Supplementary Material

Supplement to "Teacher labor markets, school vouchers, and student cognitive achievement: Evidence from Chile"

(*Quantitative Economics*, Vol. 12, No. 1, January 2021, 173–216)

Michela M. Tincani
Department of Economics, University College London

Appendix C1: Markets

To design market boundaries I analyzed mobility of parents and teachers. The design balances the trade-off in terms of sample-size and mobility across markets: a large within-market sample size yields low mobility across markets but a small number of (large) markets, whereas a large number of markets is obtained by having small within-market sample sizes with large across-market mobility.

The unique geographical configuration of Chile aided in the identification of empirically closed or nearly closed markets: the country occupies a narrow but long coastal strip, where mobility between northern and southern regions is hindered.[1] The resulting number of nearly perfectly closed markets is 18. Table C1 reports the region in which each market lies, and the number of schools in each school sector and market, which correlates to within-market sample sizes.

Michela M. Tincani: m.tincani@ucl.ac.uk

[1]With a total area of 291,933 square miles (756,102 km^2), Chile is larger than all U.S. states except Alaska and larger than all countries in the European Union. Yet, it extends 2653 miles (4270 km) from north to south, and it averages only 110 miles (177 km) from east to west.

Available at http://qeconomics.org. https://doi.org/10.3982/QE1057

Table C1. 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		2436	2061.5
Average		135.3	114.5

Note: Data source: SIMCE 2006. The number of schools in each market and sector is the average between the number of primary and of secondary schools.

Appendix C2: Additional tables and figures

Table C2. Labor supply: descriptive analysis.

	Teach	Teach in Public Sector	Work
Female	0.189	0.031	−0.175
	(0.013)	(0.021)	(0.015)
Number of children	−0.021	0.016	−0.015
	(0.006)	(0.009)	(0.007)
Age	0.008	0.020	−0.003
	(0.001)	(0.001)	(0.001)
Observations	5061	3195	5471
Pseudo R^2	0.100	0.133	0.120

Note: Data source: ELD and CASEN for 2006. Standard errors in parentheses. Results from Probit Regression of the indicated dummy dependent variable on: gender, number of children, number of children interacted with gender, age, age squared. Marginal effects reported. Inverse Probability Weights. Samples restrictions by columns: (1): individuals who work, (2): individuals who teach, (3): full sample.

Table C3. Log-wage regressions by occupational sector: descriptive analysis.

	Public	Voucher	Nonteaching
Age	0.030	0.052	0.027
	(0.012)	(0.018)	(0.017)
Age squared	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)
Female	−0.102	−0.142	−0.384
	(0.031)	(0.032)	(0.044)
Prof. certifications	0.029	0.087	0.073
	(0.031)	(0.034)	(0.042)
Graduate degree	0.051	0.093	0.204
	(0.029)	(0.036)	(0.062)
Constant	11.948	11.629	12.810
	(0.254)	(0.354)	(0.359)
Observations	1186	1217	1806
R^2	0.181	0.123	0.103

Note: Data sources: ELD and CASEN for 2006. Results from ordinary least square regressions. Standard errors in parentheses. Dependent variable: logwage. Inverse Probability Weights.

Table C4. Demand for public sector schooling: descriptive analysis.

	Value	Standard Error
Log income	−0.112	0.002
Primary	0.029	0.003
Rural	0.117	0.006
Parents' education	−0.029	0.001
Family size	0.015	0.001
Observations	100,000	
Pseudo R^2	0.091	

Note: Data source: SIMCE 2006. Results from Probit regression of the dummy for Public school on log household income, elementary school dummy, rural home dummy, parental education, family size. Marginal effects reported.

Table C5. Achievement by school sector, descriptive analysis.

	Public	Voucher
Parents' education	0.079	0.088
	(0.002)	(0.002)
Income pro-capite (100,000 CLP)	0.455	0.354
	(0.020)	(0.015)
Squared income pro-capite (100,000 CLP)	−0.079	−0.059
	(0.007)	(0.004)
Constant	−1.184	−1.105
	(0.016)	(0.018)
Observations	47,007	52,993
R^2	0.104	0.118

Note: Data source: SIMCE 2006. Results from ordinary least square regressions. All regressors reported. Dependent variable: average between Mathematics and Spanish SIMCE test scores (standardized). Standard errors in parentheses.

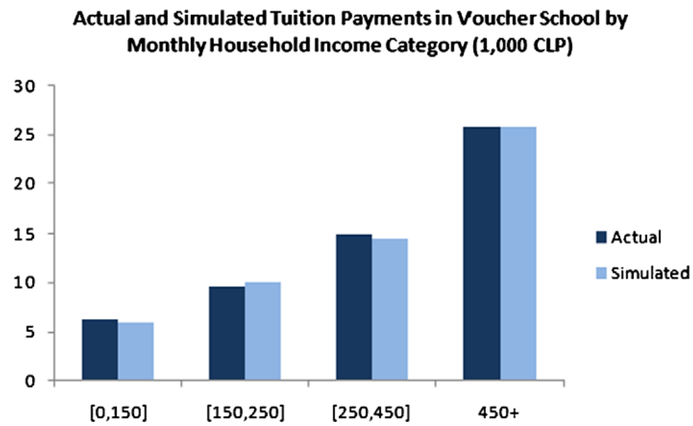

FIGURE C1. Model fit: tuition payments net of financial aid.

C2.1 *Additional estimates of structural parameters*

TABLE C6. Socioeconomic status of students by school sector, at baseline, and under the merit-based teaching reform (first counterfactual).

	Municipal	Voucher
	Prereform (Baseline)	
Parental education (yrs)	9.78	11.54
Household income (CLP)	207,249	337,811
	Post-reform (Counterfactual)	
Parental education (yrs)	10.08	11.64
Household income (CLP)	224.976	353,600

Note: Baseline values computed at the simulated baseline choice.

TABLE C7. Type proportions.

Type	Proportion	Standard Error
Households type 2	0.4790	0.1374
Households type 3	0.3240	0.1499
Potential teachers type 2	0.4240	0.0002
Potential teachers type 3	0.3790	0.0001

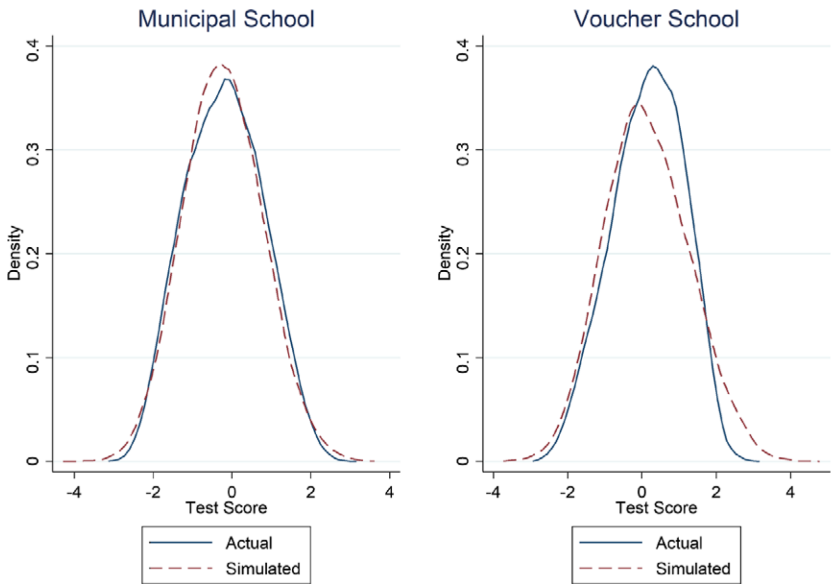

Figure C2. Model fit: test scores by school sector.

Table C8. Parameters of nonpecuniary utility of potential teachers.

	Home	Public	Voucher
Intercept type 1	−3300.0	−0.8000	−0.9590
	(0.0002)	(0.0002)	(0.0002)
Intercept type 2 minus type 1	−2150	−0.1050	0.5120
	(0.0002)	(0.0002)	(0.0002)
Intercept type 3 minus type 1	651.0	−0.1550	0.3400
	(0.0002)	(0.0002)	(0.0002)
Female	1640.0		
	(0.0002)		
Female × N children	36.00		
	(0.0002)		
Age	−17.2		
	(0.0002)		
N children	14.3		
	(0.0002)		
Has children aged 0–2	−11.9		
	(0.0002)		
Has children aged 3–6	173		
	(0.0002)		
Age squared	0.3310		
	(0.0002)		
Non-pecuniary utility from teaching if female		1.00	
		(0.0002)	

Note: Standard errors in parenthesis.

Table C9. Log of prices of teaching skills.

Market	Log of Skill Price	Standard Error
2	−0.109	0.0002
3	0.0395	0.0002
4	−0.440	0.0002
5	−0.0558	0.0002
6	−0.532	0.0002
7	−0.229	0.0002
8	−0.374	0.0002
9	−0.288	0.0003
10	−0.227	0.0002
11	−0.702	0.0002
12	−0.0627	0.0002
13	−0.0113	0.0002
14	0.216	0.0002
15	0.0371	0.0002
16	−0.596	0.0002
17	−0.456	0.0002
18	−0.147	0.0002

Note: Standard errors in parenthesis. Log of skill price normalized to 0.00 in market 1.

Table C10. Parental preference parameters.

	Value	Standard Error
Intercept of preference for Municipal, type 1	−1.1200	0.1018
Types 2 minus type 1	0.7530	0.1432
Types 3 minus type 1	−0.0758	0.1432
primaria in preference for Municipal	0.5020	0.1159
rural in preference for Municipal	0.3730	0.1128
Weight on consumption, type 1	0.1180	0.1439
Type 2 minus type	0.1870	0.1679
Types 3 minus type 1	5.570	0.1354
Log of preference shock standard deviation	−4.5200	0.1812

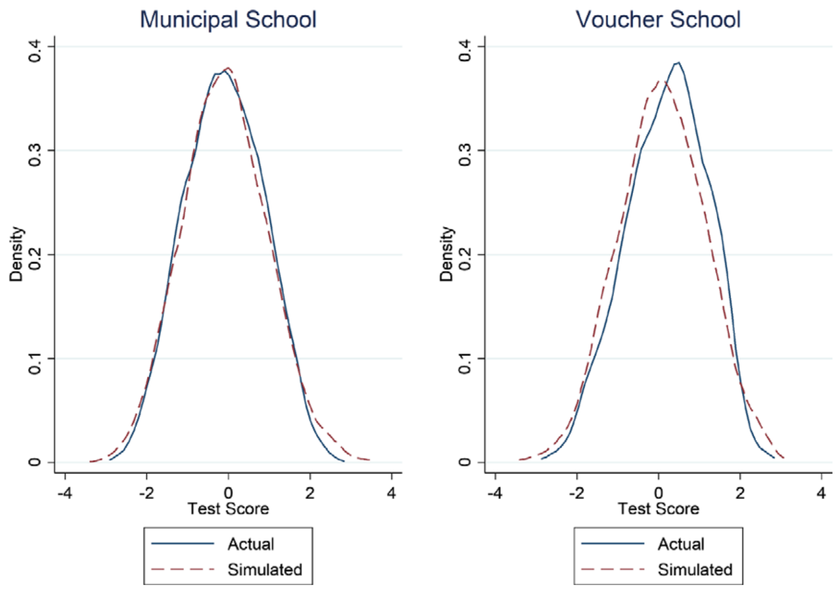

FIGURE C3. Model fit: test scores by school sector, market number 5.

TABLE C11. Fellowship assignment.

	Value	Standard Error
Intercept	0.4480	0.1431
p, price charged by school net of voucher	0.1860	0.1278
primaria	0.6670	0.1294
Family size	0.1050	0.1141
rural	−0.3250	0.1156
Monthly income	−0.0542	0.1414
Log of standard deviation of measurement error	−5.91	0.1241

TABLE C12. Profit function of voucher schools.

	Value	Standard Error
Variable cost c_1: enrollment	0.01097	0.6610
Variable cost c_2: enrollment squared	2.928e−06	0.0009
N classes per teacher, c_3	4.099e+03	2.87e+07
Log of standard deviation of shock σ_{cost}	−3.212	453e+02

Note: For an average-sized school, the estimated variable cost per student is 4.8% of the voucher amount.

Appendix C3: Governmental formulae for private school revenues

This Appendix reports the formulae in article 25 of the law *Decreto con Fuerza de Ley N^o 2, De Educacion, de 20.08.98*, fully incorporated into the model. Let p denote the tuition charged by the school, and let v be the voucher. Each household choosing the voucher sector is responsible for the payment of $p - v - f(Z_h)$ where $0 \leq f(Z_h) \leq (p - v)$ is the amount of fellowship received by a family with characteristics Z_h if the voucher sector is chosen. Z_h includes an indicator for schooling level, family size, whether the household is in a rural location, and household income. However, the amount of the per-pupil subsidy that is effectively received by the private school is not necessarily v. It decreases as the average tuition payments in the school increase. The latter depend on the tuition fee charged by the school, and on the composition of the households at the school, which determines the amount of financial aid received by the student body at the school. Formally, let $EPV(p, r)$ denote the mean tuition payments in the voucher school sector:

$$EPV(p, r) = \int \big(p - v - f(z)\big) g^z(z|V \text{ chosen}; p, v)\, dz, \tag{8}$$

where the conditional density $g^z(z|V \text{ chosen}; p, r)$ is indexed by (p, r) because it depends on parental school choice, which is a function of the prices (p, r). Below, I drop the dependence of EPV on (p, r) for simplicity. The first Government formula adjusts the amount of per-pupil voucher subsidy effectively received by the school. Accordingly, the adjusted per-pupil gross revenues ($R^g(EPV)$) in the private school sector are:

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

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

Adjusted per-pupil net revenues are different from adjusted per-pupil gross revenues because the school is also required to contribute to the financial aid budget. That is, private schools partially cover the fellowship expenses. The law provides that the per-pupil contribution to financial aid due by the private schools be

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

Therefore, the adjusted net per-pupil revenues are $\tilde{R}(p, r, E(p, r)) = R^g(EPV) - R^f(EPV)$.

Appendix C4: Constrained maximization of approximated profits

Because cumulative normal distribution functions enter the expression for profits, the function Π and its first and second derivative functions do not admit a closed form; hence, Π's critical points and curvature properties cannot be derived analytically. In estimation, I approximate profits with a function with a known closed form and I solve the constrained maximization of approximated profits at each candidate parameter value. To perform the approximation, I first evaluate numerically the true profit function at a large number of points $(p^{(s)}, r^{(s)})$. To evaluate the true profit function, I derive numerically the student enrollment and teacher supply functions: $E(p, r; v)$, $T(p, r; v)$, and $NT(p, r; v)$ at each evaluation point $(p^{(s)}, r^{(s)})$ by solving the second stage of the model, and I plug them into the profit function. I then approximate the profit function using an interpolating regression. Notice that the approximation is conditional on a realization of the variable cost shock ϵ_{cost}. For simplicity, in the following formulae I drop the dependence of the estimated regression coefficients on the cost shock. Approximation is by ordinary least squares using a cubic interpolating polynomial:

$$\hat{\Pi} = \hat{a}_1 + \hat{a}_2 p + \hat{a}_3 p^2 + \hat{a}_4 r + \hat{a}_5 r^2 + \hat{a}_6 pr + \hat{a}_7 p^3 + \hat{a}_8 r^3 + \hat{a}_9 p^2 r + \hat{a}_{10} pr^2. \tag{9}$$

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. I solve for the school's choice variables (p, r) and for the Kuhn–Tucker–Lagrange multiplier λ.

The approximated problem of the firm is the following:

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

or equivalently

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

I solve for the optimal (p^*, r^*) and for the Kuhn–Tucker–Lagrange multiplier λ^* following the procedure described in Judd (1998), Chapter 4, page 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, 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 the Newton method to solve numerically the following system of two equations in the two unknowns (r, λ):

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

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%, and the prices that maximize the approximated function are a very good approximation to the prices that maximize the true function. Specifically, tuition fees are identical because the caps are binding in both scenarios. With regards to skill rental rates, the average difference between the maximum of the approximated function and the maximum of the true function is 2.04% of the true maximum.

Appendix C5: Further model details

C5.1 *Equilibrium existence and uniqueness*

Proposition 1. *An equilibrium exists.*

Proof. An equilibrium exists because, given prices r_m and p_m, each potential teacher and each household in market m solve a discrete-choice problem that admits at least one most-preferred choice by construction: utilities are well-defined. Second, the voucher school's profit is a continuous real-valued function defined on the compact set $[0, \bar{p}] \times [0, \bar{r}] \in R^2$, where it is assumed that there exists an arbitrarily large upper bound $\bar{r}$ to the rental rate that the school can offer. The profit function attains a maximum value by the extreme value theorem. □

Proposition 2. *The equilibrium of the second-stage subgame is unique.*

Proof. The potential teacher labor supply depends only on the rental rate r, and not also on the tuition fee p: $T(r)$, $NT(r)$. Therefore, the second stage of the model is equivalent to a sequential game where teachers move first, and parents, observing the teacher distribution across sectors, choose school sectors. In the last stage, each household h has a unique most preferred alternative $e(\tilde{X}_h, k_h, \nu_h) \in \{M, V\}$ by virtue of the fact that the preference and technology shocks ν_h are continuously distributed. Therefore, no household is indifferent between two options. Similarly, in the second-last stage, each potential teacher i has a unique most preferred alternative $d(\tilde{X}_i, \psi_i, \epsilon_i) \in \{M, V, NT, H\}$ because the preference, wage and technology shocks ϵ_i are continuously distributed. Therefore, by backward induction the equilibrium of the subgame is unique. □

C5.2 *Equilibrium teacher skills by school sector*

To compute the mean teaching skills supplied to the voucher sector in each m, 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 individual i are

$$s_i = \exp\bigl(a_0(l_i) + a_1' X_i + \epsilon_i^{\text{tech}}\bigr) \tag{10}$$

with $\epsilon_i^{\text{tech}} \sim N(0, \sigma_V^2)$. That is, conditional on type, skills are log-normally distributed. Conditional on $X_i = x$, the density of teaching skills depends both on the density of the shock ϵ_i^{tech} and on the type probability ψ_{l_i}:[2]

$$f^s(s_i|x) = \frac{\psi_{l_i}}{s_i \sigma_V \sqrt{2\pi}} \exp\left\{-\frac{\bigl(\ln s_i - a_0(l) - a'x\bigr)^2}{2\sigma_V^2}\right\}.$$

The population density is obtained by integrating over the distribution of x in market m, $f_m^x(x)$:

$$f_m^s(s_i) = \int \frac{\psi_{l_i}}{s_i \sigma_V \sqrt{2\pi}} \exp\left\{-\frac{\bigl(\ln s_i - a_0(l) - a'x\bigr)^2}{2\sigma_V^2}\right\} f_m^x(x)\, dx.$$

To derive the density of teaching skills in the voucher school, define $A(q, \epsilon_i^{\text{tech}}, l_i)$ to be the subset of $\mathbb{R}^3$ that is such that if $\epsilon_i^{-\text{tech}} = [\epsilon_i^M\ \epsilon_i^{NT}\ \epsilon_i^H]' \in A(q, \epsilon_i^{\text{tech}}, l_i)$, an individual with characteristics q, shock realization ϵ_i^{tech}, and type realization l_i chooses the voucher school. Letting $\Pr_m(V)$ denote the proportion of individuals choosing sector V in market m, the density of teaching skills in sector V may be written as

$$g_m^V(s_i|\text{sector } V \text{ chosen}) = \frac{1}{\Pr\limits_m(V)} \psi_{l_i} \int_{\epsilon_i^{-\text{tech}} \in A} f_m^s(s_i) f^{-\text{tech}}\bigl(\epsilon_i^{-\text{tech}}\bigr)\, d\epsilon_i^{-\text{tech}},$$

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

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

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

The density of teaching skills in the municipal school, $g_m^M(s_i|\text{sector } M \text{ chosen})$, can be derived in a similar way.[3]

The mean skills supplied to each sector in market m are obtained using the conditional densities g_m^M, g_m^V:

$$\begin{aligned}\bar{s}_{Mm} &= \sum_{l_i} \psi_{l_i} \int s_i g_m^M(s_i|\text{sector } M \text{ chosen})\, d\epsilon_i^{\text{tech}}, \\ \bar{s}_{Vm} &= \sum_{l_i} \psi_{l_i} \int s_i g_m^V(s_i|\text{sector } V \text{ chosen})\, d\epsilon_i^{\text{tech}}.\end{aligned} \tag{11}$$

Appendix C6: List of moment conditions

I compute 607 moments, 321 pertaining to parents and 286 to potential teachers.

C6.1 *Parents' moments: Matching choices, test scores, and fellowship amounts*

I use the following categories:

- family size $nfam_h$: [2, 3], [4, 6], ≥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 moment conditions are obtained by multiplying the difference between actual and predicted outcomes by this indicator. The moment conditions are built on the following outcomes (number of moment conditions in parentheses):

- Test scores by sector and by
  - market ($18 \times 2 = 36$)
  - monthly income per capita and parental education ($10 \times 10 \times 2 = 200$)
- Fraction choosing voucher school by
  - market (18)
  - parental education (10)

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

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

- Private school tuition payments made by parents by
  - elementary school, number of individuals in the household, rurality of the residence ($2 \times 3 \times 2 = 12$)
  - monthly income (10)
  - market (18)

Total number of parents' moments: 321.

C6.2 *Potential teachers' moments: Matching choices and accepted wages*

I use the following categories:

- coarse age, age_i: $[20-30]$, $[31-40]$, $[41, 50]$, ≥ 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, ≥ 3
- number of children aged 0–2, nkids2_i: 0, ≥ 1
- number of children aged 3–6, $\text{nkids3}-6_i$: 0, ≥ 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 moment conditions build on the following outcomes (number of moment conditions in parentheses):

- Accepted wages by sector (3 working options) and by
  - age, gender, professional certifications ($3 \times 4 \times 2 \times 2 = 48$)
  - graduate degree ($3 \times 2 = 6$)
  - market ($3 \times 18 = 54$)

- Fractions in sector M, V, and NT (exclude one sector to avoid multicollinearity and hence singularity of the variance-covariance matrix of the moment conditions) by
  - professional certifications ($3 \times 2 = 6$)
  - age, gender, graduate degree ($3 \times 4 \times 2 \times 2 = 48$)
  - market ($3 \times 18 = 54$)
  - gender, number of kids ($3 \times 2 \times 4 = 24$)

  - number of kids up to 2 years of age, age ($3 \times 2 \times 4 = 24$)
  - number of kids of aged 3–6 ($3 \times 2 = 6$)

- Accepted wages in the teaching occupations (2) by finer age category ($2 \times 8 = 16$)

Total number of potential teachers' moments: 286.

Appendix C7: Details of the first step of the estimation

For simplicity, I drop the subscript from θ_{II}, and refer to the parameters of the second stage of the model, estimated in the first step of the estimation, as θ. Let y_i denote an observed outcome for individual i. Let $\Omega_i \times \{1, \dots, L\}$ denote the state space of individual i with elements (ω_i, l_i) (where $l_i \in \{i, \dots, L\}$ is the person's type). Vector ω_i contains, for example, degrees, age, gender, etc. Let $\hat{y}_i(\omega_i, \theta)$ denote the outcome predicted by the model. This outcome is replaced by the simulator:

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

obtained by drawing S simulated shocks from the model's shock distribution under parameter θ and using the model to simulate behavior, and hence an outcome for each individual, simulation, and type: $\tilde{\hat{y}}_i(\omega_i, l_i, s, \theta)$.[4] The simulated outcomes are then averaged across simulations and types. Moment conditions are constructed by taking the average of the individual difference between the actual and the simulated outcome: $m_i(\theta) = y_i - \tilde{\hat{y}}_i(\omega_i, \theta)$.

The MSM finds the vector θ that minimizes the weighted distance of the empirical moment conditions from zero:

$$\hat{\theta}_{\text{MSM}} = \arg\min_{\theta} m(\theta)'_n W_n m_n(\theta), \tag{12}$$

where W_n is a symmetric positive definite weighting matrix such that as $n \to \infty$, $W_n \to W$ in probability with W symmetric and positive definite. Vector $m_n(\theta)$ is the sample average of the individual deviations $m_i(\theta)$.

Estimation accounts for the fact that multiple datasets of different sizes are used. Consider the population moment condition based on outcome y_i:

$$E\big[\big(y_i - \hat{y}_i(\omega_i, \theta)\big) I_i(\omega_i, y_i \text{ nonmissing})\big]$$

and suppose that there are M moment conditions based on the M deviations $\{m_i^1, \dots, m_i^M\}$, with

$$m_i^m = \big(y_i^m - \hat{y}_i^m(\omega_i^m, \theta)\big) I_i\big(\omega_i^m, y_i^m \text{ nonmissing}\big).$$

[4] S is set equal to 100. The shocks drawn at each simulations are all the preference, wage, and technology shocks in the second stage of the model: ν_{hm}^{pref}, ν_{hj}, $\epsilon_{iH}^{\text{pref}}$, ϵ_i^{tech}, ϵ_{iM}, ϵ_{iNT}.

Let m_i be a vector that stacks all deviations 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 deviations computed at a parameter value θ. 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.[5] Let $n = n_A + n_B$ and premultiply the sample moments by $\frac{n}{n}$. Denote the vector of empirical moments based on a sample of size n by $m_n(\theta)$:

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

where $a_A = \frac{n}{n_A}$, $a_B = \frac{n}{n_B}$ and $I(\cdot)$ is an indicator function equal to 1 if the expression in parentheses is true.

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

$$\hat{\theta}_{\text{MSM}} = \arg\min_{\theta} m_n(\theta)' W_n m_n(\theta), \tag{13}$$

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 see the asymptotic properties of the estimator, 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 Bhattacharya (2005), who derives the asymptotic properties of the generalized method of moments with a stratified sample. The MSM estimator defined in (13) is consistent and asymptotically normal:

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

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.[6]

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

[6]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. I cannot adopt the optimal weighting matrix because the variance covariance matrix is a high-order sparse matrix that cannot be numerically inverted. 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. This matrix is easily invertible.

To estimate consistently the asymptotic variance of the estimator, I substitute V with a consistent estimate $\hat{V}$ computed at $\hat{\theta}_{\text{MSM}}$. The estimator includes a stratum correction that accounts for the sampling design.[7] The estimator of the variance covariance matrix is

$$\begin{aligned}\hat{V} = {} & \sum_{i\in A}\left(\frac{H_A}{n_A}w_i\right)^2 m_i(\hat{\theta}_{\text{MSM}})m_i(\hat{\theta}_{\text{MSM}})' \\ & + \sum_{i\in B}\left(\frac{H_B}{n_B}w_i\right)^2 m_i(\hat{\theta}_{\text{MSM}})m_i(\hat{\theta}_{\text{MSM}})' \\ & - \frac{1}{n_A}\left(\sum_{i\in A}\frac{H_A}{n_A}w_i m_i(\hat{\theta}_{\text{MSM}})\right)\left(\sum_{i\in A}\frac{H_A}{n_A}w_i m_i(\hat{\theta}_{\text{MSM}})\right)' \\ & - \frac{1}{n_B}\left(\sum_{i\in B}\frac{H_B}{n_B}w_i m_i(\hat{\theta}_{\text{MSM}})\right)\left(\sum_{i\in B}\frac{H_B}{n_B}w_i m_i(\hat{\theta}_{\text{MSM}})\right)', \end{aligned} \tag{14}$$

where $m_i(\hat{\theta}_{\text{MSM}})$ is the $M \times 1$ vector of individual-level deviations between actual and simulate outcomes computed at $\hat{\theta}_{\text{MSM}}$. To estimate consistently the matrix of moments' partial derivatives, I use

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

where the differentiation is numerical. Letting Δ_t denote a vector of the same size as the parameter vector with zeros everywhere and $\delta > 0$ as its tth element, the derivative of the mth element of $m_i(\theta)$ with respect to the tth element of θ is computed as

$$\begin{aligned}&\frac{\partial \hat{m}_i^m(\theta)}{\partial \theta_t}\bigg|_{\theta=\hat{\theta}_{\text{MSM}}} \\ &\quad = \frac{-\hat{m}_i^m(\theta+2\Delta_t) + 8\hat{m}_i^m(\theta+\Delta_t) - 8\hat{m}_i^m(\theta-\Delta_t) + \hat{m}_i^m(\theta-2\Delta_t)}{12\delta}\bigg|_{\theta=\hat{\theta}_{\text{MSM}}}. \end{aligned} \tag{15}$$

Appendix C8: Second step of the estimation: Implementing NPSML

Let $\theta_I = [c_1 c_2 c_3 \sigma_{\text{cost}}]'$ denote the vector of parameters to be estimated in the second step (i.e., the parameters from the first stage of the model), $\hat{\theta}_{\text{II}}$ the vector of estimates of θ_{II} and $\hat{r}_m$ the vector of wage rates obtained in the first step of the estimation. Imagine obtaining a sample of markets, and suppose that for each sampled market, a sample of college graduates making labor supply decisions and of students is available. Denote by $X_m = \{x_1, \dots, x_i, \dots, x_{NS}\}_{i\in m}$ and $Q_m = \{q_1, \dots, q_i, \dots, q_{NC}\}_{i\in m}$ the within-market samples of students and of college graduates, respectively. Let $(\hat{r}_m, X_m, Q_m)_{m=1,\dots,M}$ be an

[7] The correction term is derived and discussed in Bhattacharya (2005). Intuitively, ignoring the fact that observations come from two separate strata would overestimate the between-strata variances.

independently and identically distributed sample of markets. The true log-likelihood is

$$L_M(\theta_I) = \sum_{m=1}^{M} \ln l_m(\theta_I),$$

where $l_m(\theta_I)$ is the market-contribution to the likelihood, that is, the density of r computed at the observed value $\hat{r}_m$ conditional on the exogenous characteristics in the market, on θ_I and on $\hat{\theta}_{\text{II}}$: $l_m(\theta_I) = f(r|\theta_I, X_d, Q_d; \hat{\theta}_{\text{II}})|_{\hat{r}_m}$. The function $l_m(\theta_I)$ cannot be computed analytically, therefore, I approximate it using a kernel estimator based on an i.i.d. simulated sample $(\epsilon_{\text{cost}}^{ms})_{s=1,\dots,S}$ of draws from the log-normal distribution of ϵ_{cost}.

Denote by $\hat{r}_m^s(\theta_I)$ the simulated wage rate in market m given a value for θ_I, and conditional on $\hat{\theta}_{\text{II}}$, X_m, Q_m. The s superscript means that for every simulated draw s, the wage rate is derived as a solution to private school profit maximization. I estimate the likelihood $l_m(\theta_I)$ by

$$\tilde{l}_S(\hat{r}_m|X_m, Q_m, \theta_I; \hat{\theta}_{\text{II}}) = \tilde{l}_{mS}(\theta_I) = \frac{1}{Sh}\sum_{s=1}^{S} \mathcal{K}\left(\frac{\hat{r}_m - \hat{r}_m^s(\theta_I)}{h}\right),$$

where $\mathcal{K}(\cdot)$ is the 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 obtained by summing over markets:

$$\tilde{L}_{MS}(\theta_I) = \sum_{m=1}^{M} \ln \tilde{l}_{mS}(\theta_I)$$

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

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

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

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

where Ω is the asymptotic variance-covariance matrix of the exact maximum likelihood estimator:

$$\Omega = \left(-E\left[\frac{\partial^2 L_M(\theta_{I,0})}{\partial\theta_2\partial\theta_I'}\right]\right)^{-1} E\left[\frac{\partial L_M(\theta_{I,0})}{\partial\theta_I}\frac{\partial L_M(\theta_{I,0})}{\partial\theta_I'}\right]\left(-E\left[\frac{\partial^I L_M(\theta_{I,0})}{\partial\theta_I\partial\theta_I'}\right]\right)^{-1}. \tag{16}$$

References

Bhattacharya, D. (2005), "Asymptotic inference from multi-stage samples." *Journal of Econometrics*, 126 (1), 145–171. [15, 16]

Judd, K. L. (1998), *Numerical Methods in Economics*. MIT press. [9]

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 24 January, 2018; final version accepted 2 September, 2020; available online 18 September, 2020.