Intergenerational Educational Mobility within Chile[*]

Ercio Muñoz[†]
The Graduate Center, CUNY

This draft: June 2021

Abstract

I estimate intergenerational mobility (IGM) in education at a disaggregated geographic level for a cohort born within the nineties using full-count census microdata from Chile. I document wide variation across more than three hundred communes in eight measures of IGM. Relative mobility measured as one minus the correlation coefficient between children's years of schooling and parents' years of schooling ranges between 0.50 and 0.96. Relative mobility is positively correlated to the number of doctors, and negatively correlated to the ratio of students per teacher, and to labor earnings inequality, especially in the upper half of the income distribution. Using a LASSO, I find that the share of students enrolled in public schools, the numnber of students per teacher, the number of doctors, and municipal budget, among others, are the strongest predictors of IGM. In addition, I also document within country variability in how parental education affects other child's outcomes such as attending tertiary education and being mother as a teenager in the case of females.

JEL-Codes: D63, I24, J62.

Keywords: *Socioeconomic mobility, Geography, Chile, Education.*

[*]I am grateful to the participants of the micro brown bag and the dissertation seminar at CUNY for useful feedback. Pablo Vargas provided outstanding research assistance. The usual caveats apply. First draft: December 2020.

[†]Email: emunozsaavedra@gc.cuny.edu.

I Introduction

A recent literature estimates intergenerational mobility (IGM) in education within countries (see for example Alesina, Hohmann, Michalopoulos, & Papaioannou, 2020; Asher, Novosad, & Rafkin, 2018; Van der Weide, Ferreira de Souza, & Barbosa, 2020). This is an extension of the literature on intergenerational income mobility within countries initiated by Chetty, Hendren, Kline, and Saez (2014), and the intergenerational mobility in education at the country-level (see Torche, 2019, for a survey focused on developing countries).

In this paper, I contribute to this literature in three ways. First I estimate intergenerational mobility in education in Chile at the country, region, and commune level using census data for a cohort born within the nineties. I focus on one minus the correlation coefficient between years of schooling of children and parents[1] as a measure of relative mobility, but I also compute seven other measures that provide information about different aspects of educational IGM. I provide all estimates in an online data appendix for future research. Second, I show how other child's outcomes such as teenage pregnancy and tertiary education attendance that are related to parental education at the country-level also display wide variation within-country. Finally, I explore how the estimates of educational IGM are correlated with a rich set of variables related to income, geography, education, municipal budget, and other characteristics of the communes. Furthermore, I investigate by means of a lasso (least absolute shrinkage and selection operator) which correlates have the most predictive power over IGM at the level of commune.

IGM literature for Chile. Previous studies have used different household and opinion surveys (see for example, Celhay & Gallegos, 2015; Celhay, Sanhueza, & Zubizarreta, 2010; Hertz et al., 2007; Narayan et al., 2018; Neidhöfer, Serrano, & Gasparini, 2018; Nunez & Miranda, 2010; Sapelli, 2016; Torche, 2005; Van der Weide, Lakner, Gerszon Mahler, Narayan, & Ramasubbaiah, 2021) to document IGM in income, education, and other socioeconomic

[1]Throughout the paper, I will refer to the cohort of interest as "children" and I will refer to their parents or older relatives living in the same household as "parents". I will precisely define who will be considered as parent in Section II.

measures. However, they all have in common that the samples are not representative at the commune level, so they focus on country level estimates. Two exceptions are Celhay and Gallegos (2015) that also explore mobility at the regional level (the coarser administrative unit in which the country is divided), and Gutierrez, Diaz, and Villarroel (2020) that use labor earnings in the formal sector from administrative records to estimate income mobility at regional-level and across communes (the smallest administrative unit) in the Metropolitan Region.

Institutional background. Chile is an interesting case study to analyze IGM at the sub-national level. One the one hand, the country is one of the richest economies in the Latin American region and has shown important progress in poverty reduction and income per capita growth over the last three decades. On the other hand, income inequality is relatively high for OECD standards and previous research has documented high school-level stratification by socioeconomic status, which has fueled some educational reforms in the last decade. Moreover, the country is marked by the free-market reforms inherited from the military dictatorship (1973-1990). This include an universal voucher system and decentralization of the administration of public schools, which are managed by municipalities.[2]

In terms of IGM at the country level, the best evidence available at global scale (Narayan et al., 2018; Van der Weide et al., 2021) shows some interesting findings for Chile. Among the 148 countries for which there are estimates of educational mobility for the cohort born in the 1980s, the country ranks around the middle when a summary statistic of relative IGM is used. For example. when measured with the slope of a regression between years of schooling of children and parents or their correlation coefficient, although more mobile according to the latter. Similarly, absolute mobility measured as the share of children with higher education than their parents ranks the country as highly mobile. However, when a measure that aims to capture directional mobility from the bottom to the top is considered (i.e., "rags to riches"), then the country appears among the least mobiles (see Figure A1 in

[2]A recent reform started a process of centralization in 2018.

the Appendix).[3]

II Data and Methods

Data sources. I use full-count census 2017 microdata obtained from the National Institute of Statistics to compute a set of measures of IGM. In addition, I use the Unemployment Insurance System administrative data set to create income-related correlates, data from the Center for Crime Studies and Analysis (CEAD in Spanish), the Chilean Education Quality Agency, and the National System of Municipal Information to gather a set of budgetary, health, geographic, and education-related correlates. A description of the set of covariates is available in Table A1 of the Appendix.

Geography. Chile is divided into 16 regions, 56 provinces and 346 communes. The data set contains information about where the interview was conducted and the place of birth in terms of these three administrative divisions. I use the latter to assign individuals into places. In particular, my estimates of IGM are done for the entire country, by region of birth, and by commune of birth.

Education. The census data contains a variable reporting years of schooling, regardless of the track or kind of study. When I study how the educational attainment of children relates to the attainment of parents or old relatives living in the same household, I take the highest attainment among the individuals in the older generation.[4] Given the typical educational path in Chile where students start first grade at the age of 6, the average student would be able to attain at most 15 years of schooling by the age of 21. To accommodate for this, the indicators are computed using years of schooling censored at 15 for both children and parents.[5]

Measurement. I consider eight different measures that relate to different aspects of

[3] Ranked 138 among the 148 available estimates.

[4] The results are qualitatively similar if I use the average rounded to the nearest integer instead of the maximum.

[5] Similar censoring of years of schooling is used in Neidhöfer et al. (2018) with survey data to compute IGM at country level for 18 countries in Latin America.

educational IGM. The first two are derived from a simple OLS regression that relates educational attainment of children to attainment of parents. Hence, these measures come from the following specification by commune c:

$$y^y_{ic} = \alpha_c + \beta_c y^o_{ic} + \epsilon_{ic} \tag{1}$$

where y^y_{ic} is educational attainment of individual i (using a sample of individuals with ages between 21-25), y^o_{ic} is the attainment of his/her parents or older relatives cohabiting in the same household, and the parameters of interest α_c and β_c are respectively used to measure absolute and relative mobility $(1 - \beta_c)$ for commune of birth c (see Narayan et al., 2018; Torche, 2019, for a discussion about the concepts of absolute and relative mobility in education). Given that the expected years of schooling of an individual according to equation 1 depends on the average years of schooling of parents in his/her commune (in addition to the parameters α_c and β_c), I also compute average years of schooling of parents by commune as the third indicator. The fourth measure relates to the concept of absolute mobility measured as the share of children attaining more years of schooling than their parents (including ties at 15). The fifth measure corresponds to the Pearson's correlation coefficient between years of schooling of children and parents, which in contrast to the regression coefficient is not affected by the marginal distributions of educational attainment of parents and children. The last three measures address directional mobility. First, upward IGM (or "rags to riches") measured as the probability of children reaching the top quintile in the distribution of educational attainment of children in the country (approximately 15 years of schooling) if their parents were in the bottom quintile of educational attainment (approximately less than 10 years of schooling) of parents in the country.[6] Second, intergenerational low education which is the probability of an attainment in the bottom quintile of the children distribution (approximately less than 12 years of schooling) when their parents attainment is also in the bottom quintile of the parents distribution (approximately less than 10 years of schooling).

[6]The quintiles are defined by sorting individuals by attainment and solving ties randomly.

Finally, intergenerational high education, which is the probability of children attainment in the top quintile (approximately more than 14 years of schooling) when their parents attainment is in the top quintile (approximately more than 13 years of schooling).[7] The indicators are summarized in Table 1.

Table 1: Indicators of Educational Intergenerational Mobility

Indicator		Description
Absolute mobility	α	OLS estimate of intercept in Eq. 1
Relative mobility (regression coefficient)	$1-\beta$	OLS estimate of slope in Eq. 1
Average education	$\bar{Y}$	Average years of schooling of parents
Above parent	$\bar{y}^{\geq}$	Share with higher schooling than parents
Relative mobility (correlation coefficient)	$1-\rho$	Pearson correlation coefficient
Rags to riches	$P_{1,5}$	Conditional probability of top education
Intergenerational low	$P_{1,1}$	Conditional probability of bottom education
Intergenerational high	$P_{5,5}$	Conditional probability of top education

Above parent considers ties at the maximum number of years of schooling in the data as children having higher education than parents. The subscripts in the last 3 rows refer to quintiles.

Correlates of relative mobility. I correlate a measure of relative mobility $(1-\rho)$ with a rich set of local area characteristics with the aim of documenting a set of stylized facts. I do it by running regressions of relative mobility (i.e., $1-\rho$) at commune-level against each covariate:

$$1-\rho_c = \gamma_0 + \gamma_1 Z_c [+\gamma_2 W_c] + \epsilon_c \tag{2}$$

where Z_c is the covarite, W_c is average education of individuals born in the commune that are older than 24 but younger than 66. For each correlate, I estimate γ_1 with and without controlling by W_c to get a sense of how IGM is related to a given covariate above and beyond "initial conditions" of the commune in terms of educational attainment.

Predictors of relative mobility. I estimate a LASSO (least absolute shrinkage and selection operator) over the full set of covariates to select the set with the strongest predictive power on relative IGM (i.e., $1-\rho$) at the level of commune. I compute the "optimal" degree of regularization using 10-fold cross-validation and plot the coefficients path allowing the

[7] I also compute these three indicators using quintiles of the distribution of educational attainment within the region or commune instead of the country. I compare both alternatives in the Appendix.

regularization parameter to range from 0 (OLS) to infinity (where all the coefficients are set to zero).

Sample. The full count sample consists of 17,574,003 individuals. I drop those that are considered domestic service, living in collective housing, persons in transit, and individuals considered homeless, which reduces the sample to 16,673,838. The target sample to estimate mobility uses only individuals born in Chile with ages between 21 and 25, which further reduces the sample to 1,155,207. This target sample is composed by 568,231 men and 586,976 women.

Linking individuals across generations. The data set enumerate individuals into households and contains a variable that describes the relationship of each individual with the head of the household. I use this variable to link individuals with their parents or older relatives according to table 2. In addition, those living only with individuals not identified in the table are matched with other relatives, provided that these relatives are at least 15 years but less than 40 years older than them. In the end, I am able to match approximately 73% of the target sample using specific relationships to the head and an extra 6% using other relatives, reaching a final sample of 833,107 individuals.

The use of co-residents may generate bias in the estimates of intergenerational mobility as individuals who reside with their parents may systematically differ from those not residing with them. However, Narayan et al. (2018) use household survey data that contain retrospective information for a large number of countries to show that the bias is small when using individuals of age 21-25. In addition, I compare my estimates of relative mobility (based on the regression coefficient as well as the one based on the correlation coefficient) at the country level with those obtained from recent literature that use survey data with retrospective information and I get very close figures, which suggests that this bias is minor.

Table 2: Relationship to household head and identification of different generations

Relationship to the head	Generation	Relationship to the head	Generation
Grandparent	-2	Sibling	0
Parent	-1	Sibling-in-law	0
Parent-in-law	-1	Child	1
Head	0	Child-in-law	1
Spouse	0	Spouse/partner of child	1
Legal live-in partner	0	Grandchild	2
Partner	0	Others	Missing

Categories left missing are: Other relative, non-relative, domestic employee, person in collective housing, visitor, and homeless person.

III Estimates of intergenerational mobility

In this section I document the level of IGM in Chile derived from my estimates. First, I go over country-level estimates for the eight indicators of IGM described in the previous section, I explore whether there is some evidence of heterogeneity by gender, urban status and indigenous population in absolute and relative mobility, and then I go over the estimates of mobility using other outcomes. Second, I document within country mobility at the region-level using the same eight indicators, describe and map the estimates at commune-level, analyze the correlation patterns between these indicators, and finally explore within country variation in the effect of parental education on alternatives outcomes.

III.1 Country-level estimates

I first estimate intergenerational mobility in education at the country level[8] and then I explore some potential heterogeneity across sub-populations such as male versus female, urban versus rural, and indigenous versus non-indigenous people in some of the indicators. Then I estimate the relationship between parental educational attainment and other child's outcomes.

IGM in education. Table 3 summarizes the level of educational IGM using the previously described indicators estimated at the country-level with a sample that includes only

[8] Figure A2 in the Appendix displays the distribution of years of schooling for children and parents.

children with ages between 21 and 25. The most recent estimates of IGM (at least for few of these indicators) at the country-level available in the literature for Chile are for the cohort born in the 1980s and 1992-1995. Compared to the latter, I find a slightly lower relative mobility as measured by $1-\beta$[9] but practically the same level when measured with the $1-\rho$ (0.64 vs. 0.62).[10] In addition, although not constructed in the same way, the indicators of directional mobility ($P_{1,5}$, $P_{1,1}$, and $P_{5,5}$) show a consistent picture with respect to Narayan et al. (2018) results in terms of high-persistence at the top of the educational distribution and relatively low chances of reaching the top conditional on having parents at the bottom.

Table 3: IGM at country-level

Absolute mobility	α	9.576	Relative mobility	$1-\beta$	0.714
Average education	$\bar{Y}$	11.125	Above parents	$\bar{y}^{\geq}$	0.666
Relative mobility	$1-\rho$	0.642	Rags to riches	$P_{1,5}$	0.088
Intergenerational low	$P_{1,1}$	0.366	Intergenerational high	$P_{5,5}$	0.354

The table reports estimates of IGM (as described in Table 1) using a sample of individuals with age between 21 and 25 linked to parents or older relatives as explained in section II.

Figure 1a displays the average attainment conditional on parental education attainment, the relationship appears linear with a deviation only in the lowest level of parental education.[11] When this regression is estimated using sub-populations, I find higher absolute and relative mobility for women compared to men (see Figure 1b). In contrast, I do not find significant differences between rural and urban population (see Figure 1c), and between indigenous versus non-indigenous populations (see Figure 1d). Nonetheless, this does not imply that the expected educational attainment between individuals in urban/rural or indigenous/non-indigenous is the same, as can be inferred by the differences in the marginal distributions of parental educational attainment. For example, the number of parents with at least 12 years of education is greater for urban (as well as for non-indigenous) than rural

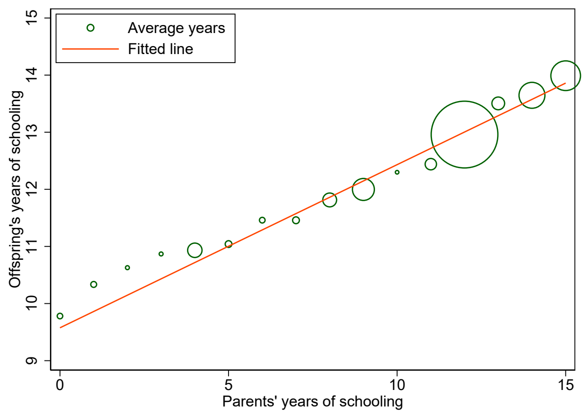
[9]For several cohorts, this difference is smaller than the discrepancy between mobility estimated using Latinobarometro and CASEN survey, as reported in Neidhöfer et al. (2018).

[10]Figure A3 in the Appendix shows the evolution over time of these indicators in the literature versus my estimates.

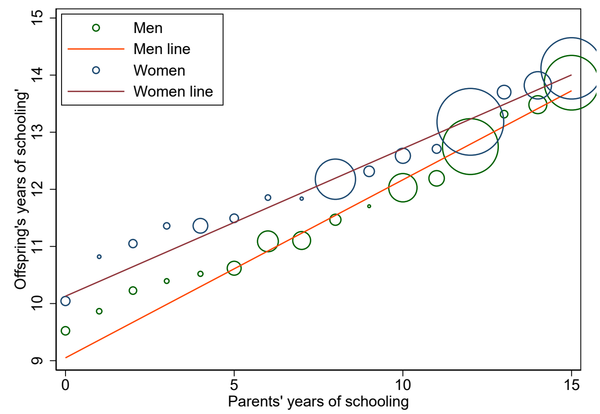
[11]Figure A4 in the Appendix displays the transition matrix between children and parental years of schooling, each of them divided into quintiles according to their respective distribution of years of schooling.

(and respectively indigenous population) population (i.e., the size of the bubbles in Figure 1 is bigger). Table A2 in the Appendix reports the eight indicators computed by subgroup, confirming these findings and highlighting some other differences between groups in other indicators.

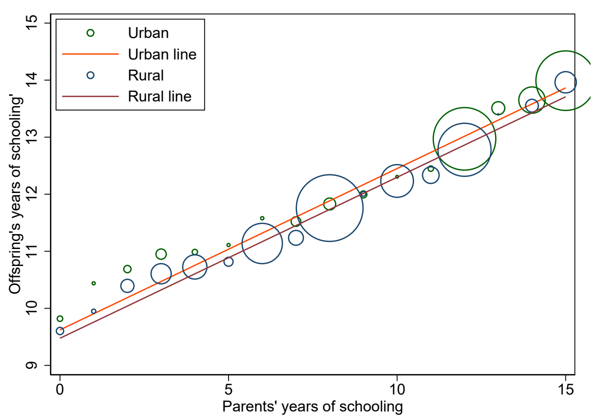
Figure 1: Country-level educational IGM



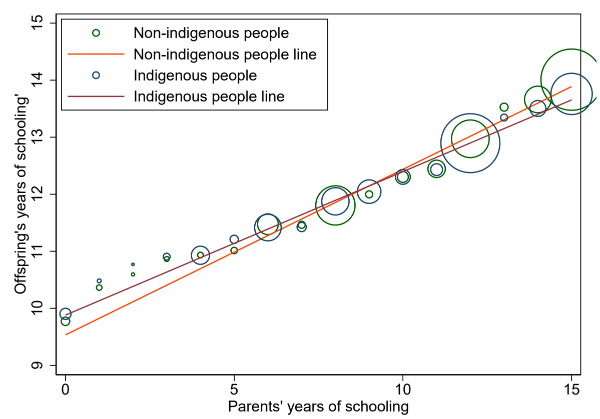
(a) All



(b) Gender



(c) Urban vs. rural



(d) Indigenous vs. non-indigenous population

The graphs display average years of schooling of children for each level of schooling of the generation above (highest years of schooling among parents and older relatives living in the same household). The sample includes only individuals with age between 21 and 25. The size of the bubble varies according to the number of individuals.

Other child's outcomes. I estimate the relationship between parental education and two additional child's outcomes: the likelihood of attending tertiary education and the likelihood of having a child while teenager in the case of women.[12] These outcomes can be measured at earlier ages than education reducing the magnitude of any potencial coresidence bias.

First, I estimate the probability of attending at least one year of tertiary education using a sample of individuals with ages between 19 and 21. Figure A3c shows this likelihood for each parental educational attainment, finding a positive slope approximately equal to 0.046 with a somewhat prominent discontinuity at 12 years of schooling and a somewhat nonlinear relationship for low values of parents' years of schooling. This contrasts with the virtually linear relationship between parental income rank and college attendance documented for the US in Chetty et al. (2014). Despite these differences and other differences in terms of measurement and concepts I find similar gaps. The gap in the likelihood of attending tertiary education for individuals with low-educated vs. highly-educated parents is around 60 percentage points while Chetty et al. (2014) documented a gap of 67.5 percentage points in the US for individuals with lowest-income vs. highest-income parents.

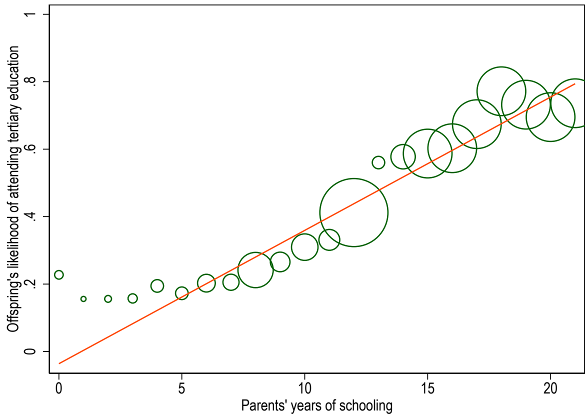
Second, I estimate the probability of becoming mother as teenager defined as having a child for females with ages between 15 and 19. Figure 2b shows this likelihood for each parental educational attainment, finding a negative relationship close to linear with a slope of -0.017. The gap between highly-educated and low-educated parents is around 20-25 percentage points (Chetty et al., 2014, documents a gap of 29.8 percentage points for highest-lowest parents' incomes).

III.2 Intergenerational mobility within Chile

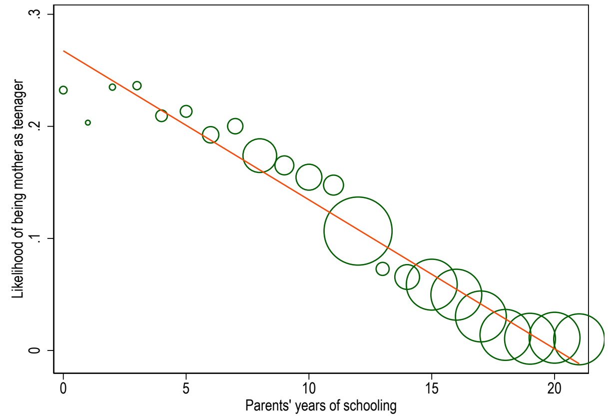
Region-level estimates. Before presenting the most disaggregated estimates, Table 4 summarizes the eight measures of interest estimated for the 16 regions of Chile. Non-negligible differences can be found across regions in most of these dimensions. For example, the chances of reaching the top quintile of the educational distribution for children with parents at the

[12] I use the same econometric specification as in Equation 1 with a different dependent variable.

Figure 2: Other child's outcomes



(a) Tertiary education attendance



(b) Teenage birth (females only)

The first plot displays the likelihood of completing at least one year of tertiary education for each level of education of the generation above (highest years of schooling among parents and older relatives living in the same household). The second plot displays the likelihood of having a child as teenager for each level of education of the generation above. The samples include individuals with age between 19 and 21 (left) and 15 and 19 (right). The size of the bubble varies according to the number of individuals.

bottom quintile (i.e., $P_{1,5}$) is more than 200% higher in the northern Arica y Parinacota region relative to Aysén region. Similarly, in terms of absolute mobility (i.e., α) there are regions with more than one year of difference, and relative mobility (i.e., $1-\beta$) is 17% higher in Arica y Parinacota than in Metropolitana de Santiago or Los Rios. When we consider relative mobility measured with the correlation coefficient $(1-\rho)$, the level in Arica y Parinacota is approximately 30% higher than in the region with the lowest value (Araucanía).[13]

Commune-level estimates. I document wide variation within country at the level of commune. Relative mobility measured as $1-\rho$, excluding places with less than 50 individuals[14], ranges between 0.50 in Cabo de Hornos, a commune located in the extreme south of the country, and 0.96 in San Pedro de Atacama, a commune located in the north. Non-negligible variation is found in all the indicators studied. Figure A6 in the Appendix

[13] Table A4 in the Appendix compares the last three measures of IGM using the distribution of educational attainment at the country level versus at the region level.

[14] Figure A5 in the Appendix displays the CDF of the sample size by commune.

Table 4: Region-level estimates of IGM Statistics

Region	α	$1-\beta$	$\bar{Y}$	$\bar{y}^{\geq}$	$1-\rho$	$P_{1,5}$	$P_{1,1}$	$P_{5,5}$
Tarapacá	9.66	0.74	11.53	0.60	0.70	0.10	0.37	0.32
Antofagasta	9.25	0.71	11.61	0.57	0.68	0.08	0.40	0.31
Atacama	9.54	0.73	10.99	0.62	0.68	0.07	0.40	0.29
Coquimbo	9.44	0.72	10.53	0.65	0.66	0.07	0.38	0.32
Valparaíso	9.61	0.72	11.23	0.65	0.68	0.09	0.35	0.34
Libertador General Bernardo O'Higgins	10.06	0.76	9.95	0.71	0.71	0.10	0.34	0.31
Maule	9.75	0.73	9.84	0.73	0.67	0.09	0.36	0.32
Biobío	10.12	0.74	10.65	0.74	0.66	0.11	0.33	0.37
Araucanía	9.58	0.71	9.88	0.75	0.61	0.07	0.38	0.37
Los Lagos	9.35	0.71	9.77	0.71	0.65	0.07	0.41	0.31
Aysén del General Carlos Ibáñez del Campo	9.38	0.76	9.59	0.65	0.73	0.05	0.44	0.23
Magallanes y de la Antártica Chilena	10.36	0.77	11.33	0.66	0.72	0.10	0.30	0.30
Metropolitana de Santiago	9.38	0.70	11.33	0.64	0.65	0.09	0.37	0.36
Los Ríos	9.46	0.70	10.18	0.72	0.62	0.06	0.38	0.34
Arica y Parinacota	10.76	0.82	11.49	0.61	0.79	0.14	0.28	0.31
Ñuble	10.02	0.74	9.86	0.76	0.68	0.11	0.33	0.36

The table reports region-level estimates of absolute mobility, relative mobility $(1-\beta)$, average parents' education, share of children with higher education than parents, relative mobility $(1-\rho)$, rags to riches, intergenerational low, and intergenerational high, respectively. A description of the measures can be found in Table 1.

shows the distributions of the commune-level estimates for the eight measures and Table A3 of the Appendix similarly reports some descriptive statistics of these estimates. For all the indicators I can find communes with levels at least 100% greater than others, in some cases several times greater.

The measures of mobility based on conditional probabilities derived from quintiles of educational attainment are constructed using the distribution of attainment at the country level for children and similarly for parents. Similar measures could be constructed using the distribution of attainment by commune. In this case, moving from the bottom to the top may require a higher number of years of schooling in some places compared to others and capture a different aspect of mobility. As an additional exercise, I compute those measures and find that $P_{1,5}$ measures constructed in both ways are highly correlated while $P_{1,1}$ is to a lesser degree but in contrast, $P_{5,5}$ is not correlated (see Figure A7 in the Appendix.).

Figure 3 maps relative mobility $(1-\rho)$ across the country. There are some regions with clusters of communes showing relatively similar levels of IGM, such as the northern regions and more heterogeneity in the center of the country. Figure A8 in the Appendix plots relative mobility dividing the map of the country into three parts, a northern region less the metropolitan region, the metropolitan region, and a southern region. These three regions have communes with relatively low and high levels of intergenerational educational mobility. However, in this map the variety IGM levels in the metropolitan region (where the highest share of the population lives) can be appreciated with more detail.

Correlations among different measures of IGM. Table 5 presents the Pearson correlation coefficients between the eight mobility statistics computed at the level of commune. I find the strongest positive correlation to be between absolute and relative mobility, both measured with $1-\beta$ and $1-\rho$. These three measures are at the same time positively correlated to above parents and rags to riches, especially absolute mobility. Intergenerational low is negatively correlated to all the other six indicators.

Table 5: Correlation among IGM statistics

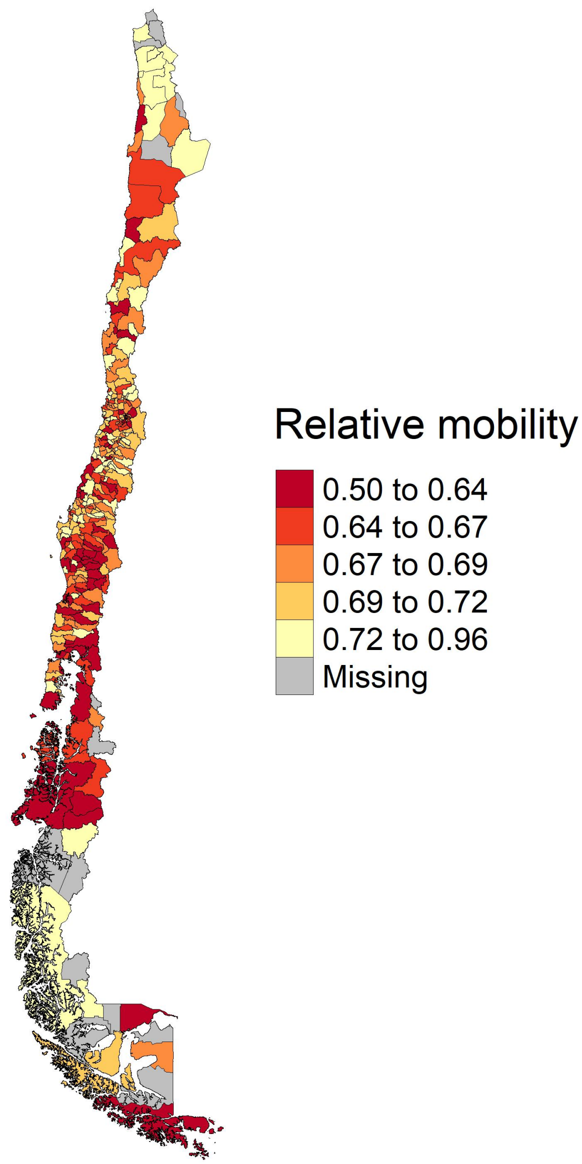
	α	$1-\beta$	$\bar{Y}$	$\bar{y}^{\geq}$	$1-\rho$	$P_{1,5}$	$P_{1,1}$	$P_{5,5}$
Absolute mobility (α)	1							
Relative mobility ($1-\beta$)	0.912***	1						
Average education ($\bar{Y}$)	-0.0175	-0.146**	1					
Above parents ($\bar{y}^{\geq}$)	0.268***	0.139*	-0.716***	1				
Relative mobility ($1-\rho$)	0.713***	0.874***	-0.0917	-0.0128	1			
Rags to riches ($P_{1,5}$)	0.478***	0.259***	0.296***	0.0604	0.228***	1		
Intergenerational low ($P_{1,1}$)	-0.730***	-0.517***	-0.207***	-0.166**	-0.380***	-0.537***	1	
Intergenerational high ($P_{5,5}$)	-0.00472	-0.141*	0.0369	0.233***	-0.140*	0.236***	-0.0715	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Other outcomes within Chile. I estimate the relationship between parental education and the two alternative outcomes described in the previous section: the likelihood of attending at least one year of tertiary education, and the likelihood of being mother as a teenager for females.

Table 6 reports these estimates at the regional-level. There is significant variation across regions in the effect of one extra year of parental schooling on the chances of attending tertiary education. Araucania shows the strongest effect (0.044) which suggest that moving

Figure 3: Intergenerational educational $(1-\rho)$ mobility within Chile



The map plots relative IGM measured as one minus the Pearson correlation coefficient (by commune) between child's years of schooling (using only age between 21 and 25) against parents' years of schooling. Educational attainment is censored at 15. Communes with less than 50 individuals are left as missing (Figure A5 in the Appendix displays the CDF of the sample size by commune).

from parents with no education to the highest level in our data increases the chances of attending tertiary education by 92 percentage points (21×0.044). A caveat to note is that this calculation may overestimate the effect in light of the non-linearity observed at the national level in Figure A3c for lower levels of parental education. If I assume that the effect is null in the first 5 years of education, then the chances increase by approximately 70 percentage points. On the other extreme, Aysén region shows the smallest average effect (0.019).

Similarly, the effect of an extra year of parents' schooling on teenage birth rates varies significantly across regions. The effect of one year goes from a fall in the likelihood of a teenage birth equal to 0.8 percentage points in Ñuble to 1.6 percentage in Antofagasta or Coquimbo. This last effect implies a gap between uneducated and highly educated parents of approximately 33.6 percentage points, which again is meaningful but may be an overestimation due to non-linearities.

Table 6: Parental education effect on other outcomes

Region	Tertiary education	Teenage birth
Tarapacá	0.038	-0.013
Antofagasta	0.038	-0.016
Atacama	0.042	-0.012
Coquimbo	0.040	-0.016
Valparaíso	0.042	-0.014
Libertador General Bernardo O'Higgins	0.028	-0.010
Maule	0.034	-0.012
Biobío	0.039	-0.013
Araucanía	0.044	-0.013
Los Lagos	0.035	-0.014
Aysén del General Carlos Ibáñez del Campo	0.019	-0.015
Magallanes y de la Antártica Chilena	0.033	-0.009
Metropolitana de Santiago	0.043	-0.015
Los Ríos	0.039	-0.013
Arica y Parinacota	0.026	-0.012
Ñuble	0.037	-0.008

The table reports of effect of an extra year of parents' schooling on the likelihood of completing at least one year of tertiary education and likelihood of having a child as teenager for females (computed using an OLS regression). The samples include individuals with age between 19 and 21 (left) and 15 and 19 (right).

IV Correlates of IGM within Chile

In this section, I study whether the measures of intergenerational mobility in education at the commune level are correlated with a rich set of variables related to income, education, budget, geography, and other characteristics at the commune-level. The definition of the variables and data sources are listed in the Appendix (see Table A1). An important caveat is that this analysis should not be interpreted as causal. The sole purpose is to document stylized facts that can later be used to model theoretically or estimate empirically the mechanisms behind local differences in IGM.

Figure A9a reports the coefficients and their respective 95% confidence intervals with or without conditioning on average education of the old generation. I find that relative mobility $(1-\rho)$ is positively correlated in a statistically significant way at the 5% to the number of doctors.[15] In contrast, relative mobility is negatively correlated in a statistically significant way at the 5% to the Gini index, 90th quantile, 95th quantile, ratio 90-10 quantiles, ratio 90-50 quantiles, and the log of students per teacher.[16] Several of these correlates are also significantly correlated with other measures of IGM. Figure A9 in the Appendix reports the results for the eight indicators of mobility.

These results suggest that in Chile higher intergenerational mobility in education is more strongly associated with lower levels of income inequality in the upper half of the income distribution. This contrasts with the findings of Corak (2019) showing that in Canada mobility (in income) is more associated to inequality in the lowest half of the income distribution. Moreover, these results are in line with the country-level evidence reported in Narayan et al. (2018) showing that income inequality is positively associated to relative mobility in education, and suggest that such a relationship may also holds within countries (see Figure A10 in the Appendix). An important caveat is that the administrative data set used to construct

[15] Budget availability, total expenditure, and number of nurses are also positively correlated but only marginally insignificant at the 5%.

[16] Secondary test scores are also negatively correlated but marginally insignificant at the 5% when conditioning on average education of older cohorts.

measures of inequality only considers the formal sector. If I correlate relative mobility with inequality in education of parents (individuals with age 40-60 at the time of the Census) in what could be considered as a "Great Gatsby curve" in education, I also find a negative relationship, which suggests that this relationship documented across countries (see Narayan et al., 2018) also holds within countries (see Figure A11 in the Appendix).

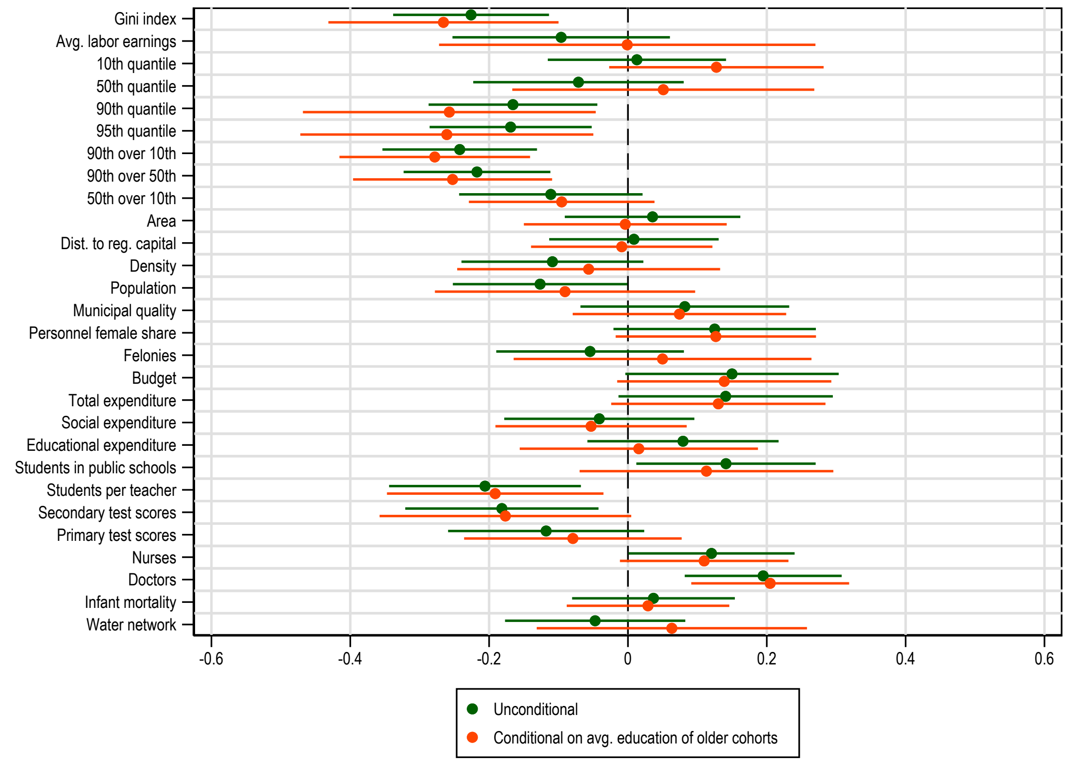
In addition, I estimate a LASSO (least absolute shrinkage and selection operator) to select via regularization the correlates that most strongly predict relative mobility $(1-\rho)$. An important caveat of this exercise is that there may be different levels of measurement error in the variables, so the chosen variables can be not only the result of predictive power but also that they are just better measured.

Figure A9b plots the entire coefficients paths derived from the LASSO allowing the penalization parameter λ to range from 0 (OLS) to infinity (where all the coefficients go to zero), highlighting only those correlates that remain non-zero after the optimal λ is used (vertical red line in the graph). I find that the set of strongest predictors is composed by the ratio 90-10 and 90-50 of labor earnings, area, municipal quality, personnel's female share, budget, social expenditure, students in public schools, students per teacher, primary test scores, nurses, doctors, water network, and educational level of the old individuals. The four strongest predictors are the share of students enrolled in public schools, the number of students per teacher, the number of doctors, and the budget, with the latter being the strongest one.

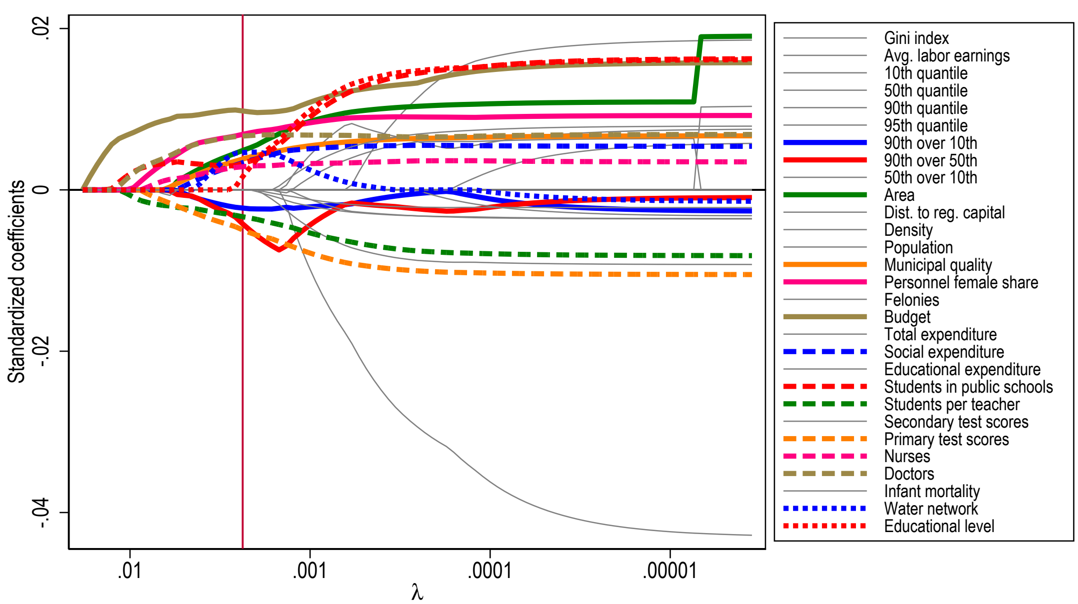
V Final remarks

In this paper, I make three main contributions to the literature on within-country intergenerational socioeconomic mobility. First, I provide estimates of intergenerational mobility in education at the country, regional, and municipal level in Chile. I document wide variation across administrative units in several measures of intergenerational mobility. Second, I doc-

Figure 4: Correlates of relative mobility $(1-\rho)$ at the commune-level



(a) Uni-variate correlations



(b) Coefficient paths from LASSO estimates ument within country variability in how parental education affects other child's outcomes such as attending tertiary education and being mother as a teenager in the case of women. The gaps between children from low- and highly-educated parents that I document are close to those previously documented for highest- lowest-income parents in the US. Finally, I show that IGM in education within Chile is correlated with labor earnings inequality, especially in the upper half of the income distribution, number of doctors in the commune, and students per teacher ratio. Moreover, I also show using LASSO that the four strongest predictors of IGM are the share of students enrolled in public schools, the number of students per teacher, the number of doctors, and the municipal budget.

References

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2020). Religion and Educational Mobility in Africa. *NBER Working Paper*.

Asher, S., Novosad, P., & Rafkin, C. (2018). Intergenerational Mobility in India: Estimates from New Methods and Administrative Data. *mimeo*.

Celhay, P., & Gallegos, S. (2015). Persistence in the Transmission of Education: Evidence across Three Generations for Chile. *Journal of Human Development and Capabilities: A Multi-Disciplinary Journal for People-Centered Development*, *16*(3).

Celhay, P., Sanhueza, C., & Zubizarreta, J. (2010). Intergenerational Mobility of Income and Schooling: Chile 1996-2006. *Revista de Analisis Economico*, *25*(2), 43–63.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553–1623.

Corak, M. (2019). The Canadian Geography of Intergenerational Income Mobility. *The Economic Journal*(uez19), 1–41.

Gutierrez, P., Diaz, J., & Villarroel, G. (2020). *Chapter 1: Intergenerational Mobility in Chile* (Unpublished doctoral dissertation). The University of British Columbia.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., & Verashchagina, A. (2007). The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends. *The B.E. Journal of Economic Analysis Policy*, *7*(2).

Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Gerszon Mahler, D., ... Thewissen, S. (2018). *Fair Progress?: Economic Mobility Across Generations Around the World.* The World Bank.

Neidhöfer, G., Serrano, J., & Gasparini, L. (2018). Educational Inequality and Intergenerational Mobility in Latin America: A New Database. *Journal of Development Economics*, *134*, 329–349.

Nunez, J. I., & Miranda, L. (2010). Intergenerational Income Mobility in a Less-Developed, High-Inequality Context: The Case of Chile. *The B.E. Journal of Economic Analysis Policy*, *10*(1).

Sapelli, C. (2016). Cap 3: La Movilidad Intergeneracional de la Educacion en Chile. In *Chile : ¿más equitativo? una mirada a la dinamica social del chile de ayer, hoy y mañana.*

Torche, F. (2005). Unequal But Fluid: Social Mobility in Chile in Comparative Perspective. *American Sociological Review*, *70*, 422–450.

Torche, F. (2019). Educational Mobility in Developing Countries. *WIDER Working Paper*(88), 1–31.

Van der Weide, R., Ferreira de Souza, P., & Barbosa, R. (2020). Intergenerational Mobility in Education in Brazil. *mimeo*.

Van der Weide, R., Lakner, C., Gerszon Mahler, D., Narayan, A., & Ramasubbaiah, R. (2021). Intergenerational Mobility Around the World. *mimeo*.

Appendices

The appendix provides additional tables and figures, and other relevant information.

Table A1 lists the set of correlates that I use together with a short description and data sources.

Table A2 reports all the indicators computed by sub-populations (male vs. female, indigenous vs. non-indigenous, and urban vs. rural).

Table A3 reports some descriptive statistics of the estimates of IGM at the level of commune.

Figure A1 plots different measures of intergenerational mobility in education at country-level highlighting where Chile falls relative to Latin America and the Caribbean and the world.

Figure A2 displays an histogram with the distributions of educational attainment of parents and children.

Figure A3 displays the evolution of mobility across birth cohorts in recent literature versus my estimate.

Figure A4 displays the transition probabilities between educational attainment of parents and children (classified into three categories).

Figure A5 shows the cumulative distribution of the sample size by commune.

Figure A6 displays the distribution of all the measures at commune-level.

Figure A7 displays scatter plots comparing indicators of mobility (that use quintiles) using country level distribution of educational attainment vs. local distribution.

Figure A8 maps the level of educational intergenerational mobility at the commune level separating the country into north, metropolitan region, and south.

Figure A10 shows a binscatter plot between relative mobility and income inequality measured with the Gini coefficient at the commune level.

Figure A9 reports the results of the correlations with a set of variables using all the measures of IGM.

Table A1: Covariates

Label	Source	Description
Gini Index	UID	Gini Index
Average earnings	UID	Average earnings in the formal sector
10th quantile	UID	10th percentile of earnings in the formal sector
50th quantile	UID	50th percentile of earnings in the formal sector
90th quantile	UID	90th percentile of earnings in the formal sector
95th quantile	UID	95th percentile of earnings in the formal sector
Ratio 90-10	UID	Ratio 90th to 10th percentile of earnings in the formal sector
Ratio 90-50	UID	Ratio 90th to 50th percentile of earnings in the formal sector
Ratio 50-10	UID	Ratio 50th to 10th percentile of earnings in the formal sector
Area	SINIM	Log of the total surface of commune
Distance to regional capital	SINIM	Log of the distance between the commune and the regional capital
Population density per km2	SINIM	Log of population density per km2 by commune
Population	SINIM	Log of commune's estimated population in June 2012
Municipal professionalization	SINIM	Share of college educated workers in the municipality
Female Share in Municipality	SINIM	Share of female workers over the total workers in personnel of the municipality
Crimes	CEAD	Log of the number of crimes with greater social connotation
Budget availability	SINIM	Log of commune's budget availability per capita
Total expenditure	SINIM	Log of commune's total expenditure per capita
Social expenditure	SINIM	Log of the commune's total expenditure in the social programs area per capita
Education expenditure	SINIM	Log of the commune'e total expenditure education programs
Students in public schools	ACE	Number of students enrolled in public schools over total enrollment
Students per teacher	SINIM	Log of students per teacher ratio in the municipal education system
Standarized test - secondary	ACE	Average score between math and language in SIMCE taken in high school
Standarized test - primary	ACE	Average score between math and language in SIMCE taken in 4th grade
Nurses by 100K inhabitants	SINIM	Log of number of nurses by 100.000 inhabitants within the commune
Doctors by 100K inhabitants	SINIM	Log of number of doctors by 100.000 inhabitants within the commune
Infant mortality rate	SINIM	Number of children under 1 year of age who die for every 1.000 live births
Water network	SINIM	Percentage of homes connected to drinking water network in the commune
Parental education	Census	Average education of individual older than 24 but younger than 66

Unemployment insurance database (UID) can be accessed at:
https://www.spensiones.cl/apps/bdp/index.php.
National system of municipal information (SINIM) can be accessed at:
http://datos.sinim.gov.cl/datos_municipales.php.
Center for crime studies and analysis (CEAD) can be accessed at:
http://cead.spd.gov.cl/estadisticas-delictuales/.
Research unit, education quality agency data (ACE) can be accessed at:
https://informacionestadistica.agenciaeducacion.cl/#/bases.
Census 2017 data can be requested from the National Institute of Statistics at:
https://www.ine.cl.

Table A2: IGM at country-level for subgroups

	Male	Female	Non-indigenous	Indigenous	Urban	Rural
α	9.049	10.129	9.535	9.881	9.622	9.476
β	0.688	0.742	0.710	0.748	0.717	0.718
$\bar{Y}$	11.126	11.123	11.260	10.247	11.336	9.203
$\bar{y}^{\geq}$	0.628	0.707	0.663	0.690	0.657	0.754
ρ	0.624	0.658	0.640	0.676	0.652	0.649
P_{15}	0.068	0.108	0.089	0.082	0.092	0.071
P_{11}	0.419	0.310	0.365	0.367	0.359	0.390
P_{55}	0.331	0.378	0.357	0.306	0.353	0.358

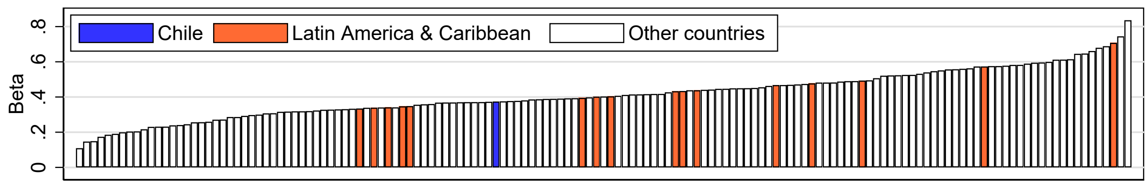
The table reports country-level estimates of absolute mobility, relative mobility $(1-\beta)$, average parents' education, share of children with higher education than parents, relative mobility $(1-\rho)$, rags to riches, intergenerational low, and intergenerational high, respectively, all computed by subgroup. A description of the measures can be found in Table 1.

Table A3: Descriptive statistics of IGM at commune-level

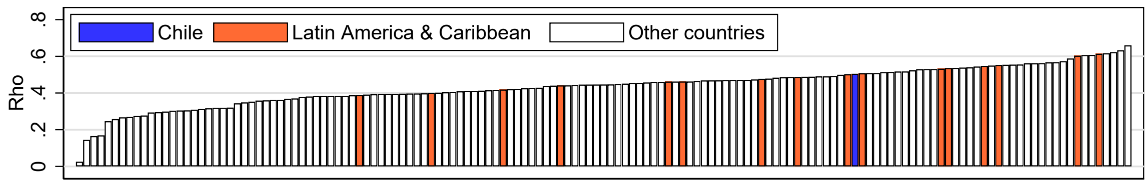
	Mean	SD	Min	Max	N
α	9.79	0.66	7.16	11.73	330
$1-\beta$	0.74	0.05	0.54	0.97	330
$\bar{Y}$	10.00	1.19	6.13	14.50	330
$\bar{y}^{\geq}$	0.71	0.07	0.48	0.90	330
$1-\rho$	0.68	0.06	0.50	0.96	330
P_{15}	0.09	0.03	0.02	0.22	312
P_{11}	0.36	0.06	0.09	0.57	313
P_{55}	0.33	0.05	0.10	0.46	190

The table reports descriptive statistics of estimates of absolute mobility, relative mobility $(1-\beta)$, average parents' education, share of children with higher education than parents, relative mobility $(1-\rho)$, rags to riches, intergenerational low, and intergenerational high, respectively, all of them at the commune-level. I omit estimates with less than 50 observations. A description of the measures can be found in Table 1.

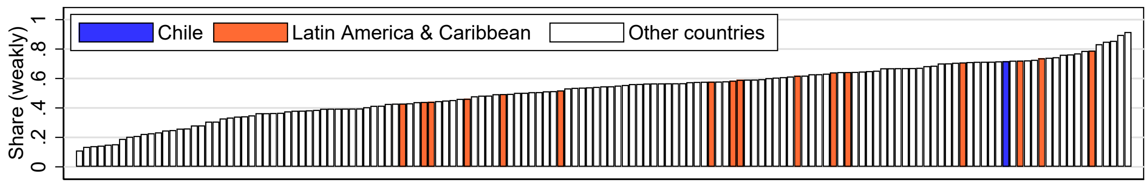
Figure A1: Chile relative to the world in terms of educational IGM



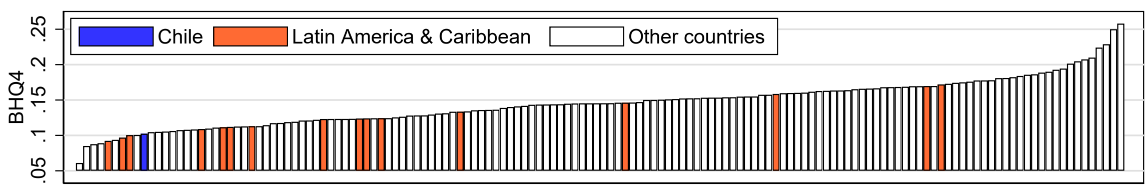
(a) Relative IGM - Intergenerational persistence (beta)



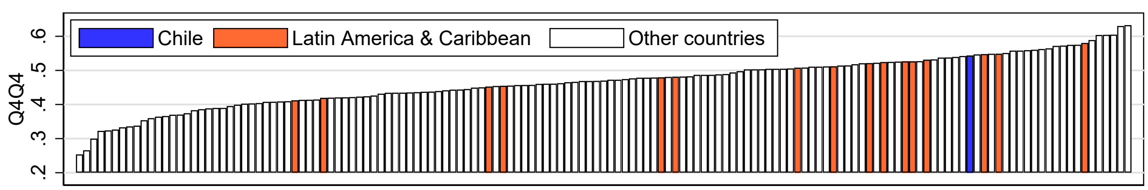
(b) Relative IGM - Correlation coefficient (rho)



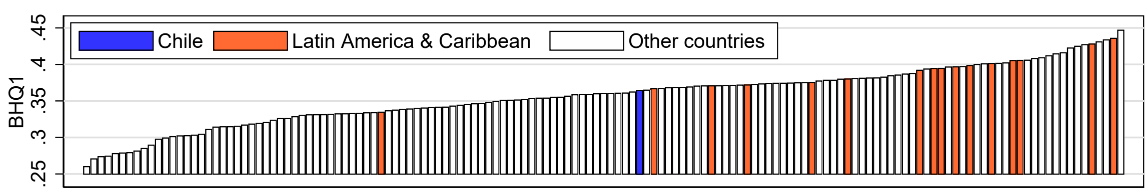
(c) Absolute IGM - Share of students with higher education than parents (share)



(d) Directional IGM - Probability child from bottom half ends up in Q4 (BHQ4)



(e) Directional IGM - Probability child from Q4 ends up in Q4 (Q4Q4)



(f) Directional IGM - Probability child from bottom half ends up in Q1 (BHQ1)

Source: Elaboration by the author with data from Narayan et al. (2018).

Figure A2: Histogram of education

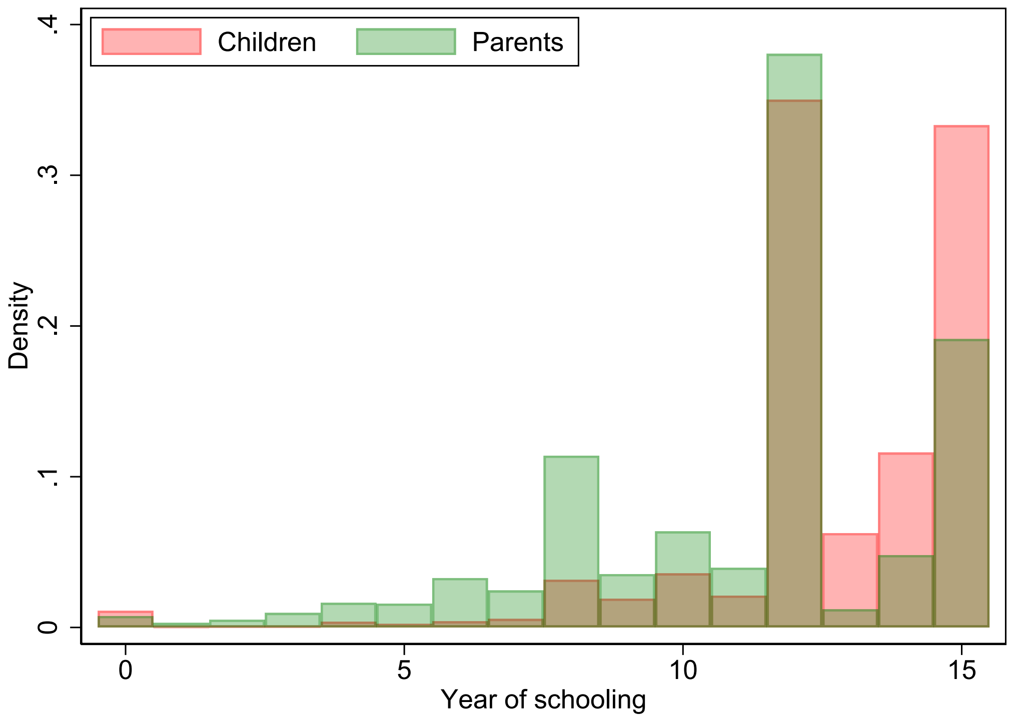
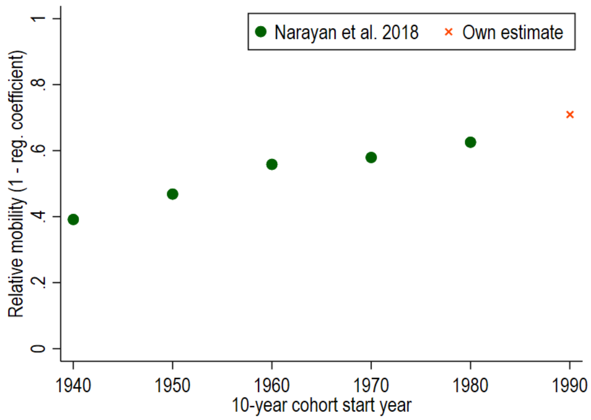
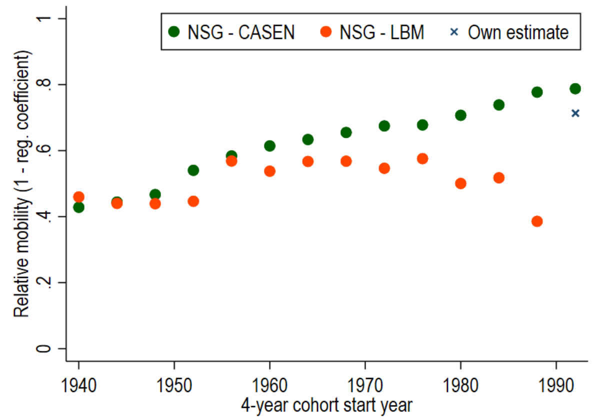


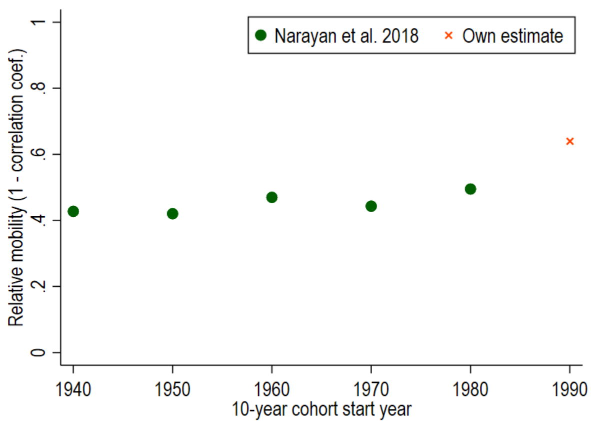
Figure A3: Own estimates versus recent literature at the country level



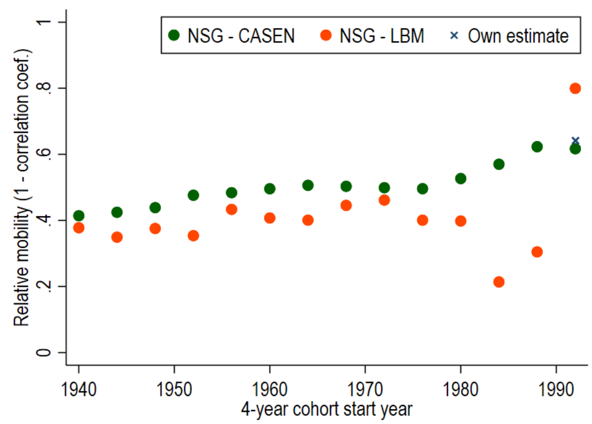
(a) Relative mobility $(1-\beta)$ from Narayan et al. (2018) and own estimate.



(b) Relative mobility $(1-\beta)$ from Neidhöfer et al. (2018) and own estimate.



(c) Relative mobility $(1-\rho)$ from Narayan et al. (2018) and own estimate.



(d) Relative mobility $(1-\rho)$ from Neidhöfer et al. (2018) and own estimate.

The figure shows estimates of intergenerational educational mobility obtained from regressing children years of schooling against parents' years of schooling, and the Pearson correlation coefficient between the same two variables. Narayan et al. (2018) uses CASEN survey while Neidhöfer et al. (2018) also uses Latinobarometro survey (LBM). The former uses 10-year cohorts, the latter uses 4-year cohorts (the most recent one is 1992-1995), and my estimate uses individuals approximately born between years 1991-1995. The last four cohorts using LBM survey contain smaller samples (831, 413, 179, and 24 observations), and hence are somewhat unreliable.

Figure A4: Transition probabilities at the country-level

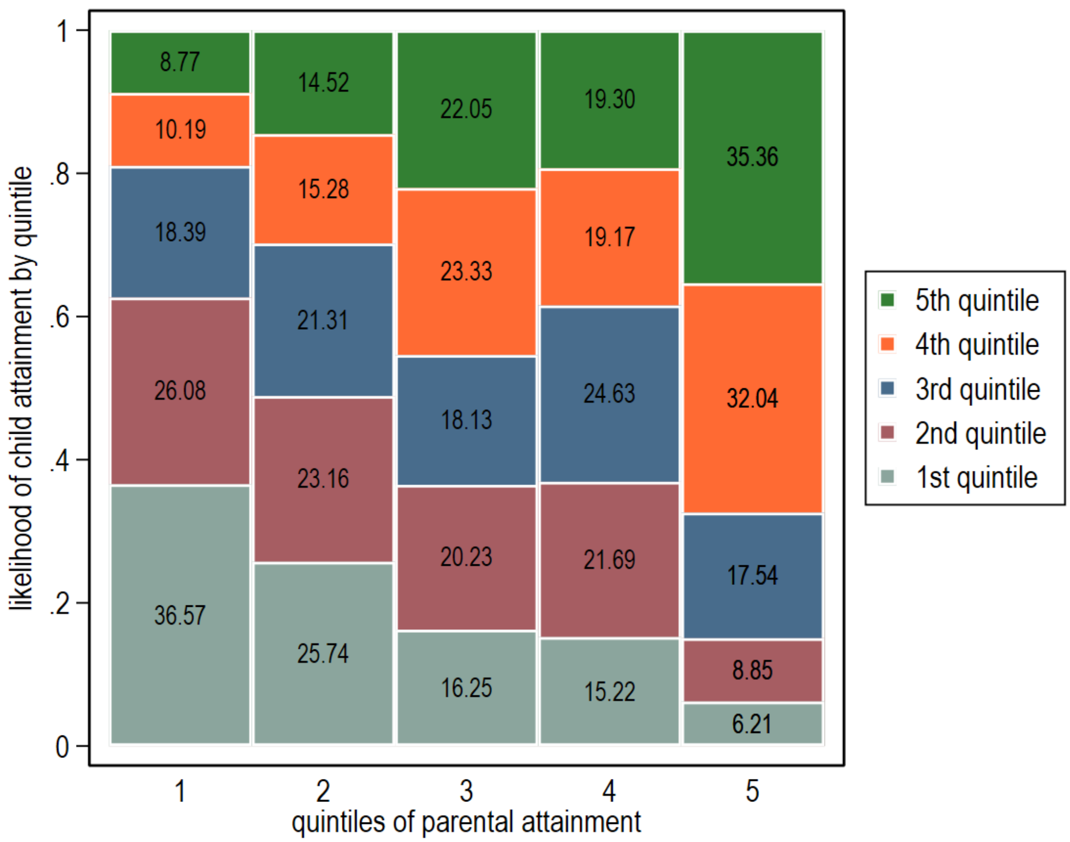


Figure A5: Cumulative distribution of the sample size at the commune level

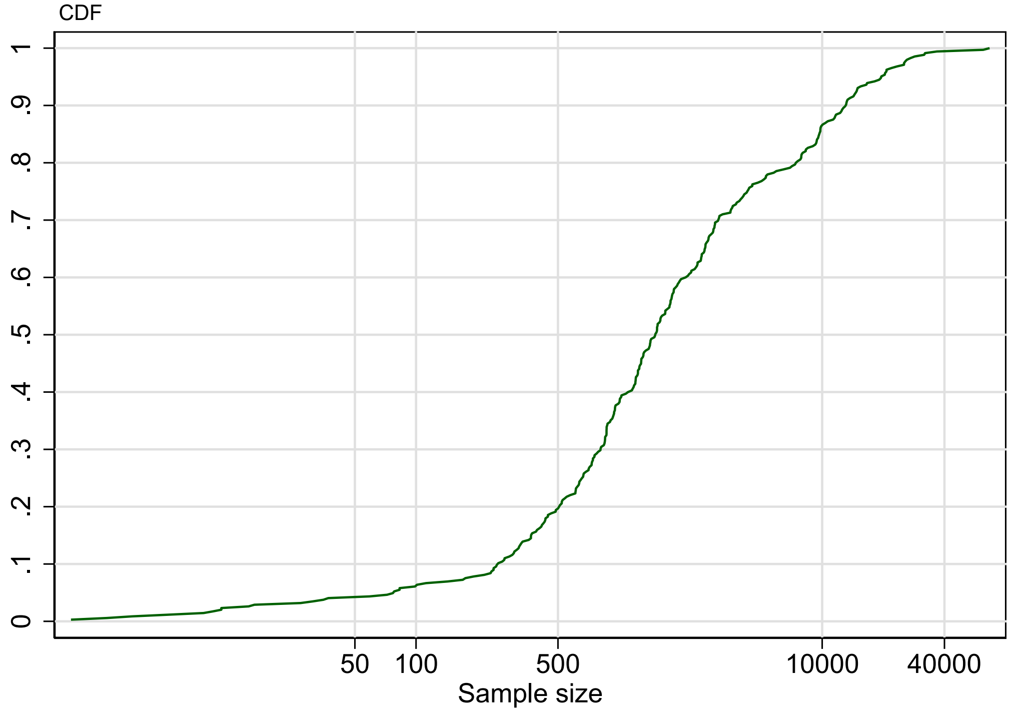
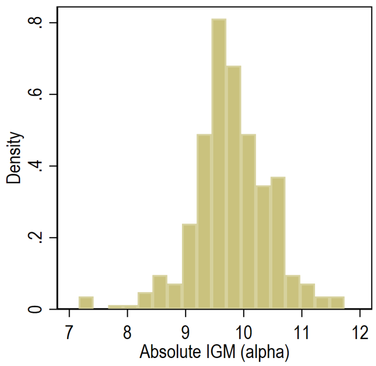
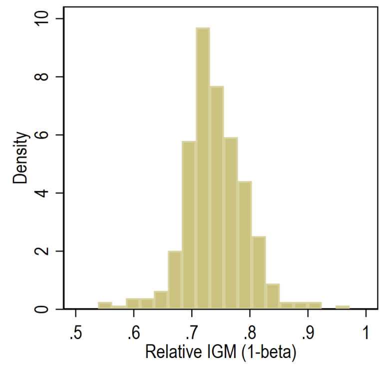


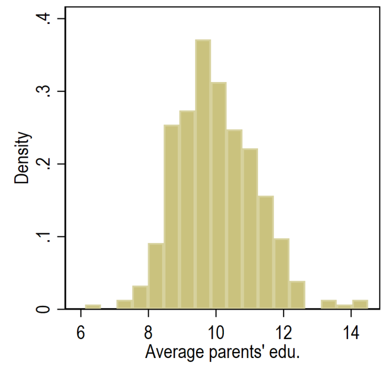
Figure A6: Distribution of commune-level estimates



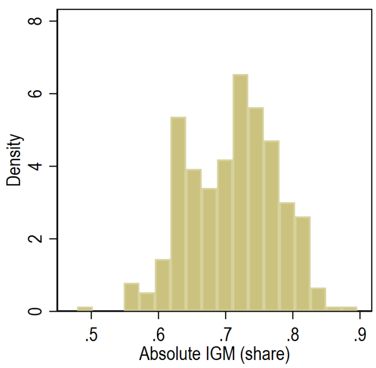
(a) Absolute IGM (α)



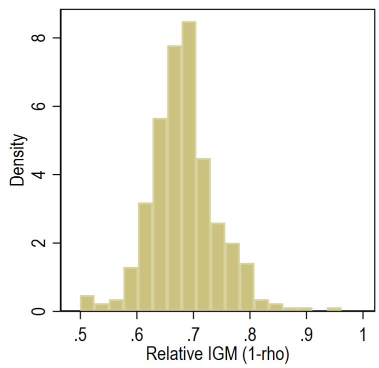
(b) Relative IGM ($1 - \beta$)



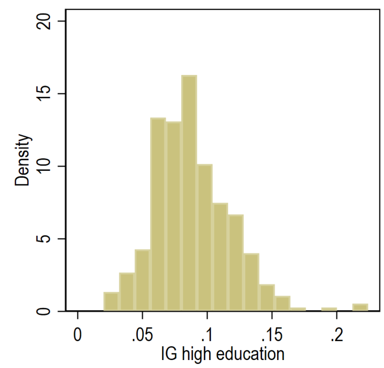
(c) Average education ($\bar{Y}$)



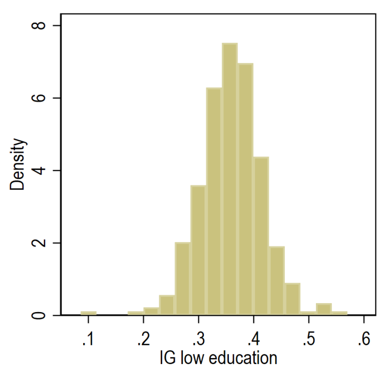
(d) Above parent ($\bar{y}^{\geq}$)



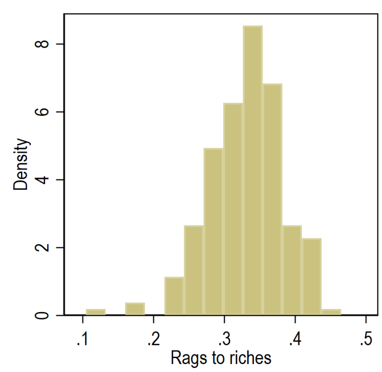
(e) Relative IGM ($1 - \rho$)



(f) Rags to riches ($P_{1,5}$)



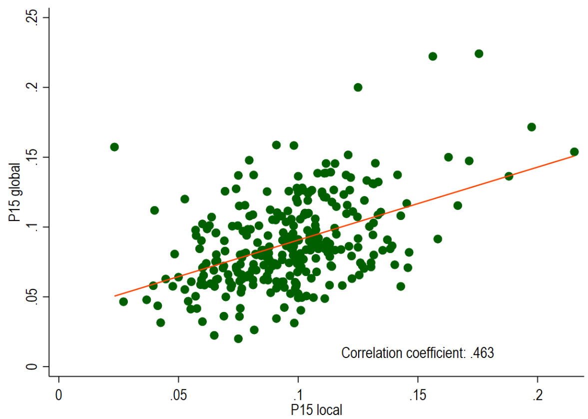
(g) Intergenerational low ($P_{1,1}$)



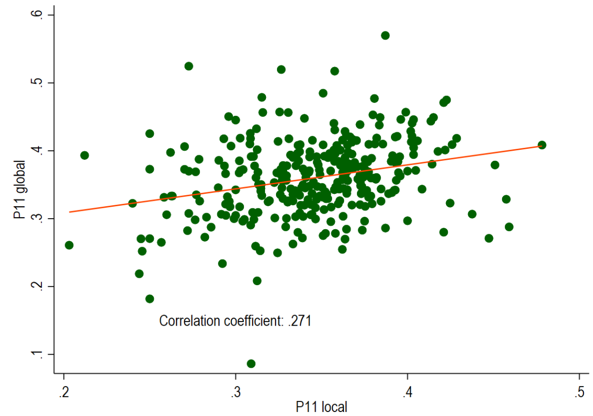
(h) Intergenerational high ($P_{5,5}$)

These histograms show the distribution of the commune-level estimates estimated with a sample of individuals of age 21-25 omitting communes with less than 50 individuals. For details about the indicators see Table 1.

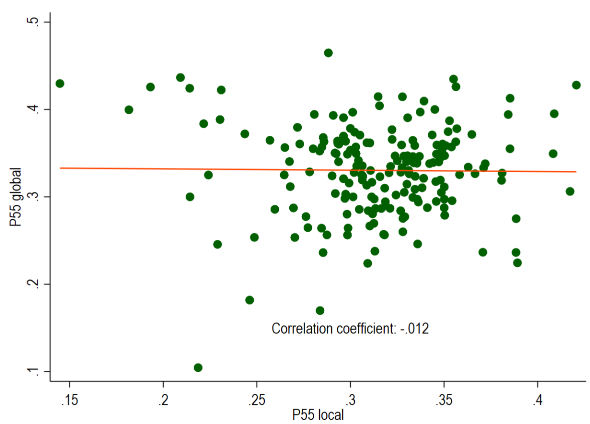
Figure A7: Comparison of indicators using country level distribution of educational attainment vs. local distribution



(a) Rags to riches ($P_{1,5}$)



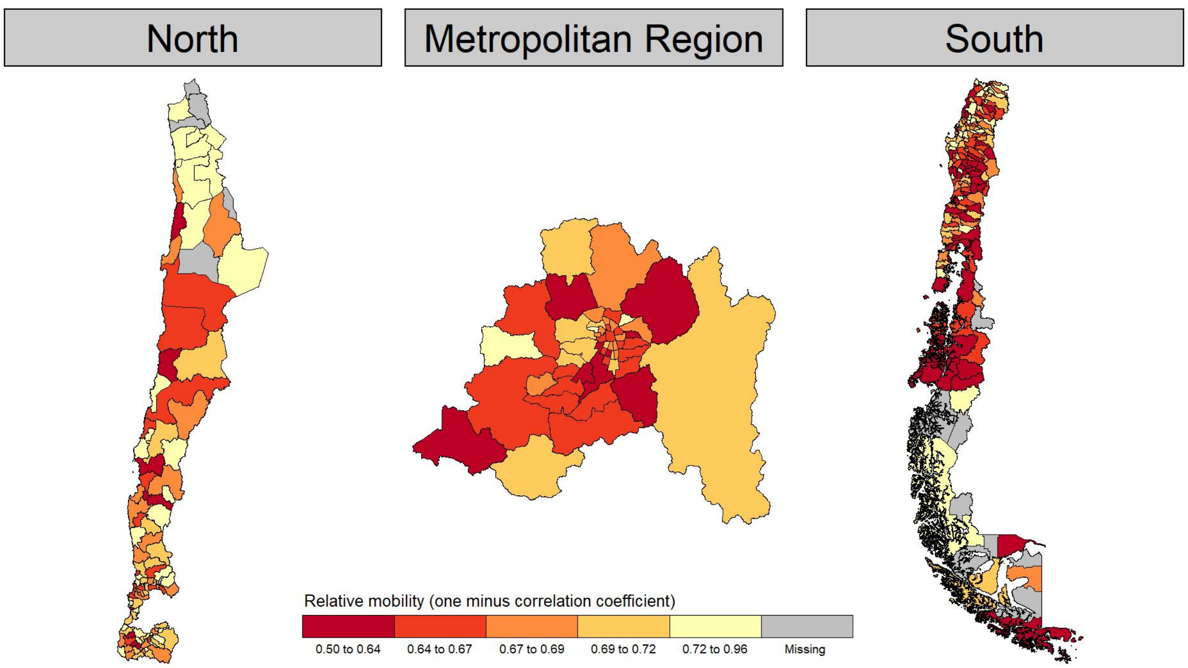
(b) Intergenerational low ($P_{1,1}$)



(c) Intergenerational high ($P_{5,5}$)

The figure compares estimates of rags to riches, intergenerational low, and intergenerational high measures computed using quintiles based on country-level educational attainment versus commune-level attainment (denoted local). Each uses a sample of individuals of age 21-25 omitting communes with less than 50 individuals. For details about the indicators see Table 1.

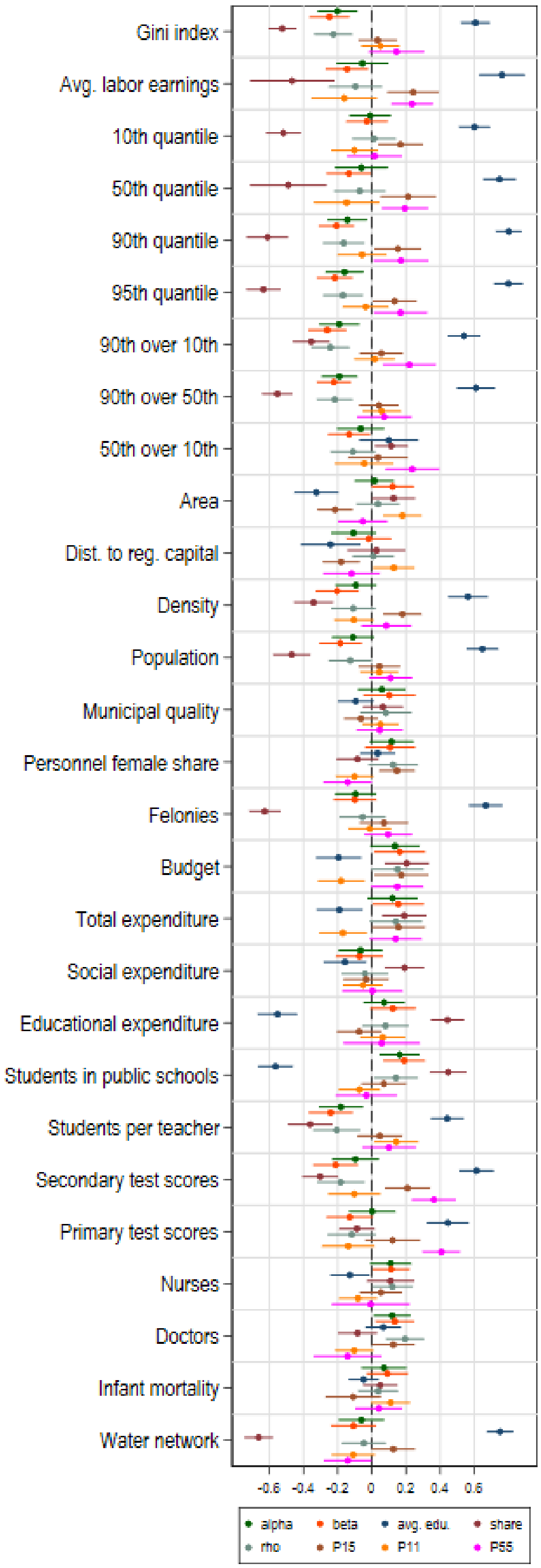
Figure A8: Intergenerational educational mobility within Chile



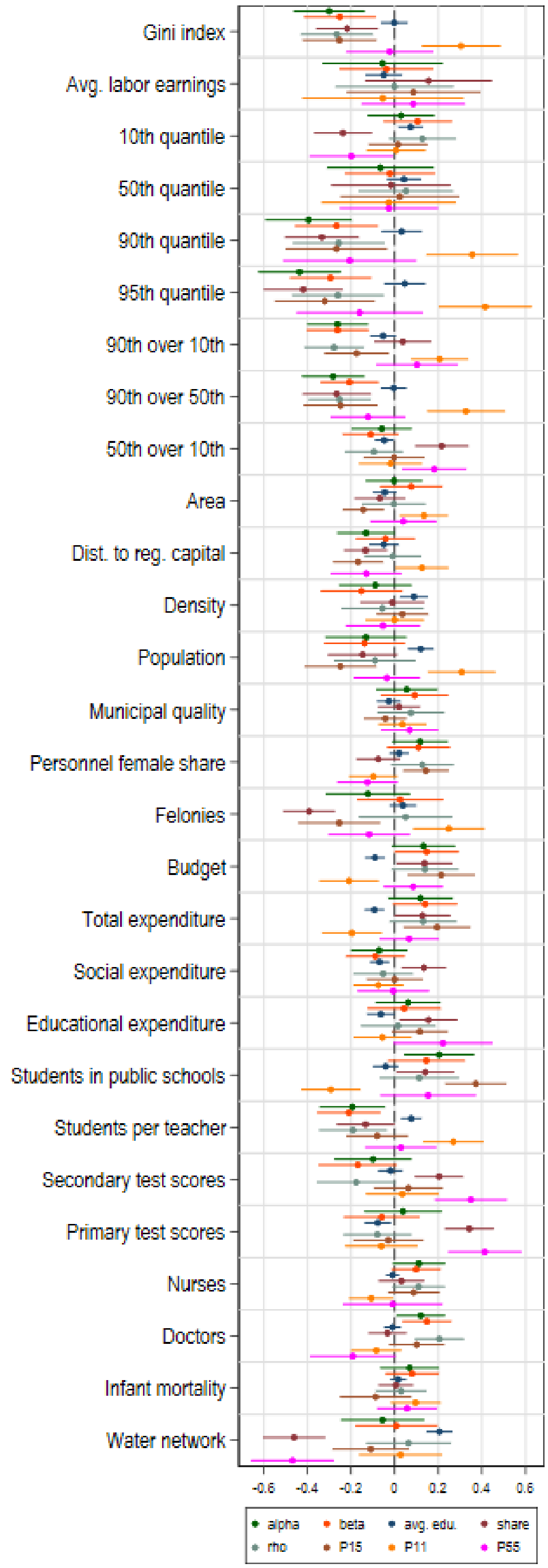
(a) Relative mobility by commune - Chile, 2017

The map plots relative IGM measured as one minus the Pearson correlation coefficient (by commune) between child's years of schooling (using age between 21 and 25) against parents' years of schooling. Educational attainment is censored at 15. Communes with less than 50 observations are left as missing.

Figure A9: Correlates of the IGM at the commune-level (all the indicators)

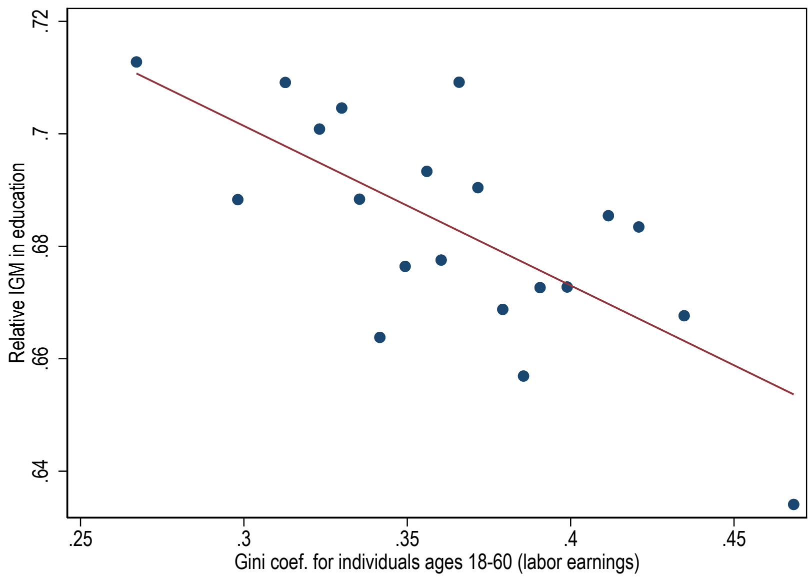


(a) Unconditional



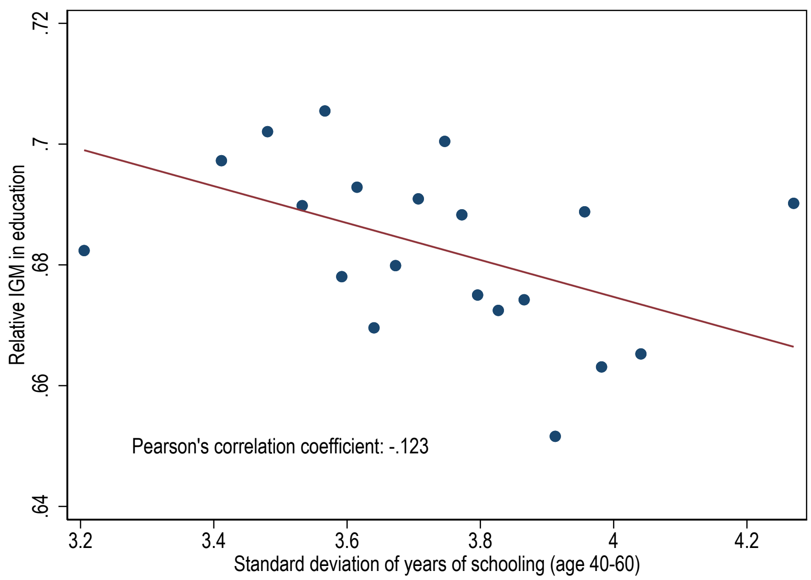
(b) Conditional on education of old cohorts

Figure A10: Intergenerational mobility in education vs. income inequality



The figure shows a binscatter plot between relative IGM (measured as one minus the Pearson correlation coefficient between child's years of schooling against parents' years of schooling) and the Gini coefficient computed using labor earnings in 2010 of individuals ages 18-60. Educational attainment is censored at 15 and the sample includes individuals with age between 21 and 25. Communes with less than 50 observations are not included.

Figure A11: Intergenerational mobility in education vs. inequality in education



The figure shows a binscatter plot between relative IGM (measured as one minus the Pearson correlation coefficient between child's years of schooling against parents' years of schooling) and the standard deviation of years of schooling computed using individuals ages 40-60 that are used as parents. Educational attainment is censored at 15. Communes with less than 50 observations are not included.

Table A4: Region-level estimates of IGM Statistics

Region	$P_{1,5}$	$P_{1,5}^{local}$	$P_{1,1}$	$P_{1,1}^{local}$	$P_{5,5}$	$P_{5,5}^{local}$
Tarapacá	0.10	0.11	0.37	0.33	0.32	0.31
Antofagasta	0.08	0.10	0.40	0.36	0.31	0.33
Atacama	0.07	0.10	0.40	0.34	0.29	0.36
Coquimbo	0.07	0.08	0.38	0.34	0.32	0.31
Valparaíso	0.09	0.10	0.35	0.36	0.34	0.33
Libertador General Bernardo O'Higgins	0.10	0.10	0.34	0.34	0.31	0.31
Maule	0.09	0.09	0.36	0.37	0.32	0.31
Biobío	0.11	0.09	0.33	0.36	0.37	0.32
Araucanía	0.07	0.07	0.38	0.39	0.37	0.36
Los Lagos	0.07	0.08	0.41	0.38	0.31	0.32
Aysén del General Carlos Ibáñez del Campo	0.05	0.09	0.44	0.35	0.23	0.32
Magallanes y de la Antártica Chilena	0.10	0.10	0.30	0.31	0.30	0.33
Metropolitana de Santiago	0.09	0.09	0.37	0.36	0.36	0.36
Los Ríos	0.06	0.06	0.38	0.37	0.34	0.35
Arica y Parinacota	0.14	0.14	0.28	0.30	0.31	0.32
Ñuble	0.11	0.09	0.33	0.35	0.36	0.27

The table reports region-level estimates of rags to riches, intergenerational low, and intergenerational high (a description of the measures can be found in Table 1). It compares measures that assign individuals into quintiles using the distribution of educational attainment at the country level with measures that use the distribution of each region (those with the superscript "local").