Received 24 July 2017
Revised 19 September 2018
Accepted 13 November 2018

The persistent effect of socioeconomic status on education and labor market outcomes

Evidence from Chile's administrative records

Juan A. Correa
Facultad de Economia y Negocios, Universidad Andres Bello, Santiago, Chile
Pablo Gutiérrez
Department of Economics, The University of British Columbia, Vancouver, Canada
Miguel Lorca
School of Business, University of New South Wales, Sydney, Australia
Raúl Morales
Department of Economics, Cornell University, Ithaca, New York, USA, and
Francisco Parro
Escuela de Negocios, Universidad Adolfo Ibáñez, Santiago, Chile

Abstract

Purpose – This paper aims to study the effect of family socioeconomic status (SES) on academic and labor market outcomes.

Design/methodology/approach – The authors used a rich data set of administrative records for test scores, individual background and adult earnings of a cohort of agents, covering a period spanning the agents' upper-secondary education and their early years in the labor market.

Findings – The authors find that students with the highest SES obtained a 1.5 standard deviations higher score in the college admission test than students who had the same academic outcomes in the eighth grade test but belong to the lowest SES. Similarly, among students that obtained the same scores in the college admission test, those with the highest SES earned monthly wages 0.7 standard deviations higher than those with the lowest SES.

Originality/value – The findings highlight that family socioeconomic background continues to influence outcomes during individuals' upper secondary education and early years in the labor market.

Keywords Human capital, Inequality, Socioeconomic status, Academic achievement

Paper type Research paper

JEL classification – D39, I24, I25

Applied Economic Analysis
Vol. 27 No. 79, 2019
pp. 62-90
Emerald Publishing Limited
2632-7627
DOI 10.1108/AEA-06-2019-0007

1. Introduction

Economic inequality has been on the rise in several regions of the world (Piketty, 2014). Agents' perceptions of the origin of inequality greatly affect their political support for different policies aiming at reducing it (Fong, 2001; Bowles *et al.*, 2002; Bowles and Gintis, 2002). For instance, societies in which individuals foresee the possibility of climbing up the income ladder could be more tolerant to inequality and less prone to support redistributive policies than other types of societies (Braun *et al.*, 2018). Therefore, the intergenerational transmission of economic inequality is a policy-relevant issue. We use administrative data for a cohort in Chile to estimate the effect of family socioeconomic status (SES) on outcomes over a period spanning the agents' upper-secondary education and early years in the labor market.

Empirical evidence documents a significant influence of family socioeconomic background on the probable economic status of the next generation (Solon, 1992, 1999, 2002; Zimmerman, 1992). Most of the literature points to the influence of parental background on children's human capital formation (Becker and Tomes, 1979, 1986)[1]. Some of the channels highlighted in the literature through which family background affects the skill formation of children are as follows: financial constraints (Becker and Tomes, 1979, 1986), peer effects (Benabou, 1996), educational policies (Schuetz *et al.*, 2008), non-cognitive skills (Bowles and Gintis, 2002) and educational investments at early stages of life (Cunha and Heckman, 2007).

We deviate from the existing literature by studying whether the effect of family socioeco nomic background continues to influence outcomes during individuals' upper secondary education and early years in the labor market. To do so, we use a rich data set of administrative records for the test scores, individual backgrounds and adult earnings of a cohort of agents in Chile. Our data contain information on the scores of a cohort of students on two different standardized tests, taken in the eighth grade and for college admission, and their earnings during their first years in the labor market. The data set also contains information on the socioeconomic background of the individual's family, which was gathered from a questionnaire answered by parents when the students took the eighth-grade test. We compare the academic outcomes in the college admission tests and labor market earnings of students who had the same outcomes in the eighth grade standardized test but who belonged to families with different SES. Analogously, we compare the labor market earnings of students who had the same outcomes in the college admission test but who were raised in families with differing SESs. Our empirical strategy relies on reduced-form regressions. We control for proxies for occupational choices and selection. In addition, we explore the effects of SES on outcomes within gender, within test score quintile and within individuals attending the same type of school. As a robustness check, we also carry out estimations using a direct measure of family income as proxy for socioeconomic background.

We find a positive and significant effect of SES on college admission test scores and adult earnings. Specifically, in our baseline specification, students with the highest SES obtained a 1.5 standard deviation higher score on the college admission test than students who had the same academic outcomes in the eighth grade test but belonged to the lowest SES. Similarly, among students who obtained the same scores on the college admission test, those with the highest SES earned monthly wages 0.7 standard deviations higher than those belonging to the lowest socioeconomic strata. Additionally, we find evidence that the effect of family SES on the earnings gap between the highest- and lowest-SES agents is increasing during the period of analysis. Finally, we find that the effect of SES on outcomes over the period spanning agents' upper-secondary education and early years in the labor market persists within school type, academic achievement quintile and gender. Overall, our evidence shows that the SES seems to have a ceaseless effect on inequality.

The rest of this paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the data. Section 4 discusses the empirical approach. Section 5 presents and discusses the results. Finally, Section 6 concludes.

2. Related literature

Three strands of the literature are related to our paper: studies that estimate the income intergenerational elasticity coefficient; studies that aim at disentangling a causal effect from intergenerational income associations; and the literature that digs deeper into the causal mechanism through which family background affects children's outcomes. In this section, we discuss these three pieces of empirical evidence, emphasizing how this paper is inserted in and contributes to this literature.

The first strand of this literature estimates the income intergenerational elasticity coefficient. Even though early research found only a small intergenerational elasticity coefficient (Blau and Duncan, 1967; Becker and Tomes, 1986), more recent studies point to a significant influence of parental income and wealth on the economic status of the next generation (Solon, 1992; Zimmerman, 1992; Mulligan, 1997; Solon, 2002). Therefore, this first piece of evidence suggests a positive influence of family background on children's outcomes.

The second strand of this literature intends to disentangle causal effects from intergenerational income associations. To do so, different approaches have been used. One approach relies on the use of data on children and parents related by adoption (Björklund *et al.*, 2006; Sacerdote, 2007). A second approach relies on differences between family members, eliminating in that way the unobserved factors they share (Blau, 1999). Other studies rely on random income shocks that reflect an effect of only income, not innate ability, to identify a causal effect of family income (Shea, 2000; Bratberg *et al.*, 2008; Oreopoulos *et al.*, 2008). Overall, parental income seems to have causal effects on children's outcomes, although these are probably much smaller than intergenerational income correlations (Börklund and Jäntti, 2011).

The third strand of this literature digs deeper into the causal mechanisms behind the observed family associations. One set of studies analyzes differences in intergenerational mobility across the earnings distribution, so as to assess the role of credit constraints. If credit constraints are an important determinant of intergenerational mobility, then earnings mobility would be the greatest among high-earning families, who are not credit constrained (Becker and Tomes, 1979, 1986). Some studies find a negative relationship between the intergenerational elasticity and parental income (Eide and Showalter, 1999; Grawe, 2004; Bratsberg *et al.*, 2005). However, there are a vast number of studies that find the opposite (Behrman and Taubman, 1990; Corak and Heisz, 1999; Schnitzlein, 2014; Raitano and Vona, 2015; Jerrim, 2016; among others). In general, the empirical studies that look at intergenerational mobility patterns across the earnings distribution do not strongly support a significant effect of short-run credit constraints on children's academic outcomes[2].

Other studies try to assess the relative importance of family income and neighborhood environment on sibling income correlations. Solon (1999) documents that neighborhood factors at the most account for the one-fifth of the factors that siblings share. Page and Solon (2003a, 2003b) extract similar conclusions. Raaum *et al.* (2006) use Norwegian census data and find that neighborhood correlations in years of schooling and in long-run earnings account for less than a third of the sibling correlations. Overall, these findings suggest that family is more important than the neighborhood.

The latter conclusion stimulated a wave of studies that try to understand what is important in the family. A first set of articles aim to distinguish between the influence of nature versus nurture factors. Evidence suggests that each factor accounts for about half of the family associations (Björklund and Solon, 2005; Sacerdote, 2007) and that the interaction between both factors can also be important (Börklund and Jäntti, 2011).

Family background can affect children's income by influencing their educational attainment. Eide and Showalter (1999) use US data and find that the intergenerational elasticity falls when years of schooling are considered as an additional variable. Blanden and Machin (2004) conclude that the higher education expansion mainly benefited children from relatively rich families. In the same line, Gregg *et al.* (2015) find that education is not so meritocratic with the role of parental income dominating that of education at the top of the distribution of earnings. Ramey and Ramey (2010) highlight the amount of time spent by parents on childcare as an alternative channel through which investment in human capital differs across families with different SESs.

It could also be argued that high-earning parents presumably have, on average, more ability than low-earning parents. If ability is transmitted from parents to children, then incomes will be persistent across generations even if parental income does not matter. Cameron and Heckman (1998) and Mulligan (1999) present evidence that suggest that the persistence of ability causes intergenerational status correlations. On the other hand, Bowles and Gintis (2002) argue that the genetic transmission of IQ appears to be relatively unimportant. Eriksson *et al.* (2005) find that around 25 per cent of the intergenerational earnings elasticity in Denmark can be accounted for by transmission through the offspring's health. Dohmen *et al.* (2006) find strong intergenerational transmission in income-enhancing personality traits, such as attitudes toward risk and trust.

A different strand of literature has highlighted the role that family ties may also play in determining occupational achievements and earnings. Granovetter (1995) and Pellizzari *et al.* (2011) suggest that the direct influence of parental background on children's labor market outcomes can occur through membership in social networks. Guell *et al.* (2007) explore the relation between mobility and the informative content of surnames in Spain. Corak and Piraino (2011) find that about 40 per cent of a cohort of young Canadian men has been employed with an employer for whom their father also worked. In EU countries, Pellizzari (2010) shows that, given the level of education, family networks affect the probability of a person finding a good job or of being employed.

Therefore, the literature mainly reports:

- positive intergenerational income associations;
- a causal effect of family background on children' s outcomes; and
- a diversity of channels through which family impacts children's outcomes.

Our paper deviates from the existing literature in two regards. First, we study whether the effect of family socioeconomic background continues to influence outcomes during individuals' upper secondary education and early years in the labor market. Second, our administrative data set allows us to directly assess later outcome gaps between individuals who belong to families with different SESs, but who exhibit the same observed academic achievement in their upper secondary education years. As far as we know, no other paper in the literature has explored this question with these types of data.

3. Data

We assemble a panel data set containing information on labor income and education outcomes for a cohort in Chile. Before further describing our data set, we briefly discuss the features of the primary and secondary Chilean education system that are relevant to our analysis. There are three types of schools. The first type consists of public schools financed by government subsidies that are transferred conditionally based on students' attendance and by additional funds received from local governments. The second type is private schools that are partially financed by public funds and are also conditional on students' attendance (voucher private schools). Lastly, there are private schools financed exclusively by parents[3]. Our data span a period where public schools operated in a different institutional setting than voucher private and private schools. First, public schools with vacancies were compelled to enroll all applicants, whereas voucher private and private schools used competitive admission processes. Besides, teachers' contracts at public schools were (and still are) regulated by a teaching statute characterized by a centralized collective bargaining process and a wage structure strongly tied to experience. Private voucher and private schools were (and still are) ruled by standard labor laws. Some of the empirical specifications used in the later analysis will rely on the schools' aforementioned classification. We further discuss our data set in detail.

Our data set contains administrative records for the test scores, individual background and adult earnings of a cohort of agents that we follow for a period of 14 years. We merged three different data sets. First, we collected information on the test scores obtained by eighth graders on the System for Measuring the Quality of Education (SIMCE) test taken in 2000. The SIMCE test is a mandatory national standardized test in Chile designed to evaluate the level of student achievement in the material taught in primary and secondary education in Math, Language, Geography and Science. It is administered annually to fourth, eighth and tenth graders[4].

We merged the first data set with the results obtained by the same students on the College Admission Test (PSU), which is taken four years after the SIMCE, that is, in 2004. The PSU is a standardized test taken to apply for admission to college. It is prepared by the Department of Evaluation, Measurement and Educational Registry (DEMRE) of the University of Chile. The PSU is used by Chilean universities belonging to the Chilean Traditional Universities Group (CRUCH) and other private universities attached to the system. The PSU is taken every year in December and consists of four tests: two mandatory, Language and Math, and two optional, Social Science and Science.

Both the SIMCE and PSU tests consist of four different subtests on the most important subjects taught in the high school: Language, Mathematics, Social Science (for the SIMCE, this takes the form of Geography) and Natural Science. We focus our analysis on students' outcomes in Language and Math – specifically, the average of the two. These subjects are mandatory, whereas Social Science and Natural Science are optional. In addition, the tests on Language and Math are the most important at the national level and have a higher weight in the application process for universities.

Additionally, we include administrative information on the earnings of those agents during their first four years in the labor market. The data on adults' earnings were obtained from the administrative register of the Unemployment Insurance System. Our data set also includes information on the socioeconomic background of the student's family, which was gathered from a questionnaire answered by parents at the time the students took the SIMCE test.

To characterize the agent's SES, we use a categorical variable that classifies the SES of the student's family into five different socio-economic groups: low, medium-low, medium, medium-high and high. These groups are classified by the Ministry of Education of Chile, based on information from four standardized variables: educational level of the father, educational level of the mother, level of household income and school vulnerability index[5]. The first three variables come from the questionnaire of the parents included in the SIMCE, and the fourth variable comes from the National Board of School Aid and Scholarships (JU-

NAEB). The last variable is computed based on the fraction of the students in the school that are in a situation of vulnerability. This is a school-level measure intended to summarize the socioeconomic environment to which the students are exposed during their primary and secondary education. Alternatively, we also present estimates directly using family income. Unlike the SES variable, family income is an individual-level categorical variable and, thus, allows us to estimate school fixed-effect models. The family income information is self-reported through questionnaires prepared by the SIMCE, collected in the days immediately before the test. Considering the collected information, the following five family income groups have been constructed: low, middle-low, middle, middle-high and high[6]. We believe that it is valuable to explore both types of socioeconomic measures, as the socioeconomic background of an agent is not determined solely by the direct influence of family income but also by the SES of the environment to which agents are the most closely exposed.

As a preliminary exploration of our data set, we present in this section some descriptive statistics. Table I describes the main data used in our empirical estimates. We observe that the SIMCE scores in both Language and Math fluctuate from slightly above 100 points to close to 400 points, with an average around 270 points. The PSU scores fluctuate between 160 and 850 points, with an average of around 500 points. Average monthly earnings fluctuate from US$5 to US$7240, with an average of US$822. We only exclude agents whose average monthly wage during 2010–2013 was US$0, as it is likely that they are not participating in the formal labor market. In any case, those agents represent a very small fraction of the total observations. Among the remaining students, 4 per cent have a low SES, whereas 13 per cent belong to the highest SES group. The rest of the students come from families with a middle-low, middle or middle-high SES.

We present in Table II the number of students taking the SIMCE test and the PSU test and the number of these individuals observed in the unemployment insurance system. As can be observed from Table II, our cohort includes 242,497 individuals that took the SIMCE test. Among these, only 100,083 are observed taking the PSU test and 75,791 are active in the formal labor market. It can also be noted from Table II that 41 per cent of those who took the SIMCE belong to the two lowest SES categories, and this percentage falls to 21 per cent

Variable	Mean	SD	Min	Max	*N*
Students' human capital					
SIMCE test score: Language	273.19	46.97	109	395	242,497
Math	274.26	46.48	118	382	242,497
Average	273.73	42.32	129	388	242,497
PSU test score: Language	490.13	107.96	167	850	100,083
Math	493.52	108.75	178	850	100,083
Average	491.83	101.20	201	840	100,083
Monthly wage (US$)	822.05	600.04	5	7240	75,791
Socioeconomic background					
Low SES	0.04	0.19	0	1	242,497
Middle-low SES	0.18	0.38	0	1	242,497
Middle SES	0.36	0.48	0	1	242,497
Middle-high SES	0.29	0.45	0	1	242,497
High SES	0.13	0.337	0	1	242,497

Source: SIMCE 2000, PSU 2004 and Unemployment Insurance Database 2010-2013
Notes: Wages were computed by taking the average monthly wage for the period 2010-2013 and are expressed in real terms. For reference, we express wages in US$, using an exchange rate of 600 Chilean pesos to US$1

Table I.
Summary statistics

when considering PSU takers. Therefore, we cannot discard the possibility of sample selection in our data. As we explain later, we rely on the two-step procedure proposed by Heckman (1979) to address the potential selection bias problem[7].

Table III characterizes the SES groups according to the type of school attended by the student, the parents' education and the family income. We observe that students from low-SES families are mainly concentrated in public schools, whereas students from high-SES families mainly attend private schools. In addition, we observe a positive gradient for parental education and family income across the SES categories. In the last row of Table III, we also observe that 42 per cent of the SIMCE takers were enrolled in public schools, 44 per cent in voucher private schools and 14 per cent in private schools.

To further inspect our data set, we also decompose the total variance of the test scores and monthly wages into three components: the within-school variance, the between-school/within-school-type variance and the between-school-type variance. The total variance of the corresponding outcome can be expressed as follows:

$$S^2 = \frac{\Sigma_k \ \Sigma_j \ \Sigma_i \left[\left(X_{i,j,k} - \overline{X_{j,k}} \right)^2 + \left(\overline{X_{j,k}} - \overline{X_k} \right)^2 + \left(\overline{X_k} - \overline{X} \right)^2 \right]}{N}, \tag{1}$$

where i is the index for the student, j for the school and k for the school type (public, voucher private or private); $\overline{X_{j,k}}$ is the average outcome of the students within school j of type k; $\overline{X_k}$ is the average outcome of students within schools of type k; and $\overline{X}$ is the total average of the

Socioeconomic group	SIMCE	PSU	Labor market
Low SES	22,347	3,629	2,795
Middle-low SES	76,227	16,989	13,517
Middle SES	85,336	34,606	27,745
Middle-high SES	40,674	29,449	21,888
High SES	17,913	15,410	9,846
Total	242,497	100,083	75,791

Source: SIMCE 2000, PSU 2004 and Unemployment Insurance Data Base 2010-2013

Table II. Number of individuals by SES

		Schools			Parents education		Family income
	Total	Public	Voucher private	Private	Mother	Father	(Monthly US$)
Low SES	3.67%	7.65%	1.06%	0.0%	7.54	7.49	176
Middle-low SES	17.78%	36.39%	5.84%	0.0%	9.11	9.36	256
Middle SES	37.60%	41.06%	44.39%	0.0%	10.97	11.38	381
Middle-high SES	28.93%	14.90%	47.46%	13.44%	12.92	13.55	647
High-income SES	13.02%	0.0%	1.24%	86.56%	15.82	17.05	2,002
Total/Average	100%	100.0%	100%	100%	11.71	12.27	637
Students	100%	42%	44%	14%	–	–	–

Source: SIMCE 2000, PSU 2004 and Unemployment Insurance Database 2010-2013
Note: Family income corresponds to the average monthly income of the families within each SES category, using an exchange rate of 600 Chilean pesos to US$1

Table III. Characterization of SES groups corresponding outcome. The first term of equation (1) represents the within-school variance; the second term represents the variance between schools of the same type; and the third term represents the variance between school types. Table IV presents the results of the variance decomposition. We observe that 70 per cent of the variance in SIMCE test scores is generated within schools, around 23 per cent is generated between schools of the same type and 7 per cent is generated between school types. Repeating the same variance decomposition for the PSU test scores, we find that about 63 per cent of the variance is generated within schools, 22 per cent is generated between schools of the same type and about 16 per cent is generated between school types. With respect to adult earnings, we observe that 82 per cent of the variance in monthly wages is generated within schools, around 11 per cent is generated between schools of the same type and 7 per cent is generated between school types.

In Tables V and VI, we present statistics that are more closely related to the empirical approach used in Section 3. Table V compares the PSU test scores for students that scored in the same quintile on the SIMCE test but differ in their SES. We observe a positive gradient

	Students	Schools	School type
SIMCE	69.9%	22.7%	7.4%
PSU	62.6%	21.8%	15.6%
Wages	81.9%	11.2%	6.9%

Source: Own elaboration based on SIMCE 2000, PSU 2004 and Unemployment Insurance Database 2010-2013
Note: Wages were computed by taking the average monthly wage for the period 2010-2013, using an exchange rate of 600 Chilean pesos to US$1

Table IV. Variance decomposition

	SIMCE quintiles				
Socioeconomic group	Q1	Q2	Q3	Q4	Q5
Low SES	3.64	4.05	4.44	4.89	5.38
Middle-low SES	3.68	4.14	4.58	4.99	5.52
Middle SES	3.85	4.32	4.70	5.12	5.67
Middle-high SES	4.14	4.46	4.94	5.35	6.00
High SES	4.61	4.98	5.33	5.74	6.39

Source: Own elaboration based on SIMCE 2000 and PSU 2004
Note: PSU test score presented in standard deviations

Table V. PSU Test score by SIMCE quintiles and SES

	PSU Quintiles				
Socioeconomic group	Q1	Q2	Q3	Q4	Q5
Low SES	1.02	1.12	1.31	1.53	2.11
Middle-low SES	1.11	1.22	1.41	1.69	1.95
Middle SES	1.22	1.32	1.44	1.73	2.13
Middle-high SES	1.27	1.35	1.47	1.79	2.38
High SES	1.61	1.63	1.65	2.16	3.14

Source: Own elaboration based on SIMCE 2000 and PSU 2004
Note: Wages presented in standard deviations

Table VI. Wages by PSU quintiles and SES within each SIMCE quintile. That is, students with a higher SES perform better on the college admission test than students from low SES, conditional on the SIMCE score quintile. Table VI includes adult earnings as the corresponding outcome. Again, within each quintile of PSU test scores, students from families with a higher SES or income have higher earnings in the labor market. Therefore, the preliminary evidence in Tables V and VI suggests that the SES affects the outcomes during the agents' upper secondary education and early years in the labor market. In the next sections, we formally analyze this hypothesis.

4. Empirical strategy

Our empirical approach consists in reduced-form regressions motivated by the following idea. Assume that the outcome of agent i at period t depends on the stock of human capital of that agent, denoted by $H_{i,t}$, and other factors summarized by a single variable $Z_{i,t}$. Denote by $O_{i,t}$ the outcome of agent i at period t. Then:

$$O_{i,t} = f\left(H_{i,t}, Z_{i,t}\right). \tag{2}$$

Additionally, the production of human capital can be described using the following function:

$$H_{i,t} = f\left(H_{i,t-1}, E_{i,t}, F_{i,t}\right), \tag{3}$$

where $H_{i,0} = A_i$, with A_i being the innate abilities, which are assumed to be constant over time, $H_{i,t-1}$ is the stock of human capital of agent i at time $t-1$, the $E_{i,t}$ are the inputs received from formal education, and the $F_{i,t}$ are the family inputs. At the same time, the inputs provided by formal education depend on the type and quality of the educational institution that a student attends, denoted by $S_{i,t}$, and the past influence of schooling:

$$E_{i,t} = f\left(S_{i,t}, H_{i,t-1}\right). \tag{4}$$

Parents make decisions on the type and quality of the educational institutions that their children attend. The stock of human capital of an individual also influences the type and quality of the educational institutions that will be attended. Therefore:

$$S_{i,t} = f\left(F_{i,t}, H_{i,t-1}\right). \tag{5}$$

Similarly, all other (non human-capital) variables that determine success in education and in the labor market can be influenced by family SES:[8]

$$Z_{i,t} = f\left(F_{i,t}\right). \tag{6}$$

We assume that the functional forms involved in equations (2) to (6) are well-represented by the following linear model:

$$O_{i,t} = \alpha_0 + \alpha_1 E_{i,t} + \alpha_2 F_{i,t} + \alpha_3 H_{i,t-1} + \alpha_4 Z_{i,t} + \varepsilon_{i,t} \tag{7}$$

$$E_{i,t} = \beta_1 S_{i,t} + \beta_2 H_{i,t-1} + \xi_{i,t} \tag{8}$$

$$S_{i,t} = \gamma_1 F_{i,t} + \gamma_2 H_{i,t-1} + \upsilon_{i,t} \tag{9}$$

$$Z_{i,t} = \lambda_1 F_{i,t} + \zeta_{i,t}, \tag{10}$$

where $\varepsilon_{i,t}$, $\xi_{i,t}$, $\upsilon_{i,t}$ and $\zeta_{i,t}$ are idiosyncratic error terms. Substituting equations (8), (9) and (10) into equation (7), we arrive at the following reduced-form linear regression:

$$O_{i,t} = \alpha_0 + \rho F_{i,t} + \theta H_{i,t-1} + \epsilon_{i,t}, \tag{11}$$

where

$$\rho = \alpha_2 + \alpha_1 \beta_1 \gamma_1 + \alpha_4 \lambda_1$$

$$\theta = \alpha_3 + \alpha_1 \beta_1 \gamma_2 + \alpha_1 \beta_2$$

$$\epsilon_{i,t} = \varepsilon_{i,t} + \alpha_1 \xi_{i,t} + \alpha_1 \beta_1 \upsilon_{i,t} + \alpha_4 \zeta_{i,t}.$$

In the reduced-form regression model described using equation (11), the coefficient ρ captures the overall effect of family SES on outcome $O_{i,t}$, conditional on the initial level of human capital. The baseline model considers a quasi-experiment that compares the outcomes at period t of agents who had the same academic achievement at some period $t-j$ but whose SESs differ. In this way, we estimate the total – direct and indirect – effect of SES on outcomes. Thus, in our baseline specification, the coefficient ρ captures the overall influence of SES on outcomes, conditional on the initial human capital stock of the student.

To implement the empirical model described by equation (11), we exploit a rich administrative data set that contains information about the test scores (at two different moments in time) and adult earnings of a cohort of agents. We run three different regressions. First, we use students' SIMCE test scores as a measure of the initial level of human capital, with the score obtained on the PSU test as the outcome. As explained in Section 3, the PSU test is taken four years after the SIMCE test. Our main variables of interest are those for family socioeconomic background. A positive and significant coefficient ρ would be evidence that family SES affects the student's outcome at the stage of upper-secondary education, conditional on the initial level of human capital[9].

In the second regression, we use the average monthly wage earned by the individual (2010-2013) as the outcome. As the measure of the initial stock of human capital, we use the test score obtained by the individual on the PSU test taken in 2004. In this case, a positive and significant coefficient ρ is interpreted as evidence that family SES influences the individual's outcome during their tertiary education and early years in the labor market. Lastly, we run a third regression where we use, again, wages as the outcome and the SIMCE test score obtained by the agent in the year 2000 as the measure of the initial stock of human capital. By comparing the coefficients ρ in the regressions using the PSU and SIMCE test scores as the measures of initial human capital, we can get some insight into the nonlinearity of the effects of SES on outcomes over the upper-secondary–labor market period.

In addition, we expand the baseline model along several dimensions. First, we include two sets of control variables to avoid problems of omitted variable bias. In the specification that considers the PSU test as the outcome variable, we include as additional covariates gender[10], the region where the school is located, and grade retention. In the wage regression our set of additional control variables includes gender, nationality, the region where the firm is located and the industrial classification of economic activity.

Second, Table II shows that while the 242,497 individuals of the cohort that we follow took the SIMCE test, only 100,083 of them took the PSU test. A similar pattern can be seen for labor activity, where we observe 75,791 individuals active in the labor market during the period 2010-2013. From Table II, we can also see that about 41 per cent of those who took the SIMCE belong to the two lowest SES categories, whereas this percentage decreases to 21 per cent when considering PSU test takers. Therefore, we may face a selection bias problem, as less skilled students from the lowest SES categories might choose not to continue to tertiary education and then, not to participate in the formal labor market. To address this potential sample selection bias problem, all the regressions are estimated using the two-step procedure proposed by Heckman (1979). As exclusion restrictions, we use preschool education attendance (which is not mandatory) when the outcome variable is the PSU test score, and civil status and parental labor activity when the outcome variable is wages.

Third, we examine the potential heterogeneity of this effect in terms of school type and gender[11]. Therefore, we present disaggregated regressions by school type and gender[12]. We also present regressions by quintile of the test score distribution to explore nonlinearities of the effect of SES on outcomes. Finally, as a check on the robustness of our results, we present alternative estimates where we explicitly include family income instead of the SES variable. Those alternative specifications will allow us to study explicitly the effects of family income, an individual-level variable, on academic and labor market outcomes.

5. Results

We first present, in Table VII, the results of the empirical model that uses the SIMCE test score as the measure of the initial stock of human capital and the PSU test score as the measure of academic outcome. As explained in Section 4, we expand the baseline model (11) by including as additional covariates a gender dummy (except in Column 2), the school's region and grade retention. We estimate this expanded model using Heckman's two-step procedure to deal with potential sample selection problem. We observe in Column (1) a positive and significant effect of SES on academic outcome. Concretely, students with the highest SES obtained a 1.5 standard deviation higher score on the PSU test than those with the lowest SES, conditional on their academic achievement four years before in the SIMCE test and the other covariates included in the model. This positive and significant effect of SES on academic outcome is observed across all of the SES categories. Moreover, we observe in Column (2) that the effect of SES in the case of women is quite similar to that observed for the total sample. Additionally, we observe in Columns (3) to (7) that the positive and significant effect of SES on academic outcome is also observed across the entire distribution of SIMCE test scores.

In Table VIII, we analyze the effect of SES within each school type. First, we observe that the positive and significant effect of SES on academic achievement is present within all school types[13]. Second, we observe that within the public schools and voucher private schools, the SES–outcome gradient seems to be slightly more pronounced than for the total sample. Specifically, Column (1) of Table VIII shows that public school students with a middle–high SES background obtained a 1.3 standard deviation higher score in the college admission test than those with the lowest SES, conditional on their initial stock of human capital. This effect is only 1.1 in the total sample. When comparing the lower SES categories, the conditional SES gradient within public schools is more similar to that observed for the total sample. In the case of voucher private schools (Column 2), we observe that, compared with the total sample, the conditional SES–outcome gradient becomes more evidently pronounced for the middle SES categories. For instance, we observe that the SES effect is 1.3

Table VII. Heckman second step regression of SES on academic outcomes in upper secondary education, by gender and SIMCE quintiles

	(1) Total sample	(2) Female	(3) Quintile 1	(4) Quintile 2	(5) Quintile 3	(6) Quintile 4	(7) Quintile 5
SIMCE (student-level)	0.916*** (0.00891)	0.918*** (0.0120)	0.320*** (0.0649)	0.579*** (0.0932)	0.628*** (0.0447)	0.940*** (0.0502)	0.874*** (0.0191)
Middle-low SES	0.225*** (0.0146)	0.208*** (0.0190)	0.0567 (0.0584)	0.136** (0.0555)	0.114*** (0.0338)	0.199*** (0.0314)	0.167*** (0.0304)
Middle SES	0.608*** (0.0184)	0.571*** (0.0236)	0.262** (0.124)	0.457*** (0.122)	0.292*** (0.0617)	0.460*** (0.0562)	0.339*** (0.0412)
Middle-high SES	1.124*** (0.0260)	1.048*** (0.0334)	0.576** (0.237)	0.937*** (0.240)	0.548*** (0.112)	0.821*** (0.0999)	0.600*** (0.0590)
High SES	1.548*** (0.0284)	1.457*** (0.0360)	0.951*** (0.305)	1.502*** (0.304)	0.978*** (0.137)	1.236*** (0.119)	0.922*** (0.0643)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	189,135	95,656	36,343	37,352	37,853	38,420	39,167
Adjusted R^2	0.644	0.646	0.114	0.167	0.186	0.226	0.443

Notes: The excluded category for the SES dummy is the lowest one; the variables SIMCE Test Score and PSU Test Score are in standard deviations; control variables are gender, school's region and grade retention; gender dummy is excluded as control variable when performing the female regression; we use preschool education attendance as exclusion restriction. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table VIII. Heckman second step regression of SES on academic outcomes in upper secondary education, by school type

	(1) Public school	(2) Voucher private school	(3) Private school
SIMCE test score (student-level)	1.097*** (0.0281)	0.927*** (0.0184)	1.042*** (0.0722)
Middle-low SES	0.280*** (0.0241)	0.264*** (0.0475)	–
Middle SES	0.644*** (0.0336)	0.813*** (0.0585)	–
Middle-high SES	1.314*** (0.0550)	1.289*** (0.0727)	–
High SES	–	1.634*** (0.0845)	0.357*** (0.0559)
Control variables	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes
Observations	106,324	67,071	15,740
Adjusted R^2	0.595	0.569	0.561

Notes: The excluded category for the SES dummy is the lowest one (the low SES category in the regression for public and voucher private schools and the middle-high SES category in the regression for private school); the variables SIMCE Test Score and PSU Test Score are in standard deviations; control variables are gender, school's region and grade retention; we use preschool education attendance as exclusion restriction. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

and 0.8 standard deviations in the middle and middle-high SES categories within the private voucher schools, respectively, whereas the analogous numbers are 1.1 and 0.6 in the total sample. In the case of the middle-low and the high SES, even though the gradient is still more pronounced than for the total sample, the difference is less significant[14]. Therefore, Table VIII suggests that the conditional SES–outcome gradient persists within school types, and it is more pronounced within public schools and private voucher schools relative to the total sample. Overall, the results of Tables VII and VIII support the idea that the SES significantly affects academic outcomes during upper secondary education.

Tables IX and X present the impact of SES on labor market outcomes. We use the score obtained on the PSU test as the measure of initial human capital and the average monthly real earnings earned by agents during their first four years in the labor market as the outcome. In this case, the additional covariates include gender, nationality, firm's region and the industrial classification of the firm's economic activity. As before, we use Heckman's two-step procedure to estimate this regression. We also disaggregate by quintiles in the human capital distribution and by school type. We observe, conditional on the initial stock of human capital, a positive and a significant effect of SES on labor market earnings. For instance, students with the highest SES earned a monthly wage 0.7 standard deviations higher than students with low SES, conditional on the academic achievement in the college admission test taken five to eight years before, and on the other covariates included in the model[15]. For the total sample, the positive and significant effect of SES on earnings is observed for all SES categories. Column (2) shows that there are no significant differences in the effect of SES on outcomes during tertiary education and the first years in the labor market between women and men, a similar result to the one obtained for the upper secondary education period. However, we observe that in the highest quintile, only students from families with the highest SES exhibit significantly greater earnings compared with students from the lowest SES.

Table X disaggregates by school type. We observe that the positive and significant effect of SES on earnings persists within each school type. In addition, we observe that, compared with the total sample, the conditional SES–earnings gradient is more stepped only within the public schools. Specifically, within public schools, individuals belonging to the middle-

Table IX. Heckman second step regression of SES on outcomes in tertiary education and the labor market, by gender and PSU quintiles

	(1) Total Sample	(2) Female	(3) Quintile 1	(4) Quintile 2	(5) Quintile 3	(6) Quintile 4	(7) Quintile 5
PSU (student-level)	0.342*** (0.00487)	0.344*** (0.00613)	0.0860*** (0.0141)	0.154*** (0.0405)	0.151*** (0.0512)	0.409*** (0.0543)	0.821*** (0.0315)
Middle-low SES	0.0396* (0.0224)	0.0435 (0.0272)	0.0610** (0.0239)	0.0603* (0.0338)	0.0325 (0.0469)	0.106 (0.0790)	−0.143 (0.159)
Middle SES	0.0515** (0.0215)	0.0764*** (0.0261)	0.138*** (0.0236)	0.0926*** (0.0325)	0.0593 (0.0447)	0.135* (0.0754)	−0.0300 (0.151)
Middle-high SES	0.109*** (0.0221)	0.152*** (0.0270)	0.163*** (0.0264)	0.123*** (0.0340)	0.123*** (0.0453)	0.187** (0.0752)	0.0877 (0.149)
High SES	0.661*** (0.0245)	0.622*** (0.0300)	0.430*** (0.0485)	0.329*** (0.0471)	0.404*** (0.0513)	0.568*** (0.0777)	0.638*** (0.150)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	100,083	54,020	20,203	19,878	20,066	19,939	19,997
Adjusted R^2	0.268	0.258	0.183	0.172	0.177	0.196	0.296

Notes: The excluded category for the SES dummy is the lowest one; the variables wage and PSU Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; gender dummy is excluded as control variable when performing the female regression; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table X.
Heckman second step regression of SES on outcomes in tertiary education and the labor market, by school type

	(1) Public school	(2) Voucher private school	(3) Private school
PSU test score (student-level)	0.245*** (0.00664)	0.311*** (0.00729)	0.657*** (0.0164)
Middle-low SES	0.0405* (0.0212)	0.0564 (0.0594)	–
Middle SES	0.0836*** (0.0212)	0.0617 (0.0556)	–
Middle-high SES	0.213*** (0.0248)	0.121** (0.0557)	–
High SES	–	0.302*** (0.0755)	0.381*** (0.0411)
Control variables	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes
Observations	40,642	42,570	16,871
Adjusted R^2	0.208	0.185	0.301

Notes: The excluded category for the SES dummy is the lowest one (the low SES category in the regression for public and voucher private schools and the middle-high SES category in the regression for private school); the variables wage and PSU Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

high SES earned wages 0.2 standard deviations higher than individuals with the lowest SES, conditional on their initial stock of human capital. This effect is 0.1 in the total sample. However, within voucher private schools and private schools, we observe that the conditional SES–earnings gradient is less pronounced than the one for the total sample.

In Tables XI and XII, we explore how SES separately affects outcomes during agents' upper-secondary education and the period spanning their tertiary education and early years in the labor market. To do so, we estimate a model where the measure for the initial human capital is the SIMCE test score and the outcome is the average monthly real wage earned by the agent in the first four years in the labor market. Among students that obtained the same scores on the SIMCE test, those with the highest SES earned monthly earnings 1.0 standard deviations higher than those of low-SES students[16]. As before, we observe that the positive and significant effect of SES on earnings persists in spite of gender, and within all quintiles and school types. In addition, the information provided by Tables IX and XI allows us to compute the fraction of the 1.0 standard deviation gap that is generated during an agent's upper secondary education, tertiary education and first years in the labor market, respectively. The earnings gap created during their tertiary education and early years in the labor market is 0.7 standard deviations, as observed in Table IX. Therefore, we can conclude that approximately 0.3 standard deviations of the 1.0 standard deviation gap are generated during their upper secondary education. Therefore, most of the gap between the highest and the lowest SESs is generated during tertiary education and the early years in the labor market. In this sense, we can say that the effect of SES on the inequality between the highest and lowest SES groups is increasing over the period of analysis.

Table XII shows that this conclusion is somewhat less marked within public and voucher private schools. We observe in Table XII that, in public schools, the earnings gap between individuals with a middle-high SES and those with the lowest SES that is generated during the period spanning upper secondary education and the early years in the labor market is 0.4. Table X shows that 0.2 of this earnings gap is created during the agent's tertiary education and early years in the labor market. By doing an analogous analysis for the voucher private schools, we can conclude that approximately half of the earnings gap between the highest and the lowest SES categories is generated during their tertiary

Table XI. Heckman second step regression of SES on outcomes in upper secondary education and the labor market, by gender and SIMCE quintiles

	(1) Total sample	(2) Female	(3) Quintile 1	(4) Quintile 2	(5) Quintile 3	(6) Quintile 4	(7) Quintile 5
SIMCE (student-level)	0.200*** (0.00249)	0.216*** (0.00337)	0.115*** (0.00982)	0.115*** (0.0206)	0.198*** (0.0241)	0.232*** (0.0251)	0.457*** (0.0173)
Middle-low SES	0.0627*** (0.00756)	0.0666*** (0.0107)	0.0536*** (0.00955)	0.0654*** (0.0119)	0.0765*** (0.0145)	0.0871*** (0.0203)	0.0746* (0.0406)
Middle SES	0.166*** (0.00768)	0.169*** (0.0107)	0.166*** (0.0105)	0.186*** (0.0124)	0.186*** (0.0146)	0.193*** (0.0199)	0.219*** (0.0389)
Middle-high SES	0.338*** (0.00911)	0.360*** (0.0124)	0.269*** (0.0182)	0.321*** (0.0175)	0.336*** (0.0180)	0.342*** (0.0221)	0.425*** (0.0396)
High SES	1.032*** (0.0119)	0.973*** (0.0157)	0.680*** (0.0403)	0.759*** (0.0344)	0.792*** (0.0292)	0.824*** (0.0283)	1.160*** (0.0419)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	212,115	102,517	39,318	40,822	42,220	43,867	45,888
Adjusted R^2	0.254	0.259	0.183	0.175	0.178	0.171	0.264

Notes: The excluded category for the SES dummy is the lowest one; the variables Wage and SIMCE Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; gender dummy is excluded as control variable when performing the female regression; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table XII. Heckman second step regression of SES on outcomes in upper secondary education and the labor market, by school type

	(1) Public school	(2) Voucher private school	(3) Private school
SIMCE test score (student-level)	0.166*** (0.00306)	0.203*** (0.00431)	0.462*** (0.0157)
Middle-low SES	0.0588*** (0.00710)	0.0799*** (0.0194)	–
Middle SES	0.166*** (0.00764)	0.183*** (0.0181)	–
Middle-high SES	0.386*** (0.0117)	0.325*** (0.0193)	–
High SES	–	0.602*** (0.0447)	0.535*** (0.0378)
Control variables	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes
Observations	118,944	74,603	18,568
Adjusted R^2	0.201	0.182	0.248

Notes: The excluded category for the SES dummy is the lowest one (the low SES category in the regression for public and voucher private schools and the middle-high SES category in the regression for private school); the variables wage and PSU Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

education and early years in the labor market. In the case of private schools, Table XII shows an earnings gap of 0.5, of which 0.4 is generated in the tertiary education–labor market period.

Table XIII presents regressions of how SES affects both the initial wage and wage growth, controlled by the PSU test as the measure of initial human capital. We can observe that students with the highest SES earned an initial wage 0.8 standard deviations higher than that earned by individuals in the lowest SES category. We can also notice that individuals with the highest SES experience a wage increase of 0.4 standard deviations compared with those in the lowest SES group. Thus, higher SES levels are related to both higher initial wages in the labor market and higher wage growth rates. This means that SES levels drive divergent income levels along the first year of work life in workers with different SES levels.

Table XIII. Heckman second step regression of SES on initial wage and wage growth

	Initial wage	Wage growth
PSU test score (student-level)	0.227*** (0.0100)	0.300*** (0.0106)
Middle-low SES	0.047 (0.0288)	0.040 (0.0310)
Middle SES	0.089*** (0.0278)	0.095*** (0.0300)
Middle-high SES	0.157*** (0.0298)	0.209*** (0.0322)
High SES	0.843*** (0.0397)	0.441*** (0.0427)
Control variables	Yes	Yes
Selection correction	Yes	Yes
Observations	100,083	100,083
Adjusted R^2	0.139	0.127

Notes: The excluded category for the SES dummy is the lowest one; the variables Initial Wage, Wage Growth and PSU Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

We also explore the family background–outcome gradient that is observed within schools by estimating school-specific fixed effects models. To do so, we exploit the family income categories collected from the questionnaires prepared by the SIMCE[17]. Family income is an individual level variable, so we can perform school fixed-effects regressions to decompose the effect driven by the family background from that driven by school characteristics. The results are presented in Table XIV. We estimate in the Appendix the empirical models of Tables VII to XIII using family income as an individual-level measure for family background. Even though we will pospone the discussion of this set of results for the end of this section, we will refer to Tables AI, AIII and AV to compare the school fixed-effects results reported in Table XIV. Column (1) of Table XIV presents the results from the model that includes the PSU test score as the outcome and the SIMCE test score as the measure of the initial stock of human capital. In Columns (2) and (3), the outcome variable is monthly earnings. We observe in Column (1) that, controlling for the idiosyncratic characteristics of the school, students from the highest income group obtained a 0.7 standard deviation greater score on the PSU college admission test than those in the lowest family income group. Therefore, the effect of family background is reduced by almost a half compared with the one estimated in Column (1) of Table AI. In Column (2), we observe that the within-school effect of family background on monthly earnings is 0.3 standard deviations higher for individuals belonging to the richest families compared with those in the poorest families. This effect is about one-third of the one observed when school-specific fixed effects are excluded (see Column 1 in Table AIII). Finally, when considering the upper-secondary–labor market period (Column 3), we observe an effect that is one-third of that observed without conditioning by school-specific fixed effects (see Column 1 in Table AV). Therefore, Table XIV shows that part of – but not all – the effect of family income is channeled through differences in the characteristics of the school that the student attends.

As a robustness check, we estimate, in the Appendix, the empirical models of Tables VII to XIII using family income as an individual-level measure for family background. Consistent with the findings derived from the models that included the SES variable, we

	(1) PSU test score	(2) Monthly wage	(3) Monthly wage
SIMCE test score (student-level)	0.874*** (0.0089)		0.171*** (0.0034)
PSU test score (student-level)		0.291*** (0.0062)	
Middle-low income	0.167*** (0.0093)	0.025 (0.0156)	0.038*** (0.0065)
Middle income	0.395*** (0.0149)	0.049*** (0.0171)	0.084*** (0.0083)
Middle-high income	0.595*** (0.0194)	0.116*** (0.0207)	0.165*** (0.0115)
High income	0.689*** (0.0217)	0.292*** (0.0311)	0.327*** (0.0189)
School fixed-effects	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes
Observations	78,098	54,891	129,136
Adjusted R^2	0.677	0.280	0.285

Notes: The excluded category for the family income dummy is the lowest one; the variables wage, SIMCE Test Score and PSU Test Score are in standard deviations; control variables are gender, school's region and grade retention in Column (1), and gender, nationality, firm's region and firm's industrial classification economic activity in Columns (2) and (3); we use preschool education attendance as exclusion restriction in Column (1); we use civil status and parental labor activity as exclusion restrictions in Columns (2) and (3). *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table XIV. School fixed-effects regression observe a positive and significant effect of family income on academic outcome, conditional on the initial stock of human capital. We also observe that this positive and significant influence of family income on academic achievement is present across the entire distribution of human capital, and school types. In addition, when using our family income variable, the magnitude of the effects for women are also similar to those found for the aggregate sample. We also find that most of the earnings gap generated during the period spanning upper secondary education and the early years in the labor market is created from the tertiary education. Finally, we observe in Table VIIA that a higher family income increases both the initial wages and wage growth, similar to the conclusion when using the SES variable in the model.

6. Conclusions

We studied the effect of family SES on the academic performance of students in the later stages of the educational process and in their early years in the labor market. We used a rich data set of administrative records for test scores, individual backgrounds and adult earnings, of a cohort of agents. Using reduced-form regressions, we find that among agents with the same stock of human capital in their upper secondary education, those with the highest SES perform better on the college admission test and earn higher wages in their early years in the labor market. Specifically, students with the highest SES obtained a 1.5 standard deviation higher score on the college admission test than students who were from the lowest socioeconomic group but had the same stock of human capital at the beginning of their upper secondary education. Similarly, among students who obtained the same scores on the college admission test, those from the highest socioeconomic group earned 0.7 standard deviations more per month during their first years in the labor market than agents with low SES.

Additionally, we find that the effect of SES on the earnings gap between the highest and the lowest SES agents is increasing during the period of analysis. Our results show that 70 per cent of the difference between the highest and the lowest SESs is generated during tertiary education and the early years in the labor market. We also find that SES affects not only initial wage but also wage growth. The existing literature mainly focuses on the effect of parents' background on children's results without decomposing the effect during different stages of children's education and labor life. Relevant literature, such as Cunha and Heckman (2007), has emphasized the importance of education investment at early stages that lowest SES students' human capital converges to the highest SES students' human capital. We complement this literature, stressing that it could also be crucial to invest in the lowest SES students at later stages of their formation, given that SES continuously affects outcome differences.

We have also performed a disaggregated analysis by gender, by quintiles of distribution of SIMCE test scores and by type of school. Our findings show that the effect of SES on outcomes for women resembles the one for the total sample during the whole period of analysis. Moreover, the SES–outcome gradient is observed across the entire distribution of SIMCE test scores and within school types. However, we found that the conditional SES–outcome gradient is more pronounced within public schools and private voucher schools relative to the total sample during the upper secondary period. For the period spanning tertiary education and the first years in the labor market, our results show that, compared with the total sample, the conditional SES–earnings gradient is steeper only within the public schools. Moreover, the estimates from the school fixed-effects models show that part – but not all – of the effect of family income is triggered through the family's choice of school.

The latter results are in line with the literature that finds that family background is an important variable that affects labor market earnings. However, these results also suggest that school choice matters. Therefore, education investment at all stages of the lowest SES students' formation is an important mechanism to limit income disparities. The higher conditional SES–wage gradient within public schools is an interesting result taking into account the fact that these types of schools allocate their resources in a more centralized form than private schools.

An interesting avenue for future research is to try to disentangle the channels through which SES affects outcomes. They could come from diverse sources. For instance, richer parents can send their children to better schools and tertiary education institutions, privately invest more in inputs at home, have better contacts in labor markets, invest more in children's health, etc. A greater understanding of the quantitative importance of such channels is important for designing public policies that reduce inequality. In this paper, we have empirically shown that whatever the most important channel is, SES seems to have a ceaseless effect over life.

Notes

1. Other studies also highlight the effects that the social connections of affluent parents in the labor market exert on the labor market outcomes of their children (Granovetter, 1995; Pellizzari *et al.*, 2011).
2. Consistent with this conclusion, Cameron and Heckman (1998) conclude that long-run family factors and not short-run credit constraints play a decisive role in explaining academic achievement, at least in the USA.
3. Non-voucher private schools are generally for-profit, whereas private subsidized schools can be either for-profit or nonprofit. Non-voucher private schools include both religious (mainly Catholic) and nonreligious schools.
4. Since 2012, the SIMCE test is also used to evaluate second and sixth graders.
5. The educational level variable is an index that ranges from 1 to 48, counting from the first year of primary education, and increases with the level of formal education completed by the individual. For example, for two parents who completed two years of tertiary education but one in a bachelor program and the other in a program leading to professional qualifications, the index will be higher for the parent who studied the bachelor program.
6. The five reported categories of monthly family income are keyed to the following cuts: US$170, US$340, US$670 and US$1650.
7. In addition, the cohort includes 121,252 women (50%) and 56,088 foreign individuals (26%). Moreover, 61% of the cohort attained a preschool education, 14% has one year of grade retention and 5% has two or more years of grade retention.
8. For instance, professional networks.
9. As we explain later in this section, the estimation of ρ might be biased because of the existence of omitted variables and sample selection. In the next paragraphs, we explain the variables and the methodology that we use to address these potential problems.
10. This is a dummy variable that takes the value 1 for female students and 0 for male students.
11. We would like to thank an anonymous referee for this useful insight.
12. We exclude the gender variable from the set of control variables when performing regressions by gender.

13. Notice that because of the segregation that exists in the Chilean school system, no public schools are classified in the high SES category, and no private schools are classified in the middle or lower SES categories. This explains the empty spaces in the tables that disaggregate by school type. The excluded category is always the lowest one. For instance, in public and voucher private schools, the excluded category is low SES, and in private schools, it is middle-high SES.
14. We also observe in Column (3) of Table VIII that students with the highest SES attending private schools obtained a 0.4 standard deviation higher score in the PSU test than those belonging to the middle-high SES (the excluded category), an effect similar to the one observed for the total sample.
15. Those earnings, then, were received five to eight years after taking the PSU test.
16. Those earnings were received 10 to 13 years after taking the SIMCE test.
17. Family income is an individual level variable, and each group is represented by at least one student in the three school types. Because of this, unlike what can be observed in Tables VIII, X and XII, there are no empty cells in Table XIV. See Section 3 for a further description of this variable.

References

Becker, G. and Tomes, N. (1979), “An equilibrium theory of the distribution of income and intergenerational mobility”, *Journal of Political Economy*, Vol. 87 No. 6, pp. 1153-1189.

Becker, G. and Tomes, N. (1986), “Human capital and the rise and fall of families”, *Journal of Labor Economics*, Vol. 4 Nos. 3/Part 2, pp. S1-S39.

Behrman, J.R. and Taubman, P. (1990), “The intergenerational correlation between children’s adult earnings and their parents’ income: results from the Michigan panel survey of income dynamics”, *Review of Income and Wealth*, Vol. 36 No. 2, pp. 115-127.

Benabou, R. (1996), “Equity and effectiveness in human capital investment: the local connection”, *Review of Economic Studies*, Vol. 63 No. 2, pp. 37-64.

Björklund, A. and Jäntti, M. (2011), “Intergenerational income mobility and the role of family background”, in Nolan, B., Salverda, W. and Smeeding, T. M. (Eds), *Oxford Handbook of Economic Inequality*, Elsevier, Amsterdam.

Björklund, A. and Solon, G. (2005), “Influences of nature and nurture on earnings variation: a report on a study of sibling types in Sweden”, in: Bowles, S., Gintis, H. and Osborne, M. (Eds), *Unequal Chances: Family Background and Economic Success*, Russell Sage Foundation, New York, NY.

Björklund, A., Lindahl, M. and Plug, E. (2006), “The origins of intergenerational associations: lessons from Swedish adoption data”, *Quarterly Journal of Economics*, Vol. 121 No. 3, pp. 999-1028.

Blanden, J. and Machin, S. (2004), “Educational inequality and the expansion of UK higher education”, *Scottish Journal of Political Economy*, Vol. 51 No. 2, pp. 230-249.

Blau, D.M. (1999), “The effect of income on child development”, *Review of Economics and Statistics*, Vol. 81 No. 2, pp. 261-276.

Blau, P. and Duncan, O.D. (1967), *The American Occupational Structure*, Wiley, New York, NY.

Bowles, S. and Gintis, H. (2002), “The inheritance of inequality”, *Journal of Economic Perspectives*, Vol. 16 No. 3, pp. 3-30.

Bowles, S., Fong, C. and Gintis, H. (2002), “Reciprocity and the welfare state”, in Kolm, S. and Mercier-Ythier, J. (Eds), *Handbook on the Economics of Giving, Reciprocity and Altruism*, Elsevier, Amsterdam.

Bratberg, E.Ø., Nilsen, A. and Vaage, K. (2008), “Job losses and child outcomes”, *Labour Economics*, Vol. 15 No. 4, pp. 591-603.

Bratsberg, E.Ø., Nielsen, A. and Vaage, K. (2005), "Intergenerational mobility in Norway: levels and trends", IZA Discussion Paper 1517.

Braun, M., Parro, F. and Valenzuela, P. (2018), "Does finance alter the relation between inequality and growth", *Economic Inquiry*, Vol. 57 No. 1, pp. 410-428.

Cameron, S.V. and Heckman, J.J. (1998), "Life cycle schooling and dynamic selection bias: models and evidence for five cohorts of American males", *Journal of Political Economy*, Vol. 106 No. 2, pp. 262-333.

Corak, M. and Heisz, A. (1999), "The intergenerational earnings and income mobility of Canadian men: evidence from longitudinal income tax data", *The Journal of Human Resources*, Vol. 34 No. 3, pp. 504-533.

Corak, M. and Piraino, P. (2011), "The intergenerational transmission of employers", *Journal of Labor Economics*, Vol. 29 No. 1, pp. 37-68.

Cunha, F. and Heckman, J. (2007), "The technology of skill formation", *American Economic Review*, Vol. 97 No. 2, pp. 31-47.

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2006), "The intergenerational transmission of risk and trust attitudes", IZA Discussion Paper 2380.

Eide, E.R. and Showalter, M.H. (1999), "Factors affecting the transmission of earnings across generations: a quantile regression approach", *The Journal of Human Resources*, Vol. 34 No. 2, pp. 253-267.

Eriksson, T., Bratsberg, B. and Raaum, O. (2005), "Earnings persistence across generations: Transmission through health?", *Mimeo*, University of Oslo.

Fong, C. (2001), "Social preferences, self-interest, and the demand for redistribution", *Journal of Public Economics*, Vol. 82 No. 2, pp. 225-246.

Granovetter, M. (1995), *Getting a Job: A Study of Contacts and Careers*, Chicago University Press, Chicago.

Grawe, N.D. (2004), "Reconsidering the use of nonlinearities in intergenerational earnings mobility as a test for credit constraints", *The Journal of Human Resources*, Vol. 39 No. 3, pp. 813-827.

Gregg, P., Macmillan, L. and Vittori, C. (2015), "Nonlinear estimation of lifetime intergenerational economic mobility and the role of education", Working Paper 15-03, Institute of Education, UCL.

Guell, M., Rodriguez-Mora, J. and Telmer, C. (2007), "Intergenerational mobility and the informative content of surnames", CEPR Discussion Papers.

Heckman, J.J. (1979), "Sample selection bias as specification error", *Econometrica*, Vol. 47 No. 1, pp. 153-161.

Jerrim, J. (2016), "The link between family background and later lifetime income: how does the UK compare to other countries?", *Fiscal Studies*, Vol. 38 No. 1, pp. 49-79.

Mulligan, C.B. (1997), *Parental Priorities and Economic Inequality*, University of Chicago Press, Chicago.

Mulligan, C.B. (1999), "Galton vs. the human capital approach to inheritance", *Journal of Political Economy*, Vol. 107 No. S6, pp. S184-S224.

Oreopoulos, P., Page, M. and Stevens, H.A. (2008), "The intergenerational effects of worker displacement", *Journal of Labor Economics*, Vol. 26 No. 3, pp. 455-483.

Page, M. and Solon, G. (2003a), "Correlations between brothers and neighboring boys in their adult earnings: the importance of being urban", *Journal of Labor Economics*, Vol. 21 No. 4, pp. 831-855.

Page, M. and Solon, G. (2003b), "Correlations between sisters and neighboring girls in their subsequent income as adults", *Journal of Applied Econometrics*, Vol. 18 No. 5, pp. 545-562.

Pellizzari, M. (2010), "Do friends and relatives really help in getting a good gob?", *Industrial and Labor Relations Review*, Vol. 63 No. 3, pp. 494-510.

Pellizzari, M., Basso, G., Catania, A., Labartino, G., Malacrino, D. and Monti, P. (2011), “Family ties in licensed professions in Italy”, A report for the Fondazione Rodolfo Debenedetti, Milan.

Piketty, T. (2014), *Capital in the Twenty-First Century*, Harvard University Press.

Raaum, O., Salvane, K. and Sørensen, E. (2006), “The neighborhood is not what it used to be”, *The Economic Journal*, Vol. 116 No. 508, pp. 200-222.

Raitano, M. and Vona, F. (2015), “Measuring the link between intergenerational occupational mobility and earnings: evidence from eight European countries”, *The Journal of Economic Inequality*, Vol. 13 No. 1, pp. 83-102.

Ramey, G. and Ramey, V.A. (2010), “The rug rat race”, Brookings Papers on Economic Activity.

Sacerdote, B. (2007), “How large are the effects from changes in family environment? A study of Korean American adoptees”, *The Quarterly Journal of Economics*, Vol. 122 No. 1, pp. 119-157.

Schnitzlein, D. (2014), “A new look at intergenerational mobility in Germany compared to the USA”, *SOEP Papers on Multidisciplinary Panel Data Research 689*, The German Socio-Economic Panel (SOEP), DIW Berlin.

Schuetz, G., Ursprung, H. and Woßmann, L. (2008), “Education policy and equality of opportunity”, *Kyklos*, Vol. 61 No. 2, pp. 279-308.

Shea, J. (2000), “Does parents’ money matter?”, *Journal of Public Economics*, Vol. 77 No. 2, pp. 155-184.

Solon, G. (1992), “Intergenerational income mobility in the United States”, *American Economic Review*, Vol. 82 No. 3, pp. 393-408.

Solon, G. (1999), “Intergenerational mobility in the labor market”, in Ashenfelter, O. and Card, D. (Eds), *Handbook of Labor Economics*, North-Holland, Amsterdam.

Solon, G. (2002), “Cross-country differences in intergenerational earnings mobility”, *Journal of Economic Perspectives*, Vol. 16 No. 3, pp. 59-66.

Zimmerman, D. (1992), “Regression toward mediocrity in economic stature”, *American Economic Review*, Vol. 82 No. 3, pp. 409-429.

Appendix A. Additional results

Table AI. Heckman second step regression of family income on academic outcomes in upper secondary education, by gender and SIMCE quintiles

	(1) Total sample	(2) Female	(3) Quintile 1	(4) Quintile 2	(5) Quintile 3	(6) Quintile 4	(7) Quintile 5
SIMCE (student-level)	0.960*** (0.00964)	0.960*** (0.0128)	0.306*** (0.0643)	0.552*** (0.0971)	0.622*** (0.0472)	0.952*** (0.0534)	0.847*** (0.0203)
Middle-low income	0.265*** (0.0101)	0.266*** (0.0132)	0.0702 (0.0618)	0.158** (0.0636)	0.0785** (0.0327)	0.150*** (0.0327)	0.0912*** (0.0192)
Middle income	0.612*** (0.0162)	0.615*** (0.0213)	0.188 (0.128)	0.369*** (0.141)	0.210*** (0.0666)	0.380*** (0.0658)	0.217*** (0.0335)
Middle-high income	0.955*** (0.0207)	0.930*** (0.0265)	0.393** (0.176)	0.681*** (0.209)	0.390*** (0.0992)	0.654*** (0.0910)	0.405*** (0.0417)
High income	1.249*** (0.0216)	1.202*** (0.0276)	0.661*** (0.175)	0.958*** (0.203)	0.746*** (0.111)	0.939*** (0.0991)	0.659*** (0.0445)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	183,462	92,733	34,770	36,157	36,843	37,508	38,184
Adjusted R^2	0.635	0.641	0.095	0.134	0.161	0.205	0.434

Notes: The excluded category for the family income dummy is the lowest one; the variables SIMCE Test Score and PSU Test Score are in standard deviations; control variables are gender, school's region, and grade retention; gender dummy is excluded as control variable when performing the female regression; we use preschool education attendance as exclusion restriction. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table AII. Heckman second step regression of family income on academic outcomes in upper secondary education, by school type

	(1) Public school	(2) Voucher Private	(3) Private school
SIMCE test score (student-level)	1.127*** (0.0280)	0.908*** (0.0156)	1.011*** (0.0663)
Middle-low income	0.286*** (0.0196)	0.266*** (0.0177)	0.122 (0.253)
Middle income	0.684*** (0.0360)	0.561*** (0.0268)	0.321 (0.229)
Middle-high income	1.059*** (0.0496)	0.812*** (0.0334)	0.500** (0.226)
High income	0.927*** (0.0624)	0.933*** (0.0388)	0.723*** (0.227)
Control variables	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes
Observations	103,090	65,397	14,975
Adjusted R^2	0.585	0.569	0.568

Notes: The excluded category for the family income dummy is the lowest one; the variables SIMCE Test Score and PSU Test Score are in standard deviations; control variables are gender, school's region and grade retention; we use preschool education attendance as exclusion restriction. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table AIII. Heckman second step regression of family income on outcomes in tertiary education and the labor market, by gender and PSU quintiles

	(1) Total sample	(2) Female	(3) Quintile 1	(4) Quintile 2	(5) Quintile 3	(6) Quintile 4	(7) Quintile 5
PSU (student-level)	0.337*** (0.00532)	0.341*** (0.00673)	0.0870*** (0.0163)	0.139*** (0.0460)	0.148** (0.0576)	0.429*** (0.0607)	0.782*** (0.0351)
Middle-low income	0.0313** (0.0137)	0.0468*** (0.0168)	0.0690*** (0.0165)	0.0350 (0.0215)	0.0240 (0.0271)	0.0928** (0.0387)	0.0920 (0.0755)
Middle income	0.0806*** (0.0143)	0.127*** (0.0176)	0.114*** (0.0188)	0.0841*** (0.0228)	0.106*** (0.0280)	0.138*** (0.0383)	0.167** (0.0720)
Middle-high income	0.221*** (0.0160)	0.249*** (0.0199)	0.196*** (0.0256)	0.182*** (0.0277)	0.202*** (0.0310)	0.276*** (0.0405)	0.304*** (0.0712)
High income	0.784*** (0.0204)	0.720*** (0.0254)	0.250*** (0.0543)	0.353*** (0.0521)	0.408*** (0.0450)	0.631*** (0.0490)	0.926*** (0.0726)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	78,392	42,443	15,377	15,497	15,852	15,828	15,838
Adjusted R^2	0.265	0.252	0.189	0.148	0.128	0.165	0.284

Notes: The excluded category for the family income dummy is the lowest one; the variables Wage and PSU Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; gender dummy is excluded as control variable when performing the female regression; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table AIV.
Heckman second step regression of family income on outcomes in tertiary education and the labor market, by school type

	(1) Public school	(2) Voucher private	(3) Private school
PSU test score (student-level)	0.252*** (0.00738)	0.298*** (0.00797)	0.615*** (0.0186)
Middle-low income	0.0510*** (0.0158)	0.0222 (0.0216)	0.216 (0.230)
Middle income	0.105*** (0.0178)	0.0755*** (0.0217)	0.264 (0.204)
Middle-high income	0.253*** (0.0233)	0.168*** (0.0240)	0.362* (0.200)
High income	0.330*** (0.0609)	0.408*** (0.0459)	0.768*** (0.200)
Control variables	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes
Observations	31,048	34,503	12,841
Adjusted R^2	0.211	0.190	0.308

Notes: The excluded category for the family income dummy is the lowest one; the variables Wage and PSU Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table AV. Heckman second step regression of family income on outcomes in upper secondary education and the labor market, by gender and SIMCE quintiles

	(1) Total sample	(2) Female	(3) Quintile 1	(4) Quintile 2	(5) Quintile 3	(6) Quintile 4	(7) Quintile 5
SIMCE (student-level)	0.218*** (0.00283)	0.232*** (0.00376)	0.116*** (0.0116)	0.127*** (0.0240)	0.195*** (0.0277)	0.258*** (0.0284)	0.445*** (0.0192)
Middle-low income	0.0771*** (0.00584)	0.0842*** (0.00783)	0.0735*** (0.00846)	0.0835*** (0.00966)	0.0995*** (0.0109)	0.0962*** (0.0141)	0.126*** (0.0249)
Middle income	0.187*** (0.00712)	0.218*** (0.00950)	0.161*** (0.0127)	0.210*** (0.0133)	0.203*** (0.0136)	0.201*** (0.0164)	0.274*** (0.0258)
Middle-high income	0.395*** (0.00928)	0.422*** (0.0122)	0.223*** (0.0208)	0.317*** (0.0206)	0.393*** (0.0198)	0.408*** (0.0207)	0.513*** (0.0278)
High income	1.022*** (0.0130)	0.969*** (0.0169)	0.290*** (0.0353)	0.534*** (0.0381)	0.629*** (0.0338)	0.804*** (0.0305)	1.279*** (0.0316)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	162,195	79,352	28,623	30,790	32,341	34,169	36,272
Adjusted R^2	0.251	0.252	0.189	0.148	0.128	0.162	0.255

Notes: The excluded category for the family income dummy is the lowest one; the variables Wage and SIMCE Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; gender dummy is excluded as control variable when performing the female regression; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table AVI. Heckman second step regression of family income on outcomes in upper secondary education and the labor market, by school type

	(1) Public school	(2) Voucher private	(3) Private school
SIMCE test score (student-level)	0.176*** (0.00349)	0.208*** (0.00482)	0.446*** (0.0177)
Middle-low income	0.0812*** (0.00609)	0.0750*** (0.0110)	0.187 (0.208)
Middle income	0.187*** (0.00835)	0.174*** (0.0121)	0.268 (0.186)
Middle-high income	0.362*** (0.0130)	0.307*** (0.0150)	0.455** (0.182)
High income	0.276*** (0.0309)	0.534*** (0.0329)	0.980*** (0.182)
Control variables	Yes	Yes	Yes
Selection correction	Yes	Yes	Yes
Observations	88,859	59,232	14,104
Adjusted R^2	0.203	0.185	0.262

Notes: The excluded category for the family income dummy is the lowest one; the variables Wage and PSU Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Table AVII. Heckman second step regression of family income on initial wage and wage growth

	Initial wage	Wage growth
PSU test score (student-level)	0.220*** (0.0110)	0.305*** (0.0119)
Middle-low income	0.068*** (0.0179)	0.039** (0.0193)
Middle income	0.143*** (0.0201)	0.137*** (0.0217)
Middle-high income	0.300*** (0.0271)	0.261*** (0.0296)
High income	0.984*** (0.0400)	0.492*** (0.0436)
Control variables	Yes	Yes
Selection correction	Yes	Yes
Observations	78,392	78,392
Adjusted R^2	0.143	0.126

Notes: The excluded category for the family income dummy is the lowest one; the variables Initial Wage, Wage Growth and PSU Test Score are in standard deviations; control variables are gender, nationality, firm's region and firm's industrial classification economic activity; we use civil status and parental labor activity as exclusion restrictions. *, ** and *** indicate significance at the 10, 5 and 1% level, respectively. Standard errors are in parentheses

Corresponding author
Francisco Parro can be contacted at: fjparrog@gmail.com