr^2.$$

(8)

Approximation is by Ordinary Least Squares.

I solve the constrained maximization of approximated profits subject to the legal cap on tuition. I derive the points that satisfy the Kuhn-Tucker conditions, and then verify that at those critical points the second-order conditions are satisfied. To find the critical points I use a combination of analytical and numerical methods, described in Appendix C. I solve for the school's choice variables $(p, r)$ and for the Kuhn-Tucker-Lagrange multiplier $\lambda$.

It is important that the approximation be good in order for the solution of the approximated problem to be close to the solution of the real problem. The $R^2$ of the approximation depends on the vector of model parameters. At the parameter estimates, the average of the $R^2$ across markets is 96.14%. The domain over which the approximation was performed proved important to obtain a good approximation. In the area of the parameters close to the estimate, profits suffer a sharp decrease as $r$ approaches zero. To capture such a decrease, I used evaluation points with $r$ close enough to zero. However, for some parameter values this implied zero supply of teachers to the voucher school. In the theoretical model, the supply of teachers is positive for all positive values of $r$. This is because the distribution of shocks has an unbounded support and because there is a continuum of potential teachers. The numerical solution, however, uses the sample of potential teachers, which has a discrete support, and a finite number of error shock draws to derive the supply of potential teachers to the voucher school. For some small $r$ and some parameter values, no individual chooses to work in the voucher school. When the simulated supply of teachers to the voucher school is zero for a positive value of $r$, I treat profits as a missing value in the interpolating regression. Whenever profits

---

[50]The lack of a closed form is due to the presence of the cumulative normal distribution function in the expression for profits. These derive from the optimal choices of parents and potential teachers.

are non-missing, they trace a smooth graph, which is well approximated by the third-degree polynomial that I use.

# 4 Data

## 4.1 Constructing a Unique Dataset

The model is estimated using a unique dataset that I constructed by combining three data sources. I used the *Encuesta de Caracterización Socioeconómica Nacional* (CASEN) dataset to identify the pool of potential entrants into the teaching profession. The CASEN survey is a nationally representative survey of the general population that was originally designed to study pension systems. I extracted a sample of $3,520$ individuals holding a college degree, which is a necessary requirement for becoming a teacher, and tracked their occupational choices and accepted wages. The occupational choice variable that I constructed identifies the same choices that enter the model: whether the individual is a private or public sector teacher, whether she is working in the non-teaching sector or whether she opted for the non-market option.[51] I extracted information on individuals' characteristics such as their age, gender, certifications and fertility information, which serve as the exogenous variables $D_i$ and $X_i$ in the model. I selected individuals who reside in one of the geographical areas under analysis (some remote rural areas were excluded because of sample size issues) and those for whom no information is missing on the exogenous variables (I assume that data are missing at random).

To augment the sample of teachers, I obtained a sample of $3,195$ teachers from the *Encuesta Longitudinal Docente* (ELD) dataset.[52] I was able to extract from ELD the same set of individuals' characteristics obtained from CASEN. For each teacher in the sample, I extracted the choice of school sector and the wage, which are the endogenous variables in the model. From the original ELD sample, I selected teachers who work in either the municipal or the private voucher schools, who teach at the primary or secondary level, who reside in one of the geographical areas used in the analysis and those for whom no information is missing on the exogenous variables.[53]

Estimation is performed over multiple markets. To make CASEN and ELD suitable for estimation, I constructed a unique market identifier that is homogeneous across datasets. Because the coding of the geographical variables in ELD and CASEN did not match, I manually recoded all geographical variables in CASEN, which are at the same level as ELD, the *comuna* or municipality, so that they matched the ELD coding. Only thanks to the homogenization of the geographical variables was I able to assign individuals to local labor markets.

Finally, I used a sample of $100,000$ students randomly selected from the restricted version of the *Sistema de Medición de Calidad de la Educación* (SIMCE) dataset. This dataset contains information on all $4^{th}$ and $10^{th}$ graders in the country.[54] In particular, SIMCE contains administrative information on students' test scores in math and Spanish, which I use to measure achievement, and information on the students' household

---

[51] Identifying public and private sector teachers involved manually selecting teachers from a broader occupational class.

[52] Refer to Appendix G to see how I account for the fact that the ELD sample is a choice-based sample.

[53] Under the assumption that data are missing at random, the sample weighted using the weights provided with the survey is representative of all teachers in the municipal or voucher schools who do not reside in one of the isolated and rural areas not included in the analysis.

[54] Selecting a sample was necessary because the cluster of computers that I used in estimation was not able to handle the entire dataset, which included $306,506$ observations.

and choice of school sector. I extracted information on household income, size and parental education, which in the model affect the choice of school and the student's achievement. Finally, because the coding of the geographical variable in SIMCE was different from the coding in ELD and CASEN, I manually recoded the geographical variable. This allowed me to use the same market identifier in all three datasets.

To compute the mass of students that enters the profit function of the voucher school in each market, I built weights on the sample of students that indicate the number of students in the overall population of students represented by a student in my sample.[55] The weights were needed because the population of students from first to twelfth grade is larger than the size of the SIMCE dataset, which contains only students from two grades.[56] To recover the right population size of students in each market, I identified a representative sample of students from the nationally representative CASEN dataset. I then used the municipality-level weights provided with the CASEN survey and the unique market identified to recover the total number of students in each market. I used this information to assign weights to the SIMCE sample by dividing the correct population size by the SIMCE sample size in each market. The weights are used to compute the correct enrollment size in the voucher school, and hence the correct profits, at every parameter iteration used in estimation.

All three data sources refer to 2006, a year for which all three datasets are available. The only other year for which all three datasets are available is 2003; however, the test scores of primary school students are not available for 2003 because only secondary school students were tested.[57]

The dataset that I constructed is unique for three reasons. First, it is representative of the entire population of students and potential teachers in the country.[58] Second, it contains information on public and private school teachers and students. Private school information is typically unavailable. Third, the unique market identifier that is homogeneous across datasets allows me to match teachers, students and schools at the market level.

## 4.2   Markets

I separated the country into eighteen markets. In determining the market boundaries I ensured that a market would be both a closed labor market and a closed education market, i.e., that the mobility across markets of both students and potential teachers would be negligible. To assess teachers' mobility, I used the ELD dataset, which has information on both a teacher's residence and work location.[59] To analyze student mobility, I obtained from the *Centro de Microdatos* of the Univerity of Chile a unique matched dataset with information on the location of the school attended by each student. This information, coupled with the geographical location of the household's residence, allowed me to assess the mobility of students across markets.

---

[55]In the model, the mass of students represents the demand for enrollment in the school, $DV$, which affects both revenues and costs.

[56]Under the assumption of no cohort differences, the students in the two grades contained in the SIMCE dataset are representative of the population of students in all grades. The only role of the weights that I constructed is to expand the SIMCE sample to compute the total numbers of students, and not to make the sample representative.

[57]Although it would have been desirable to build a longitudinal dataset, it was not feasible with these data sources.

[58]Only two rural regions were excluded for sample size reasons: Aisén in the South and Tarapacá in the North.

[59]Unfortunately the CASEN dataset contains only information on the location of the home of the potential teacher and not on the job location. I looked only at the mobility of teachers when defining local labor markets, and I assume that the mobility of non-teachers is similar. Because the markets are such that almost all the teachers work in the market in which they reside, the mobility of non-teachers would have to be considerably higher than the mobility of teachers to have empirically non-negligible effects.

Choosing the size of markets presented a trade-off in terms of the implied sample sizes: a large within-market sample size is obtained by having a small number of large markets, whereas a large across-market sample size is obtained by having a large number of small markets. Some parameters in the model are estimated off of within-market variation, whereas others, notably the four parameters of the profit function, are estimated off of across-market variation. The eighteen markets I designed attempt to strike a balance between within- and across-market sample sizes.

The unique geographical configuration of Chile aided in the determination of market boundaries. With a total area of $291,933$ square miles $(756,102$ km$^2)$, Chile is larger than all US states except Alaska and larger than all countries in the European Union, its size being comparable to that of Turkey. Yet, it extends $2,653$ miles $(4,270$ km$)$ from north to south, and it averages only $110$ miles $(177$ km$)$ from east to west. The country occupies a narrow but long coastal strip, and this hinders mobility between the northern and southern parts of the country. I exploited this fact to build closed labor and educational markets. The partition of the country into markets that I designed is such that $98.8\%$ of teachers work in the same market in which they reside, and $99.0\%$ of parents choose a school in the same market in which they reside.[60]

Chile has fifteen regions, two of which are excluded from the analysis because the sample size was too small (Aisén in the South and Tarapacá in the North). The remaining thirteen regions have been divided into eighteen markets, which means that certain regions have been divided into multiple markets. I divide a region into more markets when I do not observe much mobility between separate parts of the region. Table 1 reports the region in which each market lies. The table reports also two numbers that are intended to represent how many schools within each school sector are in the actual choice set of parents. These numbers are the averages between the number of primary and the number of secondary schools within each sector, and when doubled they give, for each sector, the total number of schools in that market. The model includes both primary and secondary school students: reporting the total number of schools within each sector would over-represent the actual number of school options that each family is faced with, because each family chooses a school only from among the appropriate school level.

In the model there are only one voucher and one municipal school per sector and per market. In reality, there are multiple schools of the same type in each market, and the number of schools that the model aggregates into a single one varies considerably across markets. In Appendix D I show that the private and public school sectors containing more schools do not necessarily contain more heterogeneous schools and, therefore, that the assumption of a unique municipal and a unique voucher school per market is equally accurate for larger markets as it is for smaller ones.

Finally, table 2 presents sample sizes by market and dataset. It also shows the population sizes of potential teachers, teachers and students in each market. Table 3 shows the shares of employment and enrollment in the private subsidized schools.[61] Enrollment share in voucher schools in the whole country is $52.99\%$ and teachers' employment is $45.16\%$, however, there is considerable across-market variation.

---

[60]I dropped from the sample the students and teachers who are observed moving across markets.

[61]These are the shares of enrollment and teacher employment that are in the voucher school as opposed to the municipal school. These statistics do not take into account private unsubsidized schools.

Table 1: Markets and Number of Schools

| Market | Region | Municipal | Voucher |
|--------|--------|-----------|---------|
| 1 | Arica and Parinacota | 51.5 | 57.5 |
| 2 | Coquimbo | 171 | 99.5 |
| 3 | Libertador G. B. O'Higgins | 101 | 61.5 |
| 4 | Atacama | 27 | 15.5 |
| 5 | Maule | 286 | 93 |
| 6 | Biobío | 193.5 | 67.5 |
| 7 | Biobío | 110 | 84 |
| 8 | Los Ríos | 128.5 | 111.5 |
| 9 | Los Lagos | 201.5 | 71.5 |
| 10 | Los Lagos | 73.5 | 54.5 |
| 11 | Antofagasta | 46 | 31.5 |
| 12 | Libertador G. B. O'Higgins | 85 | 23.5 |
| 13 | La Araucanía | 170 | 198.5 |
| 14 | La Araucanía | 20 | 32.5 |
| 15 | Región Metropolitana (Santiago) | 377 | 747.5 |
| 16 | Valparaíso | 228 | 261 |
| 17 | Biobío | 153.5 | 41 |
| 18 | Magallanes and Antártica | 13 | 10 |
| TOT | | 2,436 | 2,061.5 |
| AVERAGE | | 135.3 | 114.5 |

## 4.3 Facts about Potential Teachers

With an average monthly wage of CLP 777, 396 ($\sim$ \$1,550), a college graduate employed in a non-teaching occupation earns, on average, 62.3% more than a college graduate employed in teaching, who, on average, earns CLP479,041 ($\sim$ \$960). A substantial wage premium for non-teachers persists at all ages, reaching peaks of over 80% for individuals younger than 45. This wage premium remains large, 32.9%, when individual characteristics are accounted for. Figure 1 reports age profiles of the monthly wages of teachers and of non-teachers holding a college degree.
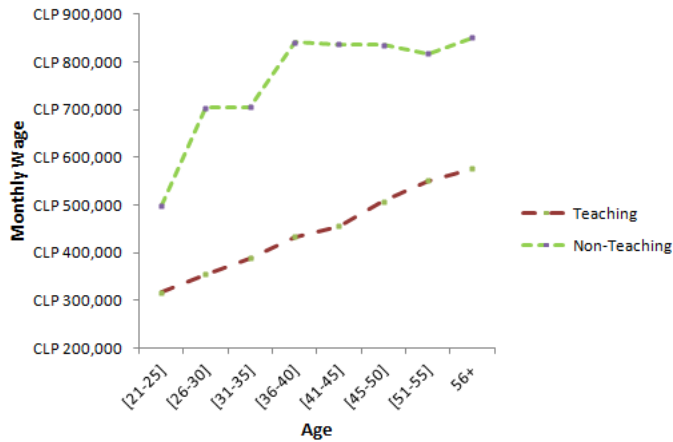


Figure 1: Average Monthly Wages of Teachers and Non-Teachers by Age

Table 2: Sample and Population Sizes by Market

| Market | CASEN | ELD | SIMCE | Pop. Pot. Teachers | Pop. Teachers | Pop. Students |
|---|---|---|---|---|---|---|
| 1 | 76 | 95 | 3,283 | 12,964 | 4,207 | 69,965 |
| 2 | 100 | 114 | 4,687 | 15,738 | 6,397 | 97,092 |
| 3 | 123 | 108 | 4,078 | 17,452 | 4,346 | 80,431 |
| 4 | 63 | 31 | 1,517 | 8,428 | 2,665 | 34,857 |
| 5 | 182 | 203 | 6,745 | 32,346 | 8,906 | 145,449 |
| 6 | 160 | 94 | 4,173 | 12,376 | 4,734 | 94,162 |
| 7 | 244 | 192 | 6,457 | 49,205 | 8,300 | 141,275 |
| 8 | 93 | 79 | 2,150 | 9,603 | 4,202 | 58,556 |
| 9 | 169 | 55 | 3,374 | 23,660 | 3,644 | 80,962 |
| 10 | 57 | 68 | 1,351 | 7,203 | 3,203 | 35,556 |
| 11 | 106 | 111 | 3,869 | 18,942 | 3,872 | 70,407 |
| 12 | 90 | 76 | 1,768 | 6,273 | 2,640 | 34,382 |
| 13 | 227 | 119 | 4,958 | 29,411 | 6,471 | 132,718 |
| 14 | 37 | 49 | 765 | 2,614 | 2,831 | 18,075 |
| 15 | 1,252 | 1299 | 36,582 | 400,904 | 41,905 | 573,628 |
| 16 | 392 | 368 | 10,631 | 61,763 | 12,989 | 137,462 |
| 17 | 120 | 100 | 2,761 | 13,752 | 4,399 | 61,227 |
| 18 | 29 | 34 | 851 | 6,653 | 1,459 | 17,787 |
| TOT | 3,520 | 3,195 | 100,000 | 729,288 | 127,170 | 1,883,991 |
| AVERAGE | 195.6 | 177.5 | 5,555.5 | 40,516 | 7,065 | 104,666 |

Table 3: Enrollment and Employment Shares in Voucher School by Market

| Market | Enroll. in V | Employment in V |
|---|---|---|
| 1 | 61.7% | 52.6% |
| 2 | 44.7% | 40.5% |
| 3 | 37.8% | 35.6% |
| 4 | 38.4% | 30.8% |
| 5 | 39.0% | 29.6% |
| 6 | 33.4% | 21.9% |
| 7 | 42.6% | 37.9% |
| 8 | 40.4% | 28.5% |
| 9 | 35.0% | 31.5% |
| 10 | 47.2% | 48.1% |
| 11 | 34.4% | 28.7% |
| 12 | 36.8% | 23.7% |
| 13 | 54.0% | 46.2% |
| 14 | 63.8% | 55.3% |
| 15 | 67.2% | 60.1% |
| 16 | 55.6% | 47.7% |
| 17 | 47.8% | 32.7% |
| 18 | 41.0% | 30.1% |

A pattern emerges when decomposing the wage premium by gender. I ran two gender-specific regressions of log-earnings on a vector of individual characteristics and a dummy for whether the individual works as a teacher or not.[62] The coefficient on the non-teaching sector dummy is 46.9% for males and 25.6% for

---

[62]The variables included in the regressions are age and dummies for professional certificates and graduate degrees.

women, indicating that the wage premium in the non-teaching sector is higher for males. To account for the fact that non-teaching occupations often involve more hours of work than teaching, I divided wages by the average number of hours worked in teaching and non-teaching occupations and used this as a measure of hourly wages.[63] Average hourly wages are 18.7% higher for non-teachers than for teachers. To decompose this premium by gender, I ran the same gender-specific regressions used for monthly wages and I found that the hourly wage premium for men is 15.6%, and it is negative 5.7% for women. This indicates that the monthly wage differences between teaching and non-teaching females is due to the fact that non-teachers work more hours, whereas a substantial part of that difference for males is due to reasons other than hours worked. Accounting for hours worked, there remains a substantial positive premium for males who choose a non-teaching occupation.

Interpreting this gap using a simple framework with a unitary skill and without non-pecuniary considerations entering labor supply decisions would favor the conclusion that the wage differential between teachers and non-teachers is representative of a teaching skill differential. A natural policy implication would then be to increase teacher wages in order to attract the more highly skilled teachers who are currently choosing better paying occupations. A more complex and realistic framework, however, would allow for different sets of skills affecting an individual's teaching ability and her productivity in an occupation different from teaching. Within such a framework, wage differentials would not be indicative of the differential in teaching skills. Moreover, a realistic framework would also allow for non-pecuniary considerations entering labor supply decisions. If, for example, higher skilled individuals choose lower paying occupations because of their non-pecuniary characteristics, the observed wage differential across occupations is not informative of the skill differential among individuals who sorted themselves into different occupations. The model in this paper allows for a distinction between teaching skills and skills that are productive in non-teaching professions, and for non-pecuniary preferences entering labor supply decisions. I do not assume *a priori* a sign for the correlation between teaching and non-teaching skills.

Comparing the wages of private voucher and municipal school teachers suggests that the wage schedules are different, as one would expect given the different treatment reserved to municipal sector wages by the law. Younger and less experienced teachers earn more in voucher schools, and this premium is reversed for older teachers and absent for more experienced teachers. Not surprisingly, teachers in the municipal sector are, on average, 8.2 years older and have, on average, 9.0 more years of experience than teachers in the private voucher sector. Wages of teachers with up to ten years of teaching experience are 14% higher in private voucher schools than in municipal schools. The wage differential disappears for more experienced teachers. Table 4 presents average wages by teaching experience and by school type in terms of 2006 Chilean pesos (1 USD=545.50 CLP).

Wage regressions show that a wage premium for young voucher school teachers remains even after including various characteristics as regressors. The coefficient on the voucher school dummy in a log-wage regression for teachers with at most ten years of teaching experience is 13.8%, but this premium disappears, and the coefficient on the voucher dummy becomes insignificant, if the regression is performed on individuals with more than ten years of teaching experience.[64]

---

[63]Hours worked are not available in the CASEN dataset, which contains the non-teaching wages. To make the teaching and non-teaching measures of hourly wages comparable, I used average hours for both.

[64]The variables included in the regressions are teaching experience, teaching experience squared, gender, possession of a professional certification, and possession of a graduate degree.

Table 4: Average Monthly Teaching Wages by Teaching Experience and Type of School

| texp | wage M | wage V | ratio wage V/wage M |
|---|---|---|---|
| (years) | (2006 CLP) | (2006 CLP) | |
| $\leq 10$ | 368,816.2 | 423,417.7 | 1.148 |
| 11-20 | 472,502 | 472,967.1 | 1.001 |
| 21-30 | 540,992 | 544,536.9 | 1.007 |
| $\geq 31$ | 585,682.5 | 583,352.7 | 0.996 |

Source: ELD 2006. The first column indicates teaching experience in years, the second and third columns contain average wages in the municipal and voucher schools, and the last column contains the ratio of the third to second column.

Work experience is not observed for all potential teachers, it is available only for the ELD sample of teachers and it therefore cannot be used in the estimation. The wage functions in the model depend on age instead of work experience, as age is present in both the ELD (teachers) and CASEN (potential teachers) datasets. In log-wage regressions like the ones mentioned above where age and age squared replace teaching experience and teaching experience squared, similar wage profiles are observed. The coefficient on the voucher school dummy is 7.6% for individuals of up to 35 years of age, and it is $-3.5\%$ for older individuals.[65] Therefore, a similar pattern is observed when age is used as a regressor in place of teaching experience. Figure 2 shows the age profiles of wages in the two teaching sectors. Municipal sector wages, which are determined by governmental formulae, increase almost linearly with age. This is not surprising as seniority is one of the main inputs into the formulae for public sector teachers' wages.
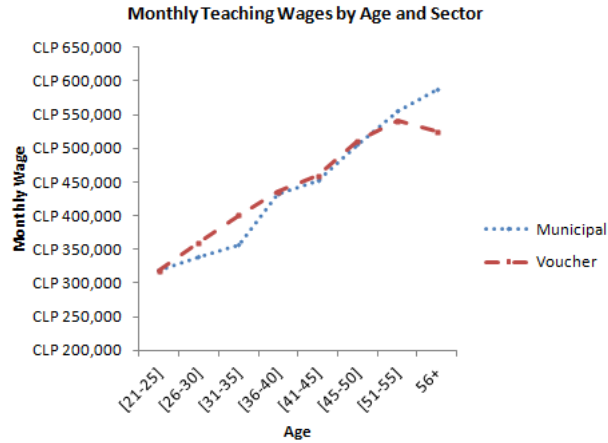


Figure 2: Average Monthly Wages of Teachers and Non-Teachers by Age

Finally, there is evidence that supports the model assumption that private schools know the underlying skills of the teacher (on which they base their wage offer). In Chile there are two types of teacher contracts: temporary and long-term. Temporary contracts are one-year contracts that can be renewed for a second year, at the end of which they are either turned into long-term contracts or not renewed. Both the municipal and the voucher school can use these two contract formats. In voucher schools 10.0% of temporary contracts are not renewed after two years. In municipal schools, on the other hand, virtually all temporary contracts

---

[65]Chileans obtain their college degree, on average, at the age of 25. A teacher who starts working right after graduation will have ten years of experience by age 35.

are turned into long-term contracts. This is consistent with voucher schools using temporary contracts as probational contracts, and learning about the true skills of the teacher during the one- to two-year probation period.

## 4.4 Facts about Parents

As other studies have also documented (see, for example, McEwan, Urquiola and Vegas (2008)), in the Chilean education system there is considerable school stratification. Table 4.4 shows average household characteristics by type of school attended computed using the 2006 SIMCE dataset. The characteristics considered are average parental education, average household monthly income, fraction of household heads reporting that they do not work and fraction of household heads reporting that they hold a job requiring low skills. Patterns similar to those presented in the table are present among virtually all the household characteristics available in the SIMCE dataset. This table suggests that it is important to understand how parents sort across the municipal and the voucher school sectors. The estimated model is capable of capturing these sorting patterns, suggesting that the motives identified in the model as driving the school choice can be considered reasonable explanations. These motives include parental preference for both consumption and achievement, direct preference for a type of school and a heterogeneous effect of schools on students with different characteristics.

| Household's characteristics | Avg in M | Avg in V | Diff |
|---|---|---|---|
| Avg parents' educ (yrs) | 9.66 | 11.92 | $2.26^{(***)}$ |
| Hh monthly income (CLP) | 169,771 | 312,320 | $142,549^{(***)}$ |
| Hh head not working (frac) | 9.08% | 4.68% | $4.40\%^{(***)}$ |
| Hh head low-skilled job (frac) | 44.50% | 22.21% | $22.30\%^{(***)}$ |

<div align="center">Source: SIMCE 2006. Three stars indicate p-value$< 0.001$.</div>

Finally, plotting raw test scores reveals a gap between municipal and voucher school students' performance. The difference in means is statistically significant and equal to 0.33 standard deviations, which is over one third of the black-white test score gap in the US, and it is larger than the gap between charter and traditional public schools in the US. The estimated model is capable of capturing this gap, and it explains it through a combination of factors: different production technologies in the two schools, different school inputs in the two schools, heterogeneous effects of schools on different students and sorting of parents on both observables and unobservables. Table 30 in the results section shows a regression of test scores on student characteristics and on a voucher school dummy: controlling for observed student characteristics the gap reduces to 0.12 standard deviations. However simulations from the model show that this gap derives entirely from sorting of students based on unobservables.

# 5 Estimation Approach

I use a two-step estimation procedure that addresses the possibility of multiple equilibria. Section 5.1 illustrates how the method accounts for multiple equilibria in estimating the second-stage parameters, and it introduces two assumptions that deal with multiplicity in the estimation of the remaining four stage-one parameters. Sections 5.2 and 5.3 describe the details of the estimations in the two steps. All numerical
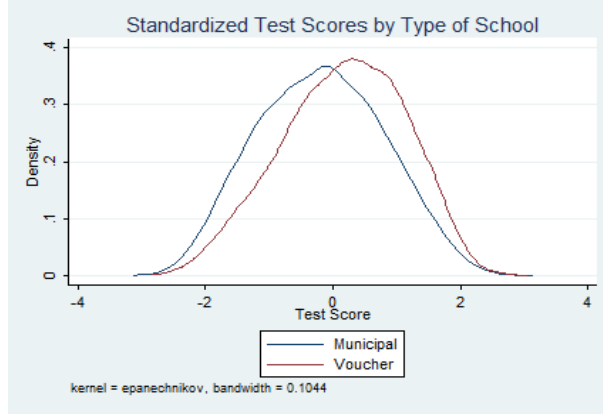
Figure 3: Source: SIMCE 2006.

optimizations are performed using APPSPACK, a derivative-free optimization software using asynchronous parallel pattern search.

## 5.1 Two-Step Procedure

I exploit three key model features to separate the estimation of parents' and potential teachers' parameters from the estimation of the profit function parameters. This separation allows me to account for multiplicity of equilibria in a way similar in spirit to the method proposed by Moro (2003).[66]

In the context of this model, multiple equilibria derive from multiple solutions to the voucher school's maximization problem. Let $\theta_{II} \in \Theta_{II}$ denote the preference and technology parameters of potential teachers and parents that enter the second stage of the model, and let $\theta_I \in \Theta_I$ denote the profit function parameters that enter the first stage of the model. The vectors $\theta_I$ and $\theta_{II}$ have no elements in common, and represent all model parameters: $\theta = [\theta_I \quad \theta_{II}]' \in \Theta$. Given $\theta_{II}$, the cost parameters $(c_1, c_2, c_3)$ and a cost shock drawn from the log-normal distribution with scale parameter $\sigma_{cost}$, more than one solution could exist to the first-stage profit maximization problem. From the point of view of the econometrician, who does not observe the cost shock draw but who knows its density, given $\theta$ there could be more than one density of solutions. Suppose, for the sake of argument, that for all parameter values and cost shocks the profit function admits two maxima, one with high tuition and skill price (H) and the other with low tuition and skill price (L). Suppose that the voucher school commits to a solution selection rule, that is, it commits to choosing L or H before the cost shock is realized. From the point of view of the econometrician, at all parameter values there exist two densities of optimal solutions, one corresponding to each solution selection rule adopted by the voucher school. Formally, let $S_I$ be the following correspondence, yielding the set of densities over the solutions to the first-stage maximization problem for each $\theta$:

$$S_I : \Theta \rightrightarrows \mathcal{F}$$

where $\mathcal{F}$ is the set of probability density functions over $[0, \bar{p}] \times [\delta, \bar{r}]$.[67]

---

[66]Similar methods have been used by Fu (2010) and Fang (2006). See Bisin, Moro, Topa (2011) for a formal discussion of the method in the estimation of models with social interactions. See also Aguirregabiria and Mira (2007).

[67]This set is restricted to the functions that are consistent with the model, in particular with a log-normal cost shock draw

29

In the second stage of the model, the parents and potential teachers observe the realized solution to the school's maximization problem. For each realized first-stage solution $(p, r)$ and second-stage parameters $\theta_{II}$, there exists a unique sorting of parents and potential teachers, i.e., a unique equilibrium action profile. Formally, let $s_{II}$ denote the following function, yielding the unique equilibrium sorting of parents and teachers in the second stage for each $(p, r)$ and $\theta_{II}$:

$$s_{II} : [0, \bar{p}] \times [\delta, \bar{r}] \times \Theta_{II} \rightarrow C_P^{[0,P]} \times C_W^{[0,W]}$$

where $P$ and $W$ are the masses of parents and potential teachers, and $C_P = \{M, V\}$ and $C_W = \{M, V, NT, H\}$ are their choice sets. $C_P^{[0,P]} \times C_W^{[0,W]}$ is the set of functions representing an action profile for parents and potential teachers.[68]

Multiplicity of equilibria in this context refers to the fact that if the profit function admits multiple optimal tuitions and skill prices for some parameters and cost shocks, the voucher school can ex-ante commit to different solution selection rules, each of which, from the point of view of the econometrician, generates a different density over the realized tuition and skill price. Because for each realized solution the second stage of the model generates a different equilibrium sorting of parents and potential teachers, if the model generates different densities over first-stage tuition and skill prices, it also generates different densities over the second-stage equilibrium.

If estimation is performed by minimizing a criterion function $Q(\cdot)$ that represents some measure of distance between the data generated by the model and the sampled data, a different criterion function corresponds to each solution selection rule. Consider, for example, a likelihood criterion function, and consider the example of a profit function with low level (L) and high level (H) optimal tuition and skill prices. If the school commits to selecting L, then observing high tuition is less likely than if the school commits to selecting H. The criterion $Q(\cdot)$, therefore, is specific to the solution selection rule adopted by the voucher school. An econometrician adopting a one-step approach would have to compute a criterion function for each possible solution selection rule, and then find the parameter vector and the solution selection rule that yield the smallest value for the criterion. Letting $s \in \{1, ..., \bar{S}_\theta\}$ denote a solution selection rule in the set of all possible solution selection rules at parameter value $\theta$, the one-step approach can be expressed as:[69]

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \left\{ \min_{s \in \{1, 2, ..., \bar{S}_\theta\}} Q(\theta | s) \right\}.$$

This procedure presents two difficulties. First, the numerical methods commonly used to minimize the criterion function require evaluating it at a large number of parameters. In this context, this requires computing the set of all possible solution selection rules at each parameter value, which is a computationally demanding task. Second, even if this were feasible, computing the criterion function conditional on a solution selection rule, $Q(\theta | s)$, would still be computationally infeasible. This is because one element of the first stage

---

and with the maximization of the specific profit function presented in the model. These functions do not have a closed form, as I explain in the section on the second step of the estimation. Skill price $r$ is restricted to belong to a compact set with arbitrary upper bound $\bar{r}$ to guarantee that a solution to profit maximization exists. This arbitrary upper bound plays no role in the estimation of $\theta_{II}$ proposed in this paper. It could instead have the undesirable consequence of affecting the estimation of $\theta_{II}$ if a one-step approach were adopted.

[68] The model generates a non-stochastic sorting of parents and teachers in the super-populations represented by $[0, W]$ and $[0, P]$. However, only data on a sample from those super-populations are available. This introduces sampling error, which is exploited in the estimation.

[69] For simplicity, I consider a countable number of solution selection rules, but this need not be the case.

solution, skill price, is not directly observed. A solution selection rule determines a joint density over tuition and skill price. In a maximum likelihood context, this joint density computed at the observed tuition serves as a measure of distance between the data generated by the model and the actual data. Skill price, however, is not observed, so that the joint density cannot be computed at an observed skill price. Skill price instead must be integrated out of the joint density.[70] This is demanding because it requires solving for the second-stage equilibrium sorting of parents and teachers for each unobserved skill price in the domain of skill price, over which the integration takes place. Moreover, if there are multiple solution selection rules, the integration must be performed multiple times, once for each joint density of tuition and unobserved skill price implied by each solution selection rule. The goal of the estimation method presented here is to avoid integrating out the unobserved skill, and the intuition is that this is possible if the realized unobserved skill price can be inferred from the data.

The method exploits three key model features. First, the equilibrium of the second stage depends on the profit function parameters $\theta_I$ only through those parameters' effect on the optimal tuition and skill price. Second, the second-stage equilibrium depends on the *realized first-stage solution* and not on the *solution selection rule* adopted by the voucher school. This means that the way that data on parents and teachers are generated is independent of the solution selection rule adopted by the voucher school. Third, for each tuition and skill price and second-stage parameters, the equilibrium of the second stage is unique. The model in Moro (2003) shares these features: in his model there is a set of parameters that affect only the set of feasible equilibria and not the way the data are generated from those equilibria. Those parameters correspond to $\theta_I$ in this paper, which affect the set of solution selection rules available to the voucher school, but not the way the data on parents and teachers are generated from the realized equilibrium.[71] Finally, in Moro's paper an equilibrium identifies a unique data generating process; in this paper, a realized solution identifies a unique process generating data on parents and teachers.

I estimate the second-stage parameters separately from the first-stage parameters, and this allows me to simulate the second-stage equilibrium only once, at the actual realized first-stage solution. In the data, tuition $p$ is observed, and it is equal to the tuition cap in all markets, so that all observed variation in tuition payments derives from the exogenous fellowship formulae. This suggests that the second-stage parameters, which include the parameters driving the price sensitivity of parents, can be estimated by conditioning on the observed (and constant across markets) $p$. Price sensitivity is identified off of exogenous fellowship variation.

The skill price, however, is not directly observed: what is directly observed is the distribution of accepted wages in the voucher school, which are the product of $r$ and individual skills. Suppose that the skill prices that generated the observed distribution of wages in all markets, $r = \{r_1, ..., r_D\}$, and the parameters $\theta_{II}$ were both identified from the observation of all available data: tuition payments, potential teachers' and parents' choices, test scores and accepted wages, i.e., suppose that there were only one vector $r$ and only one vector $\theta_{II}$ capable of generating (through the model) the true distribution of the observed data. Formally, denote by $P$ the true distribution of the observed data $T$. Denote by $P^M = \{P_\phi : \phi = (\theta_{II}, r) \in \Theta_{II} \times [\delta, \bar{r}]\}$ the distribution of the observed data implied by the model. Denoting the identified set by:

---

[70]See Appendix E, in which I use the example of the likelihood function to illustrate why unobserved skill price must be integrated out.

[71]The model in Moro (2003) is non-stochastic: it generates deterministic equilibria. In this paper, the model generates (potentially multiple) stochastic solutions to the profit function maximization: it yields (potentially multiple) densities over tuition and skill price.

$$\Phi_0(P) = \{\phi \in \Theta_{II} \times [\delta, \bar{r}] : P_\phi = P\}$$

the identification condition can be expressed by requiring that $\Phi_0(P)$ be a singleton.

From the point of view of parents and potential teachers, $r$ can be treated as a vector of parameters, as it is invariant to their own behavior. For each value of $\theta_{II}$ and $r$, and conditional on the true observed $p$, there is one and only one optimal sorting of parents and teachers, so that a criterion that estimates $\theta_{II}$ and $r$ based on the comparison between the observed and the model-generated sorting of parents and teachers is indeed a function and not a correspondence. Under the assumption of identification, the following procedure can be used to estimate the second-stage parameters:[72]

$$\min_{\theta_{II}, r} Q(\theta_{II}, r | p). \tag{9}$$

Intuitively, when the identification condition is satisfied the $r$ that generated the data is identified and the econometrician does not need to integrate out $r$ and compute the second-stage equilibrium for all possible $r$ that could have generated the data. The data and the model's restrictions reveal which first-stage realized solution underlies the data generating process.

In the first step of the estimation I estimate the $\theta_{II}$ parameters along with the unobserved component of the equilibrium, $r$, by minimizing a criterion function following the procedure in $9$.[73] Section 5.2 illustrates the estimation approach used in this step of the estimation, and it discusses the validity of the identification assumption.

The second step estimates the profit function parameters taking as given the outputs of the first step: the parameter estimates $\hat{\theta}_{II}$ and the estimated optimal skill prices $\hat{r}$ (one per market). The skill prices are treated as data, and together with the observed tuition they serve as the endogenous variables of a maximum likelihood method. Intuitively, $\hat{r}$ aide in identification of $\theta_I$ because they provide more observations on the endogenous variables, and they also allow me to avoid integrating the likelihood over all possible $r$'s in the domain. They do not, however, solve the issue of potentially multiple solution selection rules. I make two additional assumptions: first, I assume that I know the solution selection rule adopted by each school. This implies that I am able to build the correct likelihood of observed tuition and estimated skill price, which varies depending on the solution selection rule.[74] Second, I make an assumption that is needed because of the estimation technique that must be adopted to estimate $\theta_I$: even if I assume that the solution selection rule is known, it is not possible to analytically derive the density of tuition and skills and use it to analytically derive the likelihood function. This is because the profit function, which embeds the equilibrium sorting of parents and potential teachers, lacks a closed form. The technique that I use, Nonparametric Simulated Maximum Likelihood (Laroque and Salanié (1989), Fermanian and Salanié (2004)), requires nonparametrically estimating the density of tuition and skills by simulating the optimal tuition and skill prices for a large number of cost shock draws at each parameter value. I assume that if at a certain parameter

---

[72]This $Q(\cdot)$ function is a different function from the $Q(\cdot)$ function used before.

[73]My first step recovers the equilibrium while estimating a subset of the parameters of the model ($\theta_{II}$). In Moro's paper, all the model parameters are recovered in the second step: the first step estimates only the equilibrium, and the second step recovers the model parameters consistent with the estimated equilibrium.

[74]Notice that the solution selection rules adopted by the different schools need not be the same ones. Refer to Appendix E. I am currently working on an estimation approach based on first-order conditions that would not require assuming that the correct solution selection is known.

iteration and cost shock draw there are more than one solution to the profit function maximization, the solution that I numerically obtain is the solution that is being selected by the schools.

Finally, notice that these two assumptions, and in particular the assumption that the solution selection rule is known, are not needed to estimate the parameters of the second stage of the model, which represent almost all of the parameters in the model. They are estimated in the first step of the estimation, which is agnostic regarding which solution selection rule the school adopted.

## 5.2   Step One

The goal of this step is to estimate the preference, technology, and type distribution parameters of parents and potential teachers and the skill prices of teaching skills in each market.[75] Estimation is by the Method of Simulated Moments (MSM) (McFadden, 1989). The method minimizes the distance between observed outcomes and outcomes predicted by the model. The outcomes are occupational choices of college graduates, school choices of parents, wages of college graduates, test scores of children and students' fellowship amounts. A list of the moments used can be found in Appendix F.

In this section I explain how I simulate the outcomes, what type of moments I use and how I combine different datasets.

### 5.2.1   Simulating Outcomes

There are two types of outcomes: sector choices, which I refer to as *unconditional* outcomes, and what I refer to as *conditional outcomes*, for example, accepted wages conditional on a chosen sector and observed test scores conditional on a chosen school. *Conditional* refers to the fact that the outcome is conditional on an occupational sector choice or on a school choice, which are endogenous variables in the model. Because only unconditional outcomes can serve as the basis for MSM, I transform conditional outcomes into unconditional outcomes by multiplying them by a sector dummy. Formally, let $y_i$ denote an actual outcome for individual $i$. To distinguish conditional from unconditional outcomes, I add a $c$ superscript when referring to a conditional outcome, $y_i^c$. Let $\Omega_i \times \{1, ..., L\}$ denote the state space of individual $i$ and let $(\omega_i, l) \in \Omega_i \times \{1, ..., L\}$ denote an element of the state space. The state space is composed of the observed exogenous variables $\omega_i$ and of the unobserved individual's type $l$. The type is listed separately for convenience. An element of the vector $\Omega_i$ includes, for example, degrees, age, gender, etc. Let $\hat{y}_i(\omega_i, \theta)$ denote the outcome predicted by the model for individual $i$ with exogenous variables $\omega_i$. The predicted sector choice, or unconditional outcome, for individual $i$ is replaced by the simulator:

$$\tilde{\hat{y}}_i(\omega_i, \theta) = \frac{1}{S} \sum_{s=1}^{S} \sum_{l=1}^{L} Pr(l|\theta) \hat{y}_i(\omega_i, l, s, \theta)$$

obtained by drawing $S$ simulated shocks from the model's shock distribution under parameter $\theta$ and using the model to simulate optimal behavior, i.e., the sector choice for each individual, simulation, and type: $\hat{y}_i(\omega_i, l, s, \theta)$. The simulated choices are then averaged across simulations and types, resulting in a simulated fraction for each individual.

---

[75]The number of model parameters estimated in this step is 129.

The predicted conditional outcome for individual $i$ conditional on being in sector $J$ is transformed into an unconditional outcome by means of a sector choice dummy. The outcome is simulated according to:[76]

$$\tilde{y}_i^c(\omega_i, \theta) = \frac{1}{S} \sum_{s=1}^{S} \sum_{l=1}^{L} Pr_{is}(l|\theta) \hat{y}_i(\omega_i, l, s, \theta) \hat{D}^J(\omega_i, l, s, \theta)$$

where the dummy variable $\hat{D}^J(\omega_i, l, s, \theta)$ is equal to one if individual $i$ of type $l$ at simulation $s$ is predicted to choose sector $J$.[77]

### 5.2.2 Moment Conditions

To exploit exogenous data variation, I build moments constrained by varying values of the exogenous variables. Consider a moment condition conditional on some value of the exogenous variables, for example, the difference between actual, $w_i$, and predicted wage, $\hat{w}_i$, in the municipal school by gender and age. From such conditional moment conditions I derive unconditional moment conditions that are suitable as a basis for the method of moments estimation. In the wage example, I use the product of $(w_i - \hat{w}_i)$ and a dummy for whether the individual is of a certain age and gender, $I(age_i = a\tilde{g}e, fem_i = f\tilde{e}m)$.[78] Moreover, some outcomes are available only for certain types of individuals: for example, occupational choices are available only for potential teachers, and test scores are available only for students. I constrain the moments to the individuals for whom an outcome exists.[79]

Formally, let $I_i(\omega_i, y_i \text{ exists})$ be a dummy equal to one if individual $i$'s exogenous variables have value $\omega_i$ and outcome $y_i$ exists, and equal to zero otherwise. The population moment condition based on outcome $y_i$ is:

$$E\left[(y_i - \hat{y}_i(\omega_i, \theta))I_i(\omega_i, y_i \text{ exists})\right].$$

### 5.2.3 Combining Multiple Data Sources

The population of primary and secondary school students is larger than the population of potential teachers. Accordingly, the size of the sample of students (100,000) is larger than the size of the sample of potential teachers (6,715). I combine multiple samples by treating them as multiple strata. I adjust the objective function and parameter standard errors to account for the stratified sampling design and for the relative sizes of the datasets and of the relative populations of reference. Asymptotic properties are derived by letting the size of each dataset go to infinity at the same rate across datasets (Bhattacharya, 2005).

---

[76]Refer to Appendix G to see how these simulators are corrected to account for choice-based sampling in the ELD sample.

[77]Because the sector choice is an endogenous outcome of the model, the simulated dummy depends on the parameters and it introduces discontinuity in the objective function used in estimation. To guarantee asymptotic normality I assume the sufficient condition for smoothness of the limiting objective function of the MSM presented in Theorem 7.1 in Newey and McFadden (1994). Under this assumption, the results of Theorem 7.1 apply and asymptotic normality is guaranteed. Given the complexity of the model, this assumption is imposed instead of derived from the primitives of the model.

[78]In the framework of GMM, Newey (1990,1993) shows that the optimal choice of unconditional moment condition is obtained by multiplying the difference between the predicted and actual outcome by an optimal instrument. The latter is obtained as a matrix function of exogenous data and parameters which involves the derivative of the conditional moment with respect to the parameters and the inverse of the variance of the conditional moments. Adopting such an instrument would be prohibitively computationally intensive in this setting. The method used in this paper, on the other hand, is not computationally intensive and it is easy to code. A similar technique has been adopted by Behrman, Tincani, Todd and Wolpin (2011).

[79]The outcome is set to a constant (arbitrarily set to zero), which is independent of the parameters if the outcome does not exist for that individual.

Consider the population moment condition based on outcome $y_i$:

$$E\left[(y_i - \hat{y}_i(\omega_i, \theta))I_i(\omega_i, y_i \text{ exists})\right]$$

and suppose that there are $M$ moment conditions $\{m_i^1, ..., m_i^M\}$ with $m_i^m = (y_i^m - \hat{y}_i^m(\omega_i^m, \theta))I_i(\omega_i^m, y_i^m \text{ exists})$.[80] Let $m_i$ be a vector that stacks all moment conditions for individual $i$. Assume that the population is divided in two strata: the stratum of students, with mass $H_A$, and the stratum of college graduates, with mass $H_B$. The $M$ population moment conditions are:

$$H_A E_A[m_i] + H_B E_B[m_i]$$

where $E_A[\cdot]$ and $E_B[\cdot]$ represent within-stratum expectations.

Let $n_A$ be the sample size of students and $n_B$ be the sample size of potential teachers, and let $m_i(\theta)$ be the $M \times 1$ vector of empirical moment conditions computed at a parameter value $\theta$. The sample analog of the population moment conditions is:

$$H_A \frac{1}{n_A} \sum_{i \in A} w_i m_i(\theta) + H_B \frac{1}{n_B} \sum_{i \in B} w_i m_i(\theta)$$

where $w_i$ are weights provided with the datasets that are used to reweight the sample back to random sampling proportions, and that are normalized to sum to $n_A$ and $n_B$.[81] Let $n = n_A + n_B$ and pre-multiply the sample moments by $\frac{n}{n}$. Denote the vector of empirical moments based on a sample of size $n$ by $m_n(\theta)$:

$$m_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} (H_A a_A w_i m_i(\theta) I(i \in A) + H_B a_B w_i m_i(\theta) I(i \in B))$$

where $a_A = \frac{n}{n_A}$, $a_B = \frac{n}{n_B}$ and $I(\cdot)$ is an indicator function equal to one if the expression in parenthesis is true.

### 5.2.4 Asymptotic Properties of the Estimator

The method of simulated moments finds the vector $\theta$ that minimizes the weighted distance of the empirical moment conditions from zero:

$$\hat{\theta}_{MSM} = arg \min_{\theta} m_n(\theta)' W_n m_n(\theta) \tag{10}$$

where $W_N$ is an $M \times M$ symmetric positive definite weighting matrix such that as $n \to \infty$, $W_n \to W$ in probability with $W$ symmetric and positive definite.

To derive the asymptotic properties of the estimator, I let $n_A, n_B \to \infty$ with $\frac{n_A}{n} \to a_A < \infty$ and $\frac{n_B}{n} \to a_B < \infty$ as in Bhattacharaya (2005), who studies the asymptotic properties of the generalized method of moments with a stratified sample. The MSM estimator defined in (10) is consistent and asymptotically normal:

$$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, Q)$$

---

[80]Notice that the same outcome $y_i$ constrained to different values of the exogenous variables $\omega_i$ defines different moment conditions.

[81]For SIMCE observations the weights are all equal to one because the SIMCE sample is a simple random sample.

with $Q = (\Gamma' W_n \Gamma)^{-1} \Gamma' W_n V W_n \Gamma (\Gamma' W_n \Gamma)^{-1}$ and $\Gamma = E[\frac{\partial m(\theta)}{\partial \theta'}]$. $V$ is the variance covariance matrix of the moment vector.[82]

To estimate consistently the asymptotic variance of the estimator I substitute $V$ with a consistent estimate $\hat{V}$ computed at $\hat{\theta}_{MSM}$. The estimator includes a stratum correction that accounts for the sampling design.[83] The estimator of the variance covariance matrix is:

$$
\hat{V} = \sum_{i \in A} \left( \frac{H_A}{n_A} w_i \right)^2 m_i(\hat{\theta}_{MSM}) m_i(\hat{\theta}_{MSM})' + \sum_{i \in B} \left( \frac{H_B}{n_B} w_i \right)^2 m_i(\hat{\theta}_{MSM}) m_i(\hat{\theta}_{MSM})' \tag{11}
$$

$$
- \frac{1}{n_A} \left( \sum_{i \in A} \frac{H_A}{n_A} w_i m_i(\hat{\theta}_{MSM}) \right) \left( \sum_{i \in A} \frac{H_A}{n_A} w_i m_i(\hat{\theta}_{MSM}) \right)'
$$

$$
- \frac{1}{n_B} \left( \sum_{i \in B} \frac{H_B}{n_B} w_i m_i(\hat{\theta}_{MSM}) \right) \left( \sum_{i \in B} \frac{H_B}{n_B} w_i m_i(\hat{\theta}_{MSM}) \right)'
$$

where $m_i(\hat{\theta}_{MSM})$ is the $M \times 1$ vector of individual level moment conditions computed at $\hat{\theta}_{MSM}$. To estimate consistently the matrix of moments' partial derivatives, I use:

$$
\hat{\Gamma} = H_A \frac{1}{n_A} \sum_{i \in A}^{N} w_i \frac{\partial m_i}{\partial \theta'} |_{\hat{\theta}_{MSM}} + H_B \frac{1}{n_B} \sum_{i \in B}^{N} w_i \frac{\partial m_i}{\partial \theta'} |_{\hat{\theta}_{MSM}}
$$

where the differentiation is numerical. Letting $\Delta_t$ denote a vector of the same size as the parameter vector with zeros everywhere and $\delta > 0$ as its $t^{th}$ element, the derivative of the $m^{th}$ element of $m_i(\theta)$ with respect to the $t^{th}$ element of $\theta$ is computed as:

$$
\frac{\partial \hat{m}_i^m(\theta)}{\partial \theta_t} |_{\theta = \hat{\theta}_{MSM}} = \frac{\hat{m}_i^m(\theta + \Delta_t) - \hat{m}_i^m(\theta)}{\delta} |_{\theta = \hat{\theta}_{MSM}}.
$$

### 5.2.5 Identification

**Identification of $r$ and of the Teacher Labor Supply**

This section discusses the separate identification of $r$ from $\theta_{II}$, which is assumed in the two-step estimation approach. Notice that the parameters that govern the choices of potential teachers, which are a subset of $\theta_{II}$, are the parameters that govern the teacher labor supply, so that identification of those parameters permits identification of the teacher labor supply. The identification strategy used to identify the teacher labor supply follows closely Heckman and Sedlacek (1985).

---

[82]The optimal weighting matrix is the inverse of the variance covariance matrix of the moment conditions, $W_n^* = V^{-1}$. The asymptotic variance reduces to $(\Gamma' V^{-1} \Gamma)^{-1}$ when the optimal weighting matrix is used. Unfortunately I cannot adopt the optimal weighting matrix because the variance covariance matrix is a high order sparse matrix that cannot be numerically inverted in a correct way. The inverse of the variance covariance matrix must be obtained to compute the standard errors of the efficient MSM estimator. This negative result is standard in numerical methods. I adopt a weighting matrix that contains the variances of the moments on the main diagonal and zeros elsewhere.

[83]The correction term is derived and discussed in Bhattacharya (2005). It adjusts for the over-estimation that would result if one did not account for the stratified sampling design. Intuitively, ignoring the fact that observations come from two separate strata would over-estimate the between-strata variances.

The model of potential teacher labor supply decisions is an extended Roy model (Roy (1951)) of self-selection into occupations with log-normal skills. Identification in this class of models has been proved formally by, for example, Heckman and Honoré (1990). Typically, the log-price of skills is not separately identified from the constant in the log-skills. In this context, consider the voucher school wage offer, which is the product of skill price and teaching skills:

$$w_i^V = r exp(\alpha_{0V}(l_i) + \alpha'_{1V} Z_i + \epsilon_i^V).$$

The log-wage offer is:

$$ln(w_i^V) = ln(r) + \alpha_{0V}(l_i) + \alpha'_{1V} Z_i + \epsilon_i^V.$$

The model of self-selection allows to estimate unbiasedly the $\alpha$ parameters, up to the constant $ln(r)$. Assuming that the shock distribution is the same across markets, the distributional assumptions of the Roy model correct for the sample selection bias, which is due to the self-selection of individuals into occupations. To separately identify $ln(r)$ from the intercept in the log-skills, $\alpha_{0V}(l_i)$, I make the identifying assumption that the parameters affecting the skill production technology, $\alpha_V = [\alpha_{0V}(l_i) \quad \alpha_{1V}]'$, do not vary by market, whereas the skill prices vary by market. Heckman and Sedlacek (1985) refer to this assumption as the "proportionality hypothesis". As a consequence of this assumption, within-market variation in wages and labor supply decisions can be used to identify $\alpha_V$, whereas across market variation can be used to identify the skill prices $r$, after the normalization of the skill price in one market. In Heckman and Sedlacek (1985), the role of different markets is played by different years: the constant in the log-skills is assumed to be constant across years, but skill prices are assumed to vary by year.[84]

Estimates of the preference and technology parameters of potential teachers produce an estimate of the aggregate supply of teachers and of teaching skills within each market. Notice that skills are measured in terms of the normalizing constant used to normalize skill price in one market. As I argue below, the counterfactuals considered in this paper are invariant to this normalization, because they focus on the effect of a policy on cognitive achievement. The effect of a change in the supply of skills on cognitive achievement is identified and invariant to the choice of normalizing constant.

**Implications of the Normalization of the Skill Price in One Market**

All counterfactuals considered in this paper are invariant to the choice of normalization of the skill price in one market. An implication of normalization, however, is that some parameters are identified only up to the normalization. In this section I caution about the interpretation of the parameters affected by the normalization, and I explain why the counterfactuals are invariant to the normalization.

I normalize the rental rate in market one to a constant $c$. After the normalization, the following functions of the parameters are identified:

- $\alpha_{0V}(l) - log(c)$, $\quad \forall l$: teachers' type-specific constant in the log-teaching skills minus log of the normalizing constant $c$. This is equivalent to saying that teaching skills are identified only up to scale, i.e., only their ratio to $c$ is identified. Letting $\tilde{\bar{s}}^J$ denote the true mean skills in sector $J$, only $\bar{s}^J = \frac{\tilde{\bar{s}}^J}{c}$ is identified. This means that the cardinal value of the estimated skills is meaningless, because it depends on the choice of $c$; however, comparisons between different people's skills are meaningful, as the ratio

---

[84]See note 17 in Heckman and Sedlacek (1985), where the normalization is discussed.

of their estimated skills, which are subject to the normalization, is the same as the ratio of the true underlying skills. To see why skills are identified only up to scale, consider the market in which the skill price has been normalized and consider two alternative normalizations: $c$ and $c'$. Wages identify the product of true skills and true skill price. When true skill price is arbitrarily set to $c$, normalized skills are obtained as the ratio of observed wages and $c$. When the skill price is normalized to $c'$, normalized skills are obtained as the ratio of observed wages and $c'$. Intuitively, wages are observed and the product of normalized skill price and normalized skills must always be equal to observed wages.

- $cr^d, \forall d \neq 1$: skill prices in all other markets multiplied by the normalizing constant $c$. Because the parameters that govern skills are the same across markets, the normalization of the skill price in one market has implications for the identification of skills in all other markets: in all markets only the ratio of skills to $c$ is identified. Since wages in all markets identify the product of skill price and skills, only the product of skill price and $c$ is identified.

- $\beta_{1M}(l) = c\tilde{\beta}_{1M}(l), \beta_{1V} = c\tilde{\beta}_{1V}(l), \quad \forall l.$[85] Letting $\tilde{\beta}_{1M}, \tilde{\beta}_{1V}$ denote the effects of the true (and unidentified) mean skills on achievement, their product with the normalizing constant $c$ is identified. To see why, notice that from the test scores the overall effect of teachers $\tilde{\beta}_{1J}(l)\tilde{s}^J, \quad J = M, V$ is identified. However, because only the ratio $\frac{\tilde{s}^J}{c}$ is identified from the labor supply side of the model, the overall effect of teachers identifies only the product $c\tilde{\beta}_{1J}$, i.e., the product between the true effect of skills on achievement and the normalizing constant $c$. This implies that the estimated parameters $\beta_{1M}, \beta_{1V}$ should not be interpreted as the productivity of teaching skills, because they depend on the choice of $c$. In particular, they cannot be compared with the coefficients on the other inputs. However, it is meaningful to compare the relative magnitudes of $\beta_{1M}$ and $\beta_{1V}$ as their ratio is equal to the ratio of the true effects of skills on achievement.[86] Finally, notice that the normalization does not affect the counterfactuals, because the overall effect of teachers on achievement is identified. Consider two scenarios with two different normalizing constants: $c, c'$. Under $c$, $\beta_{1J} = c\tilde{s}^J$ is identified, under $c'$, $\beta'_{1J} = c'\tilde{s}^J$ is identified. The overall effect on achievement under both normalizations is the same: $\tilde{\beta}_{1J}\tilde{s}^J = c\tilde{\beta}_{1J}\frac{\tilde{s}^J}{c} = c'\tilde{\beta}_{1J}\frac{\tilde{s}^J}{c'}$. Consider a counterfactual experiment that varies the supply of skills from $\tilde{s}_0$ to $\tilde{s}_1$. Under $c$, the effect of teacher skills on test scores varies by: $\frac{\tilde{s}_1}{c}\tilde{\beta}_1 c - \frac{\tilde{s}_0}{c}\tilde{\beta}_1 c = \tilde{s}_1\tilde{\beta}_1 - \tilde{s}_0\tilde{\beta}_1$. Under $c'$, the effect of teacher skills on test scores varies by the same amount: $\frac{\tilde{s}_1}{c'}\tilde{\beta}_1 c' - \frac{\tilde{s}_0}{c'}\tilde{\beta}_1 c' = \tilde{s}_1\tilde{\beta}_1 - \tilde{s}_0\tilde{\beta}_1$.

**Identification of the Demand for Private Education: Identifying Price Sensitivity of Parents by Exploiting Exogenous Fellowship Variation**

The parameter $\tau$, the weight on consumption relative to child achievement in the utility of parents, and the concavity of the utility from consumption govern how parents change their school choice as tuition changes. The parameter $\tau$ can be identified off of exogenous tuition variation. Within each market in the model individuals are subject to the same tuition charged by the school, and variation in the tuition payments that

---

[85]In the following argument I drop the dependence of the parameters on type for ease of exposition.

[86]Estimates show that teacher skills are more productive in the municipal than in the voucher school. This is consistent with some features of the data. For example, the 2005 ELD teacher survey shows that teachers in municipal schools are more likely to report that the school provides them with sufficient instructional material (63% versus 50%), that the school building provides a comfortable and adequate environment for learning (63% versus 51%), that they have enough time to prepare classes (28% vs. 21%) and that the relationships among teachers in the school are good (80% versus 71%). This can be interpreted as evidence that the school environment (i.e., other school inputs), which I assume to be policy invariant, interacts with a teacher's skills in a way that makes the teacher more or less productive.

they must make to attend the private school derives from exogenous fellowship variation across individuals. I exploit this variation to identify $\tau$. Notice that because all schools are observed to price at the legal cap, across-market variation can be exploited too because, as within markets, it derives entirely from fellowship variation. Exogenous price variation, together with discrete choice estimation on a micro-dataset, allows me to consistently estimate the parameters of the demand for private education. Notice that the exogenous fellowship variation must be combined with the model's structure to identify the parameters of the demand for schooling because of lack of exclusion restrictions, i.e. variables entering the fellowship formula and not entering the cognitive achievement nor the parental direct preference for a type of school.[87] Finally, notice that the parameters of the fellowship formula, which I estimate along with the other parameters of the model, are identified from observations of fellowship amounts and students' characteristics, without the need to appeal to any model restriction. The fact that fellowship amounts are observed only for individuals who choose the voucher school does not introduce a selection bias issue because of the exogeneity of the fellowship formula, which depends only on observed parental characteristics.

## 5.3   Step Two

This step estimates the voucher school's cost parameters $c_1, c_2, c_3, \sigma_{cost}$. Estimation is by Nonparametric Simulated Maximum Likelihood (NPSML) (Laroque and Salanié (1989), Fermanian and Salanié (2004)). The likelihood function does not have a closed-form expression: the endogenous variables $(p, r)$ are an unknown function of the random cost shock, as they are obtained as a solution to the maximization of the profit function, which does not admit a closed form. The profit function does not admit a closed form because it includes the normal cumulative density functions that characterize the demand for enrollment and supply of teachers. Therefore, although the cost shock is known to be log-normally distributed, the density of the variables $(p, r)$ and hence the likelihood function cannot be analytically derived. The NPSML method approximates the unknown likelihood function with a kernel-based nonparametric estimator based on simulations of the endogenous variables, which in this context are the solution to the first stage of the model, $p$ and $r$. Under regularity conditions the estimator is consistent, asymptotically normal and asymptotically efficient when the number of simulations and observations goes to infinity and the bandwidth to zero.[88]

As before, $\theta_I = \begin{bmatrix} c_1 & c_2 & c_3 & \sigma_{cost} \end{bmatrix}'$ denotes the vector of parameters to be estimated in the second step (i.e., the parameters from the first stage of the model), and $\hat{\theta}_{II}$ the vector of estimates of $\theta_{II}$ obtained in the first step. In the model, the optimal prices $(p^d, r^d)$ in each market depend on the demand for enrollment and supply of teachers that the voucher school faces in the market. The demand and supply functions in each market can be fully derived from the second-stage parameters $\theta_{II}$ and from the market-specific distribution of the exogenous variables of students and college graduates. This distribution, which is not

---

[87]Intuitively, exogenous fellowship variation along with the model's structure identify $\tau$ because the exogenous fellowship formula provides information on the tuition payments that different families would have to make if they chose the voucher school, and the model's structure provides information on the different test scores that their children would have in the two schools, so that the structure and the observation of parents' school choice, student's achievement, tuition payments and household characteristics including income provide information on how parents trade off consumption for their child's achievement. Suppose, for example, that there exists only one unobserved type and that two families differ in such a way that the tuition payments they must make in the voucher school are considerably different, but the test scores that their children would obtain in the two schools are not. Suppose also that their children's test scores are higher in the private school. If both families are observed to choose the private school, this suggests a lower weight on consumption than if the family that has to pay higher tuition chooses the municipal school.

[88]The estimate of the asymptotic variance, however, must be corrected for the fact that the estimation is in two steps.

specified parametrically, is characterized by an infinite dimensional vector $v^d$.[89] The private school uses knowledge of $v^d$ and $\theta_{II}$ to derive the demand for enrollment and supply of teachers, which are the outcome of parents' and potential teachers' optimal behaviors. The relation between exogenous and endogenous variables of the first stage of the model can be represented as:

$$[p^d \quad r^d]' = \rho(v^d, \theta_I, \epsilon^d_{cost}; \theta_{II}) \quad d = 1, ..., D$$

where $[p^d \quad r^d]$ are the endogenous variables and $\epsilon^d_{cost}$ is the cost shock. Imagine obtaining a sample of markets, and suppose that for each sampled market, a sample of college graduates and students within the market is available. Denote by $X^d = \{x_1, ..., x_i, ..., x_{NS}\}_{i \in d}$ the within-market sample of students and by $Q^d = \{q_1, ..., q_i, ..., q_{NC}\}|_{i \in d}$ the within-market sample of college graduates.[90] Let $\hat{r}$ be the vector of skill prices, one for each market, estimated in the first step.[91] Let $(p^d, \hat{r}^d, X^d, Q^d)_{d=1,...,D}$ be an independently and identically distributed sample of markets.[92] The true log-likelihood is:

$$L_D(\theta_I) = \sum_{d=1}^{D} \ln l^d(\theta_I)$$

where $l^d(\theta_I)$ is the density of $(p, r)$ computed at the observed values $(p^d, \hat{r}^d)$ conditional on the exogenous characteristics in the market, on $\theta_I$ and on $\theta_{II}$: $l^d(\theta_I) = f(p, r|\theta_I, X^d, Q^d; \theta_{II})|_{p^d, \hat{r}^d}$. Because $l^d(\theta_I)$ cannot be computed in a closed form, I approximate it using a kernel estimator based on an i.i.d. simulated sample $(\epsilon^{ds}_{cost})_{s=1,...,S}$ of draws from the log-normal distribution of $\epsilon_{cost}$.

Denote $[p^{ds}(\theta_I) \quad \hat{r}^{ds}(\theta_I)]'$ by $\rho(v^d, \theta_I, \epsilon^{ds}_{cost}; \theta_{II})$, where the $s$ superscript means that for every simulated draw $s$, the optimal tuition and skill price are derived as a solution to the model's first stage. I estimate the likelihood $l^d(\theta_I)$ by:

$$\tilde{l}_S(p^d, \hat{r}^d|X^d, Q^d, \theta_I; \theta_{II}) = \tilde{l}^d_S(\theta_I) = \frac{1}{Sh^2} \sum_{s=1}^{S} \mathcal{K}\left(\frac{[p^d - p^{ds}(\theta_I) \quad \hat{r}^d - \hat{r}^{ds}(\theta_I)]'}{h}\right)$$

where $\mathcal{K}(\cdot)$ is the bivariate normal kernel and $h$ is the optimal bandwidth that minimizes the approximate Integrated Mean Squared Error, and it is such that $h \to 0$ as $S \to \infty$.

The simulated log-likelihood is

$$\tilde{L}_{DS}(\theta_I) = \sum_{d=1}^{D} \ln \tilde{l}^d_S(\theta_I)$$

and the NPSML estimator is defined as the global maximum of $\tilde{L}_{DS}(\theta_I)$:

$$\hat{\theta}_I(D, S) = \arg \max_{\theta_I \in \Theta_I} \tilde{L}_{DS}(\theta_I).$$

where $\Theta_I$ is assumed to be compact. Under regularity conditions, $\hat{\theta}_I(D, S)$ is asymptotically normal and

---

[89]One could parameterize the distribution of exogenous variables to reduce the dimension of $v^d$.

[90]The distributions of characteristics within each market are non-parametrically identified from these observations.

[91]For now, treat $\hat{r}$ as if it were data. A discussion of how the uncertainty over $r$ affects the estimation is presented in Appendix H.

[92]Think of markets as strata. I obtain an i.i.d. sample of strata.

asymptotically efficient:[93]

$$\sqrt{D}(\hat{\theta}_I(D,S) - \theta_{I,0}) \underset{S,D\to\infty}{\Rightarrow} N(0,\Omega),$$

where $\Omega$ is the asymptotic variance-covariance matrix of the exact maximum likelihood estimator:

$$\Omega = \left(-E\left[\frac{\partial^2 L_D(\theta_{I,0})}{\partial\theta_2 \partial\theta_I'}\right]\right)^{-1} E\left[\frac{\partial L_D(\theta_{I,0})}{\partial\theta_I}\frac{\partial L_D(\theta_{I,0})}{\partial\theta_I'}\right]\left(-E\left[\frac{\partial^I L_D(\theta_{I,0})}{\partial\theta_I \partial\theta_I'}\right]\right)^{-1}. \tag{12}$$

The variance in (12) assumes that the true values of the step one parameters are known, and that the true values of the endogenous skill price are observed. This variance needs to be adjusted for the fact that both $\theta_{II}$ and $r$ are estimated. Refer to Appendix H for a description of the adjustments to the variance covariance matrix.

### 5.3.1 Identification

The parameters of the profit function of the voucher school are identified off of a combination of exogenous data variation and model restrictions. The two key exogenous sources of identification are exogenous across-market variation in the conditions of the labor markets in the municipal and non-teaching sectors, and the exogenous legal cap on tuition. The first identification source is used to exogenously shift the labor supply in order to identify the labor demand. The second identification source is used to break the simultaneity deriving from the fact that the optimal price of skills depends on the optimal tuition.

The voucher school is a monopolist on the output market, and a monopsonist on the input market. The first order conditions of the profit maximization yield two optimal pricing functions, which I express in a stylized way as:

$$
\begin{aligned}
p^* &= f(DV^*, SV^*)\\
r^* &= g(DV^*, SV^*, p^*).
\end{aligned}
$$

First, notice that optimal tuition $p^*$ does not depend on optimal skill price $r^*$ because of the assumption that teachers do not care about the student body. Intuitively, the overall wage bill, and hence the school's costs, is independent of the student body that a given level of tuition attracts, because the supply of teachers is invariant to the student body. Optimal skill price $r^*$, on the other hand, depends on optimal tuition $p^*$ because who the teachers are affects the demand for private education, and hence the revenues of the school.

Second, all schools are observed to charge a tuition equal to the cap $\bar{p}$, so that the optimal pricing functions can be rewritten as:

$$
\begin{aligned}
p^* &= \bar{p}\\
r^* &= g(DV^*, SV^*, \bar{p}).
\end{aligned}
$$

---

[93]Let the number of individuals within each stratum go to infinity to non-parametrically identify $v^d$, i.e., the distribution of exogenous individual characteristics within each market.

From the estimation in the first step, the parameters that govern the demand for enrollment $DV(p, r)$ and the supply of teachers $SV(r)$ are known.[94] The parameters that enter the optimal tuition function are the parameters of the marginal cost function of the monopolist, whereas the parameters that enter the optimal skill price function are the parameters that govern the marginal revenue product of labor, i.e. the inverse derived demand for labor. Explicitly solving for the first order conditions of the profit function reveals that the same parameters enter both pricing functions. My identification strategy recovers these parameters by identifying the parameters of the optimal skill pricing function. Estimation of the optimal tuition pricing function requires more data variation than the observed data variation: tuition is never observed to vary across markets because the tuition cap is always binding. This information can yield, at most, some bounds on the parameters (optimal tuition is above the observed cap). The model structure maps the parameters of the inverse derived demand for labor into the deep parameters of the profit function.

First, notice that the fact that the tuition cap is binding breaks the simultaneity problem, so that one does not need to plug $p^* = f(DV^*, SV^*)$ into the skill price function. As a result, after plugging the constant tuition cap $\bar{p}$ into the skill price function, the parameters that relate $DV^*$ and $SV^*$ to $r^*$ are in fact those that affect the derived labor demand. To trace the demand for labor, one needs to observe exogenous shifts to the supply of labor. These exogenous shifts are given by the exogenous across-market skill price variation in the non-teaching sector and the exogenous municipal wage regional adjustments. Intuitively, these exogenous changes in non-voucher-school labor-market conditions work as instruments, in that they do not affect the demand for labor but they affect the supply of labor to the voucher school.[95]

### 5.3.2 Role of the Normalization of r in One Market

First, notice that the normalization does not affect the units in which the wage bill in the profit function is expressed: the profit function is expressed in Chilean pesos regardless of the constant used to normalize $r$ in market one. This is because the profit function contains the product of $r$ and the total supply of skills: $rTSV$. The normalization implies that what enters the profit function is $cr\frac{TSV}{c}$, which is identical to the product of the true unobserved $r$ and the true unobserved skills for every value of $c$. Therefore, the normalization does not have any consequences for the units in which the profits are expressed. In particular, revenues, wage bill and other costs are all expressed in the same units.

Second, notice that the estimates of the parameters of the profit function depend on the choice of $c$; however, the counterfactuals are invariant to the choice of $c$. The reason why the parameters depend on the choice of $c$ is that the parameters are obtained by maximizing a (simulated) likelihood based on the estimated $r$ from the first stage. As explained above, only the products $cr^d, d = 2, ..., D$ are identified. Under a different normalization $c'$, the second step likelihood would "match" different values of the skill prices: $c'r \neq cr$, therefore, the first-stage estimates are a function of $c$: $\hat{\theta}_I(c)$. Notice, however, that when I maximize the profit function to perform a counterfactual experiment, given a parameter estimate $\hat{\theta}_I(c)$ I obtain an optimal $cr$, i.e., I obtain an optimal skill price expressed in terms of the same normalization adopted in the estimation of the second-stage parameters. This means that it is meaningful to compare

---

[94]Notice that the supply of teaching skills is identified only up to a constant, I discuss what this means for counterfactuals below.

[95]Notice that because tuition is equal to the cap in all markets, all observed variation in the demand for enrollment across markets derives from the exogenous distribution of observed parental characteristics, so that endogeneity of $DV$ in the pricing equation is not an issue.

the behaviors of parents and teachers under the counterfactual and the baseline skill prices, and that this comparison is the same regardless of the normalizing $c$ used. Consider, for example, how a counterfactual $r$ affects achievement under two alternative choices of the normalizing constant $c$. Let $c_0$ and $c_1$ be two different normalizing constants. Under $r$, the mean skills supplied to the two schools are $\bar{s}^M, \bar{s}^V$ regardless of the choice of $c$. However, the numerical solution to the model yields $rc_0$ when $c_0$ is assumed in estimation, and $rc_1$ when $c_1$ is assumed in estimation, it does not yield $r$. Under $rc_0$, the mean skills supplied to, for example, the voucher school are $\frac{\bar{s}^V}{c_0}$ and the effect on student achievement is $\beta c_0 \frac{\bar{s}^V}{c_0} = \beta \bar{s}^V$. Under $rc_1$, the mean skills supplied to the voucher school are $\frac{\bar{s}^V}{c_1}$ and the effect on student achievement is the same as under the alternative choice of the normalizing constant, $c_0$: $\beta c_1 \frac{\bar{s}^V}{c_1} = \beta \bar{s}^V$. Different values of the normalizing constant imply different $\hat{\theta}_I(c)$ and different $\hat{\theta}_{II}(c)$; however, the counterfactual results are not affected.

# 6  Model Fit

The model fits the data well. I present two types of model fit by comparing the actual data to two types of simulations: those obtained using the observed tuition and skill prices estimated in the first step, i.e., the simulations conditional on the truth, and the simulations obtained by solving the entire model. I refer to the latter as unconditional simulations and to the former as conditional simulations. The unconditional simulation uses simulated tuition and wage rates, obtained as a solution to the voucher school's problem. It is reasonable to expect that the fit from simulating the entire model will be worse than the fit conditional on the observed prices. As the tables and graphs below show, only the fit of teaching wages and of the fraction of potential teachers choosing teaching worsens. Overall, however, the fit remains good.

## 6.1  Conditional Simulations

### 6.1.1  Parents

I first show the fit of test scores by school sector in the entire country in figure 4, and then I show two examples of test score fit in two markets in figures 5 and 6: an example of a better and a worse market-level fit. Table 5 shows actual and simulated test score means. As can be seen from the graphs and table, the country- and market-level fits are good.

Table 5: Actual and Simulated Average Test Scores by School Type, Entire Country

| Type | Actual | Simulated Cond. |
|---|---|---|
| Municipal | -0.1746 | -0.2176 |
| Voucher | 0.1574 | 0.1524 |

Next, I show how the model conditional on the true tuition and skill price fits choices and average tuition payments. I show the fit conditional on some exogenous variables (school level, rural status of the household, household income): the fit conditional on all the exogenous variables is comparable to the fit shown here.
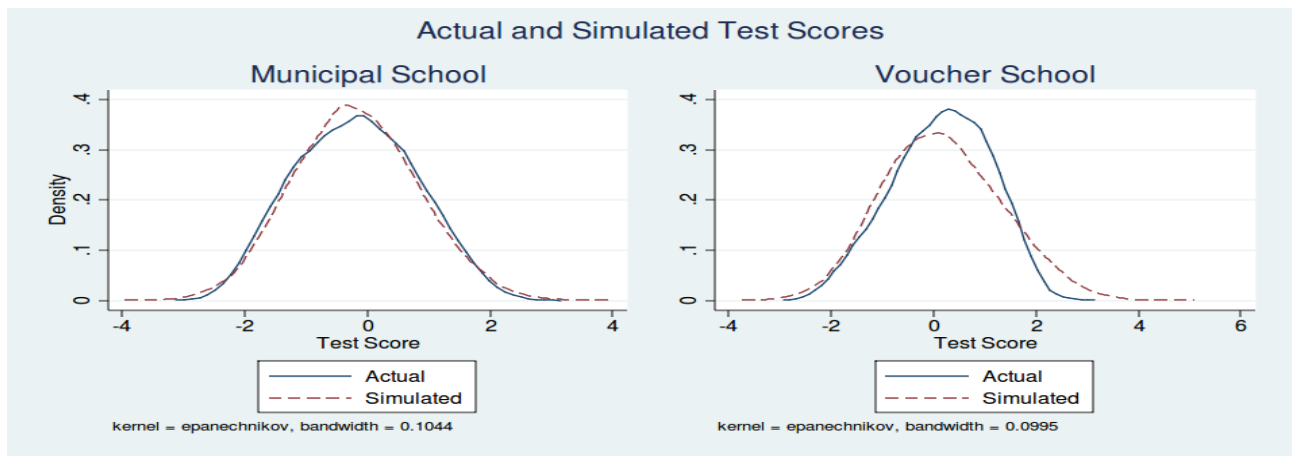
Figure 4: Conditional Test Score Simulations in the Entire Country
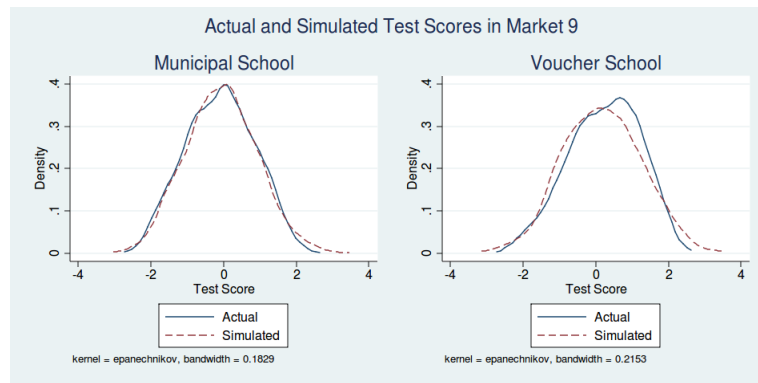


Figure 5: Conditional Test Score Simulations in Market 9
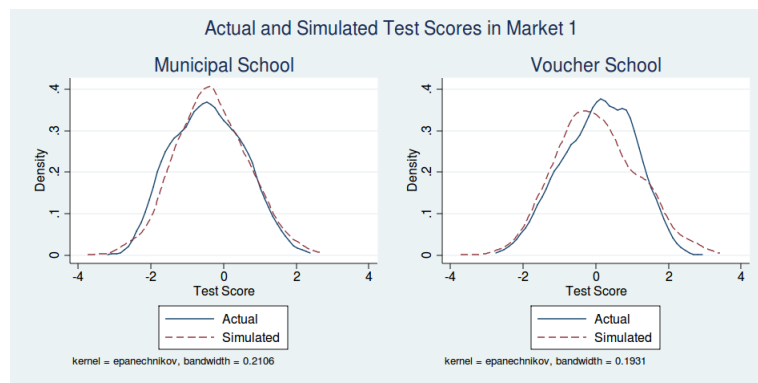


Figure 6: Conditional Test Score Simulations in Market 1

Table 6: Actual and Simulated Fractions in the Voucher School Overall and by School Level

| Type | Actual | Simulated Cond. |
|------|--------|-----------------|
| Overall | 52.99% | 52.88% |
| Primary | 50.45% | 52.07% |
| Secondary | 54.08% | 54.55% |

Table 7: Actual and Simulated Fractions in the Voucher School by Rural Status of the Household

| Type | Actual | Simulated Cond. |
|------|--------|-----------------|
| Urban | 53.95% | 55.10% |
| Rural | 29.44% | 29.83% |
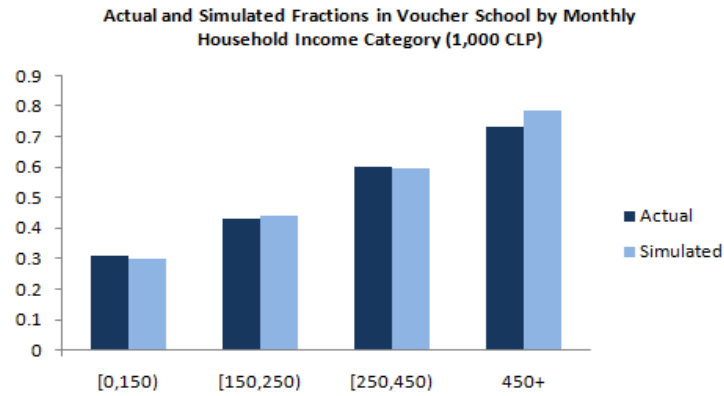


Figure 7: Conditional Choice Simulations

Table 8: Actual and Simulated Average Tuition Payments in the Voucher School (in 1,000 CLP)

| Actual | Simulated Cond. |
|--------|-----------------|
| 15.25 | 15.50 |
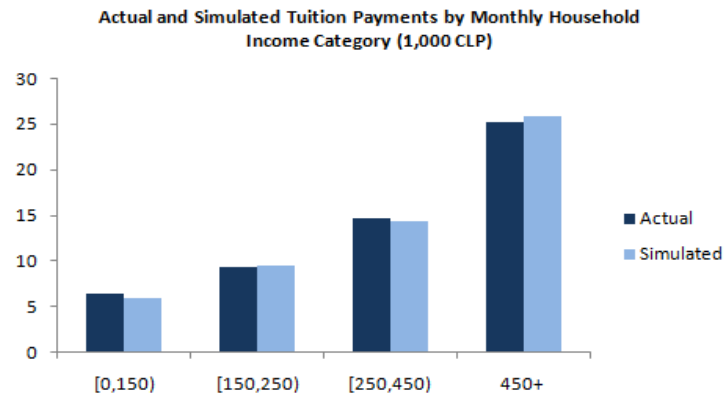


Figure 8: Conditional Choice Simulations

### 6.1.2 Potential Teachers

In figures 9 and 10 I show the fit of the distribution of accepted teaching and non-teaching wages and of the distribution of municipal and voucher school wages separately. Wages are expressed in terms of CLP 100,000 (approx. $200). [96]
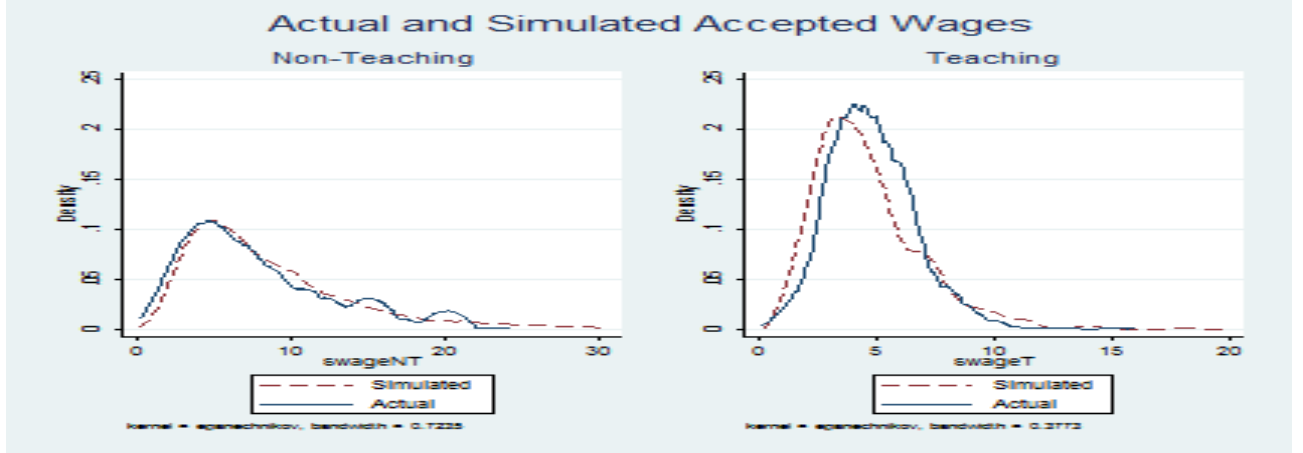


Figure 9: Conditional Choice Simulations
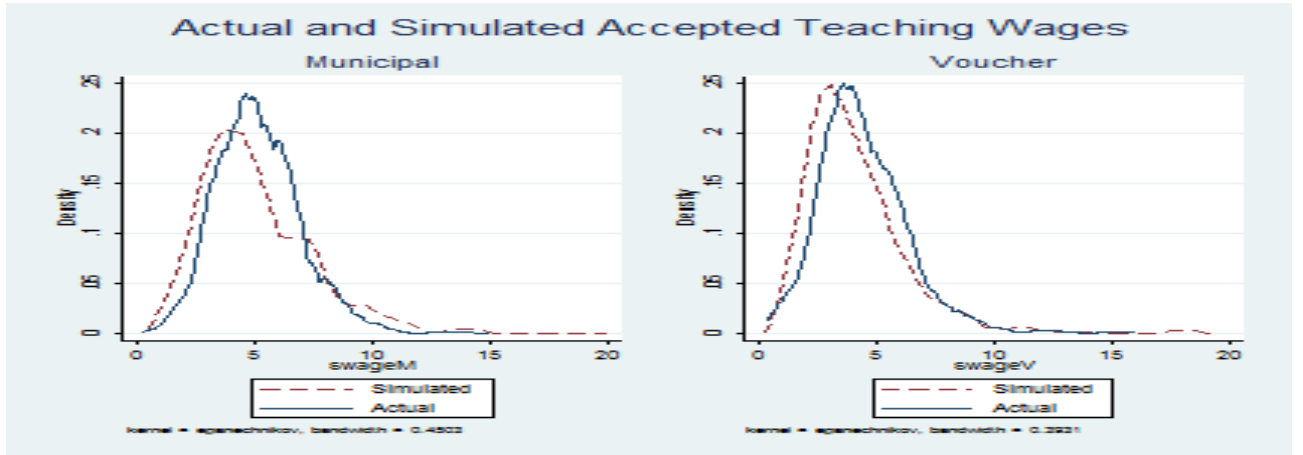


Figure 10: Conditional Choice Simulations

The model captures the age profile of accepted teaching wages, as can be seen in figure 11. In particular, simulated municipal and voucher age profiles mimic actual age profiles in that they cross once, with higher voucher school wages for younger teachers.

The model captures the fractions in each sector, both overall and conditional on all exogenous variables. I first show fractions overall, and then I show fractions in a representative market. I then show the fit of

---

[96]The fit at the market level is worse than the fit at the country level. Typically, in the class of models of self-selection into occupations, accepted wages are the hardest aspect of the data to match. The estimated model in this paper does a good job at fitting the distribution of accepted wages in the entire country.
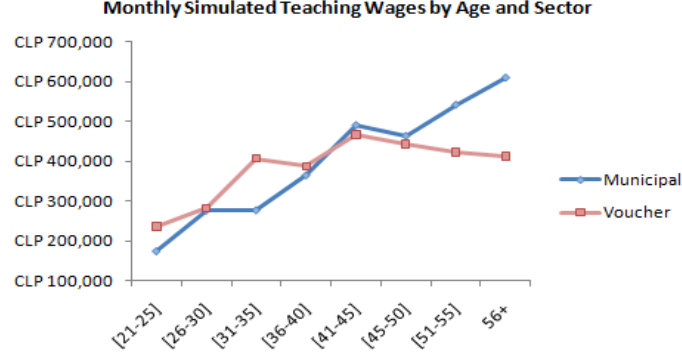
Figure 11: Conditional Choice Simulations

fractions by age and gender: the fit conditional on the remaining exogenous variables is comparable to the one shown here. Finally, tables 11 through 14 show actual and simulated occupational choices by age.

Table 9: Actual and Simulated Fractions in Each Occupation, Whole Country

| Choice | Actual | Simulated Cond. |
|---|---|---|
| Municipal | 9.48% | 10.18% |
| Voucher | 7.81% | 7.42% |
| Non-Teaching | 70.28% | 70.09% |
| Home | 12.44% | 12.32% |

Table 10: Actual and Simulated Fractions in Each Occupation, Market 16 (Valparaíso)

| Choice | Actual | Simulated Cond. |
|---|---|---|
| Municipal | 10.68% | 10.47% |
| Voucher | 9.74% | 8.89% |
| Non-Teaching | 69.86% | 67.96% |
| Home | 9.73% | 12.68% |

Table 11: Actual and Simulated Fractions in Each Occupation, Age $\leq 30$

| Choice | Actual | Simulated Cond. |
|---|---|---|
| Municipal | 2.29% | 3.34% |
| Voucher | 8.17% | 5.26% |
| Non-Teaching | 78.13% | 79.26% |
| Home | 11.42% | 12.14% |

Table 12: Actual and Simulated Fractions in Each Occupation, Age > 30 and ≤ 40

| Choice | Actual | Simulated Cond. |
|---|---|---|
| Municipal | 4.11 % | 4.09% |
| Voucher | 7.73% | 7.84% |
| Non-Teaching | 77.63% | 75.93% |
| Home | 10.53% | 12.14% |

Table 13: Actual and Simulated Fractions in Each Occupation, Age > 40 and ≤ 50

| Choice | Actual | Simulated Cond. |
|---|---|---|
| Municipal | 11.05% | 11.23% |
| Voucher | 9.21 % | 9.25% |
| Non-Teaching | 67.16% | 66.23% |
| Home | 12.57% | 13.29% |

Table 14: Actual and Simulated Fractions in Each Occupation, Age > 50

| Choice | Actual | Simulated Cond. |
|---|---|---|
| Municipal | 20.82% | 22.90% |
| Voucher | 5.82% | 6.16% |
| Non-Teaching | 57.58% | 59.47% |
| Home | 15.78% | 11.47% |



Figure 12: Conditional Choice Simulations

## 6.2 Unconditional Simulations

In this section I show the fit obtained by drawing a vector containing all model shocks and types (private school's cost shock, parents' and te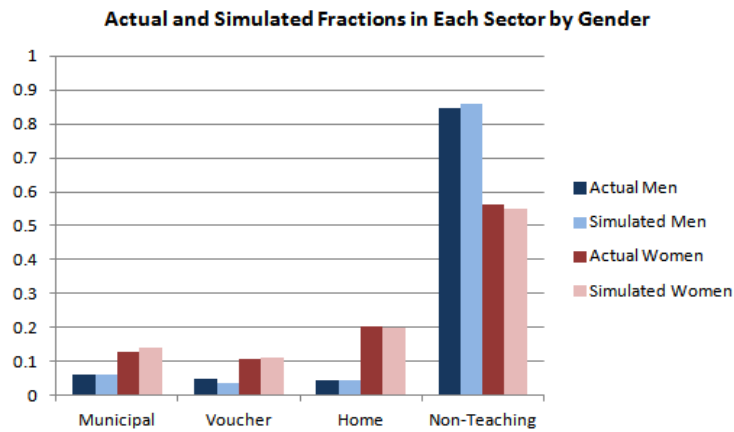achers' preference and technology shocks and types) and solving the entire model. The optimal simulated tuition in each market is binding at the price cap, as is observed in the data. The log-prices of teaching skills by market are not observed, but they are obtained from the first stage of the estimation and treated as data. Their average and standard deviations across markets are $-0.18332$ and $0.242716$. The log-prices simulated from the first stage of the model are, on average, $-0.18823$ across markets, with a standard deviation of $0.32865$. This suggests that the model is capable of reproducing the equilibrium observed in the data. It is not surprising then that the simulated behavior of parents and college graduates when tuition and skill prices are simulated is not far from actual behavior.

### 6.2.1 Parents

As before, I first show the fit of test scores by school sector in the entire country in figure 13, and then I show two examples of test score fit in two markets in figures 14 and 15: an example of a better and a worse market-level fit. Table 15 shows actual and simulated test score means. As can be seen from the graphs and table, the country- and market-level fits are good.

Table 15: Actual and Simulated Average Test Scores by School Type, Entire Country

| Type | Actual | Simulated Cond. |
|---|---|---|
| Municipal | -0.1746 | -0.2511 |
| Voucher | 0.1574 | 0.1042 |
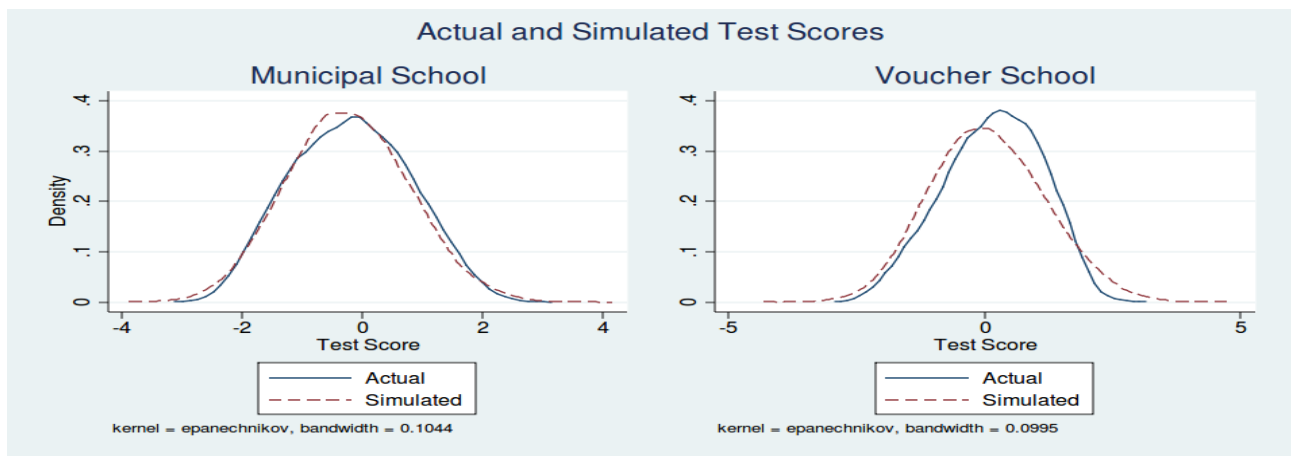


Figure 13: Unconditional Test Score Simulations in the Entire Country

Next, I show how the model fits choices and average tuition payments. As before, I show the fit conditional on some exogenous variables (school level, rural status of the household, household income): the fit conditional on all the exogenous variables is comparable to the fit shown here.
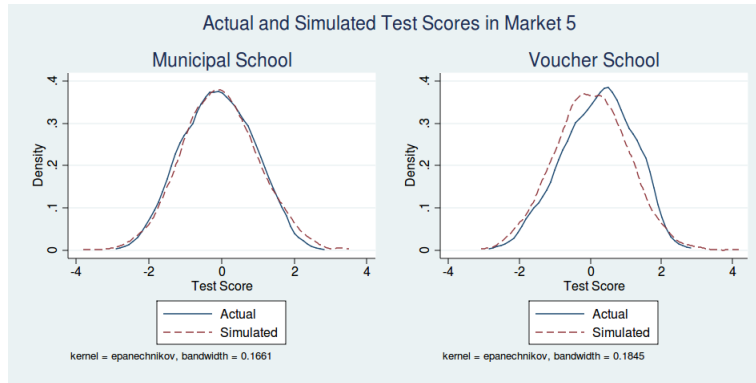
Figure 14: Unconditional Test Score Simulations in Market 5
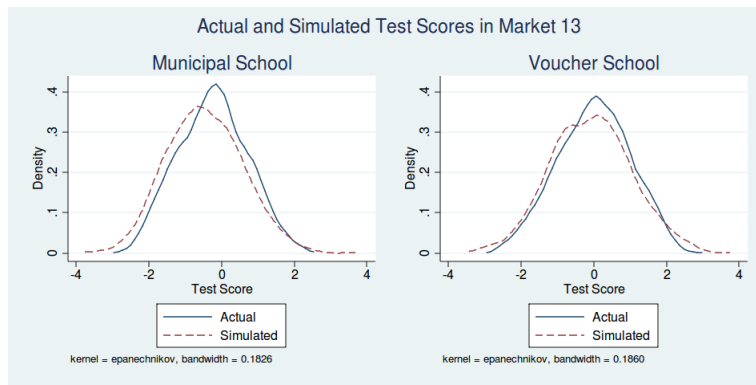


Figure 15: Unconditional Test Score Simulations in Market 13

Table 16: Actual and Simulated Fractions in the Voucher School Overall and by School Level

| Type | Actual | Simulated Uncond. |
|---|---|---|
| Overall | 52.99% | 53.46% |
| Primary | 50.45% | 51.90% |
| Secondary | 54.08% | 55.37% |

Table 17: Actual and Simulated Fractions in the Voucher School by Rural Status of the Household

| Type | Actual | Simulated Cond. |
|---|---|---|
| Urban | 53.95% | 55.46% |
| Rural | 29.44% | 30.69% |

Table 18: Actual and Simulated Average Tuition Payments in the Voucher School (in 1,000 CLP)

| Actual | Simulated Cond. |
|---|---|
| 15.25 | 15.56 |
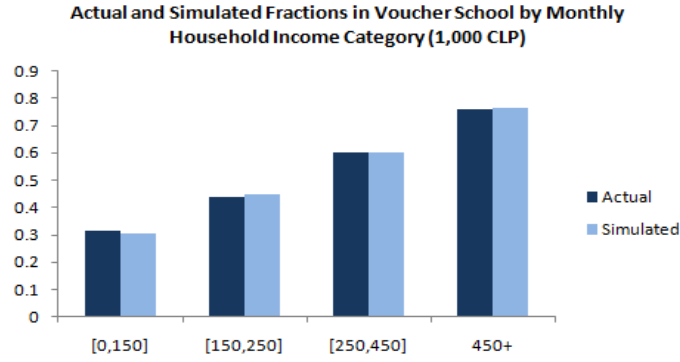
Figure 16: Unconditional Choice Simulations
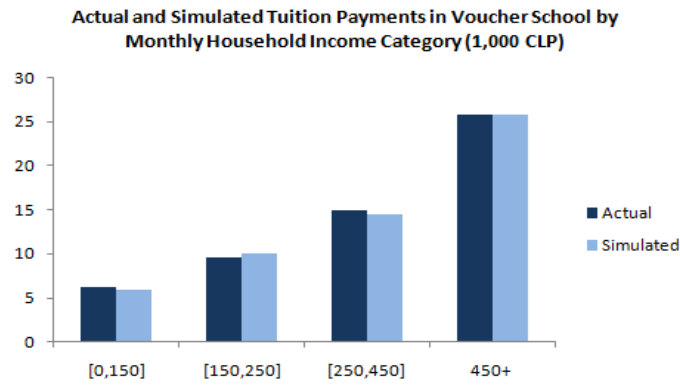


Figure 17: Unconditional Choice Simulations

### 6.2.2 Potential Teachers

As before, in figures 18 and 19 I show the fit of the distribution of accepted teaching and non-teaching wages and of the distribution of municipal and voucher school wages separately. Wages are expressed in terms of CLP $100,000$ (approx. $200).

The model captures the age profile of accepted teaching wages, as can be seen in figure 20. In particular, the wage premium for younger voucher school teachers and for older municipal school teachers is preserved.

The model captures the fractions in each sector, both overall and conditional on all exogenous variables. I first show fractions overall, and then I show fractions in a representative market. I then show the fit of fractions by age and gender: the fit conditional on the remaining exogenous variables is comparable to the one shown here. Finally, tables 21 through 24 show actual and simulated occupational choices by age.
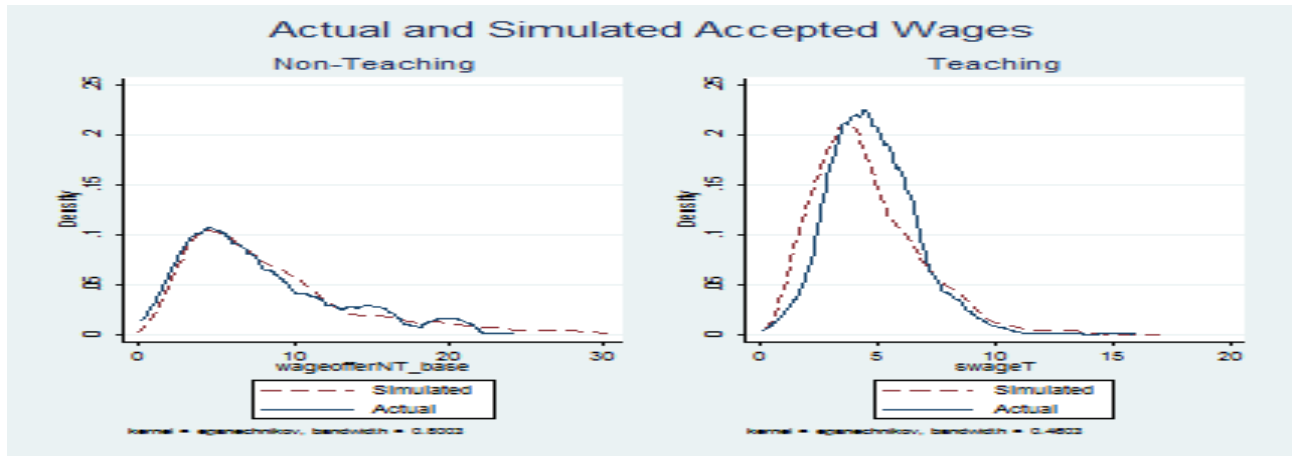
Figure 18: Unconditional Choice Simulations
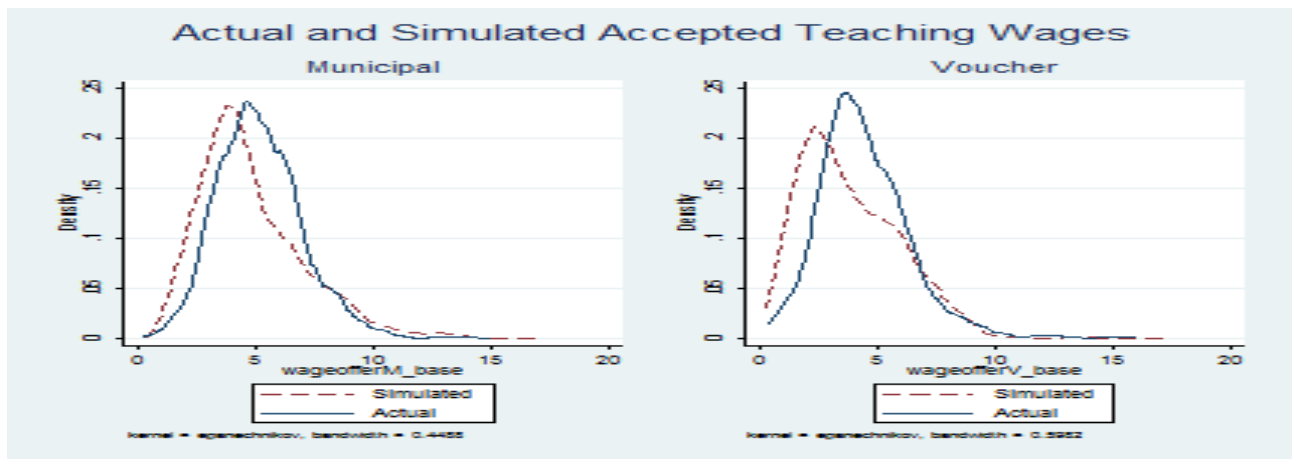


Figure 19: Unconditional Choice Simulations
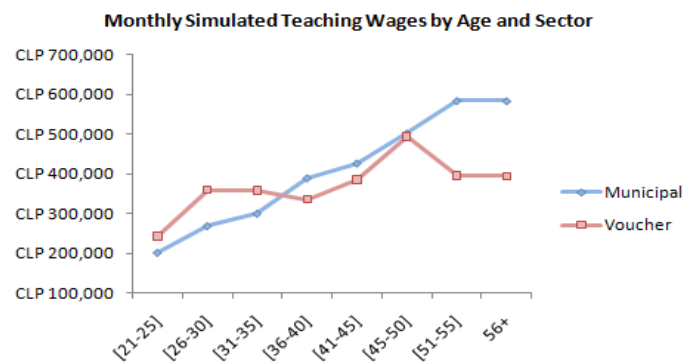


Figure 20: Conditional Choice Simulations

Table 19: Actual and Simulated Fractions in Each Occupation, Whole Country

| Choice | Actual | Simulated Cond. |
|---|---|---|
| Municipal | 9.48% | 9.52% |
| Voucher | 7.81% | 5.21% |
| Non-Teaching | 70.28% | 72.77% |
| Home | 12.44% | 12.51% |

Table 20: Actual and Simulated Fractions in Each Occupation, Market 16 (Valparaíso)

| Choice | Actual | Simulated Cond. |
|---|---|---|
| Municipal | 10.68% | 9.10% |
| Voucher | 9.74% | 4.74% |
| Non-Teaching | 69.86% | 75.55% |
| Home | 9.73% | 10.61% |

Table 21: Actual and Simulated Fractions in Each Occupation, Age $\leq$ 30

| Choice | Actual | Simulated Uncond. |
|---|---|---|
| Municipal | 2.29% | 3.27% |
| Voucher | 8.17% | 3.89% |
| Non-Teaching | 78.13% | 82.84% |
| Home | 11.42% | 10.00% |

Table 22: Actual and Simulated Fractions in Each Occupation, Age $>$ 30 and $\leq$ 40

| Choice | Actual | Simulated Uncond. |
|---|---|---|
| Municipal | 4.11 % | 7.45% |
| Voucher | 7.73% | 4.38% |
| Non-Teaching | 77.63% | 74.74% |
| Home | 10.53% | 13.43% |

Table 23: Actual and Simulated Fractions in Each Occupation, Age $>$ 40 and $\leq$ 50

| Choice | Actual | Simulated Uncond. |
|---|---|---|
| Municipal | 11.05% | 11.45% |
| Voucher | 9.21 % | 4.37% |
| Non-Teaching | 67.16% | 68.26% |
| Home | 12.57% | 15.93% |

Table 24: Actual and Simulated Fractions in Each Occupation, Age $>$ 50

| Choice | Actual | Simulated Uncond. |
|---|---|---|
| Municipal | 20.82% | 27.16% |
| Voucher | 5.82% | 2.22% |
| Non-Teaching | 57.58% | 59.38% |
| Home | 15.78% | 11.24% |

**Actual and Simulated Fractions in Each Sector by Gender**
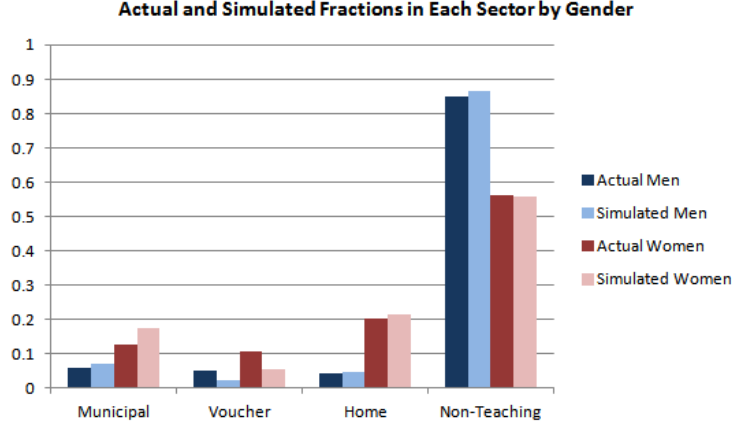


Figure 21: Unconditional Choice Simulations

# 7   Results

## 7.1   Estimation Results

I report in Appendix I the tables with the estimates and standard errors of all model parameters. In this section I describe findings that derive directly from the estimation results.

First, I find that skills are priced on average 4.43 times more in the municipal school than in the private school. This figure was found by computing the implied price of skills in the municipal school, which is the ratio of an individual's wage to her amount of skills.[97] However, and this is the second finding, teachers in the voucher school are more skilled than teachers in the municipal school. These two findings are consistent with each other because, contrary to voucher school wage offers, wage offers in the municipal school are not a function of skills. The highest skilled individuals obtain higher wage offers from the voucher school than from the municipal school, and the opposite is true for the least skilled individuals, who are, therefore, more likely to self-select into the municipal sector.[98] The high implied price of skills in the municipal sector, therefore, is due to the low supply of skills to that sector, accompanied by wages that are flat with respect to skills.

Finally, I use the parameter estimates to simulate average treatment effect (ATE), treatment effect on the treated (TT) and treatment effect on the untreated (TU) of voucher schools, using student test scores as the outcome measure. I find negative ATE, but larger positive TT and negative TU. These findings are consistent with other findings in the literature on the Chilean voucher system, notably Sapelli and Vial (2002). I find evidence of parental sorting on both observables and unobservables. In particular, both sorting on observables and unobservables explain the observed test score gap between municipal and private voucher school students.

---

[97]The implied municipal sector price of skills is identified up to scale because skills are identified up to scale. The ratio of this implied price to the price of skills in the voucher school, which is the object of interest, is identified.

[98]The implied price of skills in the municipal sector represents the price paid to the pool of teachers that chose the municipal sector under a wage offer that is flat with respect to skills. If the wage offer in the municipal school were a linear function of skills as in the voucher school, the implied skill price would attract a different pool of individuals. See the section on counterfactuals, where pay-per-skill wage schemes are considered.

### 7.1.1 Implied Skill Price in the Municipal Sector

For each municipal sector teacher I obtain the implied skill price as the ratio of her wage to her teaching skills, which are estimated up to scale. The implied skill price in the municipal school is not identified, however, its product with the normalizing constant $c$ is. Because in the voucher school the product of the skill price and the normalizing constant is identified, the ratio of the skill prices in the two sectors is identified. I compute the average ratio of municipal to voucher school skill prices in each market. This is an average because the implied skill price in the municipal school is specific to each individual: I take the average across individuals in the same market to obtain the market-specific average implied skill price, and I divide this average by the skill price chosen by the voucher school in the market.

Table 25 reports the ratio of the average implied skill price in the municipal school to the skill price in the voucher school in each market.[99] The average skill price ratio across markets is 4.43: teachers in the municipal sector, where wages are set by rigid formulae, are paid, on average, 4.43 times more per unit of skills possessed than teachers in the voucher school.[100]

Table 25: Ratio of Skill Prices in the Municipal and Voucher School

| Market | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $\frac{\tilde{r}_d^M}{r_d}$ | 3.32 | 3.99 | 3.13 | 4.80 | 3.84 | 3.88 | 5.81 | 5.16 | 3.57 |
| **Market** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** |
| $\frac{\tilde{r}_d^M}{r_d}$ | 4.94 | 4.40 | 2.90 | 5.70 | 4.69 | 5.11 | 4.71 | 5.50 | 4.31 |

### 7.1.2 Distribution of Teaching Skills by Occupation

Figure 22 plots the densities of teaching skills of teachers and of non-teachers, and it plots the densities of municipal and private sector teacher skills separately. Because skills are identified only up to scale, their value in absolute terms is meaningless, but comparisons are meaningful. It is, therefore, meaningful to compare the distribution of teaching skills by occupation. The density of teaching skills of voucher school teachers first-order stochastically dominates the densities of teaching skills in all other occupations. The densities of teaching skills of municipal school teachers and individuals currently not employed in the teaching sector are similar. To evaluate the skill difference between municipal and voucher schools in terms of student achievement, I simulated test scores in municipal schools if the mean skilled in municipal schools were equal to the mean skills supplied to voucher schools. The test scores of municipal school students would be 0.5 standard deviation higher if they were exposed to voucher school teachers.

There is a non-negligible mass of individuals with high teaching skills currently employed in a non-teaching occupation. This suggests that there is a scope for policies aimed at attracting skilled teachers into the teaching profession.

---

[99]I use the simulated skill prices obtained by simulating the private schools' behavior and not the skill prices estimated in the first step.

[100]Notice that this implied skill price was computed using accepted wages, i.e., the wages of those who choose the municipal sector. Using simulated wage offers to compute the implied skill price of all potential teachers in the market yields an across-market average of 3.07.
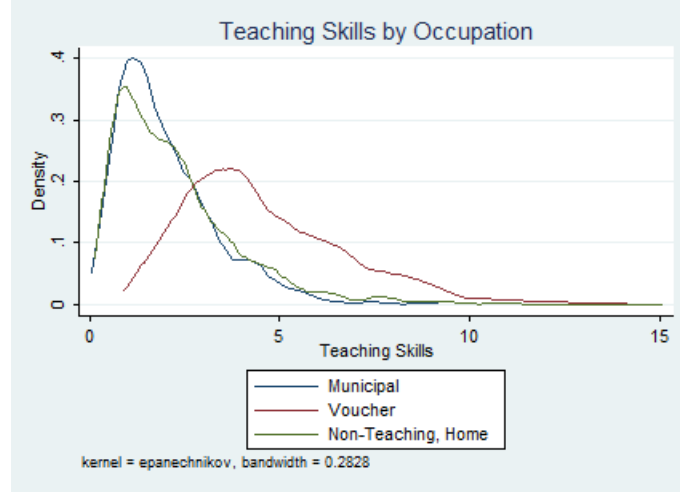
Figure 22: Teaching Skills by Occupation

### 7.1.3 The Role of Unobserved Types

**Potential Teachers**

The introduction of unobserved heterogeneity in the form of types proved determinant in fitting the data. Types affect the intercept in municipal and non-teaching log-wages and the intercept in log-teaching skills. Types also affect non-pecuniary utilities in each sector. Individuals of type one are those who possess the highest amount of teaching skills and those of type three the smallest amount, keeping all other individual level variables constant. The order in which types affect teaching skills is not reflected in municipal sector wages, where, although the lowest skilled receive the lowest wage offers, type two individuals receive wage offers that are 4.7% higher than those of individuals who share the same observables but who are of type one, and therefore better teachers. The adjustments to the wage formulae in the municipal sector, therefore, seem to not be reflective of teacher skills.

The order with which types affect non-teaching wage offers varies by market, but in general, it does not exactly reflect the order of types in terms of their teaching skills. Across-market averages reveal that, holding all other observables constant, type three individuals receive lower wage offers in the non-teaching sector, on average lower than type one's by 44.5%.[101] Type one individuals tend to be better teachers than type three, so this wage differential reflects the order in which types affect teaching skills. Individuals of type two, however, who are worse teachers than individuals of type one, receive non-teaching wage offers that are, on average, 7.0% higher than those received by type one individuals.

Table 26 reports the fractions in each occupation by type. Type three is the type most likely to choose the voucher school. Conditional on observables, type three individuals have lower teaching skills (and therefore receive lower wage offers from voucher school) than type one individuals; however, they enjoy a higher utility from choosing the voucher school. The net effect is that they are more likely than type one individuals to choose the voucher school. Similarly, type three individuals, conditional on observables, have more teaching skills than type two individuals and, therefore, receive higher wage offers, but they enjoy a lower utility

---

[101]This figure was obtained by regressing simulated wage offers in the non-teaching sector on the observables $Z_i$ that enter the wage offer function and on type dummies, using observations from all individuals in the country.

from the voucher school. The net effect of these two opposing forces is that type three individuals are more likely to choose the voucher school than type two individuals. Type three individuals are also the ones least likely to choose the non-teaching sector.[102] Finally, the model allows for sorting on unobservables, and the parameter estimates show that sorting on unobservables is an important phenomenon, as choice patterns among types, who represent subsets of the population who differ in terms of unobserved characteristics, differ.

Table 26: Choice Probabilities of Potential Teachers by Type

| Type | Prob(M) | Prob(V) | Prob(NT) | Prob(H) |
|------|---------|---------|----------|---------|
| 1 | 13.31% | 1.47% | 70.49% | 14.73% |
| 2 | 10.74% | 0.32% | 86.13% | 2.81% |
| 3 | 10.63% | 9.82% | 55.30% | 24.25% |

**Parents**

Some parameters in the cognitive achievement production function are allowed to vary by type to reflect the fact that school inputs may have different effects on different subgroups of the population of students who differ in terms of characteristics that are unobserved. Parameter estimates show that there are statistically significant differences in the way school inputs affect the achievement of different types of students, as can be seen in Appendix I and in table 27, which shows the rankings of types in the way they affect the parameters of the cognitive achievement production function in the two schools. Types in the table receive different rankings only if the type-specific parameters are statistically different from each other.

Table 27: Ranking of Types in Determining Achievement

| Parameter | Type 1 | Type 2 | Type 3 |
|-----------|--------|--------|--------|
| Intercept, Municipal | 2 | 3 | 1 |
| Coeff. on teaching skills, Municipal | 1 | 1 | 2 |
| Coeff. on parental educ., Municipal | 2 | 2 | 1 |
| Coeff. on income, Municipal | 1 | 1 | 2 |
| Intercept, Voucher | 1 | 3 | 2 |
| Coeff. on teaching skills, Voucher | 1 | 2 | 3 |
| Coeff. on parental educ., Voucher | 2 | 1 | 2 |
| Coeff. on income, Voucher | 3 | 2 | 1 |

Table 28 shows that the distribution of types in the sub-populations of voucher and municipal school children is different from the distribution of types in the entire population, which indicates that parents' sorting decisions are based also on their type.[103] This indicates evidence of sorting of parents based on unobservables.

The probability of choosing the voucher school by student type, presented in table 29, confirms that there is sorting based on the unobserved type. In particular, individuals of type one, who have the highest intercept

---

[102]Choices are driven by non-pecuniary factors as well as wages; it is, therefore, not possible to draw general conclusions in terms of potential teachers' skills and preferences based solely on knowledge of the choices made by different types.

[103]In defining the sub-populations of voucher and municipal school students, I use the simulated school choice. The population level fractions of individuals of each type are slightly different from the actual estimated type proportions because of simulation sampling errors.

Table 28: Distribution of Types in Entire Population and by School

| Type | Proportion in M | Proportion in V | Proportion in Population |
|------|-----------------|-----------------|--------------------------|
| 1 | 2.39% | 34.77% | 19.52% |
| 2 | 57.82% | 38..98% | 47.86% |
| 3 | 39.79% | 26.25% | 32.63% |

in the voucher school educational production function and who benefit the most from teacher skills in voucher schools, are very likely to choose the voucher school:

Table 29: Choice Probabilities of Parents by Type

| Type | Prob(V) |
|------|---------|
| 1 | 92.80% |
| 2 | 43.49% |
| 3 | 42.52% |

Tables 27, 28 and 29 present evidence not only that parents sort on both observable and unobservable characteristics, but also that the unobservables on which they sort affect outcomes. This implies that accounting for students' unobservable characteristics is necessary when evaluating the effect of a school on child achievement.

### 7.1.4 Treatment Effects of the Voucher School (ATE, TT, TU)

Sorting on unobservables would be inconsequential in the evaluation of production function parameters if the unobservables that affect choice did not affect outcomes. This is not the case in the choice of private versus public school in Chile, as the unobservables that affect choice have a substantial impact on outcomes. Consider the regression estimates in table 30.

Table 30: Regression of Test Scores on Parental Characteristics and School Type Dummy

| Variable | Actual | Simulated | Random Assignment |
|----------|--------|-----------|-------------------|
| $\frac{y_h}{nfam_h}$ | $0.398^{(***)}$ | $0.573^{(***)}$ | $0.627^{(***)}$ |
| | (0.012) | (0.014) | (0.016) |
| $\left(\frac{y_h}{nfam_h}\right)^2$ | $-0.068^{(***)}$ | $-0.128^{(***)}$ | $-0.144^{(***)}$ |
| | (0.003) | (0.004) | (0.004) |
| $peduc_h$ | $0.083^{(***)}$ | $0.080^{(***)}$ | $0.109^{(***)}$ |
| | (0.001) | (0.001) | (0.002) |
| $V$ | $0.121^{(***)}$ | $0.134^{(***)}$ | $-0.143^{(***)}$ |
| | (0.006) | (0.007) | (0.008) |
| Intercept | $-1.200^{(***)}$ | $-1.232^{(***)}$ | $-2.116^{(***)}$ |
| | (0.012) | (0.013) | (0.016) |

The first column shows parameter estimates from a regression of observed test scores on monthly per-capita income, monthly per-capita income squared, parental education and a dummy for the type of school. The second column shows a regression of the same variables, where I have substituted actual test scores and voucher school dummies with their simulated counterparts. Comparing the first two columns reveals that

the model fits the data well: the reduced-form parameters obtained using data simulated from the estimated model are close to the reduced-form parameters obtained from the actual data.[104] The last column shows what would happen if parents were randomly assigned to a school. I attach to each family a 50% probability of being assigned to the voucher school, and I simulate their realized school assignment and their children's test scores using the estimated structural parameters. Because school assignment is random, the students in the voucher school and those in the municipal school constitute two comparable groups, and the coefficient on the voucher school dummy is not affected by selectivity bias.[105] The results are striking: the coefficient on the voucher school dummy reduces from 13.44% of a standard deviation when parents are free to self-select into schools to −14.31% of a standard deviation under random assignment of parents to schools.[106] The average effect of a voucher school on students' outcomes is negative. This suggests that the observed test score gap that persists between private and public education even after accounting for observable students' characteristics can be attributed to the fact that students in voucher schools differ from students in municipal schools in terms of unobserved characteristics which positively affect their test scores either directly or through their interaction with school inputs, or both.

To assess the role of selection in determining achievement I quantify the treatment effect on the treated and on the untreated. Subtracting simulated test scores in the municipal school from simulated test scores in the voucher school for those who choose the voucher school, I find that the average treatment effect on the treated is considerable: the average increase in test scores from attending a voucher school as opposed to a municipal school for students who choose the voucher school is 1.18 standard deviations. In contrast, the treatment effect on the untreated, i.e., on the students who choose the municipal school, is negative and substantial, being equal to -1.63 standard deviations.[107] Figure 23 shows the p.d.f. of the simulated treatment effects in the entire population, and by treatment status. The treatment effect for any given individual is defined as the difference between the test score in the voucher school and the test score in the public school, and it is simulated from the model.

These findings are consistent with other findings in the literature, notably Sapelli and Vial (2002), who find substantially positive and significant treatments on the treated, and small and sometimes insignificant average treatment effects. The findings in this paragraph and in the previous one suggest that the sorting of students across schools has a fundamental role in determining achievement, and that unobservable students' characteristics affect both sorting and outcomes.

## 7.2  Results of the Counterfactual Policy Experiments

In this section I describe findings that derive from simulating the outcomes of alternative policy experiments. All counterfactual policy experiments assume that the government moves first, by implementing the new

---

[104]The reduced-form parameters are a function of the structural parameters.

[105]Because assignment is random, the two groups are homogeneous in terms of both observables and unobservables, hence, the coefficient on the dummy is equal to the average treatment effect and can be alternatively recovered by simulating the treatment effect for each individual and averaging across individuals. The regression approach is useful to investigate whether also the coefficients on the students' observables are subject to selection bias. Regression results suggest that unobservables only slightly affect the impact of students' observables on test scores, as the coefficients with self-selection are slightly different from the coefficients with random assignment.

[106]Average test scores in the entire nation decrease by around 70% of a standard deviation under random assignment. Notice, however, that I randomly assign to private education 50% of the population, whereas without random assignment 52.88% of the population chooses the voucher school.

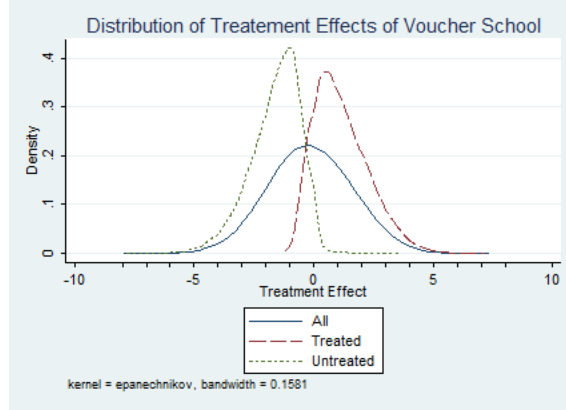[107]Subgroups are defined using simulated school choices.

Figure 23: Simulated Treatment Effects

policy, and then the two-stage problem of private schools, parents and potential teachers is solved in each market.[108]

### 7.2.1 Flat Increase in Municipal Sector Wages

I consider a wage increase in the municipal sector. The increase is accompanied by minimum competency requirements: these requirements approximate a recruiting policy such that only the most skilled teachers are hired among those willing to accept the wage offer. They are introduced because the model does not have capacity constraints. The wage increase is flat with respect to teaching skills: the municipal wage offers increase by the same amount for individuals of all skill levels. I find that an increase of CLP $90,000$ a month (approx. \$180) in wage offers yields an increase in test scores by $8\%$ of a standard deviation and it requires a modest $3\%$ increase in government spending on education.[109] The minimum competency requirement is such that only individuals with a level of skills above a certain cutoff can teach in the municipal school.[110] The cutoff used in this experiment is equal to the bottom quartile of the pre-policy distribution of teaching skills in the municipal sector.

Table 31 shows pre- and post-policy mean test scores and mean teaching skills overall and by school sector. Both mean test scores and teaching skills improve, overall and by school sector. The effect of the policy on test scores and the amount of government spending necessary to finance it depend on the reactions of potential teachers, parents and private schools. In the remainder of this section I analyze these policy responses, and I describe the implications for government spending. The baseline and counterfactual scenarios represent two different equilibria. Although I do not study transition dynamics, for ease of exposition I refer to potential teachers or parents who in the baseline equilibrium choose occupation or school A and who in the counterfactual equilibrium choose occupation or school B as individuals who "switch" from A to B.

**Policy Reaction of Potential Teacher Sorting and of Private Schools**

---

[108]All baseline figures are simulated, not actual, figures.

[109]I chose a wage increase that does not affect class size considerably. Because in the model class size does not affect achievement, the counterfactuals could under- or over-estimate the effect on achievement if class size changed considerably.

[110]In Chile a teaching certificate is required to teach. This policy experiment should be interpreted as a more stringent requirement, more tightly linked to the teacher's skills.

Table 31: Increasing Municipal School Wage Offers by CLP 90, 000: Mean Test Scores and Teacher Skills

|  | mean test scores | mean test scores in M | mean test scores in V |
|---|---|---|---|
| **Base** | -0.06 | -0.25 | 0.10 |
| **Wage Increase** | 0.02 | -0.16 | 0.18 |
|  | mean skills | mean skills in M | mean skills in V |
| **Base** | 2.70 | 2.01 | 4.92 |
| **Wage Increase** | 2.98 | 2.43 | 4.98 |

Overall, the number of teachers decreases from 118, 000 to 112, 000. The number of municipal school teachers decreases from 90, 000 to 88, 300 and the (simulated) number of voucher school teachers decreases from 28, 000 to 24, 000.[111] Table 32 shows the transitions among sectors that result from the introduction of the policy.[112]

Table 32: Increasing Municipal School Wage Offers by CLP 90, 000: Transitions Across Sectors

| Pre-Policy Choice | Post-Policy Choice | | | | |
|---|---|---|---|---|---|
|  | M | V | NT | H | Tot |
| M | 75.85% | 2.19% | 21.96% | 0.00% | 100.00% |
| V | 22.57% | 77.43% | 0.00% | 0.00% | 100.00% |
| NT | 3.62% | 0.28% | 96.10% | 0.00% | 100.00% |
| H | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% |

The policy has the effect of attracting individuals to the municipal school from both the non-teaching sector and the voucher school. A wage increase that is flat with respect to skills is unable to attract the most highly skilled teachers. Table 33 shows that, among the non-teachers, individuals with higher teaching skills are less likely to switch to the municipal sector. This is explained in the model by both pecuniary and non-pecuniary factors: the non-teaching skills of these individuals are allowed to be such that they obtain higher wage offers in the non-teaching sector, and their non-pecuniary preference for the non-teaching sector is allowed to be higher than their non-pecuniary preference for teaching. Results from model simulations, therefore, capture both factors. The table shows the fractions of individuals in the non-teaching sector who switch to the municipal sector, by skill level. Skill level is measured in terms of quartiles of the distribution of skills in the population: individuals belonging to skill quartile one are low skilled individuals, those belonging to skill quartile four are highly skilled individuals. Among individuals in the first quartile, only a few switch to the municipal sector because they are not qualified to work in that sector. Among individuals in quartiles three to four, the fractions of non-teachers who choose to switch to the municipal sector decreases as we move up in the distribution of skills.

The policy attracts the least skilled teachers from the voucher school: the mean skills of those who switch to the municipal sector are 3.06, the mean skills of those who don't are 5.42. Table 34 shows that, similarly to transitions from the non-teaching sector, among voucher school teachers, more skilled individuals are less likely to switch to the municipal school than less skilled individuals.

The transition of teachers from the voucher school to the municipal school depends on the voucher school reaction, and it is better explained through a graph. Consider figure 24. The horizontal axis represents

---

[111]The effects on class size are small. Average class size in municipal schools increases by less than one, from 10.0 to 10.8 and average class size in the voucher schools increases from 34.6 to 38.3.

[112]The fraction that stays in the municipal sector is slightly above 75%, and not exactly 75%, because of simulation error.

Table 33: Increasing Municipal School Wage Offers by CLP 90,000: Transitions From NT to M by Skill Quartile

| Skill Quartile | Frac. of NT Switching to M |
|:---:|:---:|
| 1 | 0.48% |
| 2 | 6.26% |
| 3 | 4.40% |
| 4 | 3.59% |

Table 34: Increasing Municipal School Wage Offers by CLP 90,000: Transitions From V to M by Skill Quartile

| Skill Quartile | Frac. of V Switching to M |
|:---:|:---:|
| 1 | 0.00% |
| 2 | 80.72% |
| 3 | 46.08% |
| 4 | 13.16% |

teaching skills. Voucher school wage offers are an increasing linear function of skills, whereas municipal sector wage offers do not change with skills. As a reaction to policy, voucher schools increase their wage offers: the average across markets of the price of teaching skills offered by the voucher schools increases from 0.76 to 0.83, which results in a 6.4% average increase in voucher school wage offers. They increase their wage offers because after the introduction of the public school teacher policy, it becomes harder for them to attract highly skilled teachers. In the graph, the voucher school wage offer function rotates from $w_0^V$ to $w_1^V$. Before the introduction of the wage policy, the teaching skills of those who choose the municipal sector are in the interval $[0, \bar{s}_0^T]$, after the policy individuals with skills in the interval $[\bar{s}_0^T, \bar{s}_1^T]$ switch from the voucher to the municipal school. These individuals are the worst voucher sector teachers, however, their skills are higher than the skills of the incumbent municipal school teachers, i.e., those who choose the municipal school before the policy change. The transition of the worst voucher school teachers to the municipal school increases the mean teaching skills in both school sectors.[113] Notice that had the reaction of the voucher school been of a larger magnitude, the policy could have had the effect of moving the best municipal sector teachers to the voucher school, thus lowering the mean skills in both sectors. It is, therefore, important to account for the reaction of voucher schools to correctly predict the reaction of potential teachers.

The second major movement induced by the policy is the exit from the municipal sector of individuals whose skills are below the new necessary minimum.[114] Of these individuals, a few switch to the voucher school. Their mean skills, 0.80, are lower than the mean skills of the incumbent voucher school teachers, 4.92. Therefore, they lower the mean skills in the voucher school. However, their negative effect on the composition of voucher school teachers is offset by other mobility reactions, notably by the aforementioned exit from the voucher school of the worst teachers. As a result, mean teaching skills in the voucher school on average increase.

**Effect of Policy on Parental Sorting**

---

[113]The mean skills in the municipal school increase also as an effect of the inflow of non-teachers and of the exit of the teachers from the bottom quartile.

[114]As table 32 shows, the policy also induces a minor movement of non-teachers toward the voucher school due to the increase in voucher school wages.
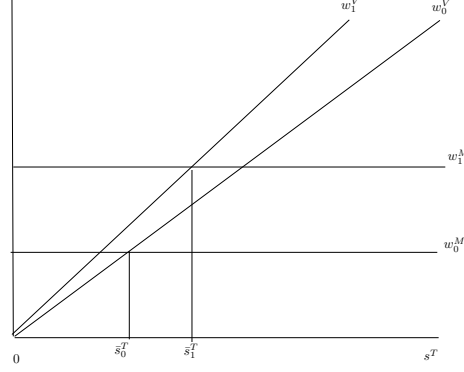
Figure 24: Role of Voucher School Wages

Overall, the number of voucher school students decreases by 49, 200. This corresponds to a three-percentage-point reduction in the voucher school share. Table 35 shows the transitions of students across school sectors. Most transitions consist of voucher school students leaving for the municipal school. Because voucher schools do not change their optimal tuitions, which remain equal to the cap, the only change affecting parents' decisions is the change in the composition of teachers across schools.

Table 35: Increasing Municipal School Wage Offers by CLP 90, 000: Transitions of Students Across Schools

| Pre-Policy Choice | Post-Policy Choice | | |
| --- | --- | --- | --- |
| | M | V | Tot |
| M | 99.82% | 0.18% | 100.00% |
| V | 4.79% | 95.21% | 100.00% |

Looking at the transition probabilities by student unobserved type and by household income reveals that less wealthy students and students of type two are more likely to switch to the municipal school. Table 36 reports the fractions of voucher school students of each type who switch to the municipal sector. Table 37 reports the fractions of voucher school students with differing household incomes who switch to the municipal school.

Table 36: Increasing Municipal School Wage Offers by CLP 90, 000: Transitions of Students by Type

| Type | Frac. Switching from V to M |
| --- | --- |
| 1 | 1.69% |
| 2 | 8.35% |
| 3 | 3.50% |

Type two individuals are more likely than other types to switch to the municipal school because, as Appendix I shows, they are the ones whose cognitive achievement benefits the most from an increase in teaching skills in the municipal sector. This is reflected in the change induced by the policy on the treatment on the treated: using the model, I simulated the pre- and post-policy difference between the test score attainable in a voucher school and that attainable in a municipal school for all students who before the policy choose the voucher school. This is a measure of the treatment on the treated. As an effect of the policy, the average treatment on the treated decreases by 11.6% of a standard deviation. However, this

Table 37: Increasing Municipal School Wage Offers by CLP 90,000: Transitions of Students by Hh Income (CLP)

| Hh Income | Frac. Switching from V to M |
|---|---|
| [0-100,000] | 6.73% |
| (100,000-150,000] | 6.02% |
| (150,000-350,000] | 4.81% |
| 350,000+ | 2.82% |

decrease is more pronounced for type two individuals: for them the decrease is 17.2% of a standard deviation; for type one individuals, who are less likely to move to the municipal school than type two individuals but more likely than type three, it is 11.6% of a standard deviation; and finally for type three students, it is only 6.1% of a standard deviation. The decrease in the treatment on the treated is constant across income levels, however, poorer families are more likely to leave as the benefits from the voucher school decrease because of the concavity in the utility from consumption.

**Government Spending**

In equilibrium, average wages of municipal sector teachers increase by 9.86%.[115] Government spending increases only by 3%. The wage increase is in part financed through the increased voucher revenues in municipal schools due to the increased attendance at those schools.[116] Table 38 shows total government spending before and after the policy and government spending by sector.[117] Government spending in municipal schools increases by $204 per student per year (PPP). The increase in voucher revenues to the municipal sector is equal to $73 per student per year (PPP), which corresponds to 36% of the wage bill increase. This means that the government must inject into the system only $204 − $73 = $131 in extra funding, which corresponds to a 3.4% increase in overall education spending.

Table 38: Increasing Municipal School Wage Offers by CLP 90,000: Government Spending Per-Student Per-Year (in PPP equivalents of USD)

| | Tot. Gov. Spending | Gov. Spend. in M | Gov. Spend. in V |
|---|---|---|---|
| **Base** | 3,860 | 2,370 | 1,490 |
| **Wage Increase** | 3,991 | 2,574 | 1,417 |

### 7.2.2 Pay-Per-Skill Municipal Sector Wages

Instead of increasing municipal school wage offers by the same amount for all individuals irrespective of their skills, I change the wage offer schedule so that it is more reflective of a teacher's skills. In particular, I let the wage offer be a linear function of a potential teacher's teaching skills: $\tilde{w}_i^M = \tilde{r}^M s_i$. I refer to this wage schedule as a pay-per-skill schedule. The skill price $\tilde{r}_d^M$ in each market $d$ is set equal to the product of a constant $c$ and $r_d^M$, the implied skill price in the municipal sector, which has been discussed in section 7.1.1.

---

[115]Because the number of municipal sector teachers slightly decreases, the municipal wage bill increases by slightly less: 8.60%.

[116]The wage increase is partly financed also through government savings deriving from the decrease in government spending to finance fellowships for private education. However, this decrease is negligible, especially if compared to the increase in voucher revenues in municipal schools.

[117]The Education at a Glance OECD report for the year covered by the sample, 2006, reports that 58.6% of government spending on education is devoted to municipal schools. Simulations from my model estimate that percentage to be 61.4%.

As before I introduce a minimum competency requirement to capture selection on the part of municipal schools in the recruiting process. I find that when $c = 0.78$ and when the cutoff is set equal to the median in the distribution of teaching skills in the entire population of potential teachers, test scores increase by 17.6%. The policy requires increasing government spending on education by 28%, and as in the previous experiment, the required increase in the wage bill is partially funded by the shift of voucher resources from the voucher to the municipal sector.

Table 39 shows pre- and post-policy mean test scores and mean teaching skills overall and by school sector. Both mean test scores and teaching skills improve overall. In the municipal sector, teaching skills and test scores improve, in the voucher sector, teaching skills and test scores worsen.

Table 39: Pay-Per-Skill Municipal Wage Offers: Mean Test Scores and Teacher Skills

|              | mean test scores | mean test scores in M | mean test scores in V |
|--------------|------------------|-----------------------|-----------------------|
| **Base**         | -0.06            | -0.25                 | 0.10                  |
| **Pay-Per-Skill** | 0.12             | 0.24                  | -0.05                 |
|              | mean skills      | mean skills in M      | mean skills in V      |
| **Base**         | 2.70             | 2.01                  | 4.92                  |
| **Pay-Per-Skill** | 3.46             | 4.13                  | 2.08                  |

As noted before, the effect of the policy on test scores and the amount of government spending necessary to finance it depend on the reactions of potential teachers, parents and private schools. In the remainder of this section I analyze these policy responses, and I describe the implications for government spending.

**Policy Reaction of Potential Teacher Sorting and of Private Schools**

Overall, the (simulated) number of teachers increases from $118,000$ to $130,000$. The number of municipal school teachers decreases from $90,000$ to $86,900$ and the number of voucher school teachers increases from $28,000$ to $42,700$.[118] Table 40 shows the transitions among sectors that result from the introduction of the policy.[119]

Table 40: Increasing Municipal School Wage Offers by CLP $90,000$: Transitions Across Sectors

| | Post-Policy Choice | | | | |
|-------------------|---------|---------|---------|---------|----------|
| Pre-Policy Choice | M       | V       | NT      | H       | Tot      |
| M                 | 31.52%  | 31.04%  | 37.44%  | 0.00%   | 100.00%  |
| V                 | 88.49%  | 11.51%  | 0.00%   | 0.00%   | 100.00%  |
| NT                | 6.48%   | 2.17%   | 91.24%  | 0.00%   | 100.00%  |
| H                 | 0.00%   | 0.00%   | 0.00%   | 100.00% | 100.00%  |

The policy attracts individuals from both the non-teaching sector and the voucher school. Table 41 shows that, unlike in the previous experiment, among the non-teachers those with higher teaching skills are more likely to switch to the municipal school. The table shows the fractions of individuals in the non-teaching sector who switch to the municipal sector, by skill level. Skill level is measured in terms of percentiles

---

[118]Average class size in municipal schools increase by less than three, from 10.0 to 10.29, and average class size in the voucher schools decreases from 34.6 to 17.5.

[119]The fraction from the second quartile that stays in the municipal sector is slightly above 0%, and not exactly 0%, because of simulation error and because the median in the distribution of skills in the municipal school is slightly above the median in the distribution of skills in the entire population, which is the minimum competency cutoff.

of the distribution of skills in the population. Individuals below the median skill level do not meet the minimum competency requirements: none of them switches to the municipal school and, therefore, they are not included in the table. Among the remaining individuals, those with higher skills switch more frequently to the municipal sector. This means that pay-per-skill wage schedules are better able to attract highly skilled individuals into the teaching profession than flat wage increases.

Table 41: Pay-Per-Skill Municipal Sector Wages: Transitions From NT to M by Skill Percentile

| Skill Percentile Interval | Frac. of NT Switching to M |
|---|---|
| [50-62.5] | 8.25% |
| (62.5-75] | 9.77% |
| (75-87.5] | 19.62% |
| (87.5-100] | 21.39% |

Unlike a flat wage increase, this policy attracts the most highly skilled teachers from the voucher school: the mean skills of those who switch to the municipal sector are 5.18, the mean skills of those who don't are 3.07. This happens in spite of the fact that the voucher school increases its wage offers substantially: the average voucher school skill price increases from 0.76 to 1.22, which results in a 73.7% average increase in voucher school wage offers.[120] The mobility of teachers from the voucher school to the municipal sector has the effect of increasing mean skills in the municipal sector and of decreasing mean skills in the voucher schools. Mean skills in the voucher school, moreover, are further decreased by the inflow of low-skilled teachers who leave the municipal sector because they fail to meet the minimum competency requirements. Their mean skills are 1.81, which is lower than the mean skills of the incumbent voucher school teachers.

**Effect of Policy on Parental Sorting**

Overall, the number of voucher school students decreases by 223,000. This corresponds to a twelve-percentage-point reduction in the voucher school share (from 51.8% to 39.9%). Table 42 shows the transitions of students across school sectors. Some voucher school students leave for the municipal school. Because voucher schools do not change their optimal tuitions, which remain equal to the cap, the only change affecting parents' decisions is the change in the composition of teachers across schools.

Looking at the transition probabilities by students' unobserved type and by household income reveals that less wealthy students and students of type two are more likely to switch to the municipal school. Table 43 reports the fractions of voucher school students of each type who switch to the municipal sector. Table 44 reports these fractions by household income.

---

[120]Voucher schools cannot retain the best teachers because the skill price in the municipal school is higher: 1.84. In counterfactuals, I assumed free exit and imposed that a voucher school exit the market if its profits are negative. Profits, however, can be recovered in estimation only up to a constant. This is because the fixed operating costs cannot be retrieved. However, an interval for those costs can be inferred. First, the fixed operating costs must be non-negative. Second, they can be at most equal to the gross profits (i.e., profits plus operating costs) that are retrieved in estimation. This is an upper bound because if the operating costs were above these gross profits, the schools that are observed in the market would be making negative profits: under the assumption of free exit this would not be possible. Once I had established an interval for the operating costs, I ran the policy counterfactual for a number of operating costs within this interval: the results of all counterfactual experiments are similar and, therefore, I report only the results for the case of zero operating costs. Finally, in simulating the voucher school response I maximized true profits and not approximated profits: I established that the optimal tuition was at the cap, and then maximized true profits with respect to the skill price. I did this because when municipal wage schemes are pay-per-skill, the approximation to the profit function of the voucher school worsens. This is because the supply of teachers is less smooth when municipal wages are a linear function of skills (intuitively, it suffers major variations when the municipal sector skill price is close to the voucher sector skill price).

Table 42: Pay-Per-Skill Municipal Sector Wages: Transitions of Students Across Schools

| Pre-Policy Choice | Post-Policy Choice | | |
|:---:|:---:|:---:|:---:|
| | M | V | Tot |
| M | 100.00% | 0.00% | 100.00% |
| V | 28.48% | 71.52% | 100.00% |

Table 43: Pay-Per-Skill Municipal Sector Wages: Transitions of Students by Type

| Type | Frac. Switching from V to M |
|:---:|:---:|
| 1 | 22.38% |
| 2 | 41.45% |
| 3 | 16.83% |

As in the previous policy experiment, type two individuals are more likely to switch because, as Appendix I shows, they are the ones whose cognitive achievement benefits the most from the increase in teaching skills in the municipal sector. Unlike in the previous experiment, however, the decrease in the effect of voucher schools on achievement is more pronounced for the less wealthy families among those who, before the policy change, choose the voucher school.[121] This means that the transitions of students from the voucher school to the municipal school are due both to income effects (the voucher school has less qualified teachers than before the policy, yet the tuition does not change) and to the detrimental effect that the decline in teacher quality has on the poorest students. Finally, test scores decrease, on average, by 26% of a standard deviation for the voucher school students who remain in the voucher school sector in spite of the worsening of its teachers. Test scores, on the other hand, increase, on average, by 49% of a standard deviation for the voucher school students who switch to the municipal school as an effect of policy, and by 59% of a standard deviation for the students who choose the municipal school both before and after the introduction the the policy.

**Government Spending**

In equilibrium, average wages of municipal sector teachers increase by 67.0%.[122] Government spending increases by 28%. The wage increase is in part financed through the increased voucher revenues in municipal schools due to the increased attendance at those schools.[123] Table 45 shows total government spending before and after the introduction of the policy and government spending by sector. Government spending in municipal schools increases by \$1,430 per student per year (PPP). The increase in voucher revenues to the municipal school is equal to \$340 per student per year (PPP), which corresponds to 23.8% of the wage bill increase. This means that the government must inject into the system only \$1,430 − \$340 = \$1,090 in extra funding, which corresponds to a 28.2% increase in overall education spending.

---

[121]The decreases in treatment effect on the treated by income categories (from the poorest to the wealthiest) are $1.57, 1.27, 1.13, 1.06$ standard deviations.

[122]Because the number of municipal sector teachers slightly decreases, the municipal wage bill increases by slightly less: 62.0%.

[123]As before, the wage increase is partly financed also through government savings deriving from the decrease in government spending to finance fellowships for private education. However, this decrease is negligible, especially if compared to the increase in voucher revenues in municipal schools.

Table 44: Pay-Per-Skill Municipal Sector Wages: Transitions of Students by Hh Income (CLP)

| Type | Frac. Switching from V to M |
|---|---|
| [0-100,000] | 38.93% |
| (100,000-150,000] | 34.44% |
| (150,000-350,000] | 27.97% |
| 350,000+ | 19.42% |

Table 45: Pay-Per-Skill Municipal Sector Wages: Government Spending Per-Student Per-Year (in PPP equivalents of USD)

| | Tot. Gov. Spending | Gov. Spend. in M | Gov. Spend. in V |
|---|---|---|---|
| **Base** | 3,860 | 2,370 | 1,490 |
| **Wage Increase** | 4,950 | 3,800 | 1,150 |

# 8  Conclusion

This work shows that wages for public school teachers that are more reflective of a teacher's skills than the current Chilean government formulae are better able to attract highly skilled individuals into municipal schools than wage increases that are equal for individuals of all skills. This work quantifies the effects of these policies not only on the composition of teachers across schools, but also on test scores. This is possible because the estimated model used as a basis for policy evaluation is able to predict the policy reaction of teacher quality, together with the effect of teacher quality on student achievement. Moreover, the model predicts the policy reactions of parents and of non-public schools. In an education system with a large-scale school choice program, like the education system in Chile, these reactions must be taken into account when evaluating the effect of a policy on the distribution of cognitive achievement. I provide examples of two public policies: a flat wage increase that improves mean test scores by 8% of a standard deviation, and the introduction of a pay-per-skill wage schedule that improves mean test scores by 18% of a standard deviation.[124]

Finally, this work shows that school choice programs have implications not only for the evaluation, but also for the design of policies targeted at teacher labor markets. In particular, the policy reaction of parents, which can be expected to be of a larger degree in education systems with more extensive school choice programs, has been shown to play a major role in the determination of the costs of a public policy. In particular, a policy that improves the quality of teachers in the municipal school would increase the demand for public education, which would increase the voucher revenues to the public sector. The increase in government spending in public schools needed to finance the public policies would, therefore, be partially financed by the increased voucher revenues to public schools. In the two examples that I analyze, the increased voucher revenues cover, respectively, over one-third and one-fourth of the policy costs.

---

[124]Both policies assume that municipal schools are able to select the best teachers among those willing to accept their wage offers.

# 9 Appendices

## Appendix A: Governmental Formulae for Private School Revenues

In this appendix I report the formulae that can be found in article 25 of the law *Decreto con Fuerza de Ley N° 2, De Educacion, de 20.08.98*. Let $\check{p}$ denote the tuition charged by the school, and let $\check{p} - voucher = p$ denote the difference between this amount and the value of the voucher. Each household is responsible for the payment of $p - f(X_h, p)$ where $f(X_h, p) \geq 0$ is the amount of fellowship received by the family. The amount of the per-pupil subsidy that is effectively received by the private school decreases as the tuition payments in the school increase. Let $EPV(p, r)$ denote the mean tuition payments in the school:

$$EPV(p,r) = \int \big(p - f(x,p)\big)g^x(x|\text{ V chosen}; p, v)dx \, EPV(p,r) = \int \big(p - f(x,p)\big)g^x(x|\text{ V chosen}; p, v)dx \tag{13}$$

where the conditional density $g^x(x|\text{ V chosen}; p, v)$ is indexed by $(p, r)$ because it is the outcome of the sorting of parents, which is a function of $(p, r)$. I drop the dependence of $EPV$ on $(p, r)$ for simplicity. The per-pupil revenues of the private school are given by:

$$g^r(EPV) = \begin{cases} \check{p} & \text{if } EPV \leq 0.5 \\ \check{p} - 10\%(EPV - 0.5USE) & \text{if } 0.5 < EPV \leq 1USE \\ \check{p} - 10\%(EPV - 0.5USE) - 20\%(EPV - 1USE) & \text{if } 1USE < EPV \leq 2USE \\ \check{p} - 10\%(EPV - 0.5USE) - 20\%(EPV - 1USE) - 35\%(EPV - 2USE) & \text{if } 2USE < EPV. \end{cases}$$

where $USE$ stands for *Unidad de Subvención Educacional*.

The school is required to partially cover the fellowship expenses. The law provides a formula for the per-pupil contribution of the school to financing the fellowships. The per-pupil expenses for financial aid due by the school are given by:

$$g^f(EPV) = \begin{cases} 5\%EPV & \text{if } EPV \leq 1USE \\ 5\%EPV + 7\%(EPV - 1USE) & \text{if } 1USE < EPV \leq 2USE \\ 5\%EPV + 7\%(EPV - 1USE) + 10\%(EPV - 2USE) & \text{if } 2USE < EPV. \end{cases}$$

Therefore, the net per-pupil revenues are $g^r(EPV) - g^f(EPV)$.

## Appendix B: Derivation of Labor Supply and School Choice

### Labor Supply

In this section I derive the proportion of potential teachers with characteristics $[Z_i \quad D_i]' = q_i$ choosing the voucher school. I do so by first deriving pairwise cutoffs, i.e., cutoffs that indicate when the voucher school is chosen in pairwise comparisons.

The voucher school is preferred to the municipal school by a potential teacher with characteristics $[Z_i \quad D_i]' = q_i$ when the shock to the municipal school wage is below a cutoff $b_M(\epsilon_i^V, l_i; q_i)$:

$$Pr(\epsilon_i^M \le \underbrace{ln(r) + \alpha_{0V}(l_i) - \alpha_{0M}(l_i) + (\alpha_V - \alpha_M)'Z_i + \mu_i^V - \mu_i^M + \epsilon_i^V}_{b_M(\epsilon_i^V, l_i; q_i)}).$$

Similarly, the voucher school is preferred to the non-teaching sector when the shock to the non-teaching sector wage is below a cutoff $b_{NT}(\epsilon_i^V, l_i; q_i)$:

$$Pr(\epsilon_i^{NT} \le \underbrace{ln(r) + \alpha_{0V}(l_i) - \alpha_{0NT}(l_i) + (\alpha_V - \alpha_{NT})'Z_i + \mu_i^V + \epsilon_i^V}_{b_{NT}(\epsilon_i^V, l_i; q_i)}).$$

Finally, the voucher school is preferred to the home option when the shock to the preference for home is below a cutoff $b_H(\epsilon_i^V, l_i; q_i)$:

$$Pr(\epsilon_i^H \le \underbrace{ln(r) + \alpha_{0V}(l_i) + \alpha_V'Z_i + \mu_i^V - \mu_i^H + \epsilon_i^V}_{b_H(\epsilon_i^V, l_i; q_i)})).$$

The proportion of potential teachers with characteristics $[Z_i \quad D_i]' = q_i$ who choose the voucher school is:

$$Pr_W^V|q_i = \sum_{l_i=1}^{\bar{L}} \psi_{l_i} \int_{\epsilon_i^V=-\infty}^{+\infty} \int_{\epsilon_i^M=-\infty}^{b_M(\epsilon_i^V, l_i; q_i)} \int_{\epsilon_i^{NT}=-\infty}^{b_{NT}(\epsilon_i^V, l_i; q_i)} \int_{\epsilon_i^H=-\infty}^{b_H(\epsilon_i^V, l_i; q_i)} f(\epsilon_i^V, \epsilon_i^M, \epsilon_i^{NT}, \epsilon_i^H) d\epsilon_i^H d\epsilon_i^{NT} d\epsilon_i^M d\epsilon_i^V$$

where $f(\epsilon_i^V, \epsilon_i^M, \epsilon_i^{NT}, \epsilon_i^H)$ is the joint density of the wage and preference shocks, which are distributed normally with a diagonal variance covariance matrix.

**School Choice**

In this section I derive the proportion of parents with characteristics $X_h = x$ who choose the voucher school. Denote by $b(k_h, \nu_h^V, \nu_{h\eta}; x)$ the cutoff value that is such that if $\nu_h^M < b(k_h, \nu_h^V, \nu_{h\eta}; x)$, the voucher school is preferred to the municipal school by parents with characteristics $X_h = x$. This cutoff is equal to:

$$
\begin{aligned}
b(k_h, \nu_h^V, \nu_{h\eta}; x) =\ & \tau(k_h)ln\left(\frac{y_h - p_h}{y_h}\right) + \beta_{0V}(k_h) - \beta_{0M}(k_h) + \beta_{1V}(k_h)\bar{s}^V - \beta_{1M}(k_h)\bar{s}^M \\
& + (\beta_{2V}(k_h) - \beta_{2M}(k_h))' x - \eta_h^M(k_h) - \nu_{h\eta}
\end{aligned}
$$

where $p_h = p - v - f(x, p)$.

## Appendix C: Solution to the Constrained Maximization of Approximated Profits

The approximated problem of the firm is the following:[125]

$$\max_{(p,r)} \hat{\Pi}$$
$$p \le \bar{p} \quad \text{w/ multiplier} \quad \lambda$$

---

[125]According to the Chilean law, the cap is defined over the average payments made at the school, $EPV$. For simplicity the model defines the cap on $p$, the difference between the tuition chosen by the school and the value of the voucher.

or equivalently:

$$\max_{(p,r)} \hat{a}_1 + \hat{a}_2 p + \hat{a}_3 p^2 + \hat{a}_4 r + \hat{a}_5 r^2 + \hat{a}_6 pr + \hat{a}_7 p^3 + \hat{a}_8 r^3 + \hat{a}_9 p^2 r + \hat{a}_{10} pr^2$$
$$p \leq \bar{p} \quad \text{w/ multiplier} \quad \lambda$$

I solve for the optimal $(p^*, r^*)$ and for the Kuhn-Tucker-Lagrange multiplier $\lambda^*$ following the procedure described in Judd (1998), Ch. 4, p.122. At the optimum, the inequality constraint is either binding or not binding. I find the set of solutions to the Kuhn-Tucker conditions under both configurations. Among the feasible solutions thus found, I select the one with the highest value of approximated profits. The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial p} = \hat{a}_2 + 2\hat{a}_3 p + \hat{a}_6 r + 3\hat{a}_7 p^2 + 2\hat{a}_9 pr + \hat{a}_{10} r^2 - \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial r} = \hat{a}_4 + 2\hat{a}_5 r + \hat{a}_6 p + 3\hat{a}_8 r^2 + \hat{a}_9 p^2 + 2\hat{a}_{10} pr = 0$$

**Case i):** when the constraint is not binding at the optimum, the Kuhn-Tucker-Lagrange multiplier is equal to zero. I set $\lambda = 0$ and I use Newton method to solve numerically the following system of two equations in the two unknowns $(p_V, r)$:

$$\begin{cases} \hat{a}_2 + 2\hat{a}_3 p + \hat{a}_6 r + 3\hat{a}_7 p^2 + 2\hat{a}_9 pr + \hat{a}_{10} r^2 = 0 \\ \hat{a}_4 + 2\hat{a}_5 r + \hat{a}_6 p + 3\hat{a}_8 r^2 + \hat{a}_9 p^2 + 2\hat{a}_{10} pr = 0 \end{cases}$$

**Case ii):** when the constraint on $p$ is binding at the optimum, $p = \bar{p}$. I use Newton method to solve numerically the following system of two equations in the two unknowns $(r, \lambda)$:

$$\begin{cases} \hat{a}_2 + 2\hat{a}_3 \bar{p} + \hat{a}_6 r + 3\hat{a}_7 \bar{p}^2 + 2\hat{a}_9 \bar{p} r + \hat{a}_{10} r^2 - \lambda = 0 \\ \hat{a}_4 + 2\hat{a}_5 r + \hat{a}_6 \bar{p} + 3\hat{a}_8 r^2 + \hat{a}_9 \bar{p}^2 + 2\hat{a}_{10} \bar{p} r = 0 \end{cases}$$

## Appendix D: Heterogeneity Across Schools in the Same Sector and Market

Table 46 shows the mean and the standard deviation of the teaching experience of teachers in each market and school sector. I interpret this standard deviation as capturing heterogeneity in school quality. The table shows that the heterogeneity of school quality is not necessarily larger in markets with a larger number of schools. For example, in the Santiago metropolitan region, which is the largest market, the standard deviations of experience in the municipal and private schools are 11.0 and 10.6, respectively. In market 12, however, which has less than a quarter of the municipal schools than Santiago has, the standard deviation of teaching experience in municipal schools ls is 12.3, which is higher than the corresponding figure in Santiago.

Similarly, in market 5, which has fewer voucher schools than Santiago has, the standard deviation of teaching experience in voucher schools is higher than in Santiago. This suggests that larger markets do not necessarily contain more heterogeneous schools and, therefore, that the assumption of a unique municipal and a unique voucher school per market is equally accurate for larger and smaller markets.

Table 46: Markets and Teaching Experience

| Market | Mean Texp M | St. Dev. Texp M | Mean Texp V | St. Dev. Texp V |
|--------|-------------|-----------------|-------------|-----------------|
| 1 | 23.1 | 10.7 | 13.5 | 9.1 |
| 2 | 20.1 | 9.0 | 16.6 | 10.9 |
| 3 | 23.9 | 11.1 | 11.7 | 8.6 |
| 4 | 19.9 | 10.1 | 9.8 | 8.3 |
| 5 | 24.5 | 11.1 | 15.9 | 11.4 |
| 6 | 25.7 | 10.2 | 12.9 | 10.6 |
| 7 | 25.2 | 9.4 | 16.1 | 9.9 |
| 8 | 27.4 | 9.5 | 11.8 | 10.4 |
| 9 | 24.3 | 8.9 | 6.1 | 7.8 |
| 10 | 25.4 | 9.1 | 7.7 | 5.8 |
| 11 | 19.7 | 11.0 | 12.5 | 9.8 |
| 12 | 27.2 | 12.3 | 20.0 | 10.4 |
| 13 | 30.1 | 8.9 | 13.2 | 10.4 |
| 14 | 27.0 | 9.4 | 17.8 | 10.7 |
| 15 | 21.2 | 11.0 | 13.8 | 10.6 |
| 16 | 22.6 | 10.1 | 15.8 | 11.5 |
| 17 | 30.7 | 9.2 | 17.8 | 10.7 |
| 18 | 20.2 | 10.4 | 13.3 | 11.0 |

## Appendix E: Integrating Out Skill Price, $r$, in a One-Step Estimation Approach

This appendix uses the example of a maximum likelihood estimation to show why in a standard one-step estimation approach unobserved $r$ must be integrated out. It also uses the likelihood example to illustrate the two-step approach adopted in the paper.

Let $p_d^o$ be the observed tuition in market $d$, and let $OS_d^o$ be the data on the sample of parents and potential teachers in market $d$. Let there be two solutions to the profit maximization, for all parameters and error shocks: $H$ and $L$. Let the solution selection rules available to the school be denoted by $s \in \{L, H\}$. Let $f^{(L)}(\cdot)$ denote the density over $(p, r)$ generated by the model when the school commits to $L$, and let $f^{(H)}(\cdot)$ be the density when the school commits to $H$. Let $f^{II}(\cdot)$ be the density associated with the second-stage sampling error: the model yields a non-stochastic equilibrium of the second stage for the super-populations of potential teachers, $[0, W]$, and parents $[0, P]$.[126] The observed sample, however, is composed of random draws from these super-populations. The choice fractions and density of wages and test scores in the super-populations implied by the model equilibrium are used to construct the choice probabilities and density of observed data that enter the likelihood. The contribution to the likelihood of market $d$ for each solution selection rule can be written as:

---

[126]I refer to these as super-populations because they are defined on a continuum. A population of individuals is composed of a countable number of individuals. A sample from this population is a finite subset of individuals.

$$l_d^{(L)} = \int_{\tilde{r}} f^{(L)}(p_d^o, \tilde{r}|\theta_I, \theta_{II}) f^{II}(OS_d^o|p_d^o, \tilde{r}, \theta_{II}) d\tilde{r}$$

$$l_d^{(H)} = \int_{\tilde{r}} f^{(H)}(p_d^o, \tilde{r}|\theta_I, \theta_{II}) f^{II}(OS_d^o|p_d^o, \tilde{r}, \theta_{II}) d\tilde{r}$$

and the likelihood is obtained as:

$$L^{(L)} = \prod_{d=1}^{D} l_d^{(L)}$$

$$L^{(H)} = \prod_{d=1}^{D} l_d^{(H)}.$$

Notice that the second-stage density $f^{II}(\cdot)$ is independent of the solution selection rule, $L, H$. Notice also that the second-stage density is not conditioned on $\theta_I$, because conditional on $(p_d^o, \tilde{r})$ the equilibrium of the second stage does not depend on $\theta_I$, so that $f^{II}(OS_d^o|p_d^o, \tilde{r}, \theta_I, \theta_{II}) = f^{II}(OS_d^o|p_d^o, \tilde{r}, \theta_{II})$.

Although the two-step approach adopted in the paper estimates $\theta_{II}$ and $r$ in the first step using the Method of Simulated Moments, for the sake of argument I show what the procedure would be under the likelihood example. The first step obtains $\hat{\theta}_{II}, \hat{r}$ by maximizing the likelihood:

$$L^{II} = \prod_{d=1}^{D} l_d(r, \theta_{II}|OS_d^o, p_d^o)$$

where $l_d(r, \theta_{II}|OS_d^o, p_d^o) = f^{II}(OS_d^o|r, \theta_{II}, p_d^o)$. No variation in $p_d^o$ is used to estimate $\theta_{II}$ and $r$ because tuitions are constant across markets (always equal to the cap) and all price variation comes from the exogenous fellowship formulae. The second step takes $\hat{\theta}_{II}$ and $\hat{r}$ as given; it assumes that the solution selection rule adopted by each school is known, for example, it is $L$ for all, and it maximizes:

$$L^{(L)} = \prod_{d=1}^{D} f^{(L)}(p_d^o, \hat{r}_d|\theta_I, \hat{\theta}_{II}).$$

## Appendix F: Moment Conditions Used in Step One

In the first step I estimate the technology and preference parameters that enter the parents' and the potential teachers' problem, along with the teaching skill prices in each market. I compute 607 moments, 321 pertaining to parents and 286 to college graduates.

### Parents' Moments: Matching Choices, Test Scores and Fellowship Amounts

The moment conditions capture how the endogenous variables vary as the exogenous variables vary. The exogenous variables of the parents' side are those that enter the fellowship formula, the production of achievement and the preference for a certain type of school. The variables that enter the fellowship formula are $primaria_h, nfam_h, rural_h, y_h$, those that enter the production of achievement are $\frac{y_h}{nfam_h}, peduc_h$ and the preference for the municipal school is affected by the variables $primaria_h, rural_h$. Achievement is affected also by school inputs, which vary by market. I use the following categories:

- family size $nfam_h$: $[2,3], [4,6], \geq 7$

- monthly income in terms of CLP100,000 $y_h$: $[0, 0.5], (0.5, 1.5], (1.5, 2.5], (2.5, 3.5], (3.5, 4.5], (4.5, 5.5],$ $(5.5, 7], (7, 9], (9, 11], > 11$

- average parental education in years $peduc_h$: $[0 - 6.5], (6.5, 8], (8, 9.5], (9.5, 10.5], (10.5, 11.5], (11.5, 12],$ $(12, 12.5], (12.5, 13], (13, 14], > 14$

- monthly income in terms of CLP100,000 divided by family size, $\frac{y_h}{nfam_h}$: $[0, 0.15], (0.15, 0.25], (0.25, 0.36],$ $(0.36, 0.45], (0.45, 0.50], (0.50, 0.70], (0.70, 0.84], (0.84, 1.13], (1.13, 1.75], > 1.75$

I partition the state of observable exogenous variables and build an indicator for whether an observation belongs to a certain element of the partition. The constrained moment conditions are then obtained by multiplying the difference between actual and predicted outcomes by this indicator, and by an indicator for the school sector in the case of test scores and tuition payments (the latter are made only by private sector parents). The constrained moment conditions use the following outcomes (number of moment conditions in parenthesis):

- Test scores constrained by a sector dummy and dummies for:

  - market (18x2=36)
  - monthly income per capita and parental education (10x10x2=200)

- Fraction choosing voucher school constrained by dummies for:

  - market (18)
  - parental education (10)
  - monthly income (10)
  - number of individuals in the family (3)
  - elementary school (2)
  - rurality of the household's residence (2)

- Tuition payments made by the household constrained by a private school dummy and dummies for:

  - elementary school, number of individuals in the household, rurality of the residence (2x3x2=12)
  - monthly income (10)
  - market (18)

Total number of parents' moments: 321. Total number of parents' parameters: 79.

**Potential Teachers' Moments: Matching Choices and Accepted Wages**

The moment conditions capture how the endogenous variables in the model vary as the exogenous variables vary. The exogenous variables of the potential teachers' side are those that enter the wage offer functions and the non-pecuniary utility from staying at home and from choosing to teach. The exogenous variables that enter the wage functions are: $age_i, fem_i, perfec_i, grad\_degree_i$ where $age_i$ is age, $fem_i$ is a gender dummy, $perfec_i$ is a dummy for whether the individual holds professional certifications, and $grad\_degree_i$ is a dummy for whether the individual holds a graduate degree (masters or Ph.D.). Wages vary by market too. The exogenous variables that enter the non-pecuniary utility are: $fem_i, nkids_i, nkids2_i, nkids3 - 6_i, age_i$ where $nkids_i$ is the number of children in the college graduates' household, $nkids2_i$ is the number of children age 0 to 2, and $nkids3 - 6_i$ is the number of children age 3 to 6. Gender affects the preference for the teaching profession.

I use the following categories:

- coarse age, $age_i$: $[20 - 30], [31 - 40], [41, 50], \geq 51$

- fine age, $age_i$: $[20, 31], (31, 36), (36, 39], (39, 45], (45, 48], (48, 52], (52, 56], > 56$

- number of children in the household, $nkids_i$: 0,1,2,$\geq 3$

- number of children aged 0-2, $nkids2_i$: 0, $\geq 1$

- number of children aged 3-6, $nkids3 - 6_i$: 0, $\geq 1$

I partition the state of observable exogenous variables and build an indicator for whether an observation belongs to a certain element of the partition. The constrained moment conditions are then obtained by multiplying the difference between actual and predicted outcomes by this indicator, and by an indicator for the occupational choice in the case of accepted wages. The constrained moment conditions use the following outcomes (number of moment conditions in parenthesis):

- Accepted wages constrained by a sector dummy (3 working options) and dummies for:

    - age, gender, professional certifications (3x4x2x2=48)
    - graduate degree (3x2=6)
    - market (3x18=54)

- Fractions in sector M, V and NT constrained by dummies for (exclude one sector to avoid multi-collinearity and hence singularity of the variance-covariance matrix of the moment conditions):

    - professional certifications (3x2=6)
    - age, gender, graduate degree (3x4x2x2=48)
    - market (3x18=54)
    - gender, number of kids (3x2x4=24)
    - number of kids up to 2 years of age, age (3x2x4=24)
    - number of kids of age 3 to 6 (3x2=6)

- Accepted wages in the teaching occupations (2) by finer age category (2x8=16)

Total number college graduates' moments: 286. Total number college graduates' parameters: 115 (+17 skill prices).

## Appendix G: Moment Conditions from the ELD Dataset, Accounting for Choice-Based Sampling

The ELD sample contains individuals who chose to teach in either the Municipal school ($M$) or the Private Voucher school ($V$). Therefore it is a selected sample based on the potential teachers' choice of becoming a teacher. When choice depends on unobservables, the distribution of unobservables in the selected sample is likely to be different from the distribution of unobservables in a random sample of potential teachers. To account for this difference, I use the following simulation technique.

Consider an outcome $y_i$, which could either be a sector choice or a wage. Let $sim = (s, l)$ be a joint draw of a shock vector, indexed by $s$, and of a type $l$. Draws are obtained from the joint density of shocks and types.[127] I simulate an outcome $\tilde{\tilde{y}}_i$ for each individual for a large number of draws of $sim$. Let $sim^*(s^*, l^*)$ be a draw that is such that individual $i$ of type $l^*$ and with shock draw $s^*$ chooses either sector M or sector V, which are the only choices observed in the ELD sample. Let $Sim^*$ be the total number of such draws. Then, the simulated outcome for individual $i$ is obtained by Crude Frequency Simulation, where only the draws $sim^*$ are used:

$$\tilde{\tilde{y}}_i(\omega_i, \theta) = \frac{1}{Sim^*} \sum_{sim^*=1}^{Sim^*} \hat{y}_i(\omega_i, sim^*, \omega).$$

## Appendix H: Adjusting the Variance-Covariance Matrix of the Second-Step Estimator

### Estimation of the Asymptotic Variance: Two-step Variance Adjustment

The variance in (12) assumes that the true values of the step-one parameters are known. In fact, the true values are not not known and the estimation uses the step-one estimates. The asymptotic variance-covariance matrix of the second-step parameters $\hat{\theta}_I$ must be corrected to account for the fact that the first-step parameters are estimated. In two-step approaches, this correction is needed whenever inconsistency of the first-step estimator implies inconsistency of the second-step estimator, which is the case in this context. Although both the first and second step can be viewed as method of moment estimators, where the second-step moments coincide with the score vector, the standard method of stacking the first- and second-step moments and deriving the variance of the second-step parameters by standard GMM arguments is complex in this context.[128] This derives from the fact that the asymptotics of the second stage is on the number of markets, whereas the asymptotics of the first stage is on the number of individuals within a market. I therefore estimate the correct standard errors by bootstrap methods.

---

[127]Shocks and types are independent random variables so that the draw $(s, l)$ can be obtained by drawing a shock from the shock density and then independently drawing a type from the type probability mass function.

[128]See Newey and McFadden (1992).

**Bootstrapping the Standard Errors of the Second-Step Parameters**

Some components of $\theta_{II}$ are market specific; denote the subset of $\theta_{II}$ that is specific to market $d$ with $\theta_{II}^{(d)}$. I adopt the following bootstrapping algorithm:

1. From the joint distribution $\{p^d, r^d, \hat{\theta}_{II}^{(d)}\}_{d=1,\ldots,D}$ draw $B$ bootstrapped samples, with replacement, of size $D$, $\left\{\{p^d, r^d, \hat{\theta}_{II}^{(d)}\}_{d=1,\ldots,D}^b\right\}_{b=1,\ldots,B}$.[129]

2. Within each bootstrapped sample $b$, obtain the NPSML estimate $\hat{\theta}_I^b$.

3. Estimate the variance of $\hat{\theta}_I$ from the sample $\{\hat{\theta}_I^b\}_{b=1,\ldots,B}$:

$$\hat{\Omega} = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}_I^b - \bar{\hat{\theta}}_I\right)^2$$

where $\bar{\hat{\theta}}_I$ is the sample mean: $\bar{\hat{\theta}}_I = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_I^b$.

**Accounting for the Variability Introduced by the Uncertainty over $r$**

The asymptotic variance-covariance matrix of the second-step estimates, estimated via bootstrap methods, does not account for the variability introduced by the fact that the skill prices $r^d$ are not observed but they are estimated in the first step. To account for such variability, I use a standard technique in the field of statistical analysis with missing data: multiple imputation. In the first step of the estimation I obtain $\hat{r} = (\hat{r}^d)_{d=1,\ldots,D}$. I then draw $M$ vectors $\hat{r}^m$ from the known (asymptotic) distribution of $\hat{r}$, where its asymptotic variance is replaced by the consistent estimate in (11). For each draw $m$ I build a dataset $p, \hat{r}^m$ and I obtain $\hat{\theta}_I^m$ as the maximum of the simulated likelihood function based on the $m^{th}$ dataset. I estimate the variance of each estimator $\hat{\theta}_I^m$ by bootstrapping. Denote by $\Omega_b^m$ the bootstrapped variance of $\hat{\theta}_I^m$. The combined estimate of $\theta_I$ is:

$$\bar{\hat{\theta}}_I = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_I^m.$$

The variability of this estimate has two components: the average within dataset variances

$$\bar{\Omega}_M^W = \frac{1}{M} \sum_{m=1}^{M} \Omega_b^m$$

and the between datasets variance

$$\Omega_M^B = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\theta}_I^m - \bar{\hat{\theta}}_I)'(\hat{\theta}_I^m - \bar{\hat{\theta}}_I).$$

The total variability is given by:[130]

---

[129] Drawing bootstrapped samples of parents and potential teachers, and obtaining $MSM$ estimates of the first-step parameters at each bootstrapped sample would be computationally prohibitive. I avoid doing so by drawing bootstrapped samples of $\hat{\theta}_{II}^{(d)}$ instead. I have confidence that this adjustment does not jeopardize consistency of the estimate of the variance-covariance matrix because of the smoothness of the moment conditions used in the $MSM$.

[130] $\frac{M+1}{M}$ is an adjustment for finite $M$.

$$\Omega^T = \bar{\Omega}_M^W + \frac{M+1}{M}\Omega_M^B.$$

## Appendix I: Estimates

**Potential Teachers**

I let the number of types be $\bar{L} = 3$ and estimate the type proportions to be 19.7% for type 1, 42.4% for type 2 and 37.9% for type 3. In the following tables I report the parameters of the wage offer functions by sector and the non-pecuniary preference parameters. The log-wage intercepts in the municipal and non-teaching sectors vary by type and location, for ease of exposition I do not report them in the following tables, the interested reader can refer to the online appendix at http://economics.sas.upenn.edu/ mtincani/jobmarket.html. In the first step of the estimation I also obtain the skill prices of teaching skills by market, which can also be found here. In estimation, wages are expressed in terms of CLP $100,000$. I use the following notation for statistical significance: three stars $(^{***})$ mean that the parameter is significantly different from zero in a two-sided Wald test at the 1% confidence level, two stars $(^{**})$ index significance at the 5% level and one star $(^{*})$ at the 10% confidence level. The absence of stars means that the p-value is above 10%.

### Parameters of the Municipal School Log-Wage Offers

| Parameter | Description | Value | Standard Error |
|---|---|---|---|
| $\alpha_{1M}$ | Coefficient on Age | $3.99\text{e-}02^{(***)}$ | 1.08e-03 |
| $\alpha_{2M}$ | Coefficient on Age Squared | -1.51e-04 | 2.69e-01 |
| $\alpha_{3M}$ | Coefficient on Female Dummy | $-1.43\text{e-}01^{(***)}$ | 2.51e-04 |
| $\alpha_{4M}$ | Coefficient on Professional Certificates Dummy | $4.25\text{e-}01^{(***)}$ | 9.41e-05 |
| $\alpha_{5M}$ | Coefficient on Graduate Degree | $4.03\text{e-}01^{(***)}$ | 9.05e-05 |
| $\log(\sigma_1)$ | Log of standard deviation of wage shock | $-1.22^{(***)}$ | 3.28e-05 |

### Parameters of the Production of Teaching Skills

| Parameter | Description | Value | Standard Error |
|---|---|---|---|
| $\alpha_{0V}^1$ | Intercept, type 1 | $6.42\text{e-}02^{(***)}$ | 6.33e-04 |
| $\alpha_{0V}^2 - \alpha_{0V}^1$ | Difference between intercept for type 2 and for type 1 | $-1.04^{(***)}$ | 4.15e-05 |
| $\alpha_{0V}^3 - \alpha_{0V}^1$ | Difference between intercept for type 3 and for type 1 | $-1.93\text{e-}02^{(***)}$ | 2.47e-03 |
| $\alpha_{1V}$ | Coefficient on Age | $8.63\text{e-}02^{(***)}$ | 4.20e-04 |
| $\alpha_{2V}$ | Coefficient on Age Squared | -1.65e-03 | 2.42e-02 |
| $\alpha_{3V}$ | Coefficient on Female Dummy | $-1.71\text{e-}01^{(***)}$ | 2.39e-04 |
| $\alpha_{4V}$ | Coefficient on Professional Certificates Dummy | $3.61\text{e-}01^{(***)}$ | 1.24e-04 |

| Parameter | Description | Value | Standard Error |
|---|---|---|---|
| $\alpha_{5V}$ | Coefficient on Graduate Degree | 2.71e-01$^{(***)}$ | 1.58e-04 |
| $\log(\sigma_2)$ | Log of standard deviation of wage shock | -8.09e-01$^{(***)}$ | 4.94e-05 |

### Parameters of the Non-Teaching Log-Wage Offers

| Parameter | Description | Value | Standard Error |
|---|---|---|---|
| $\alpha_{1NT}$ | Coefficient on Age | 1.01e-02$^{(***)}$ | 3.66e-03 |
| $\alpha_{2NT}$ | Coefficient on Age Squared | -4.00e-04 | 9.85e-02 |
| $\alpha_{3NT}$ | Coefficient on Female Dummy | -1.38e-01$^{(***)}$ | 3.14e-04 |
| $\alpha_{4NT}$ | Coefficient on Professional Certificates Dummy | -3.13e-02$^{(***)}$ | 1.19e-03 |
| $\alpha_{5NT}$ | Coefficient on Graduate Degree | 1.19e-01$^{(***)}$ | 3.29e-04 |
| $\log(\sigma_3)$ | Log of standard deviation of wage shock | -4.00e-01$^{(***)}$ | 1.33e-04 |

### Non-pecuniary Utility

| Parameter | Description | Value | Standard Error |
|---|---|---|---|
| $\mu_{0H}^1$ | Intercept preference for home, type 1 | -3.30e+03$^{(***)}$ | 1.45e-08 |
| $\mu_{0H}^2 - \mu_{0H}^1$ | Difference between intercept in preference for home of types 2 and 1 | -2.15e+03$^{(***)}$ | 1.64e-08 |
| $\mu_{0H}^3 - \mu_{0H}^1$ | Difference between intercept in preference for home of types 3 and 1 | 6.51e+02$^{(***)}$ | 5.70e-08 |
| $\mu_{1H}$ | Coefficient on gender, preference for home | 1.64e+03$^{(***)}$ | 2.33e-08 |
| $\mu_{2H}$ | Coefficient on interaction gender*number of kids, preference for home | 3.61e+02$^{(***)}$ | 1.10e-07 |

| Parameter | Description | Value | Standard Error |
|---|---|---|---|
| $\mu_{3H}$ | Coefficient on age, preference for home | $-1.72\text{e}+01^{(***)}$ | 2.17e-06 |
| $\mu_{4H}$ | Coefficient on number of kids, preference for home | $1.43\text{e}+01^{(***)}$ | 2.44e-06 |
| $\mu_{5H}$ | Coefficient on dummy for kids aged 0-2, preference for home | $-1.19\text{e}+01^{(***)}$ | 3.21e-06 |
| $\mu_{6H}$ | Coefficient on dummy for kids aged 3-6, preference for home | $1.73\text{e}+02^{(***)}$ | 2.70e-07 |
| $\mu_{7H}$ | Coefficient on age squared, preference for home | $3.31\text{e}-01^{(***)}$ | 1.30e-04 |
| $\mu_{0M}^{1}$ | Intercept preference for Municipal school, type 1 | $-8.00\text{e}-01^{(***)}$ | 5.54e-05 |
| $\mu_{0M}^{2} - \mu_{0M}^{1}$ | Difference between intercept preference for Municipal school of types 2 and 1 | $-1.05\text{e}-01^{(***)}$ | 3.26e-04 |
| $\mu_{0M}^{3} - \mu_{0M}^{1}$ | Difference between intercept preference for Municipal school of types 3 and 1 | $-1.55\text{e}-01^{(***)}$ | 2.25e-04 |
| $\mu_{0V}^{1}$ | Intercept preference for Voucher school, type 1 | $-9.59\text{e}-01^{(***)}$ | 4.52e-05 |
| $\mu_{0V}^{2} - \mu_{0V}^{1}$ | Difference between intercept preference for Voucher school of types 2 and 1 | $5.12\text{e}-01^{(***)}$ | 8.56e-05 |
| $\mu_{0V}^{3} - \mu_{0V}^{1}$ | Difference between intercept preference for Voucher school of types 3 and 1 | $3.40\text{e}-01^{(***)}$ | 1.35e-04 |
| $\mu_{0Teach}$ | Non-pecuniary utility from teaching if female | $1.00^{(***)}$ | 3.92e-05 |

| **Log of Skill Prices of Teaching Skills** | | |
|---|---|---|
| **Market** | **Estimate of Log of Skill Price** | **Standard Error** |
| 2 | $-1.09\text{e}-01^{(***)}$ | 4.32e-04 |
| 3 | $3.95\text{e}-02^{(***)}$ | 9.29e-04 |

| Market | Estimate of Log of Skill Price | Standard Error |
|:------:|:------------------------------:|:--------------:|
| 4 | -4.40e-01$^{(***)}$ | 9.94e-05 |
| 5 | -5.58e-02$^{(***)}$ | 5.82e-04 |
| 6 | -5.32e-01$^{(***)}$ | 7.61e-05 |
| 7 | -2.29e-01$^{(***)}$ | 1.66e-04 |
| 8 | -3.74e-01$^{(***)}$ | 1.01e-04 |
| 9 | -2.88e-01$^{(***)}$ | 1.14e-04 |
| 10 | -2.27e-01$^{(***)}$ | 1.60e-04 |
| 11 | -7.02e-01$^{(***)}$ | 6.28e-05 |
| 12 | -6.27e-02$^{(***)}$ | 6.39e-04 |
| 13 | -1.13e-02$^{(***)}$ | 3.30e-03 |
| 14 | 2.16e-01$^{(***)}$ | 2.01e-04 |
| 15 | 3.71e-02$^{(***)}$ | 1.08e-03 |
| 16 | -5.96e-01$^{(***)}$ | 6.81e-04 |
| 17 | -4.56e-01$^{(***)}$ | 8.16e-05 |
| 18 | -1.47e-01$^{(***)}$ | 2.90e-04 |

The log of skill price in market 1 has been normalized to 0.01.

**Parents**

I let the number of types be $\bar{K} = 3$ and estimate the type proportions to be 19.7% for type 1, 47.4% for type 2 and 32.4% for type 3. In the following tables I report the parameters of the production of achievement by sector, of the preference for the municipal school and the weight on consumption, and of the fellowship formula, including the variance of the measurement error on the fellowship. The intercepts in the cognitive achievement production functions vary by type, and some vary by geographical location, for ease of exposition I do not report them in the following tables. The interested reader can refer to the online appendix at http://economics.sas.upenn.edu/ mtincani/jobmarket.html. As above, I use the following notation for statistical significance: three stars ($^{(***)}$) mean that the parameter is significantly different from zero in a two-sided Wald test at the 1% confidence level, two stars ($^{(**)}$) index significance at the 5% level and one star ($^{(*)}$) at the 10% confidence level. The absence of stars means that the p-value is above 10%. The standard errors are in parenthesis.

**Production of Achievement in the Municipal and in the Voucher Schools**

| Parameter | Description | Municipal | Voucher |
|:---------:|:------------|:---------:|:-------:|
| $\beta_{1J}^1$ | Teachers' skills, type 1 | 3.40e-01$^{(***)}$ | 2.11e-01$^{(***)}$ |
| | | (5.89e-03) | (9.06e-03) |
| $\beta_{1J}^2 - \beta_{1M}^1$ | Deviation of type 2 from type 1, teachers' skills | 3.74e-02 | -1.95e-01$^{(***)}$ |

| Parameter | Description | Municipal | Voucher |
|---|---|---|---|
| | | (5.32e-02) | (1.08e-02e-02) |
| $\beta_{1J}^3 - \beta_{1M}^1$ | Deviation of type 3 from type 1, teachers' skills | -2.10e-01$^{(***)}$ | -2.11e-01$^{(***)}$ |
| | | (8.34e-03) | (9.66e-03) |
| $\beta_{2J}^1$ | Parental education, type 1 | 5.72e-02$^{(*)}$ | 1.03e-01$^{(***)}$ |
| | | (3.15e-02) | (2.05e-02) |
| $\beta_{2J}^2 - \beta_{2M}^1$ | Deviation of type 2 from type 1, parental education | -4.71e-02 | 1.38e-01$^{(***)}$ |
| | | (4.44e-02) | (1.31e-02) |
| $\beta_{2J}^3 - \beta_{2M}^1$ | Deviation of type 3 from type 1, parental education | 4.27e-02 | 1.19e-02 |
| | | (5.36e-02) | (1.18e-01) |
| $\beta_{3J}^1$ | Monthly income per capita, type 1 | 1.55e-01$^{(***)}$ | 9.78e-01$^{(***)}$ |
| | | (1.56e-02) | (2.22e-03) |
| $\beta_{3J}^2 - \beta_{3M}^1$ | Deviation of type 2 from type 1, monthly income per capita | 6.28e-03 | 2.33e-01$^{(***)}$ |
| | | (2.93e-01) | (8.39e-03) |
| $\beta_{3J}^3 - \beta_{3M}^1$ | Deviation of type 3 from type 1, monthly income per capita | -1.48e-01$^{(***)}$ | 2.81e-01$^{(***)}$ |
| | | (1.24e-02) | (7.60e-03) |
| $\beta_{4J}$ | Monthly income per capita squared | -4.67e-02 | -2.52e-01$^{(***)}$ |
| | | (4.07e-02) | (7.13e-03) |
| $\log(\sigma_\nu)$ | Log of standard deviation of shock | -3.22e-02 | -1.95e-01$^{(***)}$ |
| | | (6.69e-02) | (8.78e-03) |

Parents of students of type one, who are the ones who benefit the most from teaching skills, have a preference for the municipal school independent of the school's effect on achievement $(\eta^1)$ that is lower than that of type two parents, but higher than that of type three parents. They are the parents who are willing to give up the most utility from consumption to enroll their child in a private voucher school, as can be seen in table 9. As far as the direct preference for the municipal school is concerned, model simulations show that if parents did not have such a direct preference (which is estimated to be negative), the enrollment share in private voucher schools in the entire country would be lower by ten percentage points.

**Preference for Municipal School and Weight on Consumption**

| Parameter | Description | Value | Standard Error |
|---|---|---|---|
| $\eta^1$ | Intercept of preference for municipal school, type 1 | -1.12$^{(***)}$ | 1.53e-03 |
| $\eta^2 - \eta^1$ | Difference between intercept of preference for municipal school of types 2 and 1 | 7.53e-01$^{(***)}$ | 2.33e-03 |
| $\eta^3 - \eta^1$ | Difference between intercept of preference for municipal school of types 3 and 1 | -7.58e-02$^{(***)}$ | 2.56e-02 |
| $\eta_1$ | Coefficient on *primaria* in preference for municipal school | 5.02e-01$^{(***)}$ | 4.21e-03 |
| $\eta_2$ | Coefficient on *rural* in preference for municipal school | 3.73e-01$^{(***)}$ | 5.65e-03 |
| $\tau^1$ | Weight on consumption, type 1 | 1.18e-01$^{(***)}$ | 1.70e-02 |
| $\tau^2 - \tau^1$ | Difference between weight on consumption of types 2 and 1 | 1.87e-01$^{(***)}$ | 9.04e-03 |
| $\tau^3 - \tau^1$ | Difference between weight on consumption of types 3 and 1 | 5.57$^{(***)}$ | 3.26e-04 |
| $\log(\sigma_\eta)$ | Log of standard deviation of preference shock | -4.52$^{(***)}$ | 4.85e-04 |

Finally, the parameters of the exogenous fellowship formula can be found in the table below.

**Fellowship Formula**

| Parameter | Description | Value | Standard Error |
|---|---|---|---|
| $b_0$ | Intercept | 4.48e-01$^{(***)}$ | 4.99e-03 |
| $b_1$ | Coefficient on $p$, price charged by school net of voucher | 1.86e-01$^{(***)}$ | 2.45e-03 |
| $b_2$ | Coefficient on *primaria* | 6.67e-02$^{(***)}$ | 3.07e-03 |
| $b_3$ | Coefficient on family size | 1.05e-01$^{(***)}$ | 1.64e-02 |
| $b_4$ | Coefficient on *rural* | -3.25e-01$^{(***)}$ | 5.96e-03 |
| $b_5$ | Coefficient on monthly income | -5.42e-02 | 3.56e-02 |
| $\log(\sigma_{me})$ | Log of standard deviation of measurement error | -5.91$^{(***)}$ | 3.21e-04 |

**Private Voucher School**

| | Profit Function of Voucher School | | |
|---|---|---|---|
| Parameter | Description | Value | Standard Error |
| $c_1$ | Variable cost: coefficient on enrollment | 1.097e-02 | 6.61e-01 |
| $c_2$ | Variable cost: coefficient on enrollment squared | 2.928e-06 | 9.24e-04 |
| $c_3$ | Coefficient on number of classes per teacher | 4.099e+03 | 2.87e+07 |
| $\log(\sigma_{cost})$ | Log of standard deviation of shock on $c_1$ | -3.212 | 4.53e+02 |

The monetary values entering the profit function are expressed in terms of $CLP100,000$.[131]

## Appendix J: Market-Level Aggregation of Voucher Schools' Tuitions

In this appendix I discuss the assumption of a unique voucher school tuition in each market. With this assumption I abstract from competition between private schools.[132] The unique tuition in the data is obtained by aggregating the prices charged by the different schools. On the one hand, there is no obvious way of performing this aggregation, and on the other hand, the price actually charged by each school is not observed. All that is observed is how much each family pays in each school. This amount is equal to the tuition charged by the school net of the voucher and of the fellowship received by the family. I assume that the price charged by the unique fictitious school is the maximal payment observed in each market, plus the value of the voucher. The rationale for this assumption is that the family that pays the most is only receiving a voucher and no fellowship. Because I do not observe the actual amount of fellowship received, I assume that only the family that pays the most within a district receives no fellowship. As a consequence of this aggregation rule, any difference between a family's payment and this market-wide tuition is due to the fellowship, and not to differences in the prices charged by different schools.

To assess the goodness of this assumption, consider the opposite extreme case in which each school charges a different price, and assume that the price charged by each school is equal to the maximal payment observed in that school plus the voucher. My aggregating rule selects the highest of these prices. Consider the difference between the unique, market-wide aggregated price and the school-level prices. If my aggregation is correct, this difference is due to fellowships and, therefore, there should be some correlation between this difference and the variables that enter the fellowship formula. In fact, I do observe such a correlation, and moreover, the signs of the correlation are as expected. In particular, consider defining fellowships at the school level, by taking the difference of the maximal price observed in each school and the payment made by each family.

---

[131]The standard errors are unadjusted for the two-step procedure and the uncertainty over $r$. They have been computed as the standard errors of the exact maximum likelihood and they provide a lower bound to the adjusted standard errors. The adjusted standard errors are currently being bootstrapped.

[132]Notice that this assumption does not exactly correspond to assuming that voucher schools are perfectly colluding, and further that they are choosing a unique tuition. A model of perfect collusion requires maximizing the sum of the profits of all firms with respect to the choice variables, in this case tuition and unit price of teaching skills. The formula for the financing of the fellowships introduces nonlinearity of profits with respect to tuition: because the profits of individual schools are not a linear function of tuition, maximizing their sum with respect to a single tuition does not, in general, give the same solution as maximizing the profits of the fictitious single voucher school.

Regressing this quantity on the variables $primaria, rural, nfam, y$ and on the price charged by the school gives positive signs on $primaria, rural, nfam$ and a negative sign on income.[133] Regressing the difference between the unique aggregated price and the maximal price within each school on the characteristics of the parents who pay the maximal price within each school delivers coefficient estimates of the same sign and of the same relative magnitudes, suggesting that this difference might indeed be due to the fellowship.[134] Notice that this is necessary but not sufficient evidence in favor of my assumption, as the difference between aggregated price and individual schools' prices might be due to optimal behavior on the part of the individual schools, which charge more to, for example, wealthier families in more urban areas. However, the evidence is necessary because observing no correlation of the type I observe would have casted doubt on the validity of the assumption that any price differential within different families in the same market comes from the fellowships.

---

[133]The $R^2$ is 85.27%.

[134]The coefficients on $primaria, rural, nfam, y$ for the school-level fellowship are $1824, 5781, 888, -0.0211$; the coefficients on the same variable for the difference between the aggregated price and the school-level maximal prices are $717, 14862, 768, -0.017$.