"Essays on Intergenerational Mobility and Measures of Teacher's Quality in Chile"

TESIS PARA OPTAR AL GRADO DE DOCTOR EN ECONOMÍA

Alumno: Gabriel Ignacio Villarroel Neira
Profesores Guías: Juan Díaz Maureira y Jorge Rivera Cayupi

Santiago, Agosto 2021

A mis padres, Deyanira y Gabriel, y hermano, Daniel.

Abstract

The first chapter provides the first consistent estimates of intergenerational earnings mobility in Chile which is, based on administrative records that link a child's and their parent's earnings from the formal private labour sector. We estimate that the intergenerational earnings elasticity is between 0.288 and 0.323, whereas the rank-rank slope is between 0.254 and 0.275. We find significant non-linearities in the relationship between parents' and their children's earnings, where the intergenerational mobility is high in the bottom 80% of the parents' distribution but with extremely high intergenerational persistence in the upper part of the earnings distribution. In addition, we find remarkable heterogeneity in intergenerational mobility at the regional level, where Antofagasta, a mining region, is the most upwardly-mobile region. Finally, we estimate significant differences across municipalities in the Metropolitan Region, where our estimates suggest that the place of residence makes a significant difference in intergenerational mobility for children of upper-class families, while it is less relatively important for children of lower- and middle-class families.

The second chapter estimates intergenerational mobility for Chilean males using an administrative data set that links parents' and their sons' earnings. We find that intergenerational earnings elasticity (IGE) and rank-rank correlation are 0.282 and 0.239, respectively. Our IGE estimate is about half of the previous estimates for Chile that have used the Two-Sample Two-Stage Least Squares (TSTSLS) method, where parents' earnings must be imputed. We simulate a TSTSLS setting with our data and recover these past estimates. Then, we show that TSTSLS estimates have two sources of bias: a projection bias and a variance bias, which are both consequences of imputing parents' earnings via Mincer regressions. To improve IGE estimation under TSTSLS, we provide two steps to reduce these biases: parent fixed effects to

improve the Mincer equation predictions and stochastic imputation to increase the variance of predicted wages. We show that if both of these corrections are used, we can recover our original estimates. The results are closer to our measure of IGE, but only when we have a precise first stage, which requires information beyond what is usually found in household surveys. We show that rank-rank correlations estimated using the TSTSLS method are much closer to estimates that comes directly from the administrative data. Our results suggest that administrative data should be used to measure intergenerational mobility, however, when linked earnings data between parents and their children is not available, researchers should focus on rank-rank correlations for this purpose.

The third chapter investigates two measures of teacher's quality and their impact on tertiary education attendance utilizing a novel national administrative data set. The two alternative instruments that measure teachers' effectiveness for the same sample of Chilean teachers and students are: the National Teachers' Evaluation test (Evaluación Docente, ED) and the traditional value-added results approach (VA) used in the literature. We find that the correlation between the measurements of teachers' quality from the ED and the VA approach appears to be null, which could be due to differences in the dimensions of teacher quality measured (as suggested by previous studies). Our analysis also reveals that both measures, ED and VA, positively affect the probability of tertiary education attendance, corroborating that both measures are complementary in measuring teacher quality. Additionally, we show that two (portfolio and external references) out of four parts of the ED are the best predictors of graduate students' tertiary education attendance. These results suggest that the best approach for evaluating teachers should consider a combination of the VA and ED, with improved instruments measured in the ED in terms of cost and teachers' time spent in the evaluation.

Acknowledgements

I would like to thank the people who contributed to this thesis and supported me throughout my graduate studies.

A special thanks to Juan Díaz, my friend and thesis supervisor, for his ideas, advice, knowledge, and patience. Likewise to Jorge Rivera, my thesis supervisor, for his trust, contributions, and recommendations for improving this work.

I would like to thank the contribution of Pablo Gutiérrez in Chapter I. Pablo, Juan Díaz, and I are co-authors of this work. In the same way, I would like to thank the contribution of Pablo Gutiérrez and Pablo Troncoso in Chapter II. They, Juan Díaz, and I are co-authors of this work. Both works will soon be sent to international journals for their reviews.

I would also like to thank the contribution of Rafael Sánchez and Mauricio Villena in Chapter III. They, Juan Díaz, and I are co-authors of this work, which has been published in the Institute of Labor Economics IZA DP No. 14277.

I would like to thank Josefina Rodriguez, my partner, for preparing the maps and her patience in translating and editing the text of this thesis.

I am grateful for the support of my superiors in the Budget Office of Ministry of Finance of Chile, Sergio Granados, Rodrigo Cerda, and Matías Avecedo, for their trust and for allowing me to initiate and complete my graduate studies.

Finally, I thank the committee members for taking the time to review this thesis.

Contents

Chapter 1

Intergenerational mobility in Chile

1.1 Introduction

This chapter asks whether the association between parents' and their child's earnings in Chile varies with parental earnings level and children's place of residence. Chile is an interesting case study not only due to having made significant progress in its economic development in the last three decades (reaching a GDP per capita of US$ 16,143 in 2018, IMF, 2018) but also because it is one of the countries with the most unequal income distribution in the world. It has a Gini index of 0.477 points (World Bank, 2017), and the fraction of the country's total income received by the richest 10% of the population is extremely high (37.1%) when compared to the OECD average of 24.7% (OECD, 2018). Moreover, conservative estimates suggest that the share of total income that the richest 1% take is 15%, while less conservative estimates establish it at 22-26% (Fairfield and Jorrat, 2016; Flores et al. 2019).

Under what conditions an unequal society can be tolerated is a subject of long-standing debate, especially in Chile. Supporters of meritocracy argue that economic inequality can be legitimated in a society if income differences stem from differences in reward for talent, hard work and skill, but not due to luck or transmission of advantages. According to this view, income inequality should not be tolerated in

a society with less social mobility and greater transmission of privileges or disadvantages from parent to child, where children born in poverty (richness) remain in poverty (richness) in their adulthood, regardless of their skills or efforts. In part, the de-legitimization of income inequality is one of the main causes behind the social unrest that occurred in Chile in October 2019, when the perception of unfairness in the distribution of income and privileges provoked lower and middle classes to take to the streets to express their indignation with the current situation. In this context, understanding social mobility in Chile is crucial to disentangle the origins of its current levels of economic inequality.

In this chapter, we study intergenerational mobility in Chile by building an unique data set after assembling three administrative data sources[1]. We obtain information on labor earnings of children and their parents from 2002 to 2019 from the database of the Chilean government's unemployment insurance program (UIP). We link children and their parents using administrative records provided by the Civil Registry Office. We obtain the place of residence of a child when they were between 13 and 18 years old from administrative records at the Ministry of Education. To the best of our knowledge, this is the first work that uses administrative information to estimate intergenerational mobility for a non-advanced economy.

We estimate intergenerational earnings mobility at the national level. We find that it is highly non-linear in Chile, and intergenerational mobility is very high for the bottom 80 percent of the earnings distribution, and exceed the rate of intergenerational mobility in advanced countries such as the US and Canada. But earnings are also highly persistent for the upper decile of the earnings distribution, much more so than for any advanced economy. This result resembles what Bratsberg et al. (2008) find when comparing the Nordic countries with the US and UK.

We also estimate intergenerational earnings mobility at the regional level. This

---

is of particular interest for a country like Chile, where the climate and economic conditions are significantly heterogeneous across its geography. We find that the most mobile region is Antofagasta, which is a minering intensive region located in the north of the country. This result is in line with the findings for developed economies (Australia and Canada). Meanwhile, the least mobile region is La Araucanía, where about a third of the population is ethnic Mapuche (an indigenous population) - the highest proportion of any region in Chile.

Finally, we estimate intergenerational mobility across different municipalities for the Santiago Metropolitan Region. This region contains the nation's capital, Santiago, one of the cities with a better quality of life in South America. We find that Santiago is extremely heterogeneous in upward mobility, circles of poverty and circles of privilege. In particular, there is a cluster of rich municipalities where the conditional probability that a child stays in the fifth quintile given that the parent was in the fifth quintile of their earnings distribution is higher than 0.7. Those rich municipalities are quite similar in terms of upward mobility.

We also make a methodological contribution, we use for the first time tools to estimate intergenerational mobility at the top of the distribution and Kernel conditional densities. In addition, we estimate the Gatsby curve for Chile and Santiago using two measures of intergenerational mobility: absolute intergenerational mobility and relative intergenerational mobility. We show that the Gatsby curve could be valid for a persistence indicator but not for an absolute mobility indicator. This means that inequality could be related with persistence at the top instead of mobility at the bottom.

Of course, there is a vast body of literature from economists trying to learn about social mobility from administrative records in advanced economies. For the United States, there is a series of articles that are based on a project by Raj Chetty, Nathaniel Handren and others, who use administrative tax data to estimate the in-

tergenerational elasticity of income.[2] For example, the work of Chetty, Hendren, Kline and Saez (2014) studies how social mobility varies through geographic zones called community zones in the US. For Canada, the literature on intergenerational income mobility starts with the seminal work of Corak and Heisz (1999), a pioneering paper in the use of administrative data to study intergenerational mobility of income. More recently, Corak (2019) studies intergenerational mobility in Canada utilizing census data and analyzing data at various geographic levels. Europe has also produced some interesting literature in this regard. For instance, Acciari et al. (2019) use tax data to investigate how intergenerational mobility varies geographically for Italy, as do Güell et al. (2015) for social mobility at smaller geographical units in Italy, which Heidrich (2015) also does for Switzerland. Most of these works for developed countries show that disaggregated geographical measures of intergenerational mobility provide evidence of significant heterogeneities across locations that are hidden in country-level estimates.

In the case of Chile, our work does not emerge in a vacuum. Over the last two decades, some papers have made progress in understanding social mobility by using survey data. For example, Núñez and Miranda (2010, 2011) study intergenerational income mobility by using the Two-Sample Two-Stage Least Squares (TSTSLS) methodology developed by Björklund and Jäntti (1997). Sapelli (2013) provides evidence on changes in the intergenerational mobility of education through time, using several cross-section surveys. Meanwhile, Torche (2005) analyzes the intergenerational mobility of education based on survey data, and Celhay et al. (2010) focus on the study of intergenerational mobility of income and schooling for the period 1996-2006 using longitudinal surveys. The only paper that uses administrative records to capture a specific dimension of intergenerational mobility in Chile is the work of Zimmerman (2019). Based on a regression discontinuity design, this article illus-

---

[2]In this chapter, we make the distinction between earnings, for which the source is wages, and income, for which the sources are wages and financial asset income. Our study is developed with earnings due to the available dataset.

trates the lack of upward mobility by showing that studying at an elite university has a positive effect on obtaining a managerial position with high income in the labor market, but only for those with a high-level socioeconomic background who had studied at an elite private school. This study, in part, quantifies the importance of contact networks in the generation of inequality in Chile.

1.2   Development of the income intergenerational mobility literature

Are the children of the poor doomed to stay poor? Are the children of the rich destined to stay rich? How difficult is it for someone who was born poor to belong to the middle class during her adulthood? These questions have been addressed at the international level, where there is vast literature on intergenerational income mobility. Jäntti and Jenkins (2015) and Corak (2013) summarize the historical results in this literature. Corak and Heisz (1999) were the first to use high-frequency administrative data on the income of parents and children in adulthood in their seminal study on intergenerational mobility in Canada.[3] This study was so innovative and ahead of its time that it took 15 years for literature to replicate this study for other countries. In fact, thanks to the development of computer science and generalization in the use of administrative data, the literature of intergenerational mobility has been given a new lease of life. The works of Chetty, Hendren, Kline and Saez (2014), Chetty et al. (2017), and Chetty et al. (2018a, 2018b) have extensively studied intergenerational mobility in the United States using the same type of data.

Undoubtedly, the novelty of these studies is in the data used, which mainly correspond to confidential high-frequency administrative data that cover a sufficiently long period and link the income of the parents with the adult income of their chil-

---

[3]Others important studies on intergenerational mobility for Canada are Fortin and Lefebvre (1998), and Simard-Duplain and St-Denis (2020)

dren. The advantage of administrative data is that they do not have the traditional problems present in household surveys. In fact, traditional household surveys in general are not longitudinal but cross-sectional, which makes it difficult to obtain information on the income of the parent and child in adulthood. In addition, household surveys have problems such as sampling, self-reporting and non-response, and it is known that non-response rises as the respondent's income increases (Bollinger et al., 2018).

Understanding the intergenerational mobility of income in the United States has been tremendously important in understanding the generation of inequality. There is a series of articles that are based on a project by Raj Chetty, Nathaniel Handren and others, who use administrative tax data to estimate the intergenerational elasticity of income. The work of Chetty, Hendren, Kline and Saez (2014) studies geographic zones called community zones. The abovementioned investigation by Chetty and others differentiate between absolute and relative intergenerational mobility, which has been of interest to both politicians and researchers. The Canadian literature on intergenerational income mobility starts with the seminal work of Corak and Heisz (1999), pioneering in the use of administrative data to study intergenerational mobility of income. More recently, Corak (2019) studied intergenerational mobility in Canada, using census data and analyzing intergenerational mobility within Canada at a geographic level. Acciari, Polo and Violante (2019) investigate intergenerational mobility for Italy by taking tax data, also analyzing what happens geographically. Finally, this literature has also progressed in Europe, mostly based in the Nordic countries. Jäntti (2006) illustrates very well the use of these data. Also, there are the studies for Switzerland by Heidrich (2015) and Güell et al. (2015) for Italy. Both studies are at the provincial and inter-country levels.

1.2.1 Intergenerational mobility of income, the case of developing countries

Research on intergenerational mobility of income in developing countries faces additional complications. Having longitudinal data that gather parents and children is very difficult (Daude and Robano, 2015, Neidhöfer, 2019, Neidhöfer et al., 2018) due to the limitation of household surveys and/or the difficulty of accessing administrative data.

One way to address the limitations of the data is to restrict the analysis to children and parents living in the same household or to impute an income for the parents based on multiple waves of a household survey. For example, Lambert et al. (2014) studies intergenerational mobility in Senegal and Torche (2014) summarizes intergenerational mobility in Latin America from studies that have used surveys as a primary source of information.

Recently, progress has been made to investigate intergenerational mobility using census data from 26 African countries (Alesina et al., 2019) and for the regions of India, Asher et al. (2018). In this context, our research project will be pioneering in Latin America because it uses administrative data, which is the way in which the frontier literature is studying intergenerational mobility.

1.2.2 Intergenerational mobility of income, regional differences

Recent literature has concentrated on studying the regional differences that exist within countries.[4] They find that regional intergenerational income mobility behaves differently among countries. Chetty and Corak find differences among regions, where there are certain territories that have less intergenerational mobility than other parts.

---

[4]See Chetty, Hendren, Kline and Saez (2014), Chetty et al. (2018a, 2018b), Corak (2019), Güell et al. (2015), Heidrich (2015), and Connolly et al. (2018).

However, for Switzerland, Heidrich (2015) does not find many differences. In the Chilean case, Núñez and Miranda (2011) find that the intergenerational mobility of income is higher in Santiago compared to the Chilean average. Inequality has been studied at the regional level in Chile. However, how regional intergenerational mobility varies in Chile has not been studied.

1.3 Data

1.3.1 Information on labor earnings

We obtain the information on labor earnings of children and their parents from the database of the UIP in Chile. The UIP is a benefit that covers all employees in the private sector over 18 years old and with a formal contract, whether fixed-term or permanent. Column 4 of Table 1.1 provides information from the main employment survey in Chile (Encuesta Nacional de Empleo) on the proportion of private formal contract employees over total workers, which moves between 54% and 60%.

Participation in the UIP scheme is mandatory for all contracts started after September 2002 and voluntary for contracts started before that date. This means that these administrative records contain the monthly labor earnings of all employed workers over the age of 18 who initiated a work-under-contract relationship in the private sector from October 2002 to December 2019. This data set also includes the workers with labor contracts established prior to October 2002 who voluntarily joined the UIP. It is worth mentioning that this data set excludes workers with training contracts, workers under the age of 18, domestic workers, pensioners, self-employed or own-account workers, and public sector employees.

Table 1.1 provides information on the proportion of workers covered by the UIP over several years. As can be seen, due to the voluntary retroactive nature of the UIP policy, the coverage rate for private formal contract employees[5] was below 50%

---

[5]Private Formal Employees is recorded by ENE since 2010. For years 2003 to 2009 we project

in 2003 and 2004 (column 6, Table 1.1). In the following years, this coverage rate significantly increased, attaining 60.4% in average in 2005-2007, over 80% from 2011 and over 90% from 2015. The differences between the number of private formal employees in the UIP database and the ENE survey, in its first years is due to the voluntary scheme for contracts started before the 2002, while for recent years, we believe that the differences are specific to the sources of information from which it comes, administrative records and national survey.

Table 1.1 also shows information on private formal employees covered by the UIP as a proportion of the total workers, according to the ENE survey. Initially, the total workers coverage rate was 55.2% in average for the years 2003-2007, which rapidly converged to 60% in 2013. The workers not covered by the UIP in recent years is explained by public sector employees, informal workers, training contracts, domestic workers and pensioners.[6]

We must acknowledge that the low private formal contract employees' coverage rate during the first years of the data (51.8% in average in 2003-2007) is a concern for our analysis because —as explained below— it impacts how we model permanent parental earnings for our baseline sample. To assess the plausibility of our findings, we perform a robustness exercise. We frame our analysis using data for years with a higher formal contract workers' coverage rate to construct the permanent parental earnings.

---

the number of Private Formal Employees by multiplying the number Formal Employees by the proportion of Private Formal Employees recorded by 2010, 84.7%

[6]As we can see from Table 4.1, this dataset converges to a coverage rate of 77% in average in 2013-2018 of the formal employees but only to 50.2% for the total labour force. This is in part because this dataset has limited coverage for the unemployed. Sehnbruch (2006) and Ruiz-Tagle and Sehnbruch (2010) argue that this is because a large proportion of unemployed register by ENE previously worked in the informal sector.

Year	Total Workers (ENE)	Formal Employees (ENE)	Proj. Private Formal Empl.	% Proj. Private Formal Empl/Total Workers	Total UIPD	% Total UIDP/Proj. Private Formal Empl.
2003	5,786.4	3,670.5	3,108.8	53.7%	727.5	23.4%
2004	5,942.4	3,807.7	3,225.0	54.3%	1,392.0	43.2%
2005	6,171.1	3,986.1	3,376.1	54.7%	1,897.9	56.2%
2006	6,274.4	4,166.3	3,528.7	56.2%	2,276.1	64.2%
2007	6,449.4	4,360.9	3,693.5	57.3%	2,643.2	71.6%
2008	6,638.9	4,582.5	3,881.2	58.5%	2,958.8	76.2%
2009	6,594.7	4,501.9	3,812.9	57.8%	3,039.4	79.7%
2010	7,148.5	4,910.5	4,159.0	58.2%	3,313.7	79.7%
2011	7,478.8	5,143.0	4,392.9	58.7%	3,646.1	83.0%
2012	7,627.1	5,361.9	4,553.9	59.7%	3,950.8	86.8%
2013	7,785.2	5,481.4	4,669.4	60.0%	4,112.5	88.1%
2014	7,904.2	5,530.7	4,666.4	59.0%	4,190.7	89.8%
2015	8,022.8	5,651.1	4,752.5	59.2%	4,298.1	90.4%
2016	8,122.4	5,661.7	4,789.5	59.0%	4,389.7	91.7%
2017	8,276.1	5,713.0	4,766.7	57.6%	4,440.8	93.2%
2018	8,391.8	5,821.0	4,815.1	57.4%	4,618.6	95.9%

Table 1.1: Representativity of the unemployment insurance program dataset

This dataset is compared with the information of the ENE (Encuesta Nacional de Empleo) questionnaire administered by the government statistics agency in Chile (INE-Instituto Nacional de Estadisticas). All columns correspond to annual averages. Units are measured on thousands. Total Workers recorded by ENE refers to the total number of workers including: Private Formal Employees, Public Sector Employees, Informal Workers, Training Contracts and Domestic Workers. Projection Private Formal Employees is recorded by ENE since 2010. For years 2003 to 2009 we project the number of Private Formal Employees by multiplying the number Formal Employees by the proportion of Private Formal Employees recorded by 2010, 84.7%. Total UIPD is the annual averages of private formal employees in the database of the UIP.

1.3.2   Information on child-parent linkage

We link children and their parents using administrative records provided by the Civil Registry Office (CRO). In Chile, the CRO registers all births, deaths, and marriages. It is a legal requirement in Chile that all births must be registered in the CRO, each of which is backed by a birth certificate. This birth certificate contains the information on the child and the parents given at the time of registration. We use the information provided for all the birth certificates in Chile to build the pairs of children and parents included in the UIP database. In our baseline analysis, the sample of children is composed of individuals that were 28-33 years old in 2018, while the sample of parents are individuals that were 42-87 years old in 2018.

1.3.3   Measurement of earnings

Our administrative records have information on labour earnings in the formal private sector, excluding any form of capital income for the workers covered by the UIP. In our baseline sample, we measure parental earnings as the 5-year average of monthly earnings for months worked in the formal private sector between 2003 and 2007. For example, if a parent records 30 months worked within a 5-year period, the measure of earnings used is the total income in those 5 years divided by 30. In our baseline sample, we only consider parents that worked at least 6 months in the formal private sector during 2003-2007.[7]

Our measure of parental earnings excludes the zeros because a zero in our data set does not mean that the individual has no earnings, since he/she could be earning as a public employee, in the informal sector, or in the formal private sector but not covered by the UIP, especially in its earlier years.

As with the parents, we measure child earnings in our baseline sample as the

---

[7]If both parents worked in the period, we consider the average parental earnings as the sum of parental earnings divided by two, in line with Chetty, Hendren, Kline and Saez (2014) and Corak (2019).

five-year average of monthly earnings for worked months in the formal private sector between 2014 and 2018. In our baseline analysis, we consider children that worked at least six months in the formal private sector in 2014-2018. This measure of child earnings not only excludes the zeros for the same reasons as for their parental earnings, but also because children may start participating in the private formal labor market in their late 20s, giving a series of months with earnings preceded by a series of zeros corresponding to not being in the labor market.

To minimize the noise provoked by low earners due to the uncertainty surrounding the low earnings registered with the UIP, we only consider children and parents who on average earn more than half the minimum wage.[8] In our baseline sample, we have 505,524 parent-child links.

1.3.4 Comparison between unemployment insurance program dataset and ENE survey

In Chile, 29.6 percent of the population works in the informal sector. One potential issue for our dataset is that only contains information on private formal earnings. To see how different are the percentiles including all workers, we compare the earnings percentiles generated by our dataset and the Encuesta Nacional de Empleo (ENE).

---

[8]Half the minimum wage for children is $133,000 in 2019 Chilean pesos (measured from 2014 to 2018) and $103,000 in 2019 Chilean pesos for parents (from 2003 to 2007). Using CASEN 2017 information, 14.1 percent of the population were under the minimum wage.

Percentile	UIP	ENE
1%	152,889	170,614
5%	218,433	231,840
10%	263,508	250,902
25%	343,076	330,000
50%	490,707	451,624
75%	767,851	700,000
90%	1,173,052	1,003,609
95%	1,544,161	1,304,692
99%	2,371,979	2,500,000

Table 1.2: Comparison of earnings between our dataset and ENE for individuals between 28-33 years old

This dataset is compared with the information of the ENE (Encuesta Nacional de Empleo) questionnaire administred by the goverment statistics agency in Chile (INE-Instituto Nacional de Estadisticas). *W* ENE refers to the earnings percentiles for all workers – formal, informal and self employed. Units are in 2018 Chilean pesos.

Table 1.2 compares our dataset earnings percentiles with ENE dataset percentiles for 2018. We can see that percentiles are similar using the whole population and types of sector and the formal private sector.

1.3.5 Information on child residential address

We link the pairs of child and parental earnings with the residential address of the child while attending 12th grade in school. We obtain this information from administrative records provided by the Ministry of Education of Chile. If the child's residential address while attending 12th grade is not available, we use the most recently-available residential address while she was enrolled from 7th to 11th grade

in school (when the child is 13-18 years old).[9] We end up with 93.95% of the children's sample linked to their residential address.

1.4   Intergenerational mobility for Chile

We begin our empirical analysis by characterizing the relationship between parental and child earnings at the national level. We present a set of baseline estimates of relative intergenerational mobility and then evaluate the robustness of our estimates to alternative samples.

1.4.1   Traditional indicators of intergenerational mobility

Intergenerational earnings mobility

One of the most commonly used measures of intergenerational mobility is the intergenerational earnings elasticity, i.e., the effect that a 1 percent increase in the parental earnings has over their child's earnings. In our work, we estimate the intergenerational elasticity of earnings rather than of income because our dataset only contains information on wages and not on financial asset income. We measure this elasticity by estimating the following equation:

$$y_i^c = \alpha + \beta y_i^p + \epsilon_i, \tag{1.1}$$

where y_i^c is the earnings of child i in logarithms, y_i^p is the earnings of that child's parents in logarithms, and β is the intergenerational earnings elasticity. This parameter is equal to

$$\beta = \frac{cov(y_i^p, y_i^c)}{var(y_i^p)} = \rho \cdot \frac{sd(y_i^c)}{sd(y_i^p)}, \tag{1.2}$$

---

[9]We also estimate our results by making the geographic link from 5th to 12th grade. The results are similar.

	(1)	(2)	(3)	(4)
y_p	0.288***	0.297***	0.311***	0.323***
	(0.001)	(0.001)	(0.002)	(0.002)
Constant	9.506***	9.426***	9.298***	9.193***
	(0.016)	(0.018)	(0.021)	(0.027)
Observations	505,524	416,818	282,979	173,683
R-squared	0.091	0.098	0.108	0.117

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.3: OLS estimates of the intergenerational earnings elasticity for our baseline linkage

Earnings are measured as average earnings over the months where an individual reports positive earnings over the studied 5-year period. We keep individuals that appear at least 6 times with positive earnings in the dataset with average earnings greater than half of the corresponding minimum wage. Columns (1) to (4) report results for male and female children. (1) considers individuals with at least 6 months of positive earnings, (2) considers individuals with at least 12 months of positive earnings, (3) considers individuals with at least 24 months of positive earnings and (4) considers individuals with at least 36 months of positive earnings.

where ρ is the intergenerational earnings correlation, and $sd(y_i^c)$ and $sd(y_i^p)$ are the standard deviation of child and parental log earnings, respectively. To prevent any attenuation bias, we measure child and parental earnings as the 5-year average of earnings.

Table 1.3 summarizes our estimates for intergenerational earnings elasticity (IGE), i.e., the OLS estimates of the regression slope of the log child earnings on log parental earnings. Columns (1) to (4) report results for male and female children: (1) considers individuals with at least 6 months of positive earnings (our baseline sample); (2) considers individuals with at least 12 months of positive earnings; (3) considers individuals with at least 24 months of positive earnings; and, (4) considers individuals with at least 36 months of positive earnings.

Our baseline estimation for IGE equals 0.288. With our most restrictive sample

	(1)	(2)	(3)	(4)
y_p	0.300***	0.307***	0.315***	0.326***
	(0.002)	(0.002)	(0.003)	(0.003)
Constant	9.253***	9.209***	9.169***	9.086***
	(0.024)	(0.026)	(0.032)	(0.042)
Observations	222,397	178,916	116,182	68,644
R-squared	0.103	0.111	0.119	0.128

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.4: OLS estimates of the intergenerational earnings elasticity for female children

Earnings are measured as average earnings over the months where an individual reports positive earnings over the studied 5-year period. We keep individuals that appear at least 6 times with positive earnings in the dataset with average earnings greater than half of the corresponding minimum wage. Columns (1) to (4) report results for female children. (1) considers individuals with at least 6 months of positive earnings, (2) considers individuals with at least 12 months of positive earnings, (3) considers individuals with at least 24 months of positive earnings and (4) considers individuals with at least 36 months of positive earnings.

—individuals with at least 36 months of positive earnings—, this estimate equals 0.323. This means that an increase of 10 percent in parental earnings implies, on average, an increase of between 2.88 and 3.23 percent in their child's earnings[10]. For example, for child's earnings located in the median of the distribution, $490.707 Chilean pesos, an increase of 10 percent in parental earnings implies, on average, an increase of between $14.132 Chilean pesos and $15.850 Chilean pesos in their child's earnings.

Tables 1.4 and 1.5 estimate the IGE for female and male children respectively. Our results suggest that female children are slightly less intergenerationally mobile than male children.

---

[10]This estimate is lower compared with previous estimates in the Chilean literature. Núñez and Miranda (2010,2011), and Celhay et al. (2010) estimate an elasticity between 0.5 and 0.6. Our differences can be explained by the kind of data used and the method implemented to estimate IGE. Chapter 2 discusses this point in detail.

	(1)	(2)	(3)	(4)
y_p	0.282***	0.294***	0.314***	0.329***
	(0.002)	(0.002)	(0.002)	(0.003)
Constant	9.655***	9.529***	9.313***	9.175***
	(0.022)	(0.024)	(0.028)	(0.036)
Observations	283,127	237,902	166,797	105,039
R-squared	0.087	0.094	0.107	0.117

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.5: OLS estimates of the intergenerational earnings elasticity for male children

Earnings are measured as average earnings over the months where an individual reports positive earnings over the studied 5-year period. We keep individuals that appear at least 6 times with positive earnings in the dataset with average earnings greater than half of the corresponding minimum wage. Columns (1) to (4) report results for male children. (1) considers individuals with at least 6 months of positive earnings, (2) considers individuals with at least 12 months of positive earnings, (3) considers individuals with at least 24 months of positive earnings and (4) considers individuals with at least 36 months of positive earnings.

Rank-rank correlation

Another measure of intergenerational mobility that has become extremely popular is rank-rank correlation (Jureckova, 1971; Jaeckel, 1972; and Dahl and DeLeire, 2008, in the context of relative mobility). This correlation measures the effect that an increase of a percentile in the parental earnings distribution has over the child earnings distribution. One of the arguments to use rank-rank correlation is that the rankings on the earnings distribution are determined at earlier ages and are difficult to change throughout the age distribution. We measure this correlation by estimating the following equation by OLS:

$$r_i^c = \alpha_r + \beta_r r_i^p + \epsilon_i, \tag{1.3}$$

where r_i^c is the ranking of i-th child in the national distribution of child earnings by cohorts, r_i^p is the ranking of i-th child's parent on the national distribution of parental earnings, and β^r is the rank-rank correlation.[11] This correlation is an indicator of relative mobility that compares the maximum influence of parental ranking on expected child ranking. In addition, α_r is a measure of absolute mobility because it states the expected ranking that a child would have if her parents belong to the bottom of the parental earnings distribution.

---

[11]Note that we compute the ranking of the whole cohort of children and parents, regardless of whether they are linked.

Authors calculations
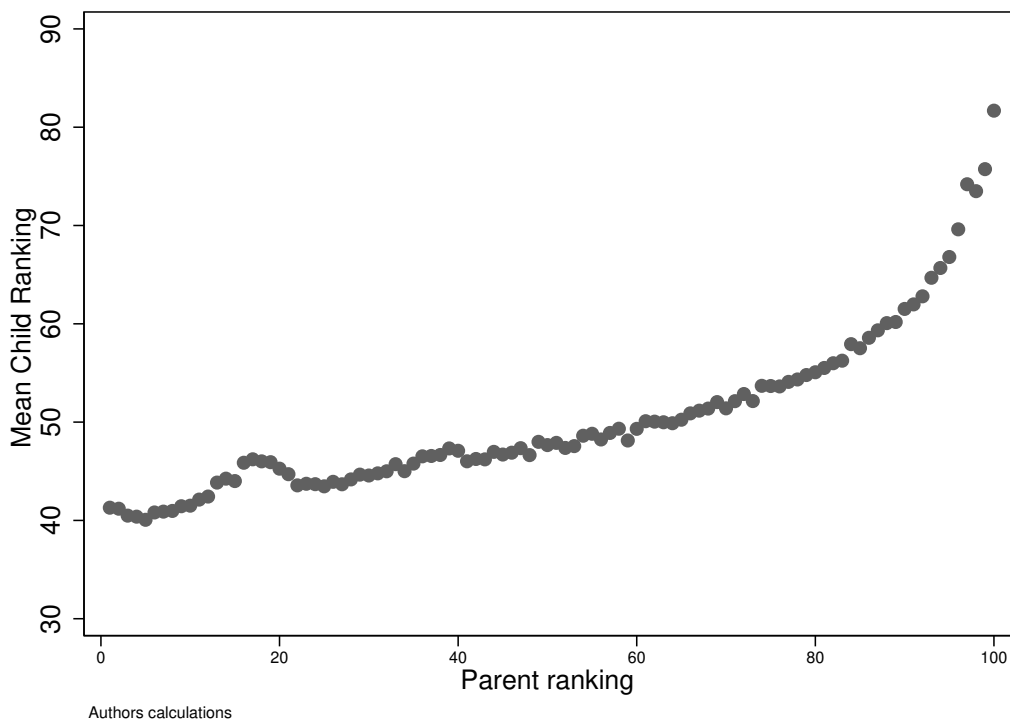
Figure 1.1: Expected child ranking conditional on parental ranking

We estimate the expected child ranking non parametrically using a simple average. Rankings were computed over the national distribution. For children we compute the cohort ranking, and for parents we compute the ranking of people 42-87 years old (in 2018).

Figure 1.1 presents a binned scatter plot of the mean percentile rank of children versus their parents' percentile rank. This graph illustrates a nonparametric estimation of the conditional expectation of a child's rank given her parents' rank ($E[r_i^c|r_i^p = p]$). As we can see, the relationship between parental ranking and child ranking is close to a linear function until the 80th parental percentile, while for parental percentiles higher than 80 it is highly non-linear with an increasing gradient as the parental ranking increases.

Table 1.6 presents our estimates for the rank-rank slope. To measure the percentile rank of the children, we consider their rankings in the distribution of child

	(1)	(2)	(3)	(4)
r^p	0.254***	0.261***	0.270***	0.275***
	(0.001)	(0.001)	(0.002)	(0.002)
Constant	37.397***	38.668***	40.859***	43.368***
	(0.080)	(0.089)	(0.110)	(0.141)
Observations	505,524	416,818	282,979	173,683
R-squared	0.064	0.068	0.073	0.078

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.6: OLS estimates of the rank-rank correlation for our baseline linkage

Earnings are measured as the average earnings over the months in which an individual reports positive earnings over 5 years. We keep individuals that appear at least 6 times with positive earnings in the dataset with average earnings greater than half of the corresponding minimum wage. Columns (1) to (4) report results for male and female children. (1) considers individuals with at least 6 months of positive earnings, (2) considers individuals with at least 12 months of positive earnings, (3) considers individuals with at least 24 months of positive earnings and (4) considers individuals with at least 36 months of positive earnings.

earnings within their birth cohorts. In the same way, we compute the percentile rank of the parents from their positions in the distribution of parental earnings in the baseline sample. Based on the child and parental percentile ranks, the rank-rank slope estimate is the OLS estimate of the regression slope of the percentile rank of a child on the percentile rank of her parents. As before, columns (1)-(4) in Table 1.6 present the results for 6 (baseline sample), 12, 24, and 36 months of positive earnings. The rank-rank correlation is between 0.254 and 0.275, that is, the maximum expected difference in child earnings rankings that depends on parental ranking is between the 25th and 28th child earnings percentiles. For example, child earnings rankings that depends on parental ranking located in the 25th of the ranking distribution, implies, on average, an increase of between 19th and 25th child's earnings percentiles[12].

---

[12]We estimate this result by $r^p - r^c$ where, $r^p = 25$, $r^c_i = 37.397 + 0.254 * (r^p = 25)$ and $r^c_i = 43.368 + 0.275 * (r^p = 25)$, respectively

	(1)	(2)	(3)	(4)
r^p	0.278***	0.285***	0.293***	0.300***
	(0.002)	(0.002)	(0.003)	(0.004)
Constant	31.669***	33.234***	35.762***	38.362***
	(0.121)	(0.138)	(0.176)	(0.233)
Observations	222,397	178,916	116,182	68,644
R-squared	0.075	0.079	0.083	0.088

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.7: OLS estimates of the rank-rank correlation for our female children

Earnings are measured as the average earnings over the months in which an individual reports positive earnings over 5 years. We keep individuals that appear at least 6 times with positive earnings in the dataset with average earnings greater than half of the corresponding minimum wage. Columns (1) to (4) report results for female children. (1) considers individuals with at least 6 months of positive earnings, (2) considers individuals with at least 12 months of positive earnings, (3) considers individuals with at least 24 months of positive earnings and (4) considers individuals with at least 36 months of positive earnings.

	(1)	(2)	(3)	(4)
r^p	0.239***	0.247***	0.258***	0.264***
	(0.002)	(0.002)	(0.002)	(0.003)
Constant	41.682***	42.504***	44.118***	46.312***
	(0.103)	(0.114)	(0.139)	(0.175)
Observations	283,127	237,902	166,797	105,039
R-squared	0.060	0.064	0.070	0.076

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.8: OLS estimates of the rank-rank correlation for our male children

Earnings are measured as the average earnings over the months in which an individual reports positive earnings over 5 years. We keep individuals that appear at least 6 times with positive earnings in the dataset with average earnings greater than half of the corresponding minimum wage. Columns (1) to (4) report results for male children. (1) considers individuals with at least 6 months of positive earnings, (2) considers individuals with at least 12 months of positive earnings, (3) considers individuals with at least 24 months of positive earnings and (4) considers individuals with at least 36 months of positive earnings.

Table 1.7 show the rank-rank correlation estimates only for female children, and Table 1.8 show rank-rank correlation estimates for male children. Comparing female and male, results show that the rank-rank correlation is higher for female children. This indicates that for females, parental ranking is more persistent than for males. In addition, absolute mobility, measured as the constant of each regression, is higher for males than for females, which means that male children of poor parents are expected to locate in a higher ranking than female children of poor parents.

Quintiles transition matrices

These child and parental earnings rankings also allow us to estimate the quintile transition probabilities. These probabilities are defined by the conditional probability that a child is in quintile m (with $m = 1, 2, 3, 4, 5$) of the child earnings distribution given that her parent is in quintile n (with $n = 1, 2, 3, 4, 5$) of the parental earnings distribution.

In the intergenerational mobility literature, there are three probabilities that are broadly studied: i) the circle of poverty, defined by the probability that, given parents who belong to the bottom quintile, the child will also belong to the bottom quintile. We denote this probability as p_{11}; ii) the circle of privilege, defined by the probability that, given parents who belong to the top quintile, the child will belong to the top quintile. We denote this probability as p_{55}; and, iii) the rags to riches, defined by the probability that, given parents who belong to bottom quintile, the child will belong to the top quintile. We call this probability p_{15}. Notice that p_{11} and p_{55} are measures of intergenerational persistence that provide evidence on transmission of disadvantages and advantages, respectively; while p_{15} is a measure of upward intergenerational mobility.

Parental	Child quintile				
quintile	1	2	3	4	5
1	0.271	0.235	0.204	0.170	0.120
2	0.236	0.235	0.213	0.186	0.130
3	0.206	0.223	0.220	0.200	0.150
4	0.171	0.193	0.215	0.223	0.198
5	0.112	0.125	0.161	0.226	0.376

Table 1.9: Transition matrix of parental earnings quintiles to child earnings quintiles

Quintiles are measured using earnings and the baseline dataset. Rows refer to parental quintile and columns to child quintiles.

Table 1.9 shows the matrix of quintile transition probabilities using our baseline sample. As can be seen in Table 1.9, p_{11} is equal to 0.271 meaning that a child whose parents belong to the bottom quintile has an observed probability of 27.1 percent of remaining in the bottom quintile; p_{55} is equal to 0.376, which means that a child whose parents belong to the top quintile has a probability equal to 37.6 percent of remaining in the top earnings quintile; and p_{15} is equal to 0.120 which means that the probability that a child whose parents belong to the bottom quintile will herself belong to the top quintile is 12 percent.

Our results suggest that there is some persistence of parental earnings because p_{55} and p_{11} are higher than 0.2, which is the value of a transition probability, assuming that parental-child transitions are random. We also find that $p_{55} > p_{11}$ meaning that persistence is higher at the top of the distribution than at the bottom. Notice that the transition probabilities of the first 4 quintiles are relatively similar and close to random transitions; however, our results reveal that the main departure from randomness occurs at the top quintile where there is a notorious intergenerational earnings persistence.

International comparison with the US and Canada

To put our analysis in perspective, we can compare Figure 1.1 with findings for the US and Canada. As reference, we use the results in Chetty, Hendren, Kline and Saez (2014) for the US, and the findings in Corak (2019) for Canada. Notice that, whereas for Chile we use earnings information, the works of Corak (2019) and Chetty, Hendren, Kline and Saez (2014) use income information.[13]
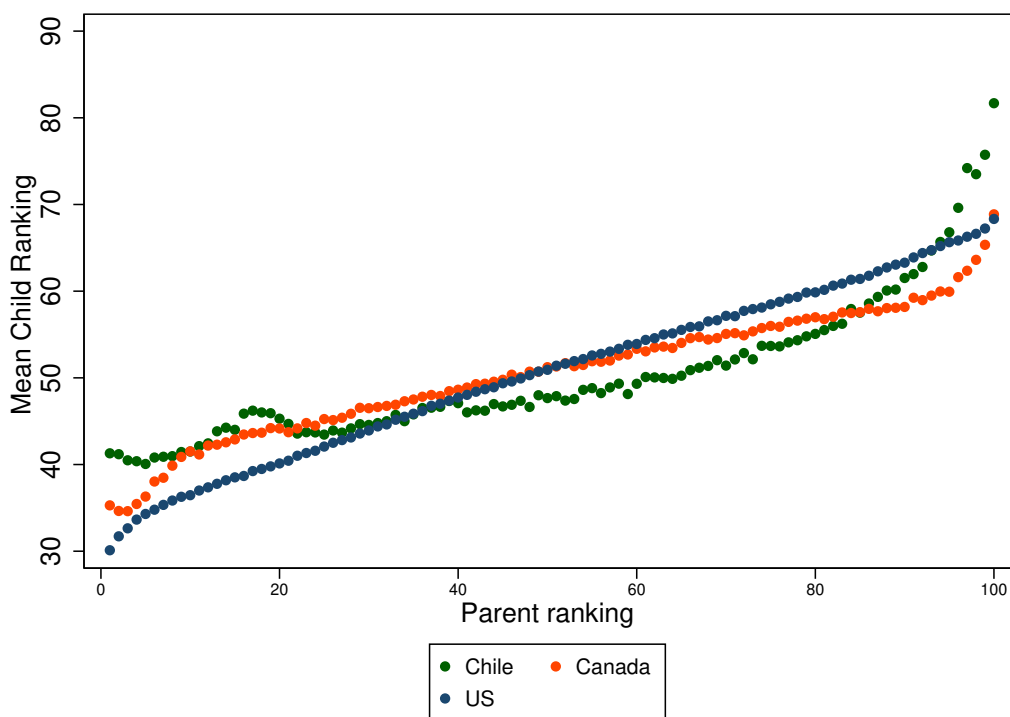
Figure 1.2: International comparison of expected child earnings ranking conditional to the parental earnings ranking

We estimate the expected child ranking non parametrically using a simple average. Rankings were computed over the national distribution. We compute the cohort ranking for children and for parents we compute the ranking of people between 42 and 87 years old (in 2018). Information for Canada is from Corak (2019) and for the US is from Chetty, Hendren, Kline and Saez (2014).

---

[13]Studies show that income is more persistent than earnings, especially at the bottom of the distribution. Thus, our results for Chile can be interpreted as a lower bound for persistence.

Figure 1.2 shows that Chile has a flatter gradient until the 80 percent in parental income/earnings. This evidence suggests that Chile is more mobile than Canada and the US in parental income/earnings until the 80th percentile. Remarkably, after the 80th parental percentile, Figure 1.2 also shows that the relationship between parental and child earnings in Chile becomes much steeper than those in the US and Canada. This graphical analysis suggests that intergenerational earnings mobility for Chile is much more non linear than the results found by the US and Canada.

1.4.2 More on non-linearities

The previous graphical analysis suggests that the relationship between parental and child earnings in Chile is highly non-linear, even more so than in the US and Canada, with the particularity of displaying significant intergenerational mobility until the 80th parental earnings quintile but a notorious degree of persistence of privileges (transmission of advantages from parent to child) at the top of the earnings distribution.

To better understand this finding, we perform two empirical exercises. First, we show the estimates of the transition probabilities for the top decile and percentiles. Second, we estimate the conditional distribution of child earnings given a parental decile (percentile), for different parental deciles (percentiles).

Decile and percentile intergenerational transition matrices

We now present decile transition probabilities. These estimates allow us to gain deeper understanding on how the child earnings distribution behaves within quintiles —especially for children with parents in the top quintile. Table 1.10 shows the matrix of decile transition probabilities.

| Parental | Child deciles | | | | | | | | | |
deciles	1	2	3	4	5	6	7	8	9	10
1	0.158	0.136	0.125	0.114	0.104	0.098	0.086	0.073	0.065	0.042
2	0.125	0.123	0.118	0.114	0.107	0.099	0.094	0.085	0.075	0.059
3	0.122	0.126	0.124	0.114	0.111	0.102	0.095	0.085	0.072	0.050
4	0.109	0.115	0.118	0.114	0.110	0.104	0.101	0.091	0.079	0.060
5	0.106	0.107	0.112	0.117	0.112	0.108	0.102	0.093	0.083	0.061
6	0.096	0.104	0.108	0.109	0.110	0.111	0.106	0.099	0.091	0.066
7	0.086	0.093	0.100	0.104	0.109	0.112	0.110	0.107	0.100	0.080
8	0.078	0.084	0.087	0.095	0.102	0.108	0.114	0.115	0.117	0.100
9	0.067	0.069	0.073	0.080	0.091	0.099	0.110	0.127	0.140	0.143
10	0.044	0.042	0.044	0.051	0.059	0.071	0.092	0.122	0.173	0.301

Table 1.10: Decile Transition Matrix

As can be seen in Table 1.10, the transition matrix —excluding the row with the 10% richest parents— shows a somewhat intergenerationally-mobile context, with all the transition probabilities roughly close to 10%, as we would expect under random transition from parent to child. However, given the parental earnings top decile, we notice that the dynamic of the transition probabilities is significantly different. For instance, the probability of persistence in privilege p_{1010} is equal to 0.3. In contrast, the probability of persistence in poverty p_{11} is close to a half of p_{1010}, suggesting that the transmission of advantages (circle of privilege) is twice as persistant as the transmission of disadvantages (circle of poverty).

We now study p_{1010} in depth by showing the probabilities associated with transitions from parental percentiles to child percentiles, for percentiles from 91 to 100. Table 1.11 summarizes this information.

| Parental	Child percentiles									
percentiles	91	92	93	94	95	96	97	98	99	100
91	0.019	0.018	0.017	0.019	0.020	0.019	0.017	0.019	0.019	0.023
92	0.020	0.017	0.019	0.020	0.019	0.020	0.018	0.018	0.022	0.021
93	0.019	0.021	0.020	0.019	0.024	0.019	0.025	0.023	0.021	0.023
94	0.021	0.022	0.021	0.022	0.028	0.025	0.022	0.021	0.024	0.025
95	0.024	0.022	0.018	0.027	0.025	0.026	0.029	0.031	0.028	0.027
96	0.026	0.025	0.026	0.026	0.028	0.026	0.032	0.034	0.033	0.039
97	0.024	0.025	0.032	0.033	0.035	0.038	0.042	0.043	0.050	0.056
98	0.024	0.023	0.026	0.029	0.035	0.038	0.042	0.042	0.045	0.057
99	0.026	0.029	0.030	0.029	0.032	0.043	0.039	0.047	0.055	0.066
100	0.027	0.035	0.035	0.029	0.045	0.051	0.056	0.060	0.068	0.105

Table 1.11: 91st to 100th parental percentile to 91st to 100th child percentile transition matrix

As can be seen in Table 1.11, the transition probabilities for children whose parents belong to the 91st to 95th percentiles of the parental earnings distribution are relatively similar, while the probability of persistence at the top percentile, $p_{100,100}$, is significantly higher compared to the rest of transition probabilities presented in Table 1.11. This means that the top percentile is even more persistent than the rest of the 10th decile. In sum, this analysis provides evidence supporting a high persistence at the top, which increases as long as parental earnings increase.

Conditional distribution of child earnings, given parental deciles

Another way to understand the association between child and parental earnings is by estimating the conditional distribution of child earnings, given parental earnings $f(y^c|y^p)$. Thus, instead of just observing a change in the mean, we can study variations in the entire distribution. To do this, we perform kernel estimations of

the conditional distribution of child earnings, given parental deciles.

Conditional log Childrens' earnings distribution on parental earnings
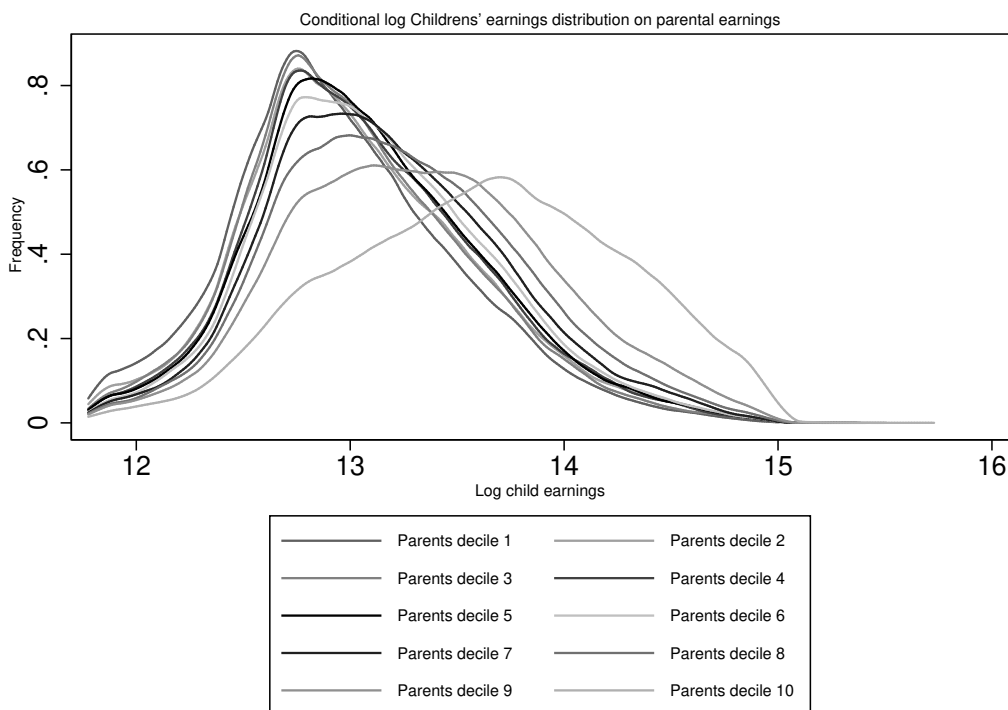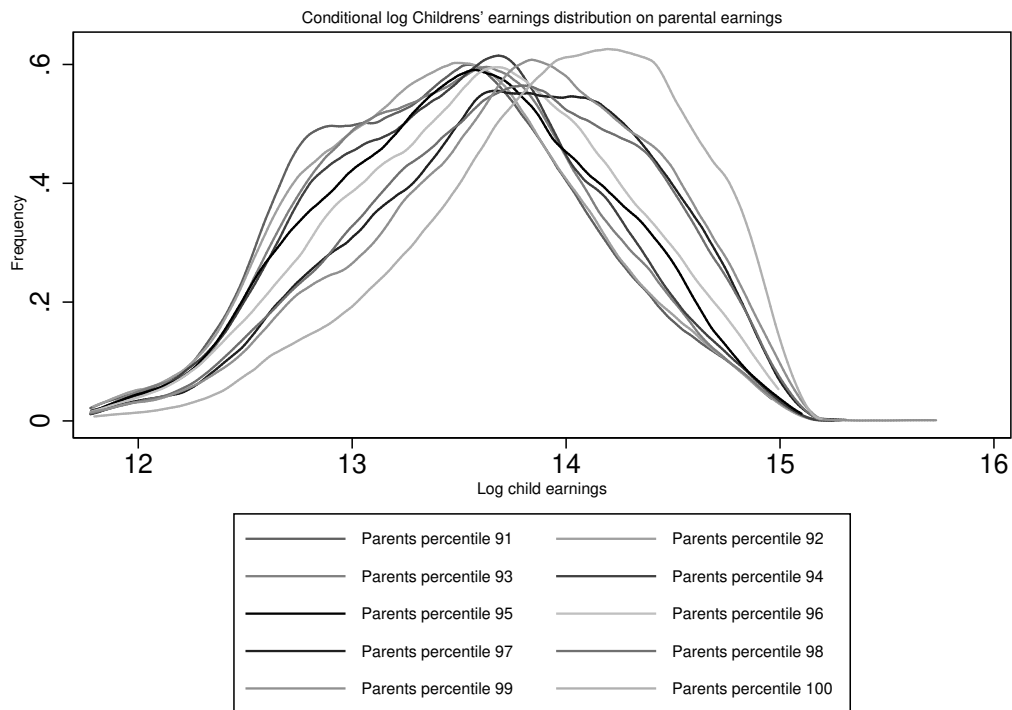
Figure 1.3: Conditional (on parental deciles) child earnings distribution

This Figure estimates conditional (on parental deciles) child earnings distribution, using kernel to estimate child earnings distribution. We use the Epanechnikov method to estimate optimal bandwith.

Figure 1.3 shows the conditional distribution of the logarithm of child earnings given that parents belong to a particular earnings decile, for earnings deciles from 1 to 10. As can be seen in Figure 1.3, roughly speaking, the conditional distributions of child earnings are unchanged between parental decile 1 and 7. After decile 8, it tends to move. Indeed, conditional on parents belonging to the top decile, the conditional distribution of log child earnings is significantly shifted to the right. This evidence is consistent with our previous findings of transmission of privileges, since it suggests that it is more likely for children whose parents belong to the top earnings decile to

obtain higher earnings. As can be also seen in Figure 1.3, the conditional distribution of log child earnings for top parental earnings has a higher variance than conditional on lower parental earnings. In sum, this analysis supports the idea that for children in the bottom and middle part of the earnings distribution, parental earnings do not affect their own distribution of earnings; however, child earnings located at the top of their distribution are dramatically affected by parental earnings.

Figure 1.4: Conditional (on parental percentiles in the top decile) child earnings distribution

This Figure estimates conditional (on parental percentiles in the top decile) child earnings distribution, using kernel to estimate child earnings distribution. We use the Epanechnikov method to estimate optimal bandwith.

Figure 1.4 presents the estimation of the conditional distribution of log child earnings, given parents that belong to a specific percentile, for percentiles from 91 to

100. As can be seen in Figure 1.4, while the conditional distribution of child earnings is quite similar for those with parents in percentiles 91 to 99, it is starkly different when we condition by parents belonging to the top 1 percent. This evidence supports our finding that the relationship between parental and child earnings is highly non-linear, even at the top parental distribution where this relationship becomes significantly more positive.[14]

1.4.3 Robustness checks

We now evaluate the robustness of our estimates of intergenerational mobility to alternative subsamples and specifications. We begin by evaluating three potential sources of bias: coverage of the dataset in initial years of the UIP, lifecycle bias, and attenuation bias.

Dataset coverage

As can be seen in Table 1.1, coverage of the unemployment insurance dataset in its first two years is less than 50% of total formal workers. To see whether this low coverage rate affects our baseline mobility estimates, we perform new estimates by considering different windows of years to measure permanent parental earnings.

---

[14]This result is in line with Zimmerman's (2019) findings. Zimmerman (2019) shows that studying in an elite college only increases the probability of belonging to the top managerial positions (obtaining higher earnings) if the student attends a top private high school, and he also shows that it is more likely that parents that belong to the top 1 percent can afford the tuition costs of private schools. Thus, Zimmerman's findings are one component of this persistence at the top where the transmission of privileges from parent to child would be through paying the tuition costs for attending a top private high school.

Parental year used	IGE	Rank-rank slope	N
2003-2007	0.288	0.254	505,524
2004-2008	0.288	0.256	550,668
2005-2009	0.287	0.260	584,770
2006-2010	0.284	0.263	607,545
2007-2011	0.283	0.268	622,339
2008-2012	0.281	0.270	632,820
2009-2013	0.280	0.272	636,640
2010-2014	0.278	0.272	638,481
2011-2015	0.280	0.275	637,808

Table 1.12: Estimations of IGE and rank-rank slope for different years where parental earnings were measured.

Table 1.12 presents IGE and rank-rank slope estimates for different windows of years to build our measure of permanent parental earnings. We can see that IGE and rank-rank slope estimates do not depend on the choice of the window of years. Specifically, IGE estimates ranges between 0.278 and 0.288, whereas the rank-rank slope is between 0.254 and 0.275.

Lifecycle bias

Prior research has shown that measuring children's income at early ages can understate intergenerational persistence in lifetime income because children with high lifetime incomes have steeper earnings profiles when they are young (Haider and Solon, 2006, Grawe, 2006, Solon 1999). To evaluate whether our baseline estimates suffer from such lifecycle bias, we can estimate the intergenerational earnings elasticity by single child cohorts. To do this, we study the effects of parental earnings on child earnings when children are 23 to 33 years old. To be consistent with the literature (Chetty, Hendren, Kline and Saez (2014); Corak, 2019), we measure the

effect of parental earnings when their children were teenagers.

Child age	IGE	Rank-rank	N
23	0.042	0.053	72,863
24	0.095	0.102	81,765
25	0.151	0.153	86,767
26	0.193	0.185	90,241
27	0.220	0.215	93,866
28	0.245	0.230	96,693
29	0.259	0.241	94,492
30	0.285	0.256	89,286
31	0.305	0.269	81,261
32	0.321	0.275	75,010
33	0.333	0.276	68,231

Table 1.13: Estimates of IGE and rank-rank slope for different child ages.

Table 1.13 shows the estimates of IGE and rank-rank slope by single child co-horts. We can see that intergenerational persistence rises as child age increases. This is consistent with Chetty, Hendren, Kline and Saez (2014). In particular, IGE is more affected by child cohorts than the rank-rank correlation, a fact that has been discussed previously in the intergenerational mobility literature.

Attenuation bias

Earnings in a single year is a noisy measure of lifetime earnings, which attenuates estimates of intergenerational persistence (Solon, 1992). To evaluate whether our baseline estimates suffer from such attenuation bias, we provide the estimates of the rank-rank slope, varying the number of years used to build our measure of permanent parental earnings.

Parental years used	IGE	Rank-rank slope	N
1	0.258	0.220	156,760
2	0.272	0.235	273,673
3	0.277	0.241	363,805
4	0.284	0.248	438,302
5	0.288	0.254	505,524
6	0.291	0.258	559,666
7	0.293	0.263	603,481
8	0.293	0.267	642,176
9	0.294	0.272	676,494
10	0.294	0.275	708,541

Table 1.14: Estimates of IGE and rank-rank slope using different years to average parental earnings.

Table 1.14 presents the estimates of the IGE and rank-rank correlations by using different numbers of years to create the permanent parental earnings. As can be seen in Table 1.14, IGE remains somewhat stable after averaging 4 years (from 0.288 to 0.294), whereas the rank-rank slope varies slightly between 0.254 and 0.275 over 4 years.

1.5 Geographic variation in intergenerational mobility: the case of Chilean regions

The previous sections suggest that the relationship between parental and child earnings varies non-linearly with parental earnings, especially with parents at the top of the earnings distribution.

Another source of variation of the relationship between parental and child earnings that has been studied in the recent literature is geographical location. The

literature has found remarkable differences in intergenerational mobility across geographies within a country. For example, Connolly et al. (2018) find that commodity booms may be important drivers of intergenerational upward mobility.[15]. In addition, Deutscher and Mazumder (2020) finds the same result for Australia. Thus, a boom of the copper price can impact directly wages and the labor market in geographies that are intensive in copper production. This finding is important for Chile because it is the main copper producer in the world by a large margin, with approximately 28% of the total world production in 2018.

1.5.1 Chilean regional context

Chile is divided into 16 regions, the first-level administrative division of the country. Each region is designated by a number —from 1 to 16— and a name. Each region is divided into provinces, the second-level administrative division. In total, there are 56 provinces, each one divided into municipalities, the third and lowest-level administrative division.[16]

In Table 4.3 in Appendix, we present current information of each region. Among the 16 regions, the Metropolitan Region (the 13th region) stands out as the most populated region in the country (in number and density), with a population of over 7.5 million in 2017 (41% of Chile's population) according to the National Institute of Statistics of Chile (INE). Significantly, this region contains the capital of Chile, the city of Santiago, which has been recognized as one of the cities with the best quality of life in South America. Based on estimates of the Central Bank of Chile (BCCh) for 2018, the Metropolitan Region produces 46% of Chile's GDP, with manufacturing,

---

[15]Connolly et al. (2018) finds for Canada that commodity-producing provinces such as Alberta and Saskatchewan, and mid-west US states, present the highest upward mobility indicators.

[16]Until 2007, there were only 13 regions geographically located from north to south of the country with their numbers in geographically sequential order, except for the Metropolitan Region, also known as the 13th region, which is located roughly in the middle of the country, between the 5th and 6th regions. In the period 2007-2017, the 14th, 15th, and 16th regions were created after dividing into two areas the 10th, 1st, and 8th regions, respectively.

services, retail, and financial services as principal economic activities. According to official estimates by the Government of Chile for 2017, 5.4% of the population of this region lived in poverty in 2017 and this region contributes with 3% of Chile's GDP, with a Gini coefficient of 0.43.

The Antofagasta Region (the 2nd region), in the northern area of the country, stands out with a production of 10% of Chile's GDP, with the mining industry — led by copper— as its principal economic activity. In fact, according to estimates of the BCCh for 2018, mining output represents 54% of regional production. This region had a population of 623,851 inhabitants in 2017 (3% of Chile's population) according to INE. This region has the highest GDP per capita in the country —over USD 25,000—, 5.1% of its population live in poverty in 2017, and its Gini coefficient is 0.41.

On the other end of the income scale in Chile, we have La Araucanía Region (the 9th region), in the southern part of Chile, which is the country's poorest region in terms of GDP per capita, with USD 6,000 per inhabitant, on average. This region contributes with 3% of Chile's GDP, with 17.2% of its population living in poverty — the highest regional poverty rate in the country. It's worth noting that a third of the region's population of 994,888 (6% of Chile's population) is of indigenous Mapuche ethnicity, which represents the highest concentration of this community (or, indeed, of any other national indigenous peoples) of any Chilean region.

1.5.2   Intergenerational earnings mobility at the regional level

To characterize the variation in intergenerational mobility across geographic areas within Chile, we permanently assign each child to a single region. We use the child's residential address while attending 12th grade in school. We obtain this information from administrative records provided by the Chilean Ministry of Education. If the residential address of the child when attending 12th grade is not available, we instead use the child's most recent residence while she was enrolled during 7th-11th grade

in school.[17]

Measures of relative and absolute mobility

We measure mobility at the regional level using the baseline sample and the definitions of parental and child earnings described in Section 2. We continue to rank both children and parents on the basis of their positions in the national earnings distribution (rather than the distribution within their regions).

---

[17]The region where a child grew up does not necessarily correspond to the region she lives in as an adult at age 28-33 in 2018.
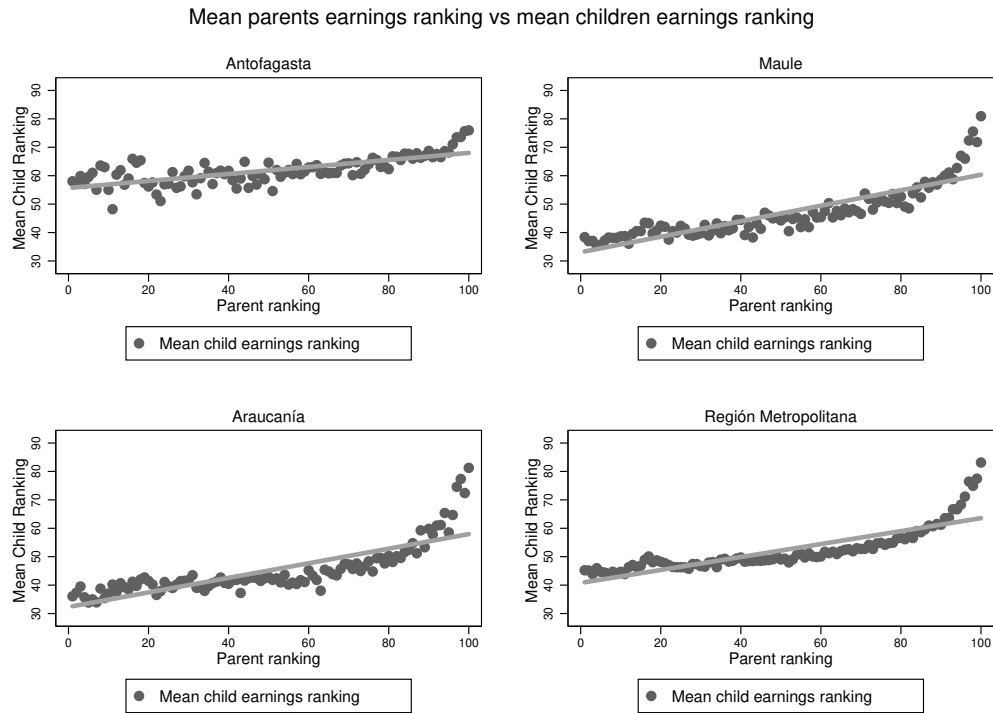
Figure 1.5: Expected child ranking conditional on parental national ranking for 4 different regions.

This Figure plots the expected child ranking conditional on parental national ranking for 4 different regions. We estimate the expected child ranking non-parametrically using a simple average. Rankings were computed over the national distribution. For children we compute the cohort ranking, and for parents we compute the ranking of people between 42 and 87 years old (in 2018).

Figure 1.5 presents a binned scatter plot of the mean child rank versus parent rank for children who grew up in the second region (Antofagasta), the seventh region (Maule), the ninth region (La Araucanía), and the Metropolitan region. As can be seen in Figure 1.5, in each region there is a linear relationship between the parental and child ranks for the bottom part of the parental earnings distribution. The higher levels of persistence at the top of the parental earnings distribution are a common characteristic of the four regions displayed in Figure 1.5. Despite this non-linearity at the top of the distribution, we rely on Chetty, Hendren, Kline and Saez (2014)

and Acciarri et al. (2020) to characterize the relationship between child rank given the parents' rank in each region using a simple linear regression. More formally, we regress child rank on parental rank by region to calculate absolute upward mobility and relative mobility by region. We define absolute upward mobility as

$$r_{abs_r} = \alpha_r + \beta_r E(r_p | r_p < 50), \tag{1.4}$$

where α_r and β_r are the intercept and the rank-rank regression slope estimated for region r, respectively. That is, the conditional expected child's position on the national earnings distribution given that her parental earnings are below the median of the national distribution. We approximate this value as $r_{abs_r} = \alpha_r + \beta_r \cdot 25$.[18] In addition, we define persistence as the conditional expectation of a child's percentile on the national earnings distribution given her parent belonging to the 10th decile. We measure this expression as $r_{per_r} = \alpha_r + \beta_r E(r_p | r_p > 90)$ and approximate it as $r_{per_r} = \alpha_r + \beta_r \cdot 95$.[19] We complement this analysis studying the three transition probabilities described in section 1.4. Specifically, we show transition probabilities p_{11} (circle of poverty), p_{15} (rags to riches), and p_{55} (circle of privilege).

---

[18]We also estimate the absolute upward mobility coefficient using a nonparametric estimation of $E(r_p | r_p < 50)$. Results remain unchanged.

[19]We also estimate $E(r_p | r_p > 90)$ nonparametrically. Results remain almost unchanged.

Region	N	β_r	α_r	r_{abs}	r_{per}	p_{15}	p_{11}	p_{55}
1	6,584	0.145	47.324	50.942	61.072	0.243	0.190	0.360
2	16,911	0.146	54.351	58.012	68.265	0.321	0.126	0.451
3	9,851	0.146	49.064	52.726	62.978	0.206	0.165	0.368
4	19,962	0.169	42.522	46.746	58.575	0.166	0.257	0.337
5	48,015	0.199	37.554	42.527	56.450	0.116	0.282	0.313
6	28,806	0.244	36.332	42.441	59.548	0.112	0.310	0.368
7	28,874	0.228	35.024	40.731	56.710	0.088	0.317	0.321
8	42,993	0.197	38.060	42.997	56.820	0.111	0.275	0.307
9	20,891	0.202	34.385	39.439	53.589	0.082	0.311	0.308
10	21,105	0.197	36.002	40.932	54.735	0.091	0.255	0.305
11	4,400	0.142	38.055	41.600	51.528	0.095	0.260	0.243
12	5,462	0.183	39.700	44.278	57.094	0.120	0.194	0.291
13	196,004	0.256	39.103	45.509	63.447	0.135	0.222	0.398
14	6,631	0.178	38.998	43.454	55.931	0.094	0.250	0.349
15	4,228	0.128	46.601	49.799	58.755	0.167	0.212	0.345
16	10,510	0.189	37.608	42.341	55.595	0.108	0.289	0.306

Table 1.15: Intergenerational mobility indicators for different Chilean regions.

As can be seen in Table 1.15, there is substantial heterogeneity across regions. For instance, the region with the highest absolute mobility is Antofagasta, where a child whose parents earn below the median national earnings level has an expected national ranking of 54.4; whereas, for La Araucanía, the same child can expect to place in the 34.4(th) percentile of the child earnings distribution. In the same way, for probability p_{11} we estimate 0.126 for Antofagasta and 0.311 for La Araucanía. In addition, we can notice something similar for the rags to riches probability. For Antofagasta, p_{15} is equal to 0.321 and for La Araucanía is equal to 0.082, thus a child who grew up in Antofagasta with a parent that belongs to the bottom quintile

is almost 4 times more likely to arrive to the top quintile than the same child who grew up in La Araucanía. Finally, persistence is also higher in Antofagasta than in La Araucanía: children with parents in the top earnings quintile are more likely to remain in the top quintile in Antofagasta than in La Araucanía.

Figure 1.6 and 1.7 present heat maps of absolute upward mobility and relative mobility for Chilean regions. We can see that the most upwardly-mobile regions are those located at the north of the country. In particular, Antofagasta is the most upwardly-mobile region. Regarding relative mobility, the least mobile region is the Metropolitan region.

Figures 1.8 and 1.9 present heat maps of circle of poverty p_{11} and circle of privilege p_{55} transition probabilities for Chilean regions. We can see that the regions most persistent in poverty are those located in the upper south area of the country, particularly El Maule and La Araucanía regions. In contrast, the most persistent regions in privileges are those located in the north and the Metropolitan region. Thus, we corroborate Conolly et al. (2018) results by providing evidence that Antofagasta, a commodity-intensive region, presents the highest upward mobility indicators.

Is there a Gatsby curve in Chile?

The Gatsby curve refers to the negative relationship between income inequality and intergenerational mobility. This relationship has been extensively explored by the literature (see for instance Corak, 2013). We use the geographical variation across regions in Chile to study the Gatsby curve.
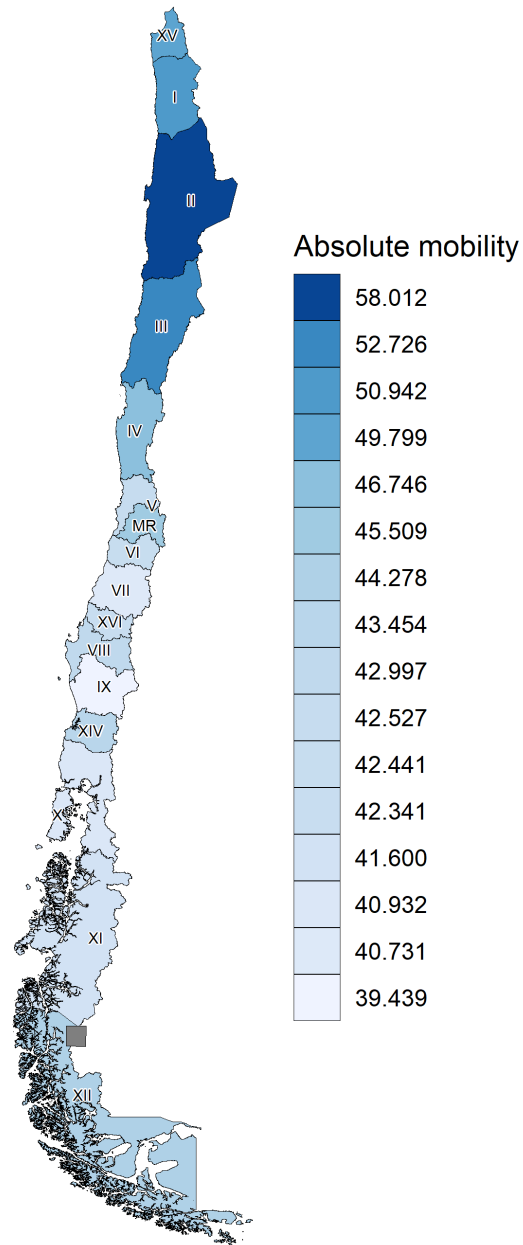
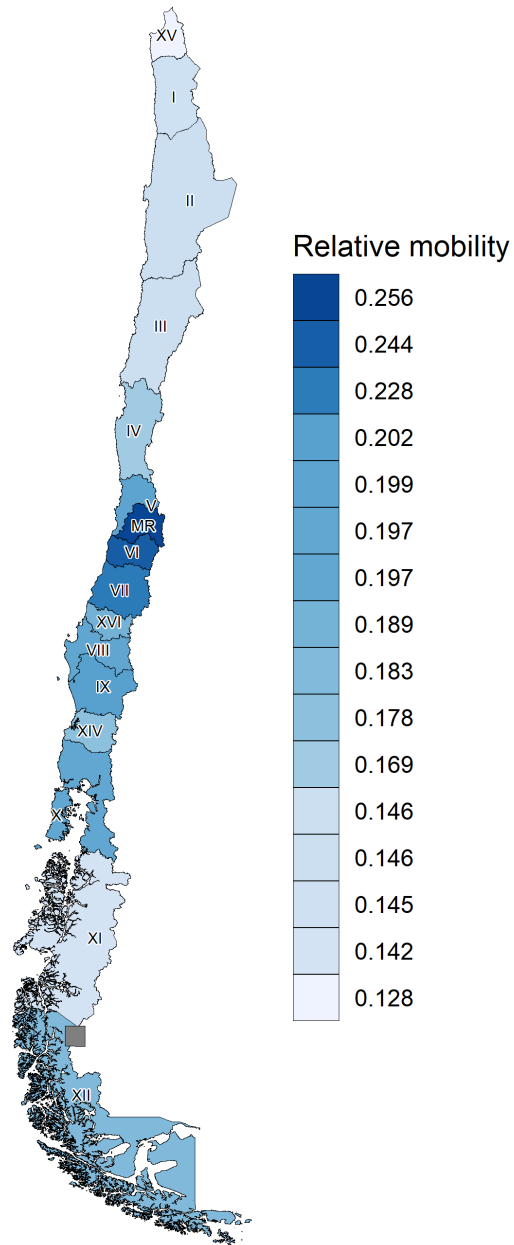**Figure 1.6:** Heat maps for absolute upward mobility in Chilean regions
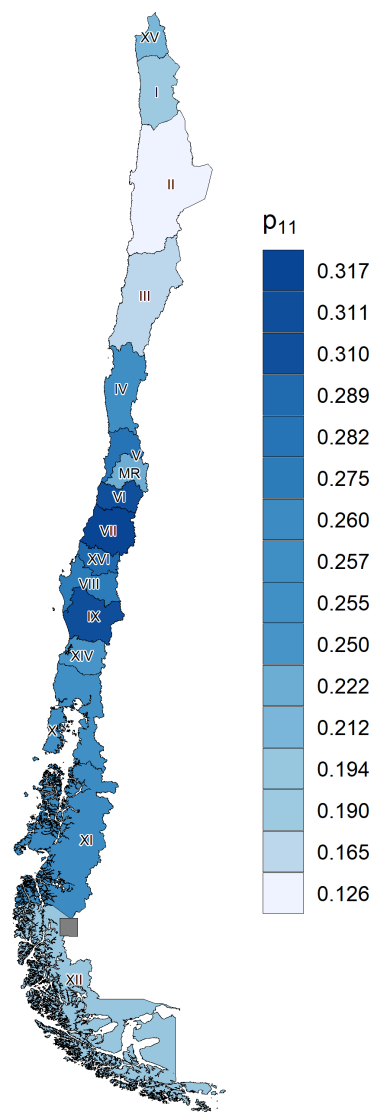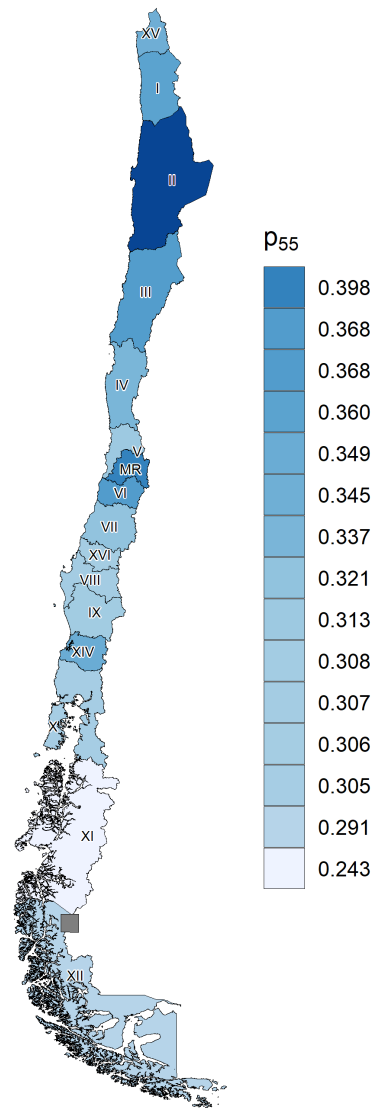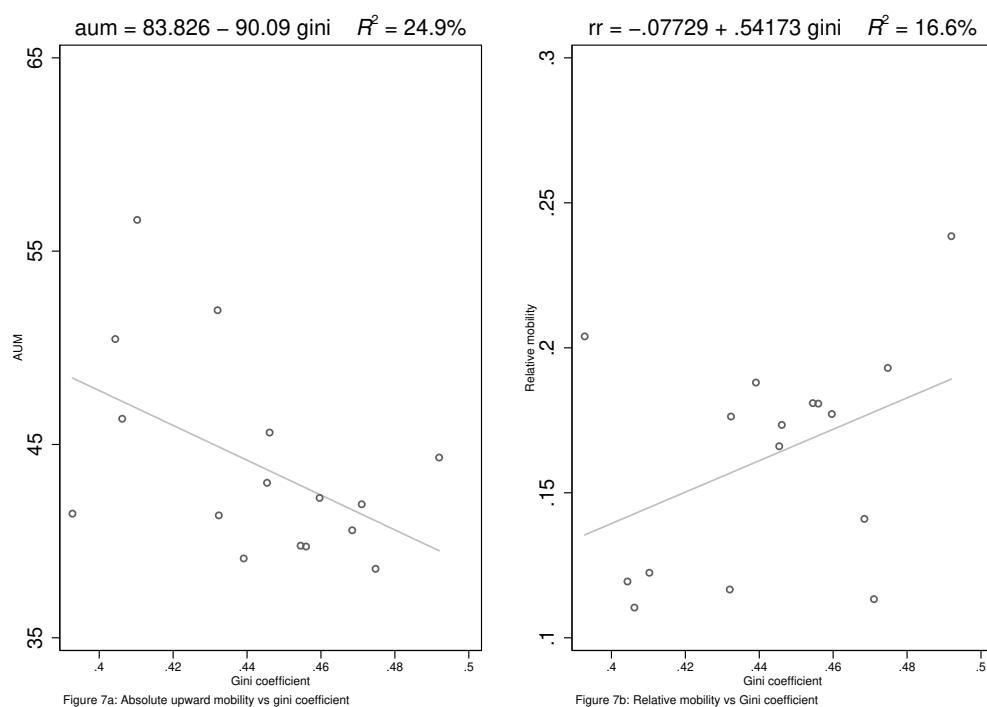
Figure 1.7: Heat maps for relative mobility in Chilean regions

Figure 1.8: Heat maps for circle of poverty p_{11} transition probability for Chilean regions.

Figure 1.9: Heat maps for circle of privilege p_{55} transition probability for Chilean regions.

Figure 1.10: Gatsby curve Chilean regions

This Figure plots the relationship between upward mobility and the Gini coefficient at the regional level. We measure the Gini coefficient using the 2017 CASEN survey, considering the total income before transfer and tax variant. Those results remain unchanged when we use other income definitions to measure the Gini coefficient.

Figure 1.10 left reports the relationship between absolute upward mobility and the Gini coefficient, while Figure 1.10 right reports the relationship between relative mobility and the Gini coefficient. As can be seen in these Figures, there is evidence of a Gatsby curve, where more unequal regions experience less intergenerational earnings mobility. This evidence suggests the existence of a vicious circle between intergenerational mobility and inequality.

1.6 Geographical variation in intergenerational mobility within the Metropolitan region

We now study the intergenerational mobility across municipalities, which are the least aggregated geographic units in Chile. We do this analysis inside the Metropolitan Region of Santiago —the finance and government center of Chile. It contributes with 40% percent of Chile's GDP, contains the capital of Santiago (the largest city in the country), and is the most densely populated region in the country, with close to 40 percent of the total population. This allows us to estimate intergenerational mobility at municipality level.

1.6.1 The Metropolitan Region

Although the Metropolitan Region of Santiago shows obvious signs of modernization, especially in the city of Santiago —which exhibits modern buildings and highways, a subway system, malls, and an extensive telecommunications network—, there are also elements that make it a residentially-segregated region, reflecting the economic inequality that characterizes the Chilean economy. Residential segregation in Santiago has its origin in several urban planning policies dating from the 1950s that tended to create residential areas for the lower classes (social housing) on the urban periphery of the city. This residential segregation intensified because of the implementation of slum eradication policies under the military dictatorship during the 1980s, where inhabitants of slum neighborhoods were relocated to social housing constructed on the periphery of the city. This policy of building social housing on the periphery continued after the return to democracy, as the proportion of social housing units in peripheral municipalities was continuously increasing and no new social housing was constructed in the upper-class municipalities.
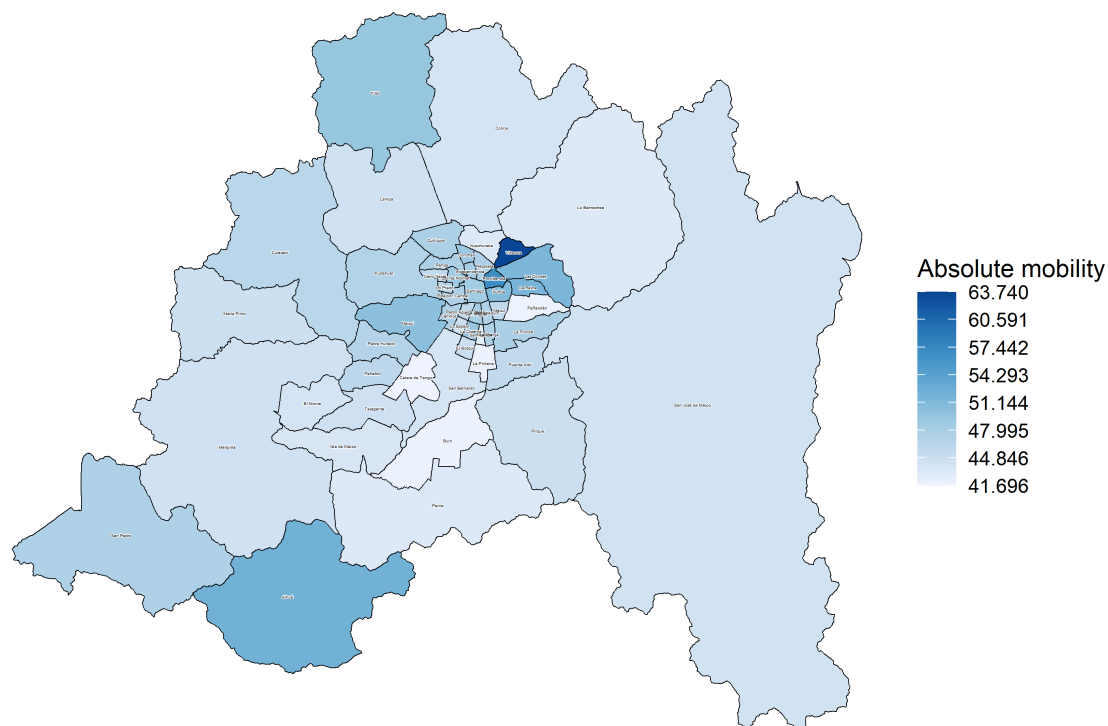
Figure 1.11: Heat maps for absolute upward mobility indicators for Metropolitan region municipalities.

1.6.2 Estimates of intergenerational mobility

We estimate the same measures of intergenerational mobility as for the regional case. Table 4.4 in Appendix summarizes these mobility measures by municipality.

Figures 1.11 to 1.14 present color maps for intergenerational earnings mobility on the metropolitan region. There is a remarkable heterogeneity across municipalities. For poor municipalities such as Cerro Navia, La Pintana and San Ramón, absolute upward mobility is not lower than 42, which means that children whose parents belong to the bottom 50 percent of the earnings distribution, are expected to locate at least in the 42th percentile of the children earnings distribution. In addition, persistence at the bottom and at the top probabilities are not to far from 0.2 which means
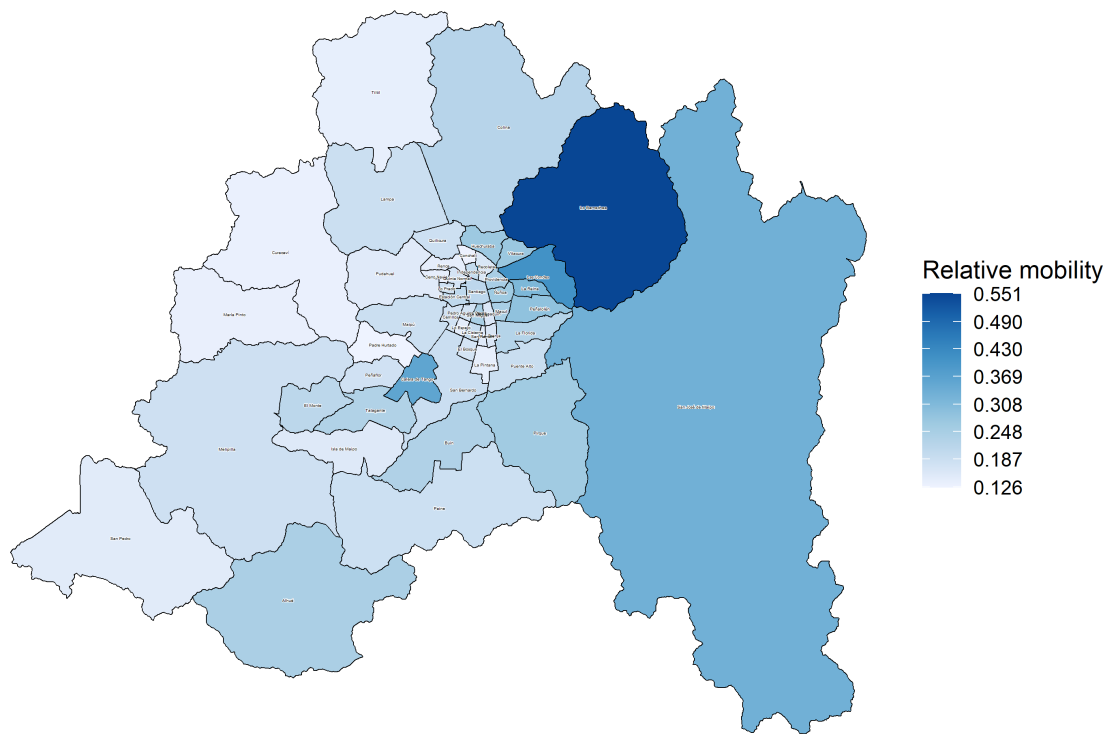
Figure 1.12: Heat maps for relative mobility for Metropolitan region municipalities.
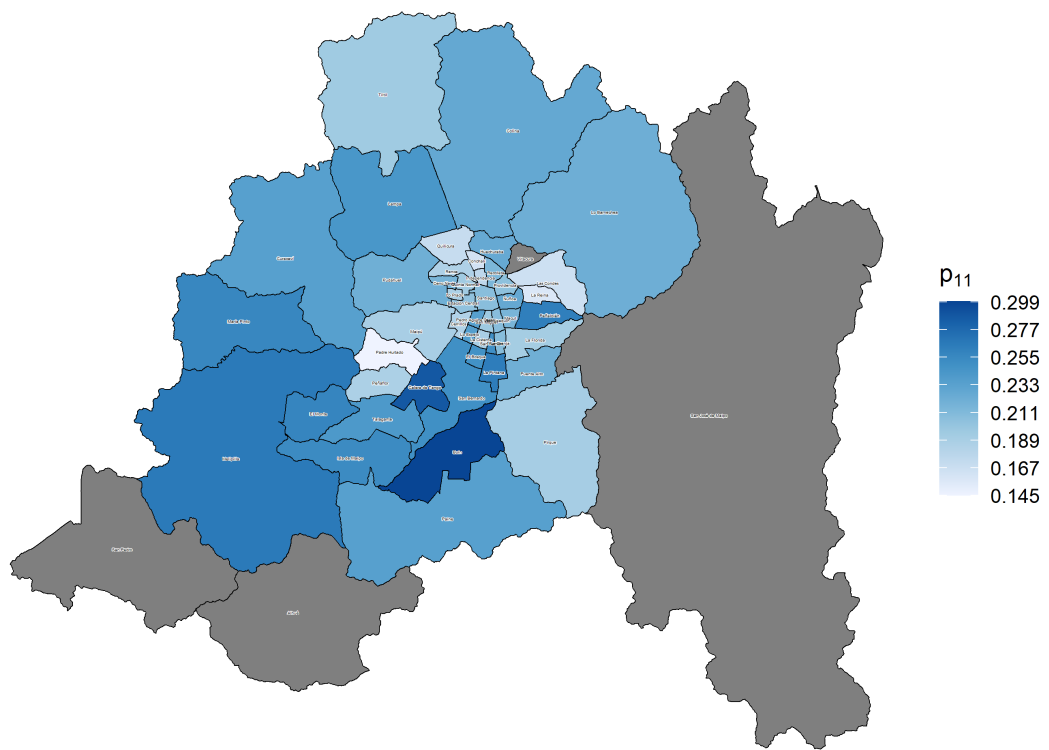
Figure 1.13: Heat maps for circle of poverty p_{11} transition probabilities for Metropolitan region municipalities.

Figure 1.14: Heat maps for circle of privilege p_{55} transition probabilities for Metropolitan region municipalities.

that there are not markedly persistence. However, the rags to riches probability is lower than 0.1.

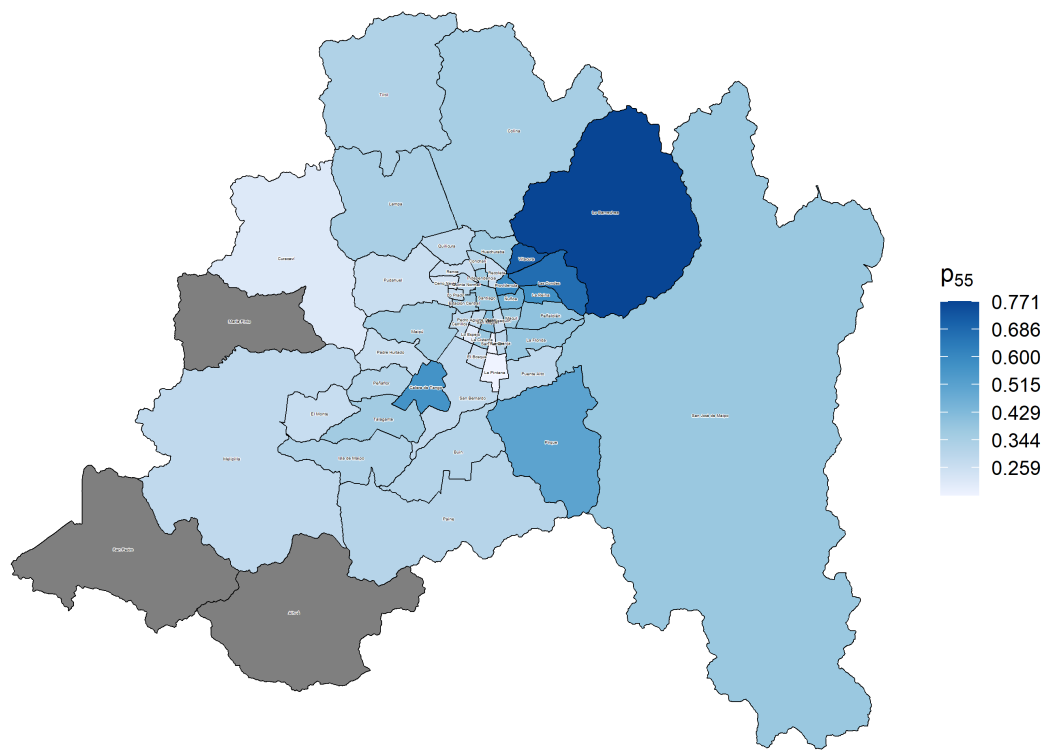On the other hand, almost all the rich municipalities in the northeast of the city, such as Las Condes, Vitacura, and Lo Barnechea, are the most persistent municipalities at the top, with probabilities of persistence of privileges, the conditional probability that a child is in the fifth quintile given that his parent is in the fifth quintile. Lo Barnechea (0.77), Las Condes (0.68) and Vitacura (0.72) have the highest circle of privilege probability of the Metropolitan region by far, the mean of which is 0.337. Thus, for a child with a parent that belongs to the highest quintile, it is highly likely that child will also be in the upper quintile.

But the differences in absolute upward mobility with more middle-class municipalities such as Ñuñoa, Santiago[20] or Maipu are relatively small. For instance, absolute upward mobility in Las Condes (51.49), is very close to Ñuñoa (51.02) and Maipu (50.29). Different is the case of Lo Barnechea, where upward mobility is very low compare to the other rich municipalities and is closer to La Pintana, which is a poor municipality. The major differences on persistence of privileges found between the rich municipalities and the rest indicates that the place of residence is an important factor to explain the high persistence at the top of the earnings distribution. One possible explanation for this finding is that social connection may play an important role on persistence of privileges.

The Gastby curve in the Metropolitan region

We can study the relationship between intergenerational mobility and inequality inside the metropolitan region.

---

[20]Santiago is the name of the city and also the name of a municipality —the latter is the statistic presented in this table. The municipality of Santiago is what inhabitants refer to as "downtown" and contains the presidential building La Moneda.
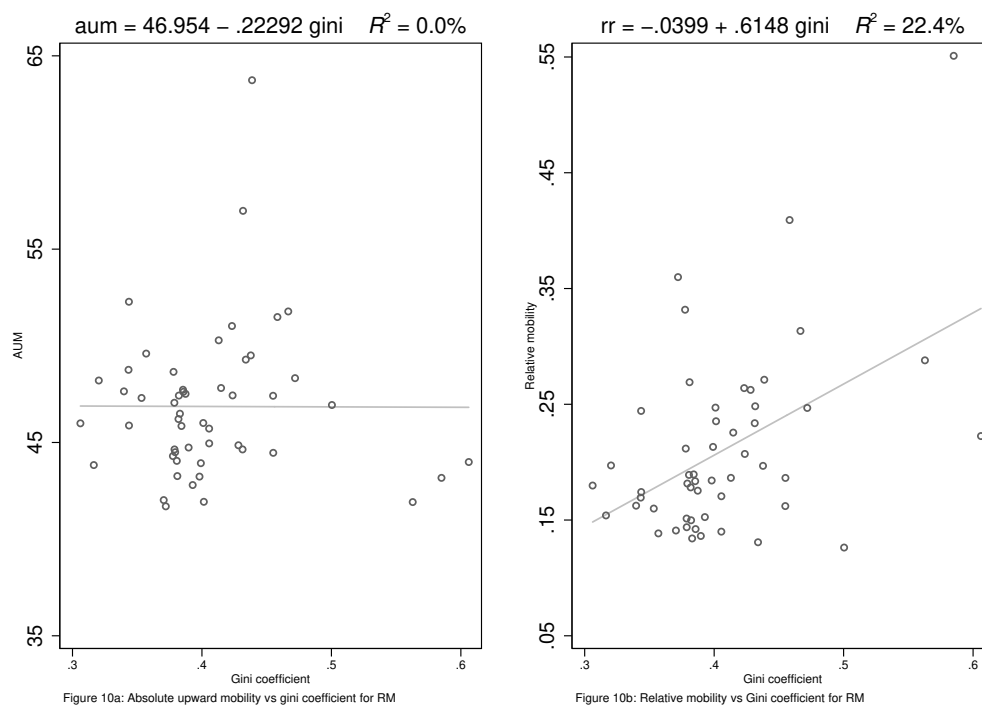
Figure 1.15: Gatsby curve Metropolitan Region municipalities

This Figure plots the relationship between upward mobility and the Gini coefficient at municipality level for the Metropolitan region. We measure the Gini coefficient using the 2017 CASEN survey, considering the "total income before transfer and tax" variant. Those results remain unchanged when we use other income definitions to measure the Gini coefficient.

Figure 1.15 shows the Gastby curve for the Metropolitan region. Comparing with Figure 1.12 can see that the intergenerational mobility and inequality relationship is more steeper for persistence than for upward mobility compared with regions. In particular, upward mobility does not strongly relate with inequality in the Metropolitan region. However persistence does strongly relate with inequality. This relationship is stronger than the regional relationship.

1.6.3   Geographic correlations and mobility across the Metropolitan region

In this section we present a correlations between mobility measures (relative mobility, upward mobility, persistence and probabilities of transition matrix) and local characteristics across Chilean municipalities in the Metropolitan region. We do not expect these correlations to be interpreted as causal relations, but we present them with the intention of being a guide in the research on the determinants of intergenerational mobility. A similar analysis has been performed by Chetty, Hendren, Kline and Saez (2014) for the U.S. and by Güell et al. (2018) and Accarci et al. (2020) for Italy.

To study the relationship between mobility and municipal socioeconomic characteristics, we start from a large set of correlates based on the literature. We use as i) measures of inequality: Gini coefficient and the share of the top 1 percent. We use as ii) demographic characteristics: proportion of immigrants, monoparental households and the proportion of people of indigenous ethnicity. We also include as iii) municipal amenities: municipal per capita expenditure and per capita square meters of green areas. Finally we include as iv) socio-economic characteristics: proportion of people with more than 18 years of schooling, proportion of students in publicly-funded schools, proportion of people with public health plans, proportion of overcrowded households, and poverty.

Indicator	β_r	r_{abs_r}	r_{per_r}	p_{11}	p_{55}	p_{15}
β_r	1.000					
r_{abs_m}	-0.106	1.000				
r_{per_m}	0.876	0.387	1.000			
p_{11}	0.103	-0.696	-0.242	1.000		
p_{55}	0.906	0.221	0.947	-0.128	1.000	
p_{15}	0.251	0.820	0.630	-0.468	0.515	1.000
Gini	0.503	-0.107	0.415	-0.045	0.436	-0.022
Share top 1 percent	0.373	-0.029	0.332	0.018	0.356	0.055
% Immigrants	0.161	0.369	0.328	-0.324	0.249	0.361
% Monoparental households	-0.288	-0.459	-0.489	0.100	-0.423	-0.538
% Indigenous	-0.601	-0.316	-0.710	0.181	-0.765	-0.512
Per capita expenditures	0.721	0.397	0.861	-0.152	0.819	0.533
Per capita M2 of green areas	0.477	-0.021	0.432	-0.083	0.375	0.076
% People with more than 18 years of schooling	0.722	0.413	0.870	-0.142	0.825	0.579
% Students in public schools	-0.282	0.030	-0.247	0.099	-0.108	-0.122
% Public health plan	-0.330	-0.797	-0.692	0.554	-0.529	-0.771
Overcrowding	-0.337	-0.437	-0.524	0.255	-0.493	-0.508
Poverty	-0.509	-0.383	-0.658	0.215	-0.636	-0.452

Table 1.16: Correlation between mobility measures and socio-economic characteristics

To measure these socioeconomic indicators we use information from the CASEN survey and the "Registro Social de Hogares" dataset.

Table 1.16 sheds lights on the relationship between inequality measures and the indices. The correlations with the Gini coefficient are positive and strong with persistence measures and relative mobility but weak and negative with upward mobility as we show in Figure 1.15. However, an alternative measure of inequality like the

share of the top 1 percent presents correlations of the same sign, but weaker than the gini index. This result it is line with evidence found by Chetty, Hendren, Kline and Saez (2014) concluding that intergenerational mobility is primarily correlated with inequality among the bottom 99% and not the extreme upper tail inequality of the form that has increased dramatically in recent decades, adding that upward mobility is strongly positively correlated with the size of the middle class.

The proportion of immigrants is positively correlated with persistence at the top and upward mobility but is negatively related to persistence at the bottom. One possible explanation is that the immigrant population reaches similar positions in time with respect to the native population, at least in the municipalities where the immigrant population decides to reside.

The proportion of monoparental households correlates negatively with upward mobility, persistence measure and persistence at the top, but positively with persistence at the bottom, i.e., children raised by a single parent may have worse outcome than these raised by two parents, indicating that the stability of the social environment may affects children's outcome more broadly.

The proportion of indigenous populations correlates negatively with persistence at the top and upward mobility but correlates positively (albeit weakly) with persistence at the bottom. This correlation could be driven by two very different channels as proposed by Chetty, Hendren, Kline and Saez (2014): One possibility is that indigenous children may have lower incomes than non-indigenous children conditional on parent income, so areas with a larger indigenous population may have lower upward mobility. Another possibility is that areas with large indigenous populations might have lower rates of upward mobility for children of all races. Unfortunately, we do not observe each individual's ethnicity in our data to distinguish both channels, but it would be interesting to study this result in depth, especially in view of our results for La Araucanía region, where both results are present: higher proportion of indigenous population and lower upward mobility.

Municipal per capita expenditure correlate positively with persistence at the top, upward mobility, and correlates negatively with persistence at the bottom. We corroborate what was found in section 1.6.2 related to the richest/poorest municipalities which are highly correlated with the income received by each municipalities.

One of the strongest correlations is the proportion of people with more than 18 years of schooling. Municipalities with more educated people tend to see more mobility, and tend to be more persistent at the top and less so at the bottom. Another high correlation is between proportion of people with public health plan and intergenerational mobility indicators. This means that the type of health that a child can benefit is a main variable that can explain intergenerational mobility in Chile. Finally, Overcrowding and Poverty, correlate negatively with upward mobility and persistence at the top but positively with persistence at the bottom.

We also notice a weak correlation between absolute upward mobility and relative mobility (-0.106). This is explained by the fact that the variance in relative mobility is higher compared to the variance in absolute upward mobility across municipalities. This means that there is more variance in persistence at the top than upward mobility. For the Metropolitan region, this finding supports the claim that, in terms of intergenerational upward mobility, where to live matters more for children from richer families than for children from middle- and lower-earnings families.

1.7 Conclusion

This is the first work that studies intergenerational mobility in Chile using administrative records. We build a data set that links parental and child earnings using information from the formal labor sector and the place of residence of children during their adolescence. Our analysis reveals that intergenerational mobility at the national level is significantly lower than what was estimated in previous research. However, intergenerational mobility is extremely non-linear. We found that mobil-

ity is very high for the bottom 80 percent of the earnings distribution but is very persistent at the upper tail of the parental and child distributions.

In addition, Chile is a highly heterogeneous country in its intergenerational mobility measures at the regional level. For instance, Antofagasta, which is a mining region, has a probability of rags to riches higher than 0.3. This result is in line with what Conolly et al. (2018) founds for the US and Canada. Meanwhile, regions like La Araucanía or El Maule have a circle of poverty probability higher than 0.3. It is worth digging a little deeper in future research to understand why those regions are so persistent in poverty.

We also find heterogeneity within the Metropolitan region, with municipalities having a circle of privilege probability higher than 0.7, and other municipalities with a circle of poverty probability closer to 0.3. We also learn that the variance of persistence at the top is higher than the variance of upward mobility. This means that the place of residence affects children of upper-earnings parents more than middle- or poor-class parents. Future research should focus on understanding the causes behind these differences. Although our work is descriptive in nature, it sheds lights on intergenerational mobility in a highly unequal country that does not belong to the advanced economies.

Moreover, we make a some methodological contributions. We use Kernel conditional densities to study intergenerational mobility at the top. Those tools help usus to show that intergenerational mobility is very persistent at the top in Chile. In addition, we differentiate the Gatsby curve for Chile and Santiago using two measures of intergenerational mobility: absolute intergenerational mobility and relative intergenerational mobility. We show that the Gatsby curve is valid for persistence and upward mobility for Chile but only for persistence for Metropolitan Region. This help us to differentiate different mechanisms that may affect intergenerational mobility for Chile.

This work builds on previous national literature and brings the state of research

up to the robustness of analysis seen among works in developed economies. As such, not only does it provide more useful information for academics; it also provides an important counterpoint to similar works from developed economies by analyzing intergenerational earnings mobility in a non-developed [o developing] economy in a way that can be contrasted with the results of that literature. We believe that, by providing a clearer picture of how intergenerational earnings mobility occurs in Chile at a regional level, this work can both inspire further research on the matter both in Chile and other developing economies. These results can also help Chilean authorities better understand how and where to apply certain related social/economic programs in order to improve their impact, as well as provide input for drawing up and discussing proposed bills affected by this study's results.

Chapter 2

How Much Should We Trust TSTSLS Estimates? Evidence from Intergenerational Mobility in Chile

2.1 Introduction

Researchers have had a growing interest in intergenerational mobility, which is often approached by measuring the relationship between children's and their parents' earnings (Black et al., 2011; Blanden, 2013; Emran and Shilpi, 2019). This relationship is a proxy for the degree of intergenerational transmissions of privileges and disadvantages from a parent to her child in society, and its understanding can help us design better social policies that favor equal access to social and economic opportunities.

The relationship between children's and their parents' earnings is commonly estimated via the intergenerational earnings elasticity (IGE) or the rank-rank correlation. The former quantifies the effect of a 1% increase in parental earnings on child

earnings, while the latter measures the impact of a 1 percentile increase in parental earnings rank on child earnings rank. On the basis of data linking children's and parent's earnings, the IGE and the rank-rank correlation can be estimated with basic ordinary least square (OLS) regressions. Previous works that estimate both measures by using administrative records for parents' and their children's earnings are Chetty, Hendren, Kline and Saez (2014) for U.S.; Corak and Heisz (1999) and Corak (2020) for Canada; Acciari et al. (2019) for Italy; and Deutscher and Mazumder (2020) for Australia.

Unfortunately, administrative records linking parents' and their children's information are still scarce, especially in developing countries. Studies on intergenerational earnings mobility in these countries have mostly relied on survey data without links between parents' and their children's earnings. In this context, the Two-Sample Two-Stage Least Squares estimator (TSTSLS), proposed by Klevmarken (1982), has been an appealing and commonly used alternative for studying intergenerational mobility.[1] The TSTSLS is based on information from two samples: one sample with information of children's earnings and their parents' characteristics excluding earnings (main sample), and other sample with information of the characteristics and earnings of a different group of parents (sample of pseudo parents). Then, the TSTSLS proceeds in two stages. In a first stage, the earnings of the parents in the main sample are imputed via a Mincer's regression, which is adjusted with the sample of pseudo parents. In a second stage, the IGE (or the rank-rank correlation) is estimated using the information of children's earnings in the main sample and their parents' imputed earnings (from the first stage).

Although the TSTSLS has been widely used to study intergenerational mobility in the absence of administrative records (Barbieri et al., 2020; Dunn, 2007; Narayan et al., 2018), to the best of our knowledge there is no evidence verifying that this

---

[1]This procedure is also called the Two-Stage Instrumental Variable (TSIV) strategy Angrist and Krueger (1992).

estimator is a good substitute for the results provided by linked parent-child earnings data. In this chapter, we fill this gap by assessing the reliability of the TSTSLS for estimating the IGE and the rank-rank correlation. We do this by exploiting a novel administrative data set from Chile that links parents' and their children's earnings, and by developing analytical results of the bias of the TSTSLS of the IGE and rank-rank correlation.

Initially, we estimate the IGE and the rank-rank correlation for Chile on the basis of administrative records. To the best of our knowledge, the work developed in this chapter is the first work doing this task. We estimate an IGE of 0.282, and a rank-rank correlation of 0.239. Our analysis reveals that our IGE point-estimate is almost half of earlier TSTSLS estimates of the IGE for Chile, which range between 0.47 and 0.54 (Núñez and Risco, 2004; Núñez and Miranda, 2010; Sapelli, 2013; and Celhay et al., 2010).[2] On the other hand, rank-rank correlations estimated using TSTSLS move between 0.379 and 0.453 (Núñez and Miranda, 2011).

We then attempt to understand the discrepancies between the findings of previous works and our results by decomposing the probability limit of the TSTSLS estimator of the IGE. Formally, we show how the TSTSLS estimator is biased from the IGE estimator based on data linking parents' and their children's earnings. Specifically, we show that the bias is driven by two sources: i) the projection bias and ii) the variance bias. The former comes from the difference between the imputed and actual parents' earnings, while the latter comes from the fact that imputed parental earnings via a Mincer equation have a variance that is distinct from the variance of actual parental earnings.

To illustrate these two sources of bias, we mimic a TSTSLS setting under different specifications of the Mincer equation using the Chilean data. When we only include education and age in the Mincer equation, we estimate that the IGE via the TSTSLS

---

[2]Additional papers that study intergenerational mobility in Chile and Latin American are Meneses, (2001), Leites et al. (2020), Torche (2005), Torche (2014), Ferreira and Veloso (2006), Jiménez (2011) and Grawe (2001).

is close to 0.52, consistent with the findings of previous works using this estimator for Chile. Then, when we add the type of contract and the economic sector to the Mincer equation, i.e., we decrease the projection bias, the estimated IGE via the TSTSLS is around 0.488, value that is still far from the actual IGE value of 0.282. In part, the previous results can be explained by the lack of validity of the traditional Mincer style equation (Mincer, 1974) as documented in the literature (Lemieux, 2006; Griliches, 1977; Heckman et al., 2006; and Lagakos et al., 2018a,b). This issue is particularly relevant in countries like Chile where unobservables play an essential role in determining earnings (Zimmerman, 2019).

To overcome this issue, we carry out an hypothetical exercise where we add a variable related to parents' unobservables to the Mincer equation (Mincer, 1996; and Heckman et al. 2003). We emphasize that this is a hypothetical exercise because this variable is calculated using the longitudinal dimension of our administrative data and therefore, it cannot be obtained with cross-sectional samples. After adding this variable, the projection bias decreases significantly and we estimate that the IGE via the TSTSLS is close to 0.340, a value that still overestimates the actual IGE value of 0.282.

However, we can still improve our IGE estimates by correcting the variance bias mentioned above. We do this by using a parsimonious technique from the multiple imputation literature (Little and Rubin, 2019; Rubin, 1996, 2004). This technique is simple and consists of including an error term when imputing parental earnings. After implementing this variance correction, we estimate that the IGE via TSTSLS is 0.276. Although this result suggests that it is possible to recover calculations based on administrative records using the TSTSLS approach, doing so would only be possible with very detailed longitudinal records, which are not available in household surveys. Without this information, we cannot say much about intergenerational mobility using this estimator. This result casts doubt on previous measures of the IGE using the TSTSLS.

Finally, we also decompose the probability limit of the TSTSLS estimator of the rank-rank correlation. We formally show that the TSTSLS estimator is biased from the estimator of the rank-rank correlation based on information that links parents' and their children's earnings. Concretely, we conclude that in this case the bias is driven by only the projection bias, and that the variance bias is not present due to the nature of the rank regressions. Then, we repeat the exercise of emulating a TSTSLS setting based on the information provided by our administrative records.

The results of this empirical exercise suggest that the estimates of the rank-rank correlation using administrative data and the TSTSLS method are similar. In particular, when we include education, age, industry, and contract type in the Mincer equation (four variables commonly included in household surveys), we estimate a rank-rank correlation of 0.222 via the TSTSLS. This means that we are underestimating by just 7% the estimates that comes directly from the administrative records (0.239). Thus, our analysis reveals that we can improve what we know about intergenerational mobility in developing countries by merely using the TSTSLS approach to estimate the rank-rank correlation instead of the IGE. This improvement might come from the fact that Mincer equations seem to be better at predicting ranking than level earnings. Problems at predicting levels come from that log earnings are a non-linear function of schooling, minimum wage' policies, and that the linear approximation may only be accurate in a stable environment where the growth in relative demand is matched by a corresponding growth in relative supply (Card and Krueger, 1992, Björklund and Kjellström, 2002; Lemieux, 2006 ; Heckman et al., 2003).

The rest of the chapter is organized as follows. Section 2.2 describes the administrative data for Chile. Section 2.3 estimates the IGE and the rank-rank correlation for Chile using the administrative data directly. Section 2.4 describes the TSTSLS estimator and presents the decomposition of its bias when estimating the IGE and the rank-rank correlation. Section 2.5 shows IGE and rank-rank correlation estimated via the TSTSLS under different specifications of the Mincer equation. Section 2.6

discusses the main results, and Section 2.7 concludes.

2.2 Administrative Data for Chile

We use the Chilean unemployment insurance program (UIP) to obtain informa-
tion on sons and their parents' labor earnings.[3] The UIP is a benefit that involves all
individuals over the age of 18 working under a fixed-term or permanent contract in
the formal private sector. Participation in the scheme is mandatory for all contracts
that started after September 2002, and voluntary for contracts began before that
date.

In 2003 and 2004, the coverage rate for private formally contracted employees was
below 50% because of the UIP scheme's optional retroactive aspect. This coverage
rate improved significantly over the following years, attaining a 60.4% average in
2005-2007, over 80% from 2011 and over 90% from 2015. The differences between
the number of private formal employees in the UIP database and the ENE survey,
in its first years is due to the voluntary scheme for contracts started before the 2002.
See Table 4.1 in Appendix for more details.

We establish parent-child linkages through administrative records provided by
the Civil Registry Office (CRO). Birth certificates issued by this agency contain in-
formation on both the child and the parents at the time of birth. Thus, we can
identify and build the pairs of children and parents included in the UIP database
through these certificates. In our baseline analysis, the sample of children is com-
posed of individuals that were 29-34 in 2018, while parents were 31-66 years old in
2007.

We measure parents' earnings from 2003 to 2007 and children's earnings from 2013
to 2018 by computing the five-year average of monthly earnings for months worked
in the formal private sector. In our baseline sample, we only consider parents and

---

[3]We focus on sons because previous works for Chile have concentrated on this group.

children that worked at least six months in the formal private sector during 2003-2007 and 2013-2018, respectively. We also exclude observations with zero earnings since such individuals could be working either as a public employee, in the informal sector, or the formal private sector but not covered by the UIP, especially in its earlier years. Moreover, we decide only to include children and parents who earn more than half the minimum wage on average in order to reduce the potential noise from low earnings observations.[4] Finally, if both parents worked in the period, we consider the average parental earnings as the sum of parental earnings divided by two, in line with Chetty, Hendren, Kline and Saez (2014) and Connolly et al. (2019).

Table 4.5 in Appendix compares the earnings percentiles of our data with those of the Encuesta Nacional de Empleo (ENE) for 2018. The ENE is a representative survey that collects information about employment and labor earnings for the whole population. Although we are not considering informal and public sector employees in our analysis, the earnings distribution from our data appears to be similar to the population's distribution as Table 4.5 suggests. In our baseline sample, we have 283,127 parent-son links. Table 2.1 shows descriptive statistics about our baseline sample. As can be seen in Table 2.1, the average log permanent income is 12.71 for parents and 13.23 for sons. The mean age for parents is 47.28, while sons are on average 32.22. Moreover, 55% of parents have less than high school, 37% complete high school, 8% have some college degree, and 51% have a permanent contract.

Finally, due to potential concerns related to the limited coverage of the UIP, especially in its initial years, we conduct several robustness exercises to assess our findings' validity. Specifically, we consider pairs of parent-child with at least 12, 24, and 36 months worked in both cases during the period from 2003 to 2007 for parents and from 2013 to 2018 for children. The results we present below remain

---

[4]Half the minimum wage for children is $133,000 in 2019 Chilean pesos (measured from 2014 to 2018) and $103,000 in 2019 Chilean pesos for parents (from 2003 to 2007). Using the National Socioeconomic Characterization Survey (CASEN) in 2017, 14.1 percent of the population earn less than the minimum wage.

	Observations	Mean	SD	p25	p50	p75
Parents, log(permanent income)	283,127	12.71	.60	12.29	12.57	13.01
Children, log(permanent income)	283,127	13.23	.58	12.81	13.17	13.60
Parents, Age in 2007	283,127	47.28	5.96	43.00	47.00	51.00
Children, Age in 2018	283,127	32.22	1.69	31.00	32.00	34.00
Parents, Less than High School	283,127	.55	.50	.00	1.00	1.00
Parents, High School	283,127	.37	.48	.00	.00	1.00
Parents, Some College	283,127	.08	.28	.00	.00	.00
Parents, Permanent Contract	283,127	.51	.46	.00	.58	1.00

Table 2.1: Descriptive Statistics[†]

Note[†]: The final sample is composed of 283,127 parents-sons links. Permanent incomes are in 2018 Chilean pesos. Parents' permanent income is calculated as a 5-year average between 2003 and 2007. Sons' earnings are calculated as a 5-year average between 2013 and 2018. Parents' education is divided into three categories, and type of contract is a dummy variable equals to one in the case of a permanent contract.

valid regardless of the sample of pairs of parents-children employed.

2.3   Intergenerational earnings mobility

In this section, we define the intergenerational elasticity and the rank-rank correlation. Then, we estimate both for Chile directly from the administrative data.

2.3.1   Intergenerational Elasticity

The following equation gives the standard empirical specification for estimating the intergenerational earnings elasticity:

$$y_i^c = \alpha + \beta \cdot y_i^p + \epsilon_i, \tag{2.1}$$

where y_i^c is the logarithm of permanent individual earnings for child i, y_i^p is the logarithm of his parent's permanent earnings, α is the constant, and ϵ_i is the error

term.[5]

The coefficient β is the "intergenerational elasticity" (IGE). This quantity is one of the most well-known intergenerational mobility measures (Corak, 2018). A lower number is interpreted as a more intergenerationally mobile population.

Using our unique administrative data with information of children's and their parents' permanent earnings, we adjust equation (2.1) by OLS, and $\hat{\beta}$ is the IGE's estimate in this case. The results are shown in Table 2.2, where column (1) considers individuals with at least six months of positive earnings. As can be seen in Table 2.2, our estimate is 0.282, meaning that a 1% increase in parents' earnings is associated with a 0.282% increase in their children's earnings. Columns (2), (3), and (4) show the estimates restricting the sample to 12, 24, and 36 months of positive earnings, respectively, with similar results.

Importantly, these numbers are much lower than the previous literature has estimated for the IGE for Chile which ranges from 0.47 to 0.57. Indeed, our estimate is close to half of the lower estimate found in the literature (Núñez and Risco, 2004; Núñez and Miranda, 2010; Sapelli, 2013; and Celhay et al., 2010). Additionally, we replicate the TSTSLS approach using the surveys CASEN 2017 for Child's earnings (main sample) and CASEN 2000 for auxiliary sample of pseudo parents. We estimated a range of the IGE between 0.40 and 0.41. Columns (1) and (4) of Table 4.6 in the Appendix presents these results.

2.3.2   Rank-rank correlation

Another manner to estimate intergenerational mobility is to use rank-rank correlations. The rank-based regression was first introduced by Jureckova (1971) and

---

[5]Solon (1992) discuss that Eq. (2.1) could not be the true income model. In particular, omitted variables could affect child earnings that are correlated with parent's income. Such as parent's education. Following Chetty et al. (2014), Connolly (2019), Corak (2018) and Acciari et al. (2019) we assume that equation (2.1) is the correct model. We also ignore any measurement error in the children's and parent's income. An et al. (2020) and Durlauf et al. (2017) discuss the linearity assumption in details.

	(1)	(2)	(3)	(4)
$\hat{\beta}$	0.282***	0.294***	0.314***	0.329***
	(0.002)	(0.002)	(0.002)	(0.003)
Constant	9.655***	9.529***	9.313***	9.175***
	(0.022)	(0.024)	(0.028)	(0.036)
Observations	283,127	237,902	166,797	105,039
R-squared	0.087	0.094	0.107	0.117

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.2: OLS estimates of the intergenerational earnings elasticity for male children[†]

Note[†]: Earnings are measured as average earnings over the months where a children-parents pair report positive earnings over the studied 5-year period. We keep individuals with average earnings greater than half of the corresponding minimum wage. Columns (1) to (4) report results for male children. Column (1) considers individuals with at least 6 months of positive earnings, (2) considers individuals with at least 12 months of positive earnings, (3) considers individuals with at least 24 months of positive earnings and (4) considers individuals with at least 36 months of positive earnings.

Jaeckel (1972). More recently, Dahl and DeLeire (2008) use the rank-rank regression in the context of relative mobility and since the paper of Chetty, Hendren, Kline and Saez (2014), rank-rank correlations has been broadly used because of its stability and robustness to life cycle measurement (Chetty et al., 2017; Mazumder, 2016; and Chetty et al., 2014).

The rank-rank correlation is estimated by running the following OLS regression:

$$r_i^c = \alpha_{rr} + \beta_{rr} \cdot r_i^p + \xi_i, \tag{2.2}$$

where r_i^c is the child i's ranking over the children earnings distribution, and r_i^p is the parent i's ranking on the parents distribution, α_{rr} is the constant, and ξ_i is the error term. Here, β_{rr} is the rank-rank correlation, which is interpreted as the impact of an increment of 1% in the parental earning ranking on the children earning ranking.

On the basis of our administrative data, we know each child's and parents' earnings ranking, so we can estimate equation (2.2) directly. Table 2.3 shows the results. As before, we have four columns to illustrate the parameters under alternative criteria to determine the sample for analysis. As can be seen in Table 2.3, the rank-rank correlation moves between 0.239 and 0.264. The estimates appear to be consistent among all columns. The only previous estimates for the rank-rank correlation via the TSTSLS are between 0.379 and 0.453 (Núñez and Miranda, 2011). However, authors have fewer than 1,000 parent-child links, so differences with our results could be driven not only because of TSTSLS but also because of lack of statistical power. The following section shows a relationship between administrative linked data and TSTSLS estimators for IGE and rank-rank correlation.

	(1)	(2)	(3)	(4)
$\hat{\beta}_{rr}$	0.239***	0.247***	0.258***	0.264***
	(0.002)	(0.002)	(0.002)	(0.003)
Constant	41.682***	42.504***	44.118***	46.312***
	(0.103)	(0.114)	(0.139)	(0.175)
Observations	283,127	237,902	166,797	105,039
R-squared	0.060	0.064	0.070	0.076

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.3: OLS estimates of the rank-rank correlation for male children[†]

Note [†]: Earnings are measured as the average earnings over the months in which an individual reports positive earnings over 5 years. We keep children-parents linkages with average earnings greater than half of the corresponding minimum wage. Columns (1) to (4) report results for male children. Column (1) considers individuals with at least 6 months of positive earnings, (2) considers individuals with at least 12 months of positive earnings, (3) considers individuals with at least 24 months of positive earnings and (4) considers individuals with at least 36 months of positive earnings.

2.4   The Two-Sample Two-Stage Least Squares estimator

The Two-Sample Two-Stage Least Squares (TSTSLS) estimator has been used to measure intergenerational mobility when matched child-parent data of earnings is unavailable. For instance, most of the estimates from developing countries use this methodology (Bloise et al., 2020; Narayan et al., 2018; Torche, 2005; Torche, 2014; Ferreira and Veloso, 2006; Jiménez, 2011; and Grawe, 2001) but researchers have also applied it to study mobility in developed countries (Barbieri et al., 2020, OECD, 2018, Narayan et al., 2018, Olivetti and Paserman, 2015). The TSTSLS estimator uses retrospective information on parents' socioeconomic background and a sample of pseudo parents to impute parental incomes via Mincer equations. Although this method has been mostly used to estimate the IGE, it can also be used to estimate the rank-rank correlation.

TSTSLS requires two samples. The main sample contains information on children's earnings and demographic information on their parents. The auxiliary sample consists of pseudo-parents, that is, earlier data from the parent's cohort with earnings and demographic information. On the basis of these two samples, the estimation proceeds in two steps. First, the auxiliary sample of pseudo-parents is used to estimate a Mincer equation by OLS:

$$y_j^{pp} = \omega' z_j^{pp} + v_j, \tag{2.3}$$

where y_j^{pp} is the earnings of pseudo-parent j, z_j^{pp} is the pseudo-parent vector of time-invariant characteristics, and v_j is the residual component. Then, after estimating ω', we estimate y_i^p as $\hat{y}_i^p = \hat{\omega}' z_i^p$ where z_i^p is the vector of characteristics for the actual parent i.

Having imputed the earnings for all parents in the main sample, the second stage

of the TSTSLS estimator consists of quantifying the relationship between $\hat{y}_i^p$ and y_i^c, where y_i^c denotes the earnings of child i in the main sample. Specifically, to estimate the IGE via the TSTSLS, we regress the log of y_i^c on the log of $\hat{y}_i^p$ by OLS, and keep the estimate of the slope, while for the rank-rank correlation based on the TSTSLS, we regress r_i^c on $\hat{r}_i^p$ by OLS, and keep the estimate of the slope, where r_i^c is the ranking of earnings of child i in the children's earnings distribution and $\hat{r}_i^p$ is the ranking of imputed earnings of parent i in the parents' imputed earnings distribution.

In the context of TSTSLS, two relevant issues with Mincer equations are (1) the lack of dispersion problem and (2) the prediction problem. The former occurs when predicted income has a lower variance than actual income. The latter occurs when the model fails to predict earnings. Both are related to omitted variables in the Mincer equation. Following Mincer (1974) the right model has as a crucial characteristic that every parameter can be individual dependent. It implies that every person might present different marginal returns of schooling, experience, and age. One equivalent approach is to run a model where estimates do not vary across people, but it has to include individual fixed effects to capture variability in returns (Heckman et al. 2013). However, individual fixed effects are often omitted in the intergenerational mobility literature because of the lack of panel data.

Figure 2.1 shows both issues when omitting individual fixed effects for Chile (Eq. (2.3)). The left panel shows the predicted parents' log income, pseudo parents' log income, and actual parents' log income. The predicted income presents a lack of variance. The distribution has points with higher frequencies than the actual distribution. One way to assess this problem is by calculating the R^2. Table 4.7 in Appendix shows that the R^2 is 0.26 if we include education, age, age square, industry, and type of contract. This implies that these variables can capture 26% of the total variation in pseudo parents' income. In addition to this, the Kolgomorov-Smirnov test shows they have different distributions. The right panel depicts the ranking of predicted income as a function of actual ranking. While the Mincer equation does

not predict income well, it might do better for rank. Although the ranking of the prediction is monotone there is still a gap with the 45-degree line. In Section 2.4.1, we show both issues mathematically when using the Mincer equations to impute the first stage of the TSTSLS.
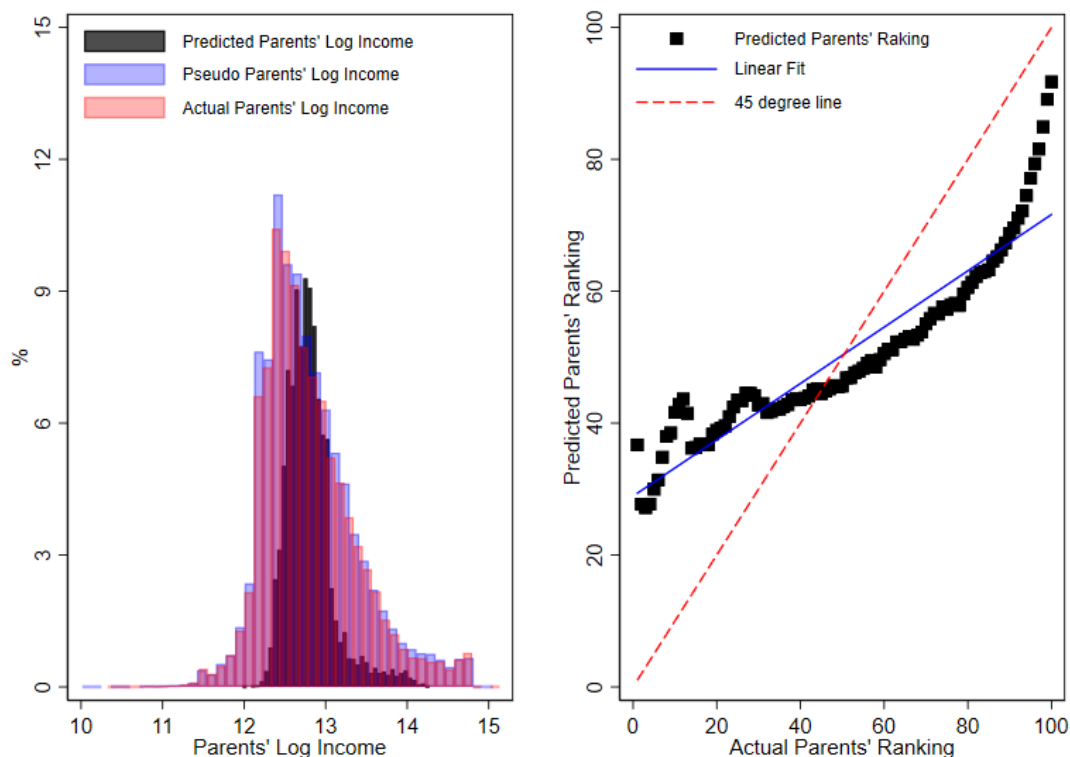
Figure 2.1: Parents' Log Income and Ranking, TSTSLS Method[†]

Note[†]: Left panel shows the distribution of the predicted parents' log income, pseudo parents' log income and actual parent' log income. Right panel depicts the ranking of predicted income as a function of the actual ranking. A linear fit line and a 45 degree line are also included. Income prediction is constructed with education, age, age square, industry and type of contract as control variables.

2.4.1   Estimating the IGE Using TSTSLS

Having obtained $\hat{y}_i^p$, we study the following relationship between y_i^c and $\hat{y}_i^p$

$$y_i^c = \alpha^{TSTSLS} + \beta^{TSTSLS} \cdot \hat{y}_i^p + \psi_i, \tag{2.4}$$

where α^{TSTSLS} is the constant, and β^{TSTSLS} is the elasticity between the imputed parent's earnings and child's earnings, while ψ_i is an error term. We estimate β^{TSTSLS} as:

$$\hat{\beta}^{TSTSLS} = \frac{c\hat{o}v\left(y^c, \hat{y}^p\right)}{v\hat{a}r\left(\hat{y}^p\right)}. \tag{2.5}$$

In this context, Solon (1992), Björklund and Jäntti (1997), Nicoletti and Ermisch (2008), and Jerrim et al. (2016) show that consistency of $\hat{\beta}^{TSTSLS}$ with respect to the true β can be obtained if either (a) the variables included in the parent characteristics z_i have no direct effect on child's earnings and (b) when the R^2 of the estimated equation (2.3) is equal to one. Typically, both conditions usually do not hold.

There are many ways to obtain $\hat{y}^p$, i.e., the imputed values of parents' income in the main sample, for example, using different specifications of the Mincer equation. Let us call $\Omega = \{\hat{y}^p(j)\}_{j=1}^\infty$ the set of imputations for y^p. Each element $\hat{y}^p(j)$ of Ω defines a different parameter, call it $\beta(j)^{TSTSLS}$ with its respective TSTSLS estimator given by $\hat{\beta}(j)^{TSTSLS}$. Importantly, we can establish a relationship between $\text{plim}\,\hat{\beta}$ and $\text{plim}\,\hat{\beta}^{TSTSLS}(j)$ in the following Proposition.

Proposition 1. $\text{plim}\,\hat{\beta}^{TSTSLS}(j)\ \forall j$ *can be greater or lower than* $\text{plim}\,\hat{\beta}$.

Proof. We know that

$$\text{plim}\,\hat{\beta} = \frac{cov(y^c, y^p)}{var\left(y^p\right)} = \frac{cov(y^c, y^p + \hat{y}^p(j) - \hat{y}^p(j))}{var\left(y^p\right)},$$

$$\text{plim}\,\hat{\beta} = \frac{cov(y^c, y^p - \hat{y}^p(j))}{var\left(y^p\right)} + \frac{cov(y^c, \hat{y}^p(j))}{var\left(y^p\right)} \cdot \frac{var\left(\hat{y}^p(j)\right)}{var\left(\hat{y}^p(j)\right)}.$$

Define $\eta(j) \equiv \frac{cov(y^c, y^p - \hat{y}^p(j))}{var(y^p)}$, and $\kappa(j) \equiv \frac{var(\hat{y}^p(j))}{var(y^p)}$. Using plim $\hat{\beta}^{TSTSLS}(j) = \frac{cov(y^c, \hat{y}^p(j))}{var(\hat{y}^p)}$ we have that

$$\text{plim } \hat{\beta}^{TSTSLS}(j) = \frac{\text{plim } \hat{\beta} - \eta(j)}{\kappa(j)}. \tag{2.6}$$

$\square$

Equation (2.6) relates the probability limit of $\hat{\beta}^{TSTSLS}$ and $\hat{\beta}$. That is, any bias generated by $E(\epsilon_i | y_i^p) \neq 0$ is going to be directly reflected on both probability limits.[6]

Given this, we can define two sources of bias between of $\hat{\beta}^{TSTSLS}$. We call $\kappa(j)$ as the variance bias and $\eta(j)$ as the projection bias. The variance bias comes from the fact that if $var\left(\hat{y}^p(j)\right) \neq var\left(y^p\right)$ then $\kappa(j) \neq 1$. Thus, variance bias exists when the variance of predicted parents' earnings is not equal to the actual variance of parents' earnings. On the other hand, the projection bias is present when $\eta(j) \neq 0$, which occurs when $cov(y^c, y^p - \hat{y}^p(j)) \neq 0$. Hence, if the difference between predicted and actual parents' earnings correlates with children's earnings, projection bias exists. Although it is often stated that $\hat{\beta}^{TSTSLS}$ is upward inconsistent (Blanden, 2013), we show in equation (2.6) that it can actually be upwards or downwards biased. See Jerrim et al. (2016) who exemplify this.

Finally, second stage standard errors will be underestimated when generated regressors are used (Murphy and Topel, 2002; Wooldridge, 2010; Inoue and Solon, 2010; Björklund and Jäntti, 1997; and Piraino, 2015) also suggest bootstrapping as a possibility to deal with the same issue.

In the next section, we show that the IGE via TSTSLS is biased in Chile. The bias on $\hat{\beta}^{TSTSLS}$ depends on which of both terms $\kappa(j)$ or $\eta(j)$ plays a stronger role.

Furthermore, including additional variables to increase the R^2 of the first-stage prediction equation may reduce the bias through $\kappa(j)$ but simultaneously increase the bias through $\eta(j)$. Hence, the total effect could be ambiguous. Jerrim et al. (2016) find similar results.

---

[6]Notice that when $E(\epsilon_i | y_i^p) = 0$, plim $\hat{\beta}^{TSTSLS}(j) = \frac{\beta - \eta(j)}{\kappa(j)}$.

2.4.2   Estimating Rank-Rank Correlations Using TSTSLS

The ranking of imputed earnings $(\hat{r}^p(j))$ is constructed over the imputed parents' earnings distribution similar to Barbieri et al. (2020) and Olivetti et al. (2018). Thus, the model is:

$$r_i^c = \alpha_{rr}^{TSTSLS} + \beta_{rr}^{TSTSLS} \cdot \hat{r}_i^p + \varsigma_i, \tag{2.7}$$

where α_{rr}^{TSTSLS} is the constant and ς_i is the error term. In this model, β_{rr}^{TSTSLS} is the rank-rank correlation under the TSTSLS approach. We can also establish a relationship between plim $\hat{\beta}_{rr}$ and plim $\hat{\beta}_{rr}^{TSTSLS}(j)$

Proposition 2. plim $\hat{\beta}_{rr}^{TSTSLS}(j)$ $\forall j$ *can be greater or lower than* plim $\hat{\beta}_{rr}$.

Proof. The probability limit can be written as

$$\text{plim} \, \hat{\beta}_{rr}^{TSTSLS}(j) = \frac{cov\left(r^c, \hat{r}^p(j)\right)}{var\left(\hat{r}^p(j)\right)} = \frac{cov\left(r^c, r^p + \hat{r}^p(j) - r^p\right)}{var\left(\hat{r}^p(j)\right)}.$$

Notice that $var\left(\hat{r}^p(j)\right) = var\left(r^p\right) = \frac{1}{12}$ because both distributions are uniform between zero and one. Therefore, the rank-rank correlation does not suffer from variance bias. Thus, the problem translates to:

$$\text{plim} \, \hat{\beta}_{rr}^{TSTSLS}(j) = \frac{cov\left(r^c, r^p\right)}{var\left(r^p\right)} + \frac{cov\left(r^c, \hat{r}^p(j) - r^p\right)}{var\left(r^p\right)},$$

$$\text{plim} \, \hat{\beta}_{rr}^{TSTSLS}(j) = \text{plim} \, \hat{\beta}_{rr} + 12 \cdot \theta(j),$$

where $\theta(j) \equiv cov\left(r^c, \hat{r}^p(j) - r^p\right)$. □

That is, there is only one source of bias, which is given by the error projection between the true parental ranking and the ranking of predicted earnings. A positive θ means that the bias from estimating parental rank increases as the child's rank increases. In summary, $\hat{\beta}_{rr}^{TSTSLS}$ will be biased when $\theta(j) \neq 0$. Nevertheless, only

imputation bias is present. Thus, the better we predict the parents' ranking, the smaller the bias. Chetty, Hendren, Kline and Saez (2014) recommend the use of rank-rank correlations rather that the IGE due to its stability and robustness to life cycle measurement. We add another advantage compared to the IGE, it eliminates variance bias in estimating intergenerational mobility using TSTSLS.

2.5 Mimicking a TSTSLS setting

Our administrative records allow us not only to estimate β and β_{rr} but also to mimic a TSTSLS estimation setting to asses its bias. Concretely, we proceed with the following simulation exercises to evaluate the bias of the IGE and rank-rank correlation estimation via the TSTSLS:

i) We take $\hat{\beta} = 0.282$ from column (1) of Table 2.2 and $\hat{\beta}_{rr} = 0.239$ from column (1) of Table 2.3 as the estimates based on administrative data.

ii) We take a random subsample Σ of 50,000 (out of 283,127 total links) parent-child links information from the baseline sample. We then randomly split this subsample into two subsubsamples, say Σ_1 and Σ_2, of 25,000 observations each. One subsubsample (Σ_1) plays the role of the auxiliary sample of pseudo-parents and, therefore, is used to estimate the Mincer equation for the pseudo-parents:

$$y^{pp} = \omega' z^{pp} + v, \tag{2.8}$$

where in z^{pp} we initially include the variables: age, age squared, industry, education type, and type of contract. We estimate ω' by OLS. We then use the parents' information in Σ_2 (the main sample) to impute $\hat{y}^p$. Let us denote this imputation as $\hat{y}^p(1)$.

iii) We compute $\hat{\beta}^{TSTSLS}(1)$, and $\hat{\beta}_{rr}^{TSTSLS}(1)$ by regressing y^c on $\hat{y}^p(1)$, and r^c on $\hat{r}^p(1)$ from Σ_2.

iv) We repeat ii)-iii) 1,000 times.

Table 2.4 shows $\hat{\beta}$ and the results of this simulation for the IGE. As can be seen in Table 2.4, $\hat{\beta}^{TSTSLS}(1)$ ranges from 0.445 to 0.532, which is close to what previous literature has estimated via TSTSLS for Chile (Núñez and Risco, 2004; Núñez and Miranda, 2010; and Sapelli, 2013). On average, $\hat{\beta}^{TSTSLS}(1)$ overestimates $\hat{\beta}$ by 73%. The results in Table 2.4 reveal that both biases are present, since $\eta(1) \neq 0$ and $\kappa(1) \neq 1$. The former causes $\hat{\beta}^{TSTSLS}$ to be upward biased, while the latter causes it to be downward biased.

Coefficient	(1) min	(2) p25	(3) p50	(4) mean	(5) p75	(6) max
$\hat{\beta}$	0.282	0.282	0.282	0.282	0.282	0.282
$\hat{\beta}^{TSTSLS}(1)$	0.445	0.479	0.487	0.488	0.497	0.532
$\eta(1)$	0.140	0.155	0.158	0.159	0.162	0.177
$\kappa(1)$	0.243	0.258	0.263	0.263	0.268	0.284

Table 2.4: Results from simulated exercise 1[†].

Note[†]: This table shows IGE when administrative data ($\hat{\beta}$) and TSTSLS approach ($\hat{\beta}^{TSTSLS}$) are used. Prediction bias ($\eta(1)$), and variance bias ($\kappa(1)$) are included. Mincer equation contains age, age squared, occupational sector, education type, and type of contract as control variables.

The low value of $\kappa(1) \equiv \frac{var(\hat{y}^p(1))}{var(y^p)}$ has been deeply discussed in the statistical literature on imputation of missing data (Little and Rubin, 2019; Rubin, 1996, 2004). Even when the regression is correctly specified the imputation will not reflect the missing data's uncertainty if the predicted variance is too small (see Figure 2.1). One way to overcome this problem is through stochastic regression imputation. Specifically, we impute parents' earnings in the main sample using stochastic regression imputation:

$$\hat{y}^p = \hat{\omega}' z^p + N(0, \hat{\sigma}_v^2),$$

where $\hat{\sigma}_v^2$ is the estimated variance of the Mincer equation fitted with the sample of pseudo-parents ($y^{pp} = \omega' z^{pp} + v$). In words, we reduce the variance bias by adding noise to the imputed earnings of the parents in the main sample. Let us denote this imputation as $\hat{y}^p(1b)$.

The results of this procedure are presented in Table 2.5. As expected, the variance bias is reduced which reflects the fact that $\kappa(1b)$ increases significantly, attaining a median value of 1.002. However, $\eta(1b) \neq 0$, so the TSTSLS of the IGE is still biased. Indeed, $\hat{\beta}^{TSTSLS}(1b)$ is a lower bound of $\hat{\beta}$ because the prediction bias remains. On average, $\hat{\beta}^{TSTSLS}(1b)$ underestimates $\hat{\beta}$ by 54%.

| | (1) | (2) | (3) | (4) | (5) | (6) |
Coefficient	min	p25	p50	mean	p75	max
$\hat{\beta}$	0.282	0.282	0.282	0.282	0.282	0.282
$\hat{\beta}^{TSTSLS}(1b)$	0.108	0.123	0.128	0.128	0.132	0.152
$\eta(1b)$	0.130	0.153	0.159	0.159	0.165	0.185
$\kappa(1b)$	0.948	0.991	1.002	1.002	1.013	1.059

Table 2.5: Results from simulated exercise 1 adding additional variance[†].

Note[†]: This table shows IGE when administrative data ($\hat{\beta}$) and TSTSLS approach ($\hat{\beta}^{TSTSLS}$) are used. Prediction bias ($\eta(1b)$), and variance bias ($\kappa(1b)$) are included. Mincer equation contains age, age squared, occupational sector, education type, and type of contract as control variables. Stochastic regression imputation is used to increase the variance of the imputed parents' income.

Reducing the projection bias is a difficult task because it requires the inclusion of variables into the Mincer equation that would increase its predictive power. The difficulty lies in that these predictors typically are not present in household surveys. To illustrate this point, in what follows we improve $\hat{y}^p$ by adding individual fixed effect as an additional predictor to the Mincer equation. We remark that this is an hypothetical exercise which cannot be implemented in an actual TSTSLS setting due to data unavailability. Specifically, we use our administrative records from 2003 to 2019 to estimate a panel Mincer equation with individual fixed effects using our baseline sample of parents. Having this fixed effect for all parents, we then use this

variable as an additional predictor when mimicking the TSTSLS setting described above.

Let us denote $\hat{y}^p(2)$ as the imputed parental earnings in the main sample via the Mincer equation augmented with the individual fixed effect and let us $\hat{y}^p(2b)$ denote $\hat{y}^p(2)$ but after correcting for the lack of variance. The results are presented in Table 2.6. We can see that $\hat{\beta}^{TSTSLS}(2)$ still overestimates $\hat{\beta}$ by 0.058 points or 20% of the value of $\hat{\beta}$. Meanwhile, the estimates that corrects the lack of variance bias, $\hat{\beta}^{TSTSLS}(2b)$, only differs in 0.007 or a 2% of $\hat{\beta}$. We can conclude that the estimated fixed effect is a very good predictor of parental earnings because the R^2 of the Mincer equation fitted the pseudo-parents is 0.81 when adding this variable as predictor (See Table 4.7). However, we emphasize that the availability of the fixed effect as a predictor to be included in the Mincer equation is an exercise that cannot be replicated in practice when implementing the TSTSLS estimator.

Coefficient	(1) min	(2) p25	(3) p50	(4) mean	(5) p75	(6) max
$\hat{\beta}$	0.282	0.282	0.282	0.282	0.282	0.282
$\hat{\beta}^{TSTSLS}(2)$	0.317	0.335	0.340	0.340	0.345	0.363
$\hat{\beta}^{TSTSLS}(2b)$	0.253	0.271	0.276	0.275	0.279	0.294
$\eta(2)$	0.00207	0.00932	0.0113	0.0113	0.0132	0.0206
$\eta(2b)$	-0.00377	0.00862	0.0114	0.0113	0.0140	0.0281
$\kappa(2)$	0.791	0.806	0.811	0.811	0.816	0.836
$\kappa(2b)$	0.974	0.994	1.001	1.001	1.007	1.030

Table 2.6: Results from simulated exercise 2[†].

Note[†]: This table shows IGE when administrative data ($\hat{\beta}$) and TSTSLS approach ($\hat{\beta}^{TSTSLS}$) are used. Prediction bias ($\eta(\cdot)$), and variance bias ($\kappa(\cdot)$) are included. Mincer equation contains age, age squared, occupational sector, education type, type of contract, and individual fixed effects as control variables. Stochastic regression imputation is used to get $\hat{\beta}^{TSTSLS}(2b)$, $\eta(2b)$, and $\kappa(2b)$.

2.5.1 The rank-rank correlation estimation

Table 2.7 shows the results for the simulated TSTSLS in regards to the rank-rank correlation. The $\hat{\beta}_{rr}$ calculated directly from the data is compared to the $\hat{\beta}_{rr}^{TSTSLS}$ under simulated exercise 1 and 2. In particular, $\hat{\beta}_{rr}^{TSTSLS}(1)$ which comes from a regression with age, age square, education, industry and type of contract underestimates $\hat{\beta}_{rr}$ by 7%, on average. We highlight this result because survey data typically includes those variables. Researchers can use them to estimate intergenerational mobility by calculating rank-rank correlation with a high level of consistency. On the other hand, $\hat{\beta}_{rr}^{TSTSLS}(2)$ which is estimated including individual fixed effects overstates the administrative records' estimate by 13%, on average. In general, both prediction biases $\theta(1)$ and $\theta(2)$ are small. This shows why the rank-rank correlation has better performance than IGE under the TSTSLS approach.

Coefficient	(1) min	(2) p25	(3) p50	(4) mean	(5) p75	(6) max
$\hat{\beta}_{rr}$	0.239	0.239	0.239	0.239	0.239	0.239
$\hat{\beta}_{rr}^{TSTSLS}(1)$	0.203	0.218	0.222	0.222	0.226	0.238
$\hat{\theta}(1)$	5.45e-05	0.00109	0.00143	0.00142	0.00175	0.00301
$\hat{\beta}_{rr}^{TSTSLS}(2)$	0.252	0.266	0.271	0.270	0.274	0.288
$\hat{\theta}(2)$	-0.00406	-0.00293	-0.00263	-0.00262	-0.00228	-0.00106

Table 2.7: Results from simulated exercise 1 and 2 [†].

Note[†]: This table shows rank-rank correlation when administrative data ($\hat{\beta}_{rr}$) and TSTSLS approach ($\hat{\beta}_{rr}^{TSTSLS}$) are used. Prediction bias (θ), is included. Mincer equation contains age, age squared, occupational sector, education type, type of contract under model (1). Moreover, model (2) also includes individual fixed effects as control variables.

2.5.2 TSTSLS: IGE vs Rank-Rank

Figure 2.2 shows the average IGE and rank-rank correlation under different prediction equations. TSTSLS tends to overestimate IGE and underestimate the rank-

rank correlation that comes directly from the administrative data. As can be seen, extra control variables improve IGE estimates only slightly. However, when individual fixed effects are incorporated, the IGE estimated though TSTSLS tends to be closer to the estimates that come from administrative data. Its inclusion improves the IGE estimation by 53% (reducing the bias from 73% to 20%). However, for rank-rank estimates there is no such improvement. Indeed, the gains are almost zero and comparing mean values might even worsen the bias, moving from an underestimation of 7.1% to an overestimation of 13%. The reason comes from the fact that θ passes from being positive to negative. This implies that as long as the child's ranking increases, the error correction between estimated and actual parental earnings decreases. Then, we have better predictions in the top part of the distribution than in the bottom. That is why a model with individual fixed-effects may overstate the estimates based on administrative data.

In the IGE estimation, we can see that the stochastic regression only improves the result when we include the fixed effects. The intuition behind this lies in the the low prediction power of Mincer equation; the improvements of variance bias are not as substantial as the problems that come from the prediction bias.

Figure 4.1, 4.2, and 4.3 in Appendix shows the whole distribution of $\hat{\beta}^{TSTSLS}$ and $\hat{\beta}_{rr}^{TSTSLS}$ obtained from our experiment. Extra control variables seem to improve the estimates (at least getting closer to the administrative correlation) only under an individual fixed effect for IGE. On the other hand, a rank-rank correlation through TSTSLS might only need typical control variables and an R^2 near to 0.26 to get consistent estimates.

From this analysis, we can conclude that i) having strong predictors on the Mincer equation is not a sufficient condition to obtain unbiased estimates, and ii) the stochastic regression procedure is only useful when the first stage of the TSTSLS method is good enough and iii) our discrepancies with the previous literature on intergenerational mobility in Chile can be explained by the inaccuracy of the traditional
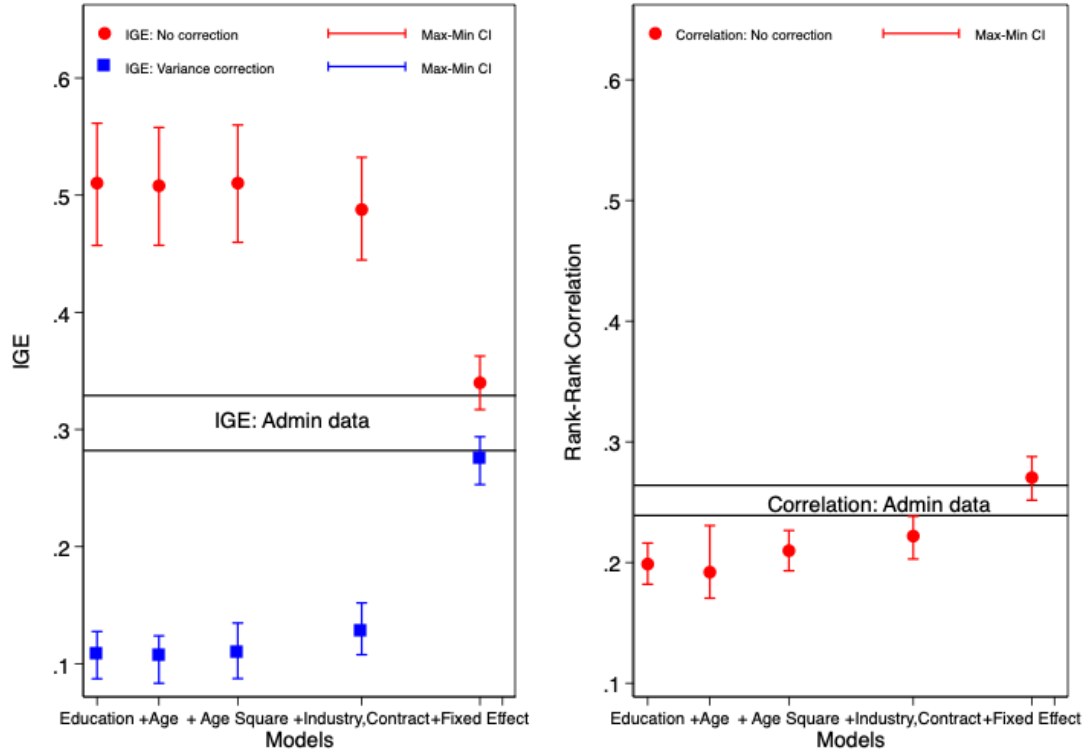
Figure 2.2: IGE and Rank-Rank Correlation: Alternative set of variables[†]

Note[†]: Average IGE and rank-rank correlation values are shown when alternative set of controls are included in the Mincer equation. The average is calculated taking the simple mean over the 1,000 replications. The max-min confidence interval is also included for each alternative model. Red circles depict averages without the variance correction while blue squares illustrate the results after controlling by that. Black lines show the minimum and the maximum values for $\hat{\beta}$, and $\hat{\beta}_{rr}$ on Table 2.2, and Table 2.3, respectively.

uncorrected Mincer equation to explain earnings. As we show here, an inadequate projection for y_i^p implies that under TSTSLS, IGE may be very far from the IGE estimated using administrative data while the rank-rank correlation estimates are very close. Indeed, our TSTSLS estimate for rank-rank correlation only underestimates the administrative data by 7%. To understand intergenerational mobility without administrative data in developing countries like Chile, it is more accurate to use rank-rank correlation rather than the IGE. To improve the IGE estimates,

we require better predictors for y^p. However, this is a significant task, especially in developing countries where earnings/income are usually determined by unobservable covariates such as social capital, non-cognitive skills, or neighborhood.

2.6 International Comparison

There is a rich literature on the relationship between intergenerational mobility and inequality, known as the Gatsby Curve (Corak, 2013). Figure 2.3 depicts the IGEs estimates for selected OECD countries using both TSTSLS procedure and administrative linked data. The graph compares the Gini Index on the x-axis to the IGE on the y-axis. The Gini index measures the extent to which the distribution of income among individuals within an economy deviates from a perfectly equal distribution. The blue circles show IGE estimates using administrative linked data for the US, Canada, Sweden, Netherlands, Norway, Finland, Denmark, Italy, and Australia (Corak and Heisz, 1999; Lindahl, 2008; Carmichael et al., 2020; Jantti et al., 2006; Deutscher and Mazumder, 2020; Chetty, Hendren, Kline and Saez, 2014; and Acciari et al., 2019). The red squares show TSTSLS estimates from survey data (Corak, 2013; and Narayan et al., 2018). The black triangle is our estimate. We then fit linear regressions of IGE on the Gini index for both direct administrative data and TSTSLS survey estimates. We can see that the gradient from administrative data is lower than the gradient of TSTSLS using survey data. It sheds light that TSTSLS estimation can give misleading results for intergenerational mobility and shows a stronger relationship between inequality and intergenerational mobility.

For Chile, we show TSTSLS overestimates IGE because one bias causes it to be bigger, but another smaller. Indeed, the variance bias tends to be more important because having an R^2 close to 0.2 implies that $\left(\text{plim}\, \hat{\beta} - \eta(j) \right)$ is multiplied approximately by 5.[7] After adding individual fixed-effect, we improve our prediction

---

[7]The R^2 is the ratio between the variance of predicted parents' income and pseudo parents income. If we assume that actual parents and pseudo-parents income have the same variance,
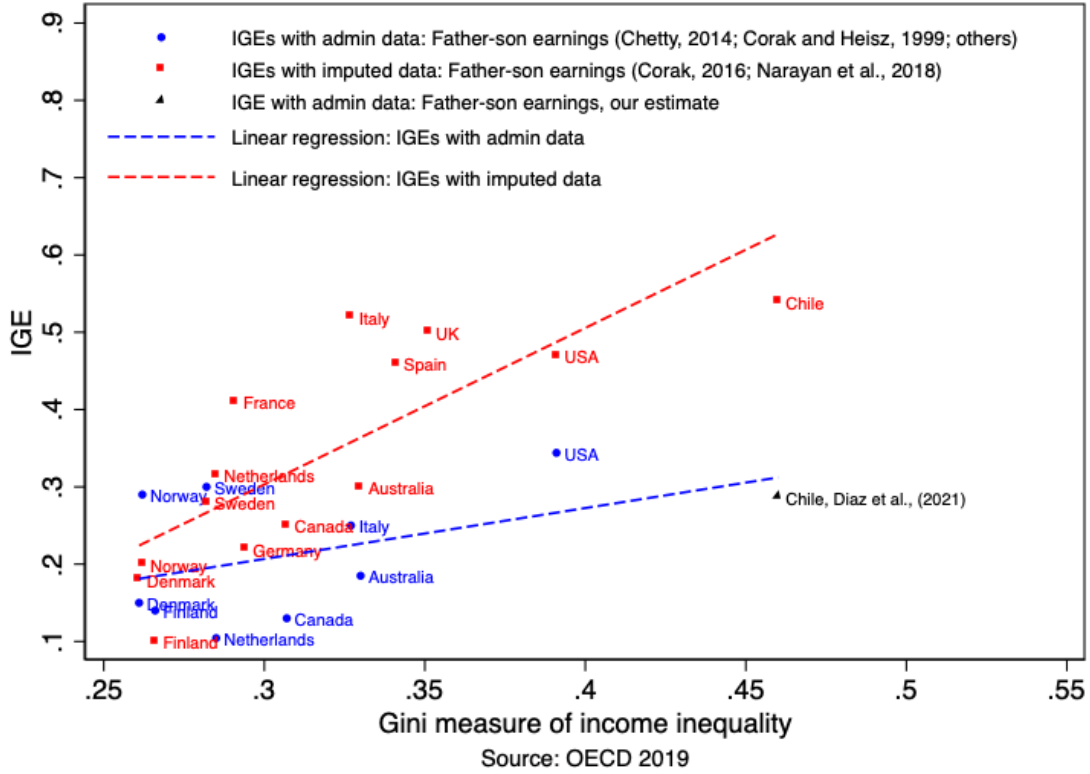
Figure 2.3: Gatsby Curve for selected OECD countries[†]

Note[†]: Gini index and IGE estimates are shown in the x-axis and y-axis, respectively. Estimates obtained from administrative data are represented by red squares (Corak and Heisz, 1999; Lindahl, 2008; Carmichael et al., 2020; Jantti et al., 2006; Deutscher and Mazumder, 2020; Chetty, Hendren, Kline and Saez, 2014; and Acciari et al., 2019). Blue circles show estimates that come from survey data (Corak, 2013; and Narayan et al., 2018). The black triangle is our estimate. Linear regressions of IGE on GINI inequality for both direct administrative data and TSTSLS survey estimates.

significantly (See Figure 4.4). This brings an IGE much closer to the administrative data' estimate. The same reason may explain the discrepancies among different countries. For instance, for Italy, IGE estimates through TSTSLS are around 0.5 (Barbieri et al., 2020; Mocetti, 2007; and Piraino, 2007). When using administrative data to link parents and children, the estimated IGE falls to around 0.25 (Acciari et

because they are drawn from the same distribution $\kappa(j) \equiv \frac{var(\hat{y}^p(j))}{var(y^p)} = R^2$ and plim $\hat{\beta}^{TSTSLS}(j) = \frac{\text{plim } \hat{\beta} - \eta(j)}{\kappa(j)}$.

al., 2019). Although Chile and Italy have different economic realities and levels of income inequality, the ratio between estimates from the TSTSLS and administrative data are approximate 2. The fact that TSTSLS tends to overestimate IGE seems to be common across countries, the US, Canada, and Australia have ratios of 1.36, 1.92, and 1.62, respectively (Corak and Heisz, 1999; Chetty, Hendren, Kline and Saez, 2014; Acciari et al., 2019; Corak, 2006; Narayan et al., 2018). The magnitude of TSTSLS bias is implicit to each country due to the Mincer equation's predictive power. Then, previous estimates using TSTSLS and, more importantly, the partial ordering across countries might be biased. That is why estimates from administrative linked data may reduce that bias modifying the levels and the relative positions. We believe that our procedure to improve TSTSLS method could improve what we know about intergenerational mobility when administrative data is not available because until this work, all estimates of intergenerational mobility in Latin America have relied on the TSTSLS estimator (Dunn, 2007; Ferreira and Veloso, 2006; and Grawe, 2004).[8]

2.7   Conclusions

In the absence of administrative data to link parents' and their children's earnings, measuring intergenerational mobility is a challenging task. In this paper, we show that relying on estimates from the commonly used TSTSLS leads to biased measures of intergenerational mobility. Unfortunately, in most developing counties such administrative data are not available and thus the literature on intergenerational mobility in these countries have relied on TSTSLS methods providing unreliable results.

To understand the magnitude of this problem, we estimate the IGE and rank-rank correlation using administrative data for Chile. These are the first estimates

---

[8]The exception being Leites et al. (2020), who study intergenerational mobility for Uruguay using administrative records.

that use linked earnings data for sons and their parents. We find that the IGE is 0.282 and rank-rank correlation is 0.239. We then estimate these measurements through TSTSLS. We show that the TSTSLS estimate for IGE is about twice of our administrative data estimate, while the rank-rank estimate is slightly lower. To the best of our knowledge, this is the first paper that compares the results provided by TSTSLS and adminsitrative records.

We show that the TSTSLS bias can be mitigated by improving the prediction power of the Mincer equation and using stochastic imputed regression to better match the spread of earnings. However, we show that these steps are very important for the IGE estimate, but less relevant for the rank-rank correlation, which remains with a relatively low bias, regardless of the specification of the Mincer equation. Our analysis suggests the use of rank-rank correlations as the measure of intergenerational mobility when data limitations are present.

Chapter 3

Effects of Teachers' Quality on Tertiary education Attendance: Evaluation Tests and Value-Added

3.1 Introduction

The existing literature on teacher effectiveness suggests that, after controlling for school and class variables, teachers are more important to student learning than any other factor (Rivkin, et al. 2005). However, knowledge about what works in teacher effectiveness is less well understood, as teachers significantly vary in their ability to improve students' performance (Hanushek and Rivkin, 2010; Grossman et al., 2013). Teacher effectiveness measures have been introduced to differentiate teacher performance better, such as teacher evaluation instruments based on elaborated protocols and teacher value-added measures (VAs).

Regarding the former, they are a set of teacher evaluations based on protocols (such as CLASS, MQI, PLATO, and FFT) that have been developed around the world to improve teachers' effectiveness based on the idea that teacher evaluation can be a way to improve teachers' performance, either by making it possible to provide

them with useful feedback or by creating incentives to implement better practices (Isoré, 2009; Taylor and Tyler, 2012). Other authors, such as Wyness et al. (2018), have questioned teachers' evaluations, arguing that teacher observation and feedback cannot solve the policy maker's problem of vast variations in teacher effectiveness in the absence of teacher incentives and nonpeer feedback and thus cannot be used to reduce differences in teacher effectiveness significantly. Additionally, teachers' evaluations take many different forms across the world and vary greatly in terms of resources involved per teacher. For example, observation protocols used in research and practice, such as the FFT, the MQI, or CLASS, all broadly capture a teacher's effectiveness in delivering quality instruction. However, each evaluates teachers on different subjects and classroom practices. These characteristics have led to no consensus on what a good evaluation system should be and how intensive it should be (Isoré, 2009; OECD, 2013a and 2013b; Jackson et al., 2014).

On the other hand, value-added measures circumvent the need to identify specific teacher characteristics related to quality and shift the focus to identifying overall teacher contributions to learning. However, it introduces additional complications and has sparked an active debate, especially about validity and reliability measures (see, for example, Chetty et al., 2014a and 2014b; Rothstein, 2010; Koedel et al., 2015; Bau and Das, 2020, among others). The traditional value-added approach rests on the assumption of selection in observables, which depends on the available information since the method consists of isolating the contribution that each teacher has in the test score of their students from the residualization of the test score once eliminating the effects of the rest of the observable variables that affect academic performance in addition to the teacher's quality. The challenges are data availability, measurement precision and bias, and dimensionality of what is understood by teaching quality. Issues that have generated debate because of the consequences of these results on teachers, such as dismissal from school.

Despite this intense and ongoing debate and as suggested by Hanushek and Rivkin

(2010), the key policy question is whether the value of even flawed value-added measures could advance the current system of personnel decisions that relies on limited information about teacher effectiveness. To answer this question, it is important to understand that good teaching is multidimensional and that each instrument of evaluation may vary in what dimension of teacher performance they measure. To date, the most commonly used instruments in the US and abroad have been value-added and teacher evaluation, and in particular, a combination of both, mainly because it is not completely clear what dimension of teaching they are truly measuring (Chin and Goldhaber, 2015).

Few studies have investigated the relationship between these two instruments, concluding that the relationship between the two is "modest" or "weak" (e.g., Kane et al., 2013; Lynch et al., 2017). In particular, the usual correlation between these two measures ranges between 0.1 and 0.3. These findings contradict what many scholars and practitioners might expect because theory and intuition suggest that teachers' strong instructional practices should improve student test performance. In part, this low association is because both measures differ in how valid they are in capturing good teaching (Chin and Goldhaber, 2015; Harris, 2012). Other authors have pointed out that given the multidimensionality of teaching quality, since the multiple measures approach helps create a composite that is more representative of stakeholders' value, validity and reliability would improve (Darlin-Hammond et al., 2015; Harris, 2012). To the best of our knowledge, no comparison of their medium-run effects has been made between these two instruments.

In this chapter, we contribute to the literature by measuring teachers' quality using ED and VA for the same sample of teachers and estimating the effect of both measures on the probability of tertiary education attendance of their students in Chile. We do this by using a novel education data set constructed by merging several Chilean administrative data sets for 2011-2017. Our results suggest that teachers' quality based on ED is not related to those based on the VA approach, with a

correlation between these two measures of -0.02. Our results also suggest that both instruments are helpful predictors of tertiary education attendance: An standard deviation (SD) increase of 1 in a teacher's true VA test score in a single grade increases the probability of tertiary education attendance by 0.6-0.7 percentage points, i.e., an increase of 1.3%-1.5% with respect to mean tertiary education attendance in public schools. An SD increase of 1 in a teacher's evaluation in a single grade increases the probability of tertiary education attendance by 1.7-1.9 percentage points, i.e., an increase of 3.5%-4.1% with respect to mean tertiary education attendance in public schools, where two (portfolio and external references) out of four parts of the ED are the best predictors for graduate students' tertiary education attendance.

Our results are consistent with the argument of the multidimensionality of teaching quality. They suggest that a combination of several pieces of information, such as those provided by value-added and teacher evaluation, should be used to assess teachers' quality. However, given that teacher evaluation is expensive and time-consuming, additional instruments that increase both validity and reliability are not worth the cost beyond a certain point. Although this recommendation of combining ED and VA would be novel for a developing country, it is not new since these elements have been used in most states in the US and several developed countries (Finland, England, France, Canada, Portugal, Singapore, among others). The challenge for poor and less developed countries is twofold. To assess teacher quality, teachers should invest in technological systems to measure and store information on teachers and their students. They should design teachers' evaluation instruments based on their contexts, as well as evidence related to each instrument's effectiveness regarding this type of evaluation.

The structure of the chapter is as follows. Section 2 presents a summary of the discussion regarding value-added models and classroom observation protocols, while section 3 describes Chile's institutional background. Section 4 presents and describes the data sets used in our analyses as well as some summary statistics.

Section 5 presents the empirical approach and results for value-added methodology, while section 6 does the same for teacher evaluation methodology. Section 7 shows a comparison of both previous results. Finally, section 8 concludes.

3.2 Literature Review

Some teacher effectiveness measures have been introduced to measure teacher quality, such as value-added models (VA), teacher evaluation based on elaborated protocols, and student surveys[1]. Critics of the use of VA for assessing teachers' effectiveness have pointed out several concerns. The most prominent is the potentially biased estimates obtained by VA (i.e., lack of validity). Rothstein (2010) conducted falsification tests and showed that standard value-added models suggest implausible and large future teacher effects on past student achievement and conclude that the assumptions underlying common value-added models are substantially incorrect and raise concerns about the potential for bias due to selection on unobservable student characteristics[2]. Darling-Hammond (2015) also criticized VA models arguing that VA is useful only if several assumptions hold, such as i) student learning is well measured by tests, ii) students are randomly assigned to teachers within and across schools, and iii) individual teachers are the only contributors to students' learning over the period used for measuring gains. She states that these assumptions rarely hold in real life, as suggested by Rothstein (2010)'s study. Angrist et al. (2017) test for VA bias using a procedure that asks whether VA estimates accurately predict random assignment achievement consequences to specific schools (i.e., lotteries), finding that conventional VA model estimates are biased.

---

[1]Due to the heterogeneity of teacher evaluations, we use the term teacher evaluation to include teacher test, classroom observation, student feedback, portfolio and other practices that may be included in the evaluation.

[2]He develops a falsification test for several widely used value-added models based on the idea that future teachers cannot influence students' past achievement. He finds that teachers' scores have large effects on lagged students' gains.

On the other hand, as Koedel et al. (2015) explain, several subsequent studies raise concerns about Rothstein's tests and how he interprets his results (see, for example, Goldhaber and Chaplin, 2015 and Guarino et al., 2015). They show theoretically and through simulations that the test proposed in Rothstein will often falsify VA models that are unbiased and fail to falsify VA models that are biased. Along the same line, Kane et al. (2008) and Kane et al. (2013) use a random assignment experiment to evaluate various nonexperimental methods for estimating teacher value-added. They find that teacher value-added is a significant predictor of student performance under random assignment. In nonexperimental data, those that control for lagged student test scores yield unbiased predictions, and those that further control for mean classroom characteristics yield the best prediction accuracy. Thus, they conclude that selection on unobservable characteristics is small (i.e., not significantly different from zero).

Chetty et al. (2014a) developed a quasi-experimental design that exploited teacher turnover at the school and grade level for identification. They found that changes in student test scores were consistent with what would be expected when teachers changed schools based on the teacher's prior VA scores. Furthermore, they examine the scope for teacher VA bias, showing that teacher value-added from models that include students' lagged test scores is not meaningfully biased by student-teacher sorting along observed or unobserved dimensions. Bacher-Hicks et al. (2017) provide evidence similar to Chetty et al.'s (2014a) results in that value-added measures are unbiased predictors of teacher impact on student achievement following random assignment. Similarly, Bau and Das (2020), with data from Pakistan, showed that VA measures produce unbiased and reliable estimates of teacher quality in developing countries.

The validity of teacher evaluation measures is less clear. For example, the MET study[3] finds positive effects of classroom observations under some circumstances;

---

[3]The Measures of Effective Teaching (MET) is a three-year experimental study designed to

however, the classroom observation included in the teacher evaluation of the MET study involved highly trained observers, something unlikely in everyday school settings. Furthermore, as Harris (2012) documents, there is older evidence that suggests that classroom observations can be influenced by factors unrelated to performance, such as age and race. Additionally, it seems likely that classroom context will affect observation measures (e.g., it may be difficult to make valid comparisons between the classroom management skills of a teacher who has emotionally impaired students and is subject to frequent disruptions, to the skills of a teacher whose students are less disruptive). Furthermore, with teacher evaluation, if the observer knows the teacher personally, we might worry that she is hampered in her ability to judge that teacher's performance objectively. Additionally, some school principals and other observers may simply not know what to look for when assessing classroom practice; this, too, can introduce bias.

Further complications for teacher evaluation validity are the differences in their design, protocols, timing, format, intensity, and resources used in different US states and countries. For example, there are several protocols for teacher evaluation (e.g., FFT, the MQI, or CLASS), and each of them evaluates teachers on different subjects and classroom practices (Chin and Goldhaber, 2015). Even if the protocol is the same between teacher evaluations, the number of instruments used may differ (classroom observation, written tests , principal or peer evaluation, among others). Even the intensity or the format may differ. For instance, the evaluation program in Cincinnati public schools, which has been studied extensively, has been found to have a positive effect; however, it involves a total budget of approximately US $7,500 per teacher evaluation, while on the other hand, a system of light-touch evaluations, such as the one in France, appears to be much more cost-effective. Briole and Maurin (2019) study secondary school teacher evaluations in France and estimate the evaluation system's cost at approximately US $110 per teacher per year. Regarding the format,

---

determine how best to identify and promote great teaching.

some teacher evaluations in the US included physical visits for classroom observation. At the same time, Chilean evaluations are scheduled at a well-known advance date so that teachers can prepare a great class for that day. These issues may explain the heterogeneous results of empirical studies that analyze teacher evaluation.

On the other hand, some studies find positive effects, such as Pianta et al. (2006), who investigate the relationship between teacher scores in the Classroom Assessment Scoring System (CLASS) and students' social, emotional, and academic achievements, finding that emotional support is as important as the quantity of math instruction to improve math achievement. Kane et al. (2011) present estimates of the relationship between specific classroom practices, as measured by the Cincinnati Public Schools' Teacher Evaluation System (TES), and student achievement gains[4]. The authors find that a 1-point increase in the overall classroom practices score predicts student achievement growth of between 0.10 and 0.14 standard deviations in math and reading, respectively. Furthermore, they find that among students assigned to different teachers with similar overall classroom practice scores, math achievement will grow more for students whose teacher is relatively better than their peers in classroom management. Specifically, for math, the classroom environment's coefficient relative to instructional practices is approximately 0.08 standard deviations.

Other authors have questioned the usefulness of teacher evaluations based on protocols. Wyness et al. (2018) conclude that two of the most common instruments used in teacher evaluations (teacher observation and feedback) cannot solve the policy maker's problem of the vast variations in teacher effectiveness. The authors suggest three potential reasons why they found different results than the studies mentioned above. First, Cincinnati teachers (used in previous studies) were formally scored, with the results carrying explicit consequences, including impact on promo-

---

[4]The TES scoring rubric is based on Charlotte Danielson's Enhancing Professional Practice: A Framework for Teaching (1996).

tions, tenure, and potential nonrenewal of the teacher's contract. Wyness et al.'s
(2018) study did not have such consequences and was designed purely to improve
teacher performance through discussion and feedback. Second, previous research in-
volved filming observed classes and being observed by nonpeer experts. This may
have had an unintended effect of encouraging students to behave differently when
filmed and observed, affecting test scores. It may have also resulted in more accuracy
in the observation process in Cincinnati, with teachers able to refresh their memory
of what they observed after the fact. Moreover, having an external expert there pro-
viding feedback may have resulted in more informative, accurate, and transparent
feedback to teachers. Third, Wyness et al.'s (2018) study took place under experi-
mental conditions, which overcomes issues typical of quasi-experimental studies. For
example, in Wyness et al. (2018), there is no difference in the characteristics of
treated versus untreated teachers; in the Cincinnati setup, younger teachers were
evaluated first, which could result in upward bias of the results if younger teachers
have higher growth in value-added than older teachers. Because of these results,
Wyness et al. (2018) pointed out that teacher observation may not be effective un-
less coupled with incentives and external evaluators and cannot be used to reduce
differences in teacher effectiveness significantly.

Apart from the bias (validity) ongoing debate, there is a second important concern
that stresses that value-added estimates are unstable (i.e., unreliable) when based
on a relatively small number of students, thus requiring several classes of students
to reduce the measurement error (Hanushek and Rivkin, 2010). Regarding this
point, several studies provide evidence about the stability of estimated teacher value-
added over time and across schools and classrooms (see, for example, Aaronson et
al., 2007; Chetty et al., 2014a; and Glazerman et al., 2013). Furthermore, Koedel
et al. (2015) point out that one factor that affects the stability of value-added
estimates is whether the VA includes fixed effects for students or schools. This is
because adding these layers of fixed effects narrows the identifying variation used to

estimate teacher value-added, which can increase imprecision in the estimation[5]. The correlations of adjacent-year value-added measures estimated from models without school or student fixed effects range between 0.47 and 0.64 (Goldhaber and Hansen, 2013), while those from models with student fixed effects are 0.29 (McCaffrey et al., 2009) and those with school fixed effects are between 0.18 and 0.33 (Goldhaber and Hansen, 2013). Moreover, the exclusion of student or school fixed effects should not bias the coefficients, as shown by Chetty et al. (2014a)[6]. Similarly, the MET study reports stability for teacher value-added measures of approximately 0.3 to 0.5 when three years of data are used. Critics of value-added measures argue that limited stability against using value-added. However, when the alternative is used for comparison, VA models do not perform too poorly. As Harris (2012) states, a single classroom observation has lower stability than a value-added measure, but a combination of four classroom observations yields higher stability of approximately 0.65. Thus, it depends on what the VA is compared to.

For the US, the MET project demonstrates that it is possible to identify great teaching by combining classroom observation protocols (e.g., CLASS, PLATO, and FFT), student achievement gains (VA), and student surveys. One important recommendation of this project is that many school districts in the US currently require a single school administrator's observations. Thus, the project recommends averaging observations from more than one observer, such as another administrator in a school or a peer observer. This suggestion regarding the use of several instruments to capture teacher quality has been followed for several US states and developed countries, which has generated the appearances of a few recent studies that try to study the association between these two ways of capturing teacher quality, finding that the correlation between the two is weak (see Bell et al., 2012; Grossman, Loeb, Cohen,

---

[5]In models with student fixed effects, estimates of teacher value-added are identified by comparing teachers who share students, while in models with school fixed effects, the identifying variation is restricted to occur only within schools (Koedel et al., 2015).

[6]They show that VA models without student and school fixed effects have no statistically significant biased estimates as long as they control for lagged students' test scores.

and Wyckoff, 2013; Hill, Kapitula, and Umland, 2011; Kane et al., 2011). These weak results and the limited budget constraints of poor and less developed countries make the cost-benefit analysis of these instruments very relevant, not to discard one instrument but to accordingly prioritize if needed.

The determinants of the weak correlation between the instruments are not clear, mainly because true teacher quality is unobservable. However, there are some attempts to explain it. Chin and Goldhaber (2015) use simulations to study three possible scenarios that result in weak correlations between value-added and teacher evaluations. The first is that one or both measures could provide unreliable estimates of one or more dimensions of teacher quality due to sampling error. The second is that teacher quality may be multidimensional, and the measures provide reliable estimates of different dimensions of teacher quality. The third is that one or more of the measures may be invalid because they do not provide a reliable estimate of any dimension of teacher quality. They find that the simulations did not allow them to rule out any of the scenarios for the weak correlations seen in prior research. Chaplin et al. (2014) analyze the case of the Pittsburgh Public Schools teacher evaluation system, which includes teacher observation based on protocols, value-added measures, and student surveys. They find that i) all three measures can differentiate among teachers and ii) correlations among measures suggest they are valid and complementary.

3.3   Institutional Background

3.3.1   General Structure of the Chilean Educational System

The Chilean educational system consists of four educational levels: preschool (0 to 4 years old), primary school (kindergarten and 1st to 8th grade), secondary school (9th to 12th grade), and tertiary education. Primary school and secondary school (high school) are mandatory, while preschool and tertiary education are optional.

There are 11 subjects in each mandatory level, including mathematics, language, and science. The grading scale for each subject corresponds to a numeric scale that ranges from 1.0 to 7.0. A grade of 4.0 or higher and a minimum attendance of 85% are required to pass a subject. For being promoted to the next grade, students must pass all subjects; however, a student could still be promoted if she fails one subject but has an average grade across all subjects of 4.2 or higher or if she fails two subjects but has an average grade across all subjects of 5.0 or higher.

It is important to note that students' grades are not used to determine teacher quality or give the educational community any incentive but are only used for grade promotion or admissions to other schools or tertiary education.

Since the decentralization of the educational system in 1980, the Chilean educational experience has been one of the most extreme cases of the introduction of market-oriented reforms at a national level (e.g., universal parent school choice, voucher schools, copayment system, standardized measurements, and multiple for-profit and not-for-profit private schools). The Chilean educational system considers three kinds of administrative alternatives: public establishments under municipal administration (i.e., public schools); private subsidized establishments funded by a voucher system and administered by the private sector (i.e., voucher schools); and private fee-paying establishments funded and administered by the private sector. Regarding students' distribution, approximately 40% of students are in public schools, 8% are in private, non-voucher (i.e., fee-paying) schools, and 52% are in private voucher schools.

The introduction of the voucher system gave parents complete freedom to choose schools for their children. Essential for this decision was introducing a standardized census-type test for all schools and students in the country. This test is known as the SIMCE and covers mathematics and language. The mere existence of this national test, along with the fact that schools' results are made public, introduces an element of competitive pressure into the system, as parents have objective indicators of results

to assess educational school outcomes.

Unlike voucher schemes implemented in other countries, private voucher and non-voucher schools in Chile can choose their students; however, public schools are prohibited from choosing, except in those cases where the demand for seats exceeds availability. This scheme, where private schools can select students, generates a positive sorting of students from high- and middle-income families into private schools and most vulnerable students into public schools (as found by Contreras et al., 2010). One of the reasons for this is that private voucher schools in Chile can operate for profit and may, therefore, select students who are less expensive to educate. This resulted in high segregation between private and public schools in Chile (Valenzuela et al., 2013).

Regarding job contracts, teachers in public schools are governed by the Teacher Statute, which is legislation that includes a centralized collective-bargaining process, wages based on uniform pay-scales[7] with special bonuses for training, experience, and working under difficult conditions, and strong restrictions on dismissals. On the other hand, private schools (in both voucher and non-voucher schools) operate as firms, and their teachers come under the labor code, such as all other private-sector workers.

In the context of a market-oriented educational system and parental school choice, school quality becomes crucial. The OECD's historical data suggest that Chile's student learning outcomes have been considerably below the OECD average. Furthermore, students' results differ considerably across the socioeconomic groups and type of school attended. In this context, the government accorded significant importance to teacher evaluation and generated conditions to establish a national evaluation test as described below.

---

[7]This is maintained prior to the reform of the Professional Teaching System (Sistema Profesional Docente in Spanish), which takes effect in July 2017 and grants differentiated payments using as a factor, among others, the result of the teacher evaluation.

3.3.2 Teachers Evaluation Test

The Evaluación Docente (ED) was introduced in 2003 as a pilot program. Since 2004, it has been mandatory for public school teachers and optional for nonpublic school teachers. The Ministry of Education is responsible for the implementation of this test.

To evaluate teacher effectiveness, a technical committee composed of the Chilean Ministry of Education, the Association of Municipalities, and the National Teachers Union established the first set of effective teaching standards based on the Framework For Teaching (FFT) protocol presented by Charlotte Danielson. This set of standards is summarized in four domains, known as the Marco para la Buena Enseñanza (MBE) (see Table 4.8). Each of these domains has several criteria (19 in total), and each criterion contains several descriptors (71 in total). Finally, each descriptor is disaggregated into observable elements of teaching practices that are measured by four different instruments on the teachers' evaluation test (ED). These instruments are a) self-assessment, b) peer interviews, c) external references, and d) portfolio. The mapping between the instruments and the observable elements of the descriptors is established using a double-entry matrix. Some descriptors are evaluated with a few instruments, while others are evaluated with several instruments. As we mentioned above, this is helpful, as there is a considerable agreement regarding the use of several instruments to evaluate teacher effectiveness due to single instruments hardly capturing all aspects of teacher quality (Grossman et al., 2013; Manzi et al., 2011).

The average annual cost per teacher for the period of analysis (2012-2018) is US $400, covering approximately 20% of the total number of teachers in the public sector, which includes preparation of the instruments, application, review, and payment for peer interviewers, among others. The total cost for 2019 was US $10.2 million for the coverage of 21 thousand teachers, i.e., an average annual cost of US $461[8].

---

[8]The information corresponds to the execution of the Desarrollo Profesional Docente, assignment that consolidates the resources for the teacher evaluation obtained from the Budget Office, http://www.dipres.gob.cl/598/w3-channel.html. The number of teachers evaluated corresponds to

3.3.3 Teachers' Classification

Each descriptor has a language that describes performance at each level of the rubric: Distinguished, Proficient, Basic, and Unsatisfactory, with an evaluator assigning the respective scores of 4, 3, 2, and 1 to these rubric levels.

For each instrument (except the portfolio), the score is calculated by averaging all descriptors associated with each criterion and then averaging all the criteria to obtain each domain's score. Finally, the score of the instrument is simply the average score of the domains. Therefore, all domains have the same weight in the final score of each of these instruments.

In the case of the portfolio, the MBE domains are operationalized in different dimensions. There are 7 dimensions in the portfolio (A, B, C, D, F, G, and H), as presented in Table 4.10. The score of the portfolio is obtained by averaging the scores of all dimensions. Each dimension has the same weight in the final score of the portfolio. Each teaching practice is evaluated with a score between 1-4 (from unsatisfactory to distinguished) to obtain the dimension's score.

Each of these four instruments is weighted (in different ways depending on how many times a teacher has taken the test) to obtain the final score (which we will use in this chapter to measure the impact of TET on tertiary education attendance). For first-time takers, Portfolio weights 60%, Peer Interview 20%, Self-Evaluation 10%, and External References 10%. With the weighted overall final score, teachers are evaluated in one of the following categories: i) unsatisfactory (1-1.75 points), ii) basic (1.75-2.5 points), iii) proficient (2.5-3.25 points), and iv) distinguished (3.25-4 points).

It is important to point out that the ED test is taken every four years for teachers who were classified as distinguished or proficient. Those classified as basic must take training plans to overcome their weaknesses (funded by the Ministry of Education) and retake the test in two years. For those classified as unsatisfactory, the test must

---

information obtained from the page of the Ministry of Education, http://datosabiertos.mineduc.cl/.

be taken the following year with a different weighting scheme. In this case, Portfolio weights 80%, Peer Interview 10%, Self-Evaluation 5%, and External References 5%.

If the teacher is again classified as unsatisfactory, the person is fired. If the teacher is instead classified as Basic in the second test, the teacher will have a new opportunity the following year. The only exemptions from the ED are teachers who are close to retirement (three years or less from retirement age).

The main benefits of distinguished and proficient are that these categories allow teachers to have preferential access to professional development opportunities, such as visiting overseas, promotions, academic seminars, and becoming workshop tutors. Furthermore, they have access to the possibility of increasing their wages, known as the Asignación Variable por Desempeño Individual (AVDI) and the Asignación de Excelencia Pedagógica (AEP), depending on their performance on a written test about discipline and pedagogical knowledge.

3.3.4 Classroom Observations: Instruments

Self-Evaluation

This instrument consists of a series of questions that have the goal of making teachers reflect on their pedagogical practices, evaluating the quality of their relationship with the students and their parents, recognizing the quality of their performance in the classroom, and identifying their strengths, weaknesses, and need for professional development. Teachers also receive an example of the very specific rubric used for this instrument. Each teacher must evaluate each of the questions with the rubric (distinguished, proficient, basic, or unsatisfactory) and submit them online after it is completed.

An example of the general concept used to construct the rubrics (common for all the indicators in all the instruments) is:

1. Unsatisfactory: means that the evaluated teacher presents clear weaknesses in

their performance evaluated by the indicator.

2. Basic: means that the evaluated teacher fulfills the expected performance, although in an irregular way.

3. Proficient: means a correct performance, a level that fulfills what is requested by the indicator, although not exceptional. This is the expected performance.

4. Distinguished: means that the teacher has a clear and consistently better performance concerning what is expected by the indicator.

Peer Interview

This instrument consists of an interview by a peer, another teacher who works under similar conditions (known as Evaluador Par (EP)). The EP is specially trained for this task. Training includes the application of the interview and the rubric to ensure a standardized evaluation. Questions included in this instrument are based on the MBE mentioned above and can be classified into three parts: i) general information about the evaluated teacher and the interviewer, ii) questions regarding performance in the classroom, and iii) questions regarding the context under which the performance is done. The EP should register every answer, and immediately after the interview, the EP must evaluate each of them with the rubric (distinguished, proficient, basic, or unsatisfactory) and submit them online after its completion. This interview takes approximately 60 minutes.

External References

This instrument consists of an external evaluation from the teacher's two hierarchical superiors (who in general are the principal of the school and the chief of the pedagogical unit of the school (UTP)). This instrument is a very precise and structured guide based on the MBE that includes several questions about the evaluated teacher's performance. For each question, the principal and the chief of the UTP sep-

arately must assess the performance of the evaluated teacher using the same rubric of the two previous instruments (unsatisfactory, basic, proficient, and distinguished) using the values 1-4, respectively, and submit each score online after its completion. The results for this instrument are obtained by averaging both scores.

Portfolio

This instrument is an evaluation where teachers must present evidence of their pedagogical practice. The portfolio must be done in the teacher's subject (and the one registered in ED). Teachers have 12 weeks for the elaboration of the following material separated into two modules:

Module 1: It includes two products. First, the design and implementation of an 8-hour pedagogical unit. In this product, teachers must develop clear and specific goals for the pedagogical unit and its classes. Additionally, they have to describe each of the classes implemented within the unit, indicating the date, duration, realized activities, and used resources, among other things. Additionally, teachers must answer some questions about a) their experience implementing the unit and b) their pedagogical performance in the classroom.

Second, teachers must present an evaluation of the student's learning in that pedagogical unit. The goal is to gather information regarding what students have learned in that unit. If, for example, a written test was used to evaluate students, the teacher must send a copy of that test with the correct answers or the appropriate criteria used to evaluate each test's answer. Additionally, teachers must answer questions related to a) their experience in applying the evaluation and b) their analysis and use given to the results (such as students' feedback and the teacher's feedback for improving their practice, among others).

Module 2: Includes only one product: a complete 40-minute video recording

(without cuts or interruptions) of one of the classes the evaluated teacher usually works with. The teacher then completes a form with information relative to the recorded class. Additionally, the evaluated teacher must attach a photocopy of the resources used in that class.

For the portfolio, each evaluated teacher receives a very detailed instruction manual to elaborate on each of the requested products so that every teacher delivers the requested outputs in a standardized format. The class's recording is the responsibility of a trained cameraman, and it is free of charge for the evaluated teacher. Additionally, there are several measures in place to ensure sound and image quality. The cameraman has specific instructions so that the recording must clearly show the teacher and the students to reflect what happens in the classroom.

From these two modules, 7 dimensions are obtained (see Table 4.10). Each includes the indicators of teaching practices with rubrics based on the MBE. Later, the modules are evaluated by specially trained peers designated by the Ministry of Education. In their evaluations, these peers must follow very strict rules to guarantee an objective and blind process, including a blind double evaluation and score recalibration in rare cases, among several other processes (see Chapter 2 of Manzi et al., 2011 for full details).

3.4 Data

In this chapter, we build a new and unique longitudinal data set based on an administrative database of educational records of 3.4 million Chilean students. For each student, we have nine years of data, from 4th year of primary school to 12th year (last grade) of secondary school, as well as detailed records of all the teachers.

The data were collected from administrative records of the Education and Finance Ministries of Chile. The resulting data set covers the period 2011-2017 and includes (i) all students who were enrolled in the Chilean educational system between 2011

and 2016, for whom we add their tertiary education records for the period 2012-2017; (ii) all teachers who were instructing in the Chilean educational system between 2011 and 2016; and (iii) an identifier that allows us to link students with their teachers in each grade and subject for the period 2011-2016.

The administrative records used to build the longitudinal data set that we use in this chapter and the variables included in it are described below.

School history information of students for the period 2011-2016

This data set has 6 years of data for 3,415,100 students enrolled in any of the 9 grades that we look at (4th primary year to the last year of secondary school). Importantly, this data set contains the census of students in the Chilean school system (public and private), which means an average of 234 thousand students by cohort.

These administrative records were provided by the Chilean Ministry of Education and included reliable information on the grades that students obtained throughout its academic history between 2011 and 2016. This information allows us to know students' final grades from a specific year-school-grade-class-subject perspective. For our analysis, we focus on language (Spanish) and mathematics. Moreover, the administrative records contain information on the final grade for each school year, the final student's situation (promoted to the next grade, repetition of grade or school transfer), gender (1=woman), vulnerability condition (1=if the student belongs to the most vulnerable 40% of the population), geographic location of the school, administrative dependency (public, private voucher or private non-voucher), and rurality condition (1=urban).

Based on this information, we construct variables that indicate school and classroom sizes and their composition in terms of sex, vulnerability proportion, and a

grade average of the previous and contemporary classes.

Information on teachers for the period 2011-2016

We use two sources of administrative records about teachers. The first corresponds to the yearly teacher census, which includes a characterization of the whole teacher universe in the system in each period of study. In total, we follow 84,719 unique teachers. This database has information on the year-school-grade-class-subject in which the instructor taught, allowing us to link it with each student's grades. Furthermore, it includes data on each teacher's characteristics, such as gender (1=woman), age (years), pedagogical hours, and their role in the school, which can be teaching or other (e.g., managerial duties).

Importantly, the teachers' administrative records can be linked with the teachers' evaluation tests described in section 3.3.2. Since these evaluations are mandatory for the public system, we have a full record of public school teachers' results for 2011-2016.

Information on tertiary education enrollment

We use an administrative database that provides information on the enrollment of students in tertiary education institutions. This data set allows us to identify each high school graduate's tertiary education enrollment status, i.e., whether they enrolled after graduation or not. Additionally, for the students who entered tertiary education, the database indicates which type of institution they enrolled in, differentiating between universities (that offer high-level professional and technical degrees), technical formation centers (that offer only high-level technical degrees), and professional institutions (that offer professional and technical studies that do not lead to academic degrees). Both modalities, technical formation centers and professional institutions will be called Vocational Education.

Based on students who graduated from high school between 2012 and 2017, 1,163,343 students are represented in the database through 7,988,659 observations. We observed the entire academic school record for each student who graduated. Of these, 615,977 students were enrolled in tertiary education immediately after graduating, e.g., an attendance rate of 53%.

Table 3.1 summarizes the final database that results after merging all the information described above. This panel data set has 25,286,565 observations for 6 years (3.4 million students with one observation per year for the 2012-2016 period). Half of the students have data for their grades in the language subject, and the other half have data for mathematics. Additionally, students are distributed across 9,590 schools, of which 84,719 teachers taught.

Variables of Dataset	Observations
Years	6
Grades (by year)	9
Subject (by year)	2
Students' Cohort (by year, grade and subject)	234,135
Total Observations	25,586,565
Unique Students	3,415,010
Schools	9,590
Public Schools	5,270
Subsidized private Schools	3,862
Private Schools	458
Teachers	84,719

Table 3.1: Summary Statistics for Dataset.

Table 3.2 presents the descriptive statistics of the variables studied. The mean classroom size is 32.9 students per class, while the mean number of students per school is 233.5, with the distribution between girls and boys being equitable. Moreover, 42% of students are vulnerable, and 41% of the schools are in rural areas. Regarding grades, the mean grade is 5.1 (standard deviation of 0.86), and 6% of students present grade repetition at some point in their schooling life.

As seen in Table 3.2, teachers present a larger share of women than men (71% vs. 29%), their mean age is 42.9 years old (12.1 years standard deviation), and teaching is the primary function of most of them.

3.5 Teachers' Value-Added Estimation

3.5.1 Conceptual Framework

To estimate the value-added of a teacher, we follow the methodology developed by Chetty et al. (2014a) and (2014b), as well as the contributions of Kane and Staiger (2008), Rothstein (2010) and Kane et al. (2013).

Essentially, the teacher's value-added is defined as each teacher's contribution to their students' academic performance. If teachers' and students' class allocation were random, estimating this contribution would not be challenging. However, it is well known that allocation is not random in Chile but rather highly selective (Lara et al., 2010; Meckes and Bascopé, 2012; and Mizala and Torche, 2012, for positive sorting of teachers between schools, Toledo and Valenzuela, 2015; and Canales and Maldonado, 2018, for sorting within-school assignment). For example, parents choose their children's schools, schools select which students can enroll in them, and school directors choose which professors the school hires and assigns them to a specific class. Therefore, if students with certain characteristics (higher skills, better grades in previous years, previous school, higher-income group, and religious beliefs) are systematically assigned to a specific type of teacher (longer experience, better performance, men instead of women, among others), not adjusting for students and teachers' characteristics will bias our value-added estimations.

To identify the professor's contribution to their students' academic achievements, one could adjust the estimations for a substantial number of measured variables before the interaction between a teacher and a student. In other words, one can assume that the teacher-student match is random conditional on a set of observables.

Variable	Mean	SD	Observations
Student Data:			
Class Size	32.94	10.40	25,286,565
Test Score	5.13	0.86	25,284,795
Language	*5.20*	*0.78*	*12,631,865*
Math	*5.06*	*0.92*	*12,652,930*
Public Schools	*5.07*	*0.87*	*9,889,014*
Subsidized Private Schools	*5.11*	*0.84*	*13,426,498*
Private Schools	*5.57*	*0.83*	*1,969,283*
Test Score SIMCE	258.41	53.46	5,672,936
Repeating grade	0.06	0.23	25,286,565
Female	0.50	0.50	25,286,563
Vulnerability Condition	0.42	0.49	25,286,565
Schools Data:			
Grade Size	239,299	13,855	54,115
School Size	233.5	295.8	54,115.0
Urban	0.41	0.49	54,115
Teacher Data:			
Female	0.71	0.45	264,250
Primary Function	0.89	0.31	264,250
Age (years)	42.94	12.13	256,936
Outcome data:			
Students Graduated from High-School			7,988,659
Students Attending Tertiary Education	53.85	49.85	7,988,659
Students Attending Vocational Education	20.08	40.06	7,988,659
Students Attending University	33.77	47.29	7,988,659
Students Attending Public University	10.53	30.70	7,988,659
Students Attending Top-3 University	1.54	12.31	7,988,659

Table 3.2: Summary Statistics for Sample Used to Value-Added Model.

More precisely, these variables will control students' characteristics, students' parents' characteristics, other teachers' and other schools' contribution to the students' academic performance, schools' student selection, directors' teachers' selection, and assignment to a class. The plausibility of this assumption depends on the quality of the variables that we control.

Under the assumption that is conditional on a set of observables, teacher-student allocations are random, and the value-added estimation is unbiased. It can be described in four stages, detailed below:

First stage: teachers' contribution to students' grades

We observe student i, who during year t is assigned to classroom c from grade g of school sch. For simplicity, let us assume that $c = c(i,t) = c(i,t,sch,g)$. Additionally, in year t, professor j teaches in school sch, in grade g and class c, for the subject s, language and mathematics. To facilitate understanding of the model, we will assume that each professor j teaches in only one grade and only one subject per year, i.e., that each instructor is assigned only to one classroom $(j = j(c(i,t)) = j(c(i,t,sch,g)))$[9].

Consider that the final grade of a student is Q_{it}, which, as noted in the previous sections, ranges from 1.0 to 7.0 for the whole school system. We standardize according to year and grade, such that for each year degree, it has a mean of zero and variance of one, obtaining A_{it}^*.

Following the methodology developed in Chetty et al. (2014a), controlling for the previous grade score and adding a fixed effect per teacher, we estimate the following equation:

$$A_{it}^* = \alpha_j + \gamma A_{(it-1)}^* + \beta X_{it} + \mu_{jt} + \epsilon_{it} \tag{3.1}$$

Let the residual student test score after removing the effect of observable char-

---

[9]The estimates incorporate the cases in which a teacher takes classes for more than one grade in a given school year.

acteristics:

Rearranging 3.1, we verify that:

$$A_{it} = A_{it}^* - \gamma A_{(it-1)}^* - \beta X_{it} = \mu_{jt} + \epsilon_{it} \tag{3.2}$$

The above is estimated separately for each subject (language and mathematics) and according to level (primary and secondary), where A_{it}^* corresponds to a standardized grade score, α_j the fixed effect per teacher, $A_{(it-1)}^*$ the standardized grade score in the immediately preceding grade. These are introduced using a cubic polynomial for the previous language and mathematics grades and interacts with the student's grade level. The vector X_{it} contains information on the student-level characteristics, including gender and grade repetition in the contemporary and previous grades, average attendance in the contemporary grade, and a set of discrete variables by year and grade.

Namely, the residuals of the grades, A_{it}, eliminating the effect of observable characteristics X_{it}, including the previous grade, would be our best predictor of the value-added of teachers, as long as it is fulfilled that $Cov(e_{it}, \hat{\mu}_{jt}) = 0$, a subject that we will deal with later.

It is important to highlight that, in the recovery of the residuals of the previous regression, these incorporate the estimated coefficients of the fixed effect of the teacher, since otherwise, we will be underestimating the impact of the value-added by residualizing the grade, including the effect that the teacher has. Viewed another way, estimating the regression without the inclusion of a fixed teacher effect would overestimate the impact of X_{it} on A_{it}^*, since the correlation between X_{it} and μ_{jt} would be different from zero. To exemplify the above, let us imagine that a student changes schools that coincides with taking classes from a high value-added teacher. In this case, if in our set of observables, we include variables related to the school and do not include a fixed effect per teacher, part of the teacher's improvement could be attributed to the school's characteristics.

Considering that we obtain unbiased teacher quality estimators, we move on to the methodology's second stage.

Second stage: Predictor of student outcomes.

From the set of values A_{it} associated with each student i in period t, we calculate the average associated with the respective teacher j in year t.

$$\overline{A}_{Jt} = \frac{1}{n} \sum_{i \in j} A_{it} \tag{3.3}$$

Performing the same procedure for other years, we obtain the vector $A_j^{-t} = (\overline{A}_{j1}, ..., \overline{A}_{jt-1})$, which corresponds to the average of the residues of professor j in different periods of t.

Chetty et al., (2014a) proposed the best linear predictor was $\overline{A}_{Jt}$, based on earlier-than-contemporary information, i.e., a prediction of $E(\overline{A}_{Jt}|A_j^{-t})$, which can be written as:

$$\hat{\mu}_{jt} = \Psi \overline{A}_{jt-1} \tag{3.4}$$

Where $\Psi = \frac{Cov(A_{jt}, \overline{A}_{jt-1})}{Var(\overline{A}_{jt-1})}$ corresponds to the coefficient that minimizes the sum of the squared errors of the prediction of academic results. The above is obtained from a linear regression between $\overline{A}_{jt}$ and $\overline{A}_{jt-1}$, as presented in equation 3.4.

The coefficient Ψ, also known as shrinkage, is aimed at correcting the temporal variation that the quality of the teacher may have. Intuitively, if a teacher's results do not change from one year to another, Ψ should have a value close to one. On the other hand, if the results change considerably, this factor will be close to zero; therefore, the extreme values will be taken to zero.

For the estimation of equation 3.4, we will consider a vector A_j^{-t}, with information for all years except the one for which we are making the prediction. Following Chetty et al., (2014a) we use information from all periods, both before and after, to predict

the value-added of period t. The values of the respective covariances between A_{it} and A_j^{-t}, vary according to the subject and level at which we are making the estimate, but in general, these range from 0.69 for $Cov(A_{it}, A_{it-1})$ to 0.46 for $Cov(A_{it}, A_{it-4})$ in the case of language and from 0.31 to 0.18 for mathematics. The covariances in primary education are higher in both cases.

Figures 3.1a and 3.1b present the distribution of value-added for each of the teachers. For the case of primary levels, the standard deviation is 0.15 for language and 0.13 for mathematics, while for the case of secondary levels, the standard deviation is 0.24 for language and 0.16 for mathematics. Both results are similar to those obtained by Chetty et al. (2014a) for the case of primary levels but higher for the case of secondary levels.

(a) Primary School

(b) Secondary School
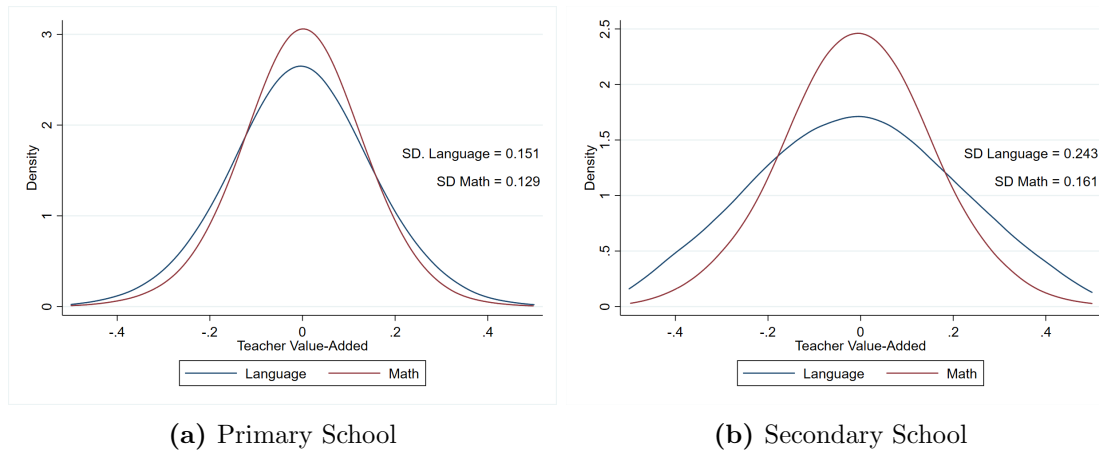
Figure 3.1: Distribution of Teachers' Value-Added by Subject

For the case of primary levels, the standard deviation is 0.151 for language and 0.129 for mathematics, while for the case of middle levels, the standard deviation is 0.243 for language and 0.161 for mathematics.

Having obtained an expression for the value-added of teacher j in period t, $\hat{\mu}_{jt}$, the next step is to check if it is a good predictor of the students' grades.

Third stage: Prediction Bias of the Value-Added Estimator.

We could estimate the following regression of the results corrected for observables for period t, A_{it}, with respect to $\hat{\mu}_{jt}$, which was constructed from different information in year t. Therefore, the following equation will correspond to the predictive potential in the student's results for period t but without considering information from this period.

$$A_{it} = \alpha_t + \lambda \hat{\mu}_{jt} + \xi_{it} \tag{3.5}$$

Where regression of A_{it} in $\hat{\mu}_{jt}$ includes controls by level, subject, and their interaction.

If the students were randomly assigned to period t, we have $E(e_{it}|\hat{\mu}_{jt}) = 0$ from equation 3.2; thus, the coefficient λ would measure the relationship between the true effect of teacher μ_{jt} and the estimator $\hat{\mu}_{jt}$. Additionally, under random assignment, it is satisfied that $\lambda = \frac{Cov(A_{it},\hat{\mu}_{jt})}{Var(\hat{\mu}_{jt})} = \frac{Cov(\mu_{it},\hat{\mu}_{jt})}{Var(\hat{\mu}_{jt})}$.

Taking the above result, we define the degree of bias of $\hat{\mu}_{jt}$ as $B(\hat{\mu}_{jt}) = \frac{Cov(e_{it},\hat{\mu}_{jt})}{Var(\hat{\mu}_{jt})} = 1 - \lambda$

If $B(\hat{\mu}_{jt}) = 0$, $\hat{\mu}_{jt}$ gives us an unbiased estimator for the prediction of teacher quality, and therefore an improvement of the value-added $\hat{\mu}_{jt}$ has the same causal effect on grades as an improvement of the true value-added μ_{jt}, which is of the same magnitude.

Under the assumption of stationarity of μ_{jt}, from the regression 3.5, we obtain a coefficient of λ equal to one per construction. The above is corroborated in column 1 of Table 3.3, which presents a coefficient of 1.004, with a 95% confidence level where the standard errors are calculated considering the cluster at the cohort-school level, to adjust for the correlation that occurs for students in the same classroom and the one associated with multiple observations per student[10]. In the case of Chetty et al.

---

[10]Chetty et al. (2014a) find that clustering according to these variables provides a more conservative and computationally manageable confidence interval than clustering according to classroom

(2014a), their results from their baseline are 0.998.

Variables	(1)	(2)	(3)
	Score in t	Score in t	Score in t
Teachers' Value-Added	1.004	0.999	1.019
	(0.000)	(0.000)	(0.000)
Baseline controls	x	x	x
Vulnerability conditions control		x	
Year t-2 test score control			x
Observations	16,758,760	16,758,760	11,750,202
R-squared	0.0687	0.0682	0.078

Table 3.3: Baseline Model Results and Selection on Exclude Observables.

Each column reports coefficients from an OLS regression, with standard errors clustered by school-cohort and p-value in parentheses. The regressions are run on the sample used to estimate the baseline VA model, restricted to observations with a non-missing leave-out teacher VA estimate. There is one observation for each student-subject-grade-school year in all regressions. Teacher VA is scaled in units of student test score standard deviations and is estimated using data from classes taught by the same teacher in other years. Teacher VA is estimated separately for each subject (language and mathematics) and according to level (primary and high school) and using the baseline control vector, which includes at student level: score in the immediately preceding grade, introduced using a cubic polynomial for the previous language and mathematics grades, interacted in turn with discrete variables of the grade in question; gender and grade repetition in the contemporary and previous grade, and average attendance in contemporary grade. Additionally, a set of discrete variables is added per year and per grade. In each columns, the dependent variable is the student's test score in a given year and subject. In Column 2, we add in the estimation of the Teachers' Value-Added the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. In Column (3) we add in the estimation of the Teachers' Value-Added twice-lagged test scores.

Figure 3.2 presents the conditional means of the residual of the grades of year t within quantiles constructed from the prediction of the value-added for period t with information from $t-1$. As expected, considering that our estimator $\hat{\mu}_{jt}$ corresponds to the best linear prediction of $\overline{A}_{Jt}$, we have a practically unitary slope.

and student.
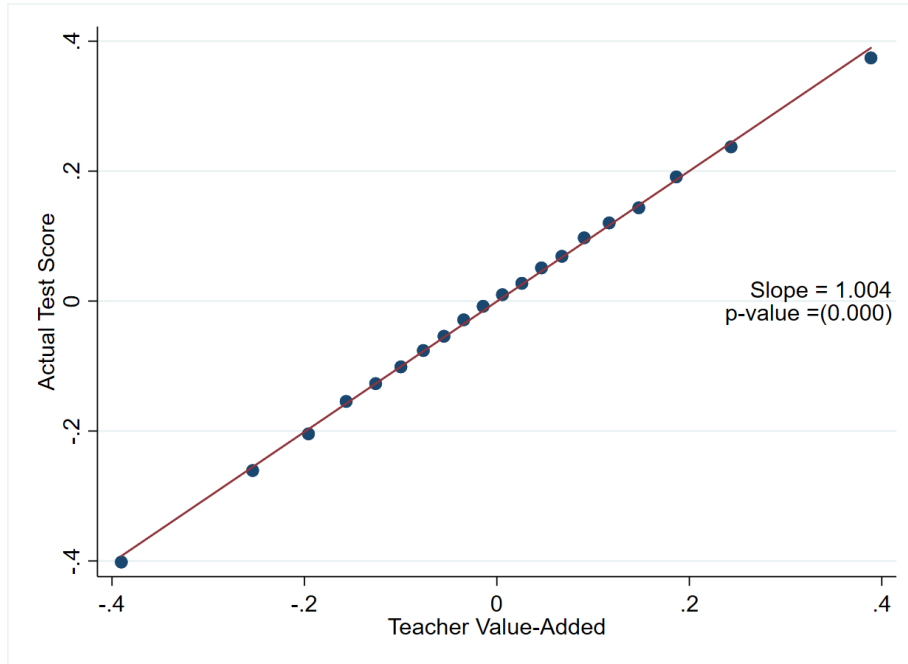
Figure 3.2: Effect of Teachers' Value-Added on Actual Scores.

This Figure are constructed using the sample used to estimate VA model, 16,758,760 observations. This plot correspond to the regression in column 1 of Table 3.3. To construct this binned scatter plot, we first residualisze the actual test score with respect to the baseline control vector (detailed in the note of Table 3.3) separately within each subject and using within-teacher variation to estimate the coefficients. Then divide the VA estimates $\hat{\mu}_{jt}$ into twenty equal-sized groups (vingtiles) and plot the means of the actual test score residuals within each bin against the mean value of $\hat{\mu}_{jt}$ within each bin. The line shows the best linear fit estimated on the underlying micro data using OLS

Finally, returning to equation 3.5, we had that only in case the assignment was random between teachers and students would we have the certainty that $Cov(e_{it}, \hat{\mu}_{jt}) = 0$; however, in case there are nonobservable variables that are determining the assignment between teachers and students, we would have that our estimator $\hat{\lambda}$ would be different from one. An indirect way to check this assignment would be to add new variables in our estimation of the residual test score but not to consider them in the construction of our estimator, $\hat{\mu}_{jt}$.

Fourth stage: Selection on excludes observables and no observables.

Let us imagine that observable variables of the student are determining the assignment between students and teachers, i.e., the school follows the same assignment rule for teachers and students in forecast period t as in previous periods, for example, according to the socioeconomic condition of the student's family or based on lagged test score gains. Adding any of these variables at the time of calculating the residual of the test score should explain an important part of the results and therefore lead to different results when regressing this new residual, A'_{it}, in our value-added, $\hat{\mu}_{jt}$.

In this case, we estimate the following equation, a variant of equation 3.1, adding the observable variable Z_{it}:

$$A'_{it} = A^*_{it} - \gamma A^*_{(it-1)} - \beta X_{it} - \rho Z_{it} \tag{3.6}$$

Later, we regressed A'_{it} in $\hat{\mu}_{jt}$, as in equation 3.5, again including controls by level, subject, and interaction. The observable variables that we will include separately will be the condition of the vulnerability of the students (1=if the student belongs to the most vulnerable 40% of the population) and the grade after the previous one, A^*_{it-2}. This last one is if the ordering between students and teachers is made by the management teams of the establishments from the students' previous results.

The results of estimates of value-added considering equation 3.6 and then replicating equation 3.5 are presented in columns 2 and 3 of Table 3.3 and show that the estimation is practically unaltered. Specifically, for the students' vulnerability condition, the coefficient is 0.999, with a 95% confidence level. For the case of including the score after the previous one, it is 1.019 with a 95% confidence level. This is corroborated in Figures 3.3a and 3.3b, which repeats what was done in Figure 3.2 and shows an adjustment practically equal to our base model.

We can conclude that selection due to the students' socioeconomic status is quite
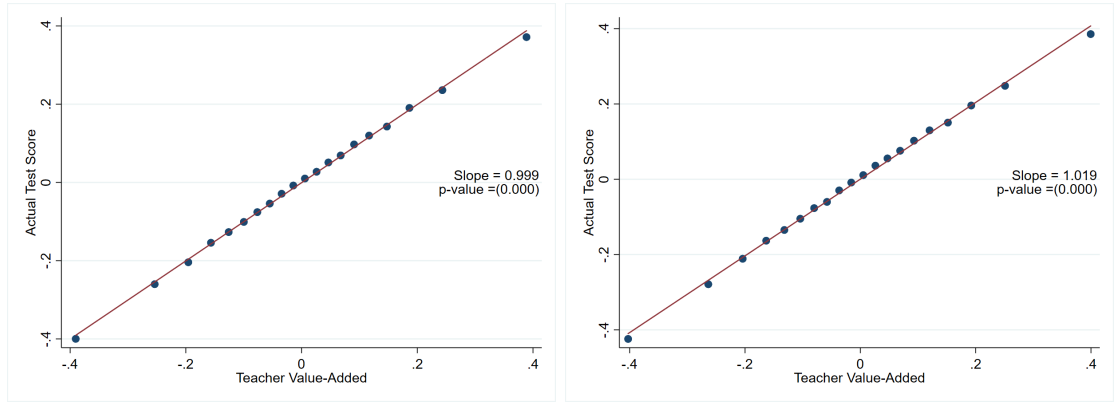
(a) Added includes Two Lags of Test Score   **(b)** Added includes Vulnerability Conditions

Figure 3.3: Effect of Teachers' VA on Actual Scores

These Figures are constructed using the same procedure explained in note of Figure 3.2. The two panels are binned scatter plots of actual test scores vs teacher VA including condition of the vulnerability of the students for his estimation (panel a) and including the twice-lagged test scores (panel b).

marginal, mainly because the student comes from a nonvulnerable (vulnerable) family regularly obtains good (bad) results and will present good (bad) results in the contemporary grade and the previous ones, which is largely captured by including the grade of the immediately previous one for control.

In the case of the impact that selection from previous results would have, our explanation again points in the same direction. Much of the variation observed in the grades after previous ones are captured by the set of controls that we include in our regression when estimating the grade residual.

The results presented are consistent with those found by Chetty et al. (2014a), in the sense that, in the case of parental characteristics, in cases measured mainly through household income, it finds values for $\hat{\lambda}$ of 0.996 (and for the case of including the score after previous, 0.976).

The above does not rule out the possibility that students are sorted to teachers based on unobservable characteristics orthogonal to the Z_{it} variables. For this purpose, we replicate the quasi-experiment realized by Chetty et al. (2014a), after considering the impossibility of conducting a random experiment as Karen and

Staiger (2008) and Kane et al. (2013).

This quasi-experiment exploits teacher turnover between schools and classes from one year to the next. A good way to understand this method's idea is to exemplify it, such as Chetty et al. (2014a), considering a school with three 8th grade classrooms (the last grade in the primary level). Suppose one of the teachers leaves the school in 2012 and is replaced by a teacher whose VA estimate in mathematics is 0.3 higher. Assume that the distribution of unobserved determinants of scores e_{it} does not change between 2011 and 2012. If forecast bias $B = 0$, this change in teaching staff should raise average 8th-grade math scores in the school by $0.3/3 = 0.1$. More generally, we can estimate B by comparing the change in mean scores across cohorts to the change in mean VA driven by teacher turnover, provided that student quality is stable over time.

We estimate the degree of forecast bias B by regressing changes in mean test scores across cohorts on changes in mean teacher VA:

$$\Delta A_{sch,gt} = a + b\Delta Q_{sch,gt} + \Delta\chi_{sch,gt} \tag{3.7}$$

Where A_{sgt} denote the mean value of A_{it} for students in school sch in grade g in year t and define the change in mean residual scores as $\Delta A_{sch,gt} = A_{sch,gt} - A_{sch,gt-1}Q_{sch,gt}$ denote the (student-weighted) mean of $\hat{\mu}_{jt}^{-(t,t-1)}$ across teachers in school sch in grade g. We define the change in mean teacher value-added from year $t-1$ to year t in grade g in school sch as $\Delta Q_{sch,gt} = Q_{sch,gt} - Q_{sch,gt-1}$.

The coefficient b in equation 3.7 identifies the degree of forecast bias as defined in equation 3.5 under the following identification assumption that changes in teacher VA across cohorts within a school grade are orthogonal to changes in other determinants of student scores, $Cov(\Delta Q_{sch,gt}, \Delta\chi_{sch,gt}) = 0$.

In general, student sorting at an annual frequency is minimal because of the costs of changing schools. Considering that families would be unlikely to change their school simply because a single teacher leaves or enters a given grade, we believe

in this assumption's plausibility.

Table 3.4 presents the results of the quasi-experiment detailed previously. We estimate six alternative value-added models, reporting correlations with baseline value-added estimates in column 1, and forecast bias for each model, defined as $B = 1 - \lambda$, in column 2. In this case, the first row presents the baseline model results, which estimates a forecast bias of 5.4%. These results are similar to those obtained by Chetty et al. (2014a) for the specification that includes only a lagged test score of 4.8%. In the second row, following the same three stages detailed above, we use all the baseline controls but omit teacher fixed effects and obtain a forecast bias of 0.3%. The next row adds information on lagged cross-subjects through a cubic polynomial interacting with the grade, marginally increasing the forecast bias by 6.8%. In row 4, we replicate our baseline specification by adding information on the same variables as our baseline but considering the averages of the classes and the school; in this case, the bias remains in ranges similar to our baseline model, 6.6%. Row 5 removes all controls related to the test score from the baseline specification, leaving only non-score controls at the student level; in this case, we confirm how relevant it is to include these controls in the estimate to obtain unbiased forecast estimates since the bias increases to 40.1%. The last row drops all controls except grades and year fixed effects, showing a forecast bias of 46.5%.

Two conclusions are obtained from this last exercise. First, although our second specification exhibits a lesser bias than our baseline model, it is because this method exploits variation both within and across teachers to identify the coefficients on the control vector and thus can understate teacher effects by overattributing test score growth to covariates if there is sorting, which is why we consider the specification including teacher fixed effects as our baseline model. Second, we corroborated that controlling for prior student-level test scores is fundamental to obtaining unbiased value-added estimates. As explained by Chetty et al. (2014a), one potential explanation for this result is that classroom assignment in large schools is made primarily

Specification	Correlation with Value-Added Baseline Estimate	Quasi-Experimental Estimate Bias %
(1) Value-Added Baseline	1.000	5.4%
		(0.000)
(2) VA Baseline without Teacher Fixed Effect	0,997	0.3%
		(0.000)
(3) VA Baseline and student's lagged score in other subject	0,992	6.8%
		(0.000)
(4) VA Baseline and Class and School scores	0,958	6.6%
		(0.000)
(5) Non-sore Controls	0,868	40.1%
		(0.000)
(6) No controls	0,788	46.5%
		(0.000)

Table 3.4: Comparison of Forecast Bias Across Value-Added Models.

In this table we estimate six alternative Value-Added models reporting correlations with the baseline Value-Added estimates in column 1. In column 2, we report quasi-experimental estimates of forecast bias for each model, defined as 1 minus the coefficient in a regression of the cross-cohort change in scores on the cross-cohort change in mean teacher VA. The regressions are run on the sample used to estimate the baseline VA model, restricted to observations with a non-missing leave-out teacher VA estimate. All models are estimated separately by school level and subject; the correlations and estimates of forecast bias pool VA estimates across all groups. Each model only varies the control vector used to estimate student test score residuals in equation 3.2; the remaining steps of the procedure used to construct VA estimates are the same for all the models. Model 1, baseline model, includes at student level: score in the immediately preceding grade, introduced using a cubic polynomial for the previous language and mathematics grades, interacted in turn with discrete variables of the grade in question; gender and grade repetition in the contemporary and previous grade, and average attendance in contemporary grade. Additionally, a set of discrete variables is added per year and per grade. Model 2 uses all of the baseline controls but omits teacher fixed effects. Model 3 uses all of the baseline controls adding a cubic in lagged cross-subject scores (for the case of language we adding mathematics and vice versa), interacted with the student's grade level. Model 4 uses all of the baseline controls and includes the same variable at student level at class and school level (mean of the student variables). Model 5 removes all controls related to test scores from the baseline specification, leaving only non-score controls at the student level. Finally, Model 6 drops all controls except grade and year fixed effects.

on the basis of prior-year test performance and its correlates.

Reviewing the robustness of our results, conditional on the available observable variables and considering the quasi-experiment's results, we have that even when there are no random assignments to teachers in forecast year t, by including prior student-level test scores, the estimate exhibits minimal predictive error.

3.5.2   Impact of Value-added in Educational Outcomes

This section estimates the impact of value-added on student outcome variables once they leave school. Following the previous notation, let us consider that Y_{it}^* corresponds to the tertiary education attendance (attending vocational education or university) of student i during their first year after graduation. We are interested in measuring the impact of value-added on this outcome variable. The following linear specification is proposed.

$$Y_{it}^* = \alpha + \tau m_{jt} + \varepsilon_{it} \tag{3.8}$$

Where the variable m_{jt} corresponds to $m_{jt} = \mu_{jt}/\sigma_\mu$, normalized teacher value-added j, such that the τ coefficient of equation 3.8 represents the reduced form of the impact of an increase of one standard deviation of teacher value-added for a given year, or grade, on tertiary education attendance. For more detail on the formalization of this reduced parameter's interpretation, see Appendix A of Chetty et al. (2014b).

It should be noted that τ will correspond to the value-added impact, measured through the students' grades, on their future tertiary education attendance. In other words, a teacher may affect students' attendance in ways other than those associated with their score, such as their confidence or aptitude when applying for university.

Assuming that the value-added is unbiased, as detailed in section 3.5.1, we cor-

roborate $\frac{Cov(\mu_{jt},\hat{\mu}_{jt})}{Var(\hat{\mu}_{jt})} = \frac{Cov(m_{jt},\hat{m}_{jt})}{Var(\hat{m}_{jt})}$, and we can verify from equation 3.8 that:

$$\frac{Cov(Y_{it},\hat{m}_{jt})}{Var(\hat{m}_{jt})} = \tau\frac{Cov(m_{it},\hat{m}_{jt})}{Var(\hat{m}_{jt})} + \frac{Cov(\varepsilon_{it},\hat{m}_{jt})}{Var(\hat{m}_{jt})} = \tau + \frac{Cov(\varepsilon_{it},\hat{m}_{jt})}{Var(\hat{m}_{jt})} \qquad (3.9)$$

provided that unobserved determinants of attendance, ε_{it}, are orthogonal to teacher VA estimates $\hat{m}_{jt}$.

For estimation, and similar to the calculation of the value-added, we residualize the dependent variable, Y_{it}^*, including the baseline covariates and teacher fixed effect:

$$Y_{it} = Y_{it}^* - \hat{\beta}^Y X_{it} \qquad (3.10)$$

Note that again, in recovering the residual from the previous equation, these must incorporate the teacher fixed effect. Otherwise, we will underestimate the value-added's impact by not incorporating the teacher's effect on the respective dependent variable.

Finally, we estimate the following regression:

$$Y_{it} = \alpha + \tau\hat{m}_{jt} + \varepsilon_{it} \qquad (3.11)$$

The interpretation of $\hat{\tau}$ will be the effect that an increase of one standard deviation in teacher quality, as measured by grades, has on their graduating students' tertiary education attendance. This can be considered a direct impact of the variation of the value-added on the probability of tertiary education attendance since our results of $\hat{\tau}$ are close to 1 ($\hat{\tau} = 1.004$).

Table 3.5 presents the results for the estimates of value-added through equation 3.11 for the entire educational system in the rates of tertiary, vocational, university, and top-3 university attendance and their respective robustness as detailed below.

The first column of each educational outcome specification is the result of equa-

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Tertiary Education Attendance			Vocational Education Attendance		
Teachers' Value-Added	1.905	1.829	1.761	-1.488	-1.471	-1.604
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean Dep. Var.	54.35	54.35	54.28	20.16	20.16	20.14
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Observations	5,331,323	5,331,323	3,583,633	5,331,323	5,331,323	3,583,633
VA Over Mean Dep. Var	3.5%	3.4%	3.2%	-7.4%	-7.3%	-8.0%

Variable	(7)	(8)	(9)	(10)	(11)	(12)
	University Attendance			Top-3 University Attendance		
Teachers' Value-Added	3.393	3.300	3.365	0.483	0.480	0.540
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean Dep. Var.	34.19	34.19	34.44	1.88	1.88	2.269
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Observations	5,331,323	5,331,323	3,583,633	5,331,323	5,331,323	3,583,633
VA Over Mean Dep. Var	9.9%	9.7%	9.8%	25.7%	25.6%	23.8%

Table 3.5: Teachers' Value-Added Outcomes.

Each column reports coefficients from an OLS regression between the residual of dependent variable and Teacher VA using the baseline control vector, with standard errors clustered by school-cohort and p-value in parentheses. Columns 1-3 use an indicator for tertiary education attendance; columns 4-6 use an indicator for vocational education attendance; columns 7-9 use an indicator for university attendance; columns 10-12 use an indicator for Top-3 university attendance. In the first column of each outcome-specification, we residualize each dependent variable using the baseline control vector detailed in note of Table 3.3. In the second column of each outcome-specification, we use the baseline control vector adding the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. In the third column of each outcome-specification, we use the baseline control vector adding the twice-lagged test scores. The last row of each table corresponds to the ratio between the impact of the Teacher's Value-Added on the average of the dependent variable.

tion 3.11 using our baseline controls. For the case of tertiary education attendance, the effect is that a 1 SD increase in a teacher's true VA test score in a single grade increases the probability of tertiary education attendance by 1.9 percentage points relative to a mean tertiary education attendance rate of 54.4%, which means an increase of 3.5% in mean tertiary education attendance in the regression sample. The null hypothesis that teacher VA does not affect tertiary education attendance is rejected with a $p - value < 0.001$.

For the case of vocational education attendance, the effect is that a 1 SD increase in a teacher's true VA test score in a single grade diminishes the probability of vocational education attendance by 1.5 percentage points relative to a mean vocational education attendance rate of 20.2%. For university attendance, the effect in a single grade is 3.4 percentage points, relative to a mean university attendance rate of 34.2%, i.e., an increase of 9.9% for mean university attendance in the regression sample almost triples if we consider any tertiary education institution type in column 1. Finally, column 10 presents the result for top-3 universities, where a 1 SD increase in a teacher's true VA test score in a single grade increases the probability of attendance at these universities by 0.5 percentage points relative to a mean top-3 university attendance rate of 1.9%, implying a 25.7% mean for top-3 university attendance in the regression sample. In all the above cases, the null hypothesis that teacher VA does not affect tertiary education attendance is rejected with a $p - value < 0.001$.

Taking the previous results, we can infer that the effect is a 1 SD increase in a teacher's true VA test score in a single grade has a general positive effect. However, the composition of this is not, affecting almost 3 times more than the effect on the average of the dependent variable in the case of universities. This effect is 2.5 times higher within this group if we consider top-3 universities, leading us to infer that the higher the institution's quality, the more important the professor's quality is in determining a student's enrollment.

We evaluate this estimate's robustness to alternative control vectors in the second

and third columns of each educational outcome specification, replicate the specification with the baseline control vector, and add the student's vulnerability condition and twice-lagged test scores, respectively (equation 3.6). Again, the coefficient does not change appreciably. Both variables are strong predictors of tertiary education attendance rates even conditioned on the baseline controls, with a $p-value < 0.001$. Hence, despite controlling for these variables, they do not significantly affect the estimates of τ, supporting the identification assumption of selection on observables.

Figures in 3.4d plot the residual of each educational attendance rate for students in school year t against $\hat{m}_{jt}$. To construct this binned scatter plot, a nonparametric representation of the conditional expectation function is used. We divide the VA estimates $\hat{m}_{jt}$ into twenty equal-sized groups (vingtiles) and plot the mean of the attendance residuals in each bin against the mean of $\hat{m}_{jt}$ in each bin. Finally, we add back the mean attendance rate in the estimation sample to facilitate the scale's interpretation. The regression coefficient and standard error are reported in this, and all subsequent figures are estimated on the class-level, with standard errors clustered by school cohort.

As the objective of this research is to study the effects of two measures on educational variables, and one of them, teacher evaluation, is applied only to public schools, Table 3.6 replicates the previous results but only to this subsample of teachers.

The results are in the same direction, although more moderate, than those obtained for the entire educational system. Specifically, the effect is that a 1 SD increase in a teacher's true VA test score in a single grade increases the probability of tertiary education attendance of graduate students who attended public schools by 0.6 percentage points relative to a mean tertiary education attendance rate of 47.1%, which means an increase of 1.3% in mean tertiary education attendance in the regression sample. In the case of vocational education attendance, the effect is not nonzero ($p-value < 0.181$), as is the case for top-3 universities ($p-value < 0.134$). Uni-

(a) Tertiary Education Attendance

(b) Vocational Education Attendance

(c) University Attendance

(d) Top-3 Universities Attendance
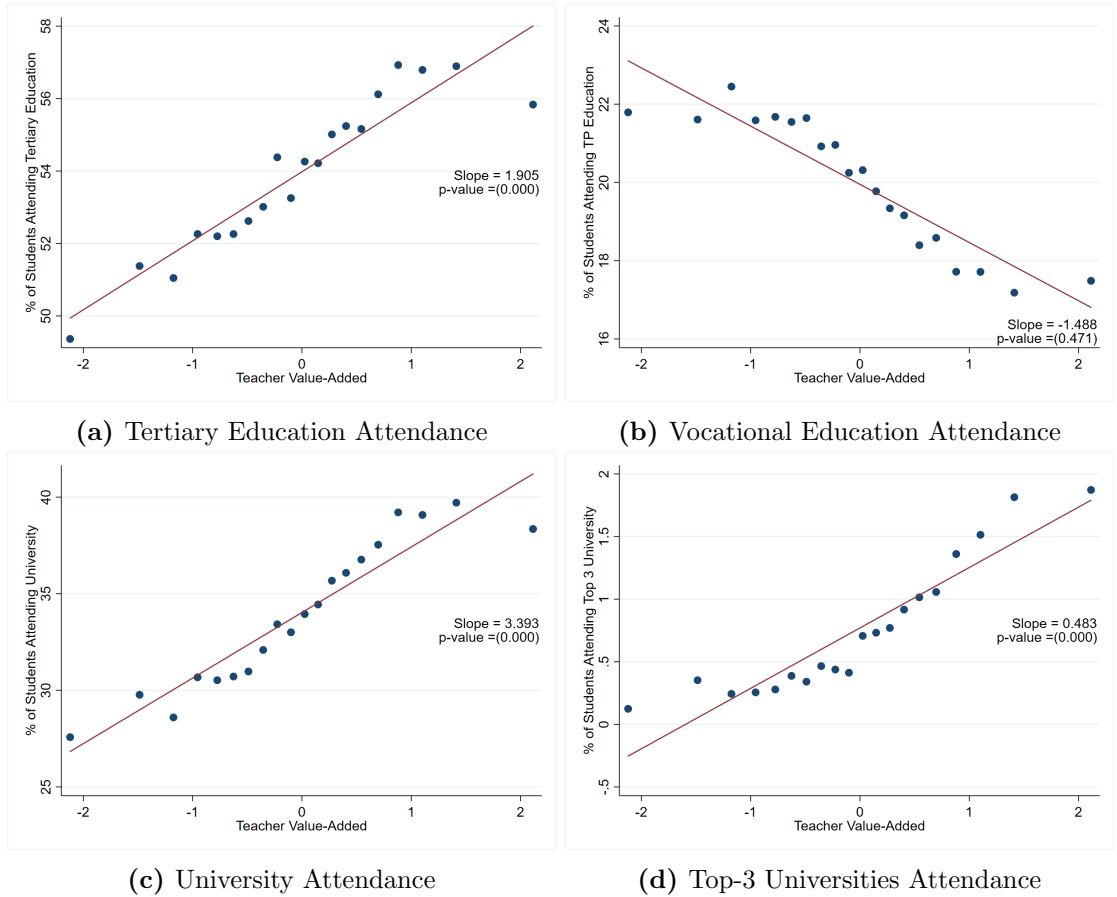
Figure 3.4: Conditional mean of Teachers' VA on Educational Outcomes

Panels (a) to (d) are binned scatter plots of tertiary education attendance rates, vocational education attendance rates, university attendance rates and top-3 university attendance rates vs. normalized teacher VA $\hat{m}_{jt}$. These plots correspond to the regressions in the first column of each outcome-specification of Table 3.5 and use the same sample restrictions and variable definitions, considering 15,568,432 observations. To construct these binned scatter plots, we first residualize the dependent variable with respect to the baseline control vector separately within each subject, using within-teacher variation to estimate the coefficients. We then divide the VA estimates $\hat{m}_{jt}$ into twenty equal-sized groups (vingtiles) and plot the means of the dependent variable residuals within each bin against the mean value of $\hat{m}_{jt}$ within each bin. Finally, we add back the unconditional mean of the dependent variable in the estimation sample to facilitate interpretation of the scale. The solid line shows the best linear fit estimated on the underlying micro data using OLS.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Tertiary Education Attendance			Vocational Education Attendance		
Teachers' Value-Added	0.626	0.603	0.587	-0.222	-0.217	-0.419
	(0.008)	(0.010)	(0.021)	(0.181)	(0.189)	(0.028)
Mean Dep. Var.	47.12	47.12	47.79	22.98	22.98	22.99
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Observations	1,815,678	1,815,678	1,216,653	1,815,678	1,815,678	1,216,653
VA Over Mean Dep. Var	1.3%	1.3%	1.2%	-1.0%	-0.9%	-1.8%

Variable	(7)	(8)	(9)	(10)	(11)	(12)
	University Attendance			Top-3 University Attendance		
Teachers' Value-Added	0.847	0.819	1.006	0.116	0.116	0.144
	(0.019)	(0.021)	(0.010)	(0.134)	(0.136)	(0.142)
Mean Dep. Var.	24.15	24.15	24.80	1.24	1.24	1.55
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Observations	1,815,678	1,815,678	1,216,653	1,815,678	1,815,678	1,216,653
VA Over Mean Dep. Var	3.5%	3.4%	4.1%	9.4%	9.4%	9.3%

Table 3.6: Teachers' Value-Added Outcomes in Public Schools.

Each column reports coefficients from an OLS regression between the residual of dependent variable and Teacher VA using the baseline control vector for public schools, with standard errors clustered by school-cohort and p-value in parentheses. Columns 1-3 use an indicator for tertiary education attendance of students graduated from public schools; columns 4-6 use an indicator for vocational education attendance; columns 7-9 use an indicator for university attendance; columns 10-12 use an indicator for Top-3 university attendance. In the first column of each outcome-specification, we residualize each dependent variable using the baseline control vector detailed in note of Table 3.3. In the second column of each outcome-specification, we use the baseline control vector adding the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. In the third column of each outcome-specification, we use the baseline control vector adding the twice-lagged test scores. The last row of each table corresponds to the ratio between the impact of the Teacher's Value-Added on the average of the dependent variable for students graduated from public schools.

versity attendance increases by 0.85 percentage points if teachers' true VA test score in a single grade increases by 1 SD, that is, an increase of 3.5% in mean university attendance.

Figures 4.11a to 4.11d in Appendix plot the residual of each educational attendance rate for graduate students who attended public schools in year t against $\hat{m}_{jt}$, as explained above.

To contextualize our results, moving a student from a teacher in the fifth to the ninety-fifth percentile of the true VA test score distribution would lead to increases the probability of tertiary education attendance of graduate students who attended public schools by 2.21 percentage points in a single grade[11]. Figure 3.5 illustrates the estimated effects of moving a student from a teacher in the fifth to the twenty-fourth (+1 SD), sixtieth (+2 SD) and ninety-fifth (+3.5 SD) percentile of the true VA test score distribution, respectively, on the probability of tertiary education attendance.

---

[11]A teacher who is at the ninety-fifth percentile is 3.5 standard deviation better than one at the fifth percentile. Therefore, the effect is 3.54 x τ, i.e., 3.54 x 0.626 = 2.21
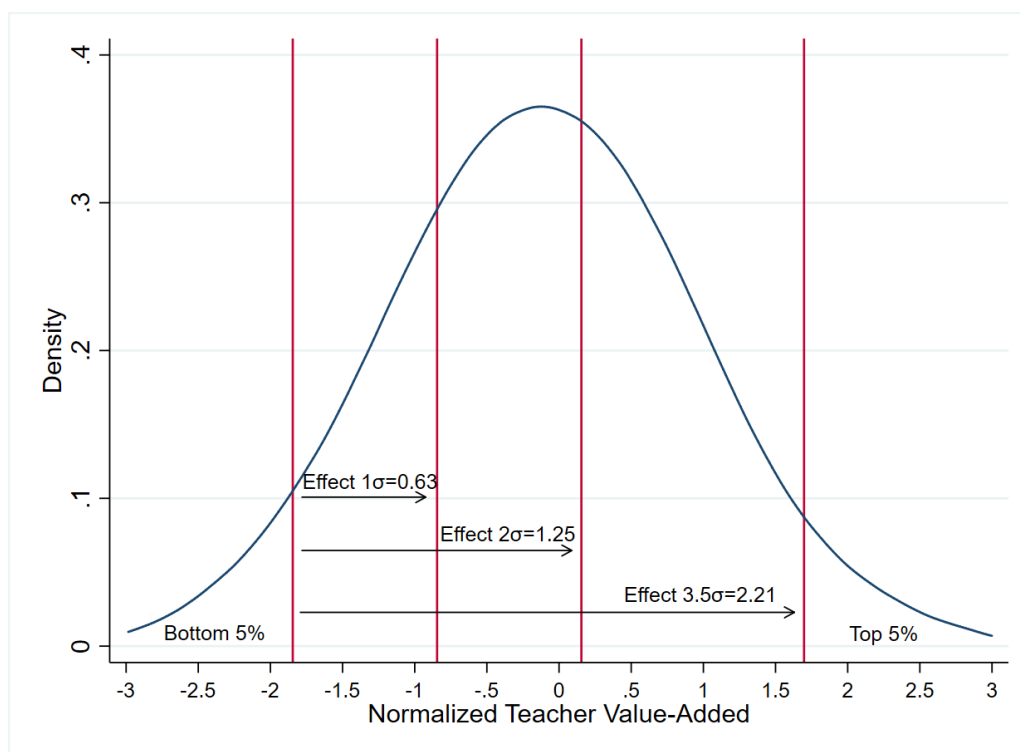
Figure 3.5: Effects of moving a student from Teachers' Value-Added in the fifth percentile of the true VA test score distribution in public schools.

Distribution of normalized Teachers' Value-Added for language in public schools . The arrows in the figure correspond to the effect of moving to a student from a teacher in the fifth to the twenty-fourth, sixtieth and ninety-fifth percentile of the true VA test score distribution. Its calculation corresponds to $1 \times \tau = 0.626$, $2 \times \tau = 1.25$, and $3.54 \times \tau = 2.21$, respectively.

3.6 Methodology for Estimating the Impact of Teacher Evaluation

To identify causal effects, the assumption of no unmeasured confounders must hold. This means that unobservable variables should be uncorrelated with the treatment variable and the variable of interest (potential outcome) (Rubin, Stuart and Zanutto, 2004). In our case, the assumption would imply that there are no unmeasured variables that affect the treatment (i.e., teaching evaluation) and potential outcome (i.e., student performance). Unfortunately, we do not have a randomized experiment to ensure the assumption just mentioned. Thus, as observational data

are used in our study, several potential sortings between and within schools might complicate our analysis as they challenge the no unmeasured confounders assumption, as has been explained during this investigation.

To address these concerns, we follow two approaches. The first of these, based on previous research (Kane et al., 2008, Kane et al., 2011, Briole, 2019, among others), attempts to quantify teacher evaluation impacts on certain outcomes through a linear estimation controlling for the potential biases mentioned above. For the second approximation, we will use a strategy similar to that used to estimate the impact of value-added on educational outcomes developed in section 3.5.1.

3.6.1 First Approach for Teacher Evaluation

For the first approach, we estimate the following equation:

$$Y_{it}^* = \alpha + \gamma A_{it-1}^* + \pi X_{it} + \theta ED_{jt} + \epsilon_{ijt} \tag{3.12}$$

When Y_{it}^* represents the tertiary educational attendance of graduating student i at time t. Similar to value-added estimates on educational outcomes, we include A_{it-1}^*, which corresponds to standardized test scores in the immediately preceding grade, and is introduced using a cubic polynomial for the previous language and mathematics grades, and interacts with the student's grade level. The vector X_{it} contains the information for student-level characteristics, including gender and grade repetition in the contemporary and previous grades, average attendance in the contemporary grade, and a set of discrete variables by year and grade. ED_{jt} is the normalized weighted total score of teacher evaluation[12], ED_{jt}/σ_{ED}. The error term is ϵ_{ijt}. Given this model, we are interested in the value of the coefficient θ of equation 3.12, which, in the case of correct identification, should capture the impact of an increase of one standard deviation of Teacher's Evaluation for a given year, or

---

[12]The original variable is a continuous variable that takes values between 1 and 4 depending on the teacher evaluation result.

grade, on tertiary education attendance.

In our estimate, we do not add a teacher fixed effect since this effect is expressed directly from the coefficient θ, which accompanies the variable ED. Incorporating a teacher fixed effect would underestimate the ED's effect since part of the teacher's effectiveness, not captured through the ED, would be absorbed through this effect.

Similarly, we do not include school or student fixed effects. Although the inclusion of school fixed effects is appealing because of its potential to reduce bias in teacher value-added estimates, the best practices to date suggest that their practical importance is limited (Chetty et al., 2014a and Koedel et al., 2015). In particular, Koedel et al. (2015) point out that once students' lagged performance has been included, adding these layers of fixed effects narrows the identifying variation used to estimate teacher value-added, which can increase the imprecision in estimation. Further evidence supports this view; Chetty et al. (2014a) show that value-added models with students' lagged test scores but without school and student fixed effects produce teacher value-added estimates with no significant bias. Similarly, Kane et al. (2008) and Kane et al. (2013) show that teacher value-added models without school and student fixed effects perform well versus experimental studies[13]. To address within-school sorting, we control the classroom averages of student characteristics (as suggested by Altonji and Mansfield, 2014).

Table 3.7 presents the results of our first methodology to evaluate the impact of teacher evaluation on the rates of tertiary, vocational, university, and top-3 university attendance estimates through equation 3.12.

The first column of each educational outcome specification uses baseline controls used in value-added estimates. The second column adds controls associated with the vulnerability condition of students and twice-lagged test scores. The third column adds class and school controls through cubic in-class means of prior-year test scores

---

[13]The intuition of this is that past test scores act as proxies for the unobserved heterogeneity; hence, the inclusion of student fixed effects is not particularly useful.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Tertiary Education Attendance			Vocational Education Attendance		
Weighted Total Score	1.670 (0.000)	1.367 (0.000)	1.253 (0.000)	-0.418 (0.000)	-0.457 (0.002)	-0.212 (0.071)
Mean Dep. Var.	47.27	47.92	47.27	22.97	22.97	22.97
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions and Year t-2 Test Score		x			x	
Class and School Controls			x			x
Observations	1,805,012	1,165,653	1,805,012	1,805,012	1,165,653	1,805,012
Over Mean Dep. Var	3.5%	2.9%	2.7%	-1.8%	-2.0%	-0.9%

Variable	(7)	(8)	(9)	(19)	(20)	(21)
	University Attendance			Top-3 University Attendance		
Weighted Total Score	2.088 (0.000)	1.824 (0.000)	1.464 (0.000)	0.225 (0.000)	0.235 (0.000)	0.217 (0.000)
Mean Dep. Var.	24.30	24.95	24.30	1.22	1.57	1.22
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions and Year t-2 Test Score		x			x	
Class and School Controls			x			x
Observations	1,805,012	1,165,653	1,805,012	1,805,012	1,165,653	1,805,012
Over Mean Dep. Var	8.6%	7.3%	6.0%	18.5%	15.0%	17.8%

Table 3.7: Teachers' Evaluation Test Outcomes - Lineal Regression.

Each column reports coefficients from an OLS regression between the dependent variable and ED, with standard errors clustered by school-cohort and p-value in parentheses. Columns 1-3 use an indicator for tertiary education attendance of students graduated from public schools; columns 4-6 use an indicator for vocational education attendance; columns 7-9 use an indicator for university attendance; columns 10-12 use an indicator for Top-3 university attendance. In the first column of each outcome-specification, we use the baseline control vector detailed in note of Table 3.3. In the second column of each outcome-specification, we use the baseline control vector adding the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. In the third column of each outcome-specification, we use the baseline control vector adding the twice-lagged test scores. The last row of each table corresponds to the ratio between the impact of the Teacher Evaluation on the average of the dependent variable for students graduated from public schools.

in each subject (language and mathematics). Each interacted with the grade and class means of all the other individual covariates.

For tertiary education attendance, a 1 SD increase in a teacher's evaluation in a single grade increases the probability of attendance of graduate students who attended public schools by 1.25-1.67 percentage points, depending on the specification, relative to a mean tertiary education attendance rate, which means an increase between 2.7% and 3.5% of mean tertiary education attendance in the regression sample.

In the case of vocational education, a 1 SD increase in a teacher's evaluation in a single grade decreases the probability of vocational education attendance of graduate students who attended public schools by 0.21-0.42 percentage points, depending on the specification, relative to a mean vocational education attendance rate, which means a decrease between 1.8% and 0.9% of mean attendance in the regression sample.

For the university analysis, we found that a 1 SD increase in a teacher's evaluation in a single grade increased the probability of university education attendance of graduate students who attended public schools by 1.46-2.09 percentage points, depending on the specification, which means an increase between 6.0% and 8.6% in mean university attendance in the regression sample. If we focus on top-3 universities, the effect is between 0.22 and 0.26 percentage points, depending on the specification, i.e., an increase between 17.8% and 18.5% of mean attendance in the regression sample.

In all the above cases, the null hypothesis that teacher VA does not affect tertiary education attendance is rejected with a $p-value < 0.001$, except for the specification in column 6, which presents a $p-value = 0.071$.

Similar to the results for the case of value-added, we observed a positive impact, at a general level, of a 1 SD increase in a teacher's evaluation in a single grade of 3.5% of mean attendance in the regression sample; however, this effect is heterogeneous depending on the type of tertiary education. From the above, we can infer that the

better the quality of the institutions in which graduate students enroll, the greater
the influence that a good teacher, measured through ED, can have on those results.

As detailed in sections 3.3.3 and 3.3.4, the variable used in equation 3.12, ED_{jt},
corresponds to a weighted total score based on 4 instruments. We are interested in
verifying the impact of each instrument separately on attendance to tertiary educa-
tional. To do this, we estimate an alternative presentation of equation 3.12:

$$Y_{it}^* = \alpha + \gamma A_{it-1}^* + \pi X_{it} + \sum_{v=1}^{4} \rho_v Instrument_{vjt} + \varepsilon_{ijt} \tag{3.13}$$

where we include the four instruments unweighted of teacher evaluation in the
same equation: self-evaluation, peer interview, external references, and portfolio.
Each instrument is normalized $Instrument_{vjt}/\sigma_{Instrument}$, such that coefficient ρ_v
of equation 3.13, capture the impact of an increase of one standard deviation of
Instrument Teacher's Evaluation for a given year, or grade, on tertiary education
attendance.

From regression, we want to know which instrument explains in a better way the
probability of tertiary educational attendance. It should be noted that the weighted
total score corresponds to a weighted average in a different way for each instrument;
however, in these regressions, we include the normalized score by year and grade of
each instrument, not considering the respective weights used in the weighted total
score of equation 3.12.

Again, the identification assumptions are the same as previously considered, the
covariance between each instruments and ε_{ijt} is zero, i.e., $Cov(\varepsilon_{ijt}, Instrument_{vjt})$.

Tables 3.8 and 3.9 present these results. As in the previous cases, for each educa-
tional outcome specification, two columns are added with the robustness associated
with the inclusion of the observable variables: second column add vulnerability con-
dition of students and twice-lagged test scores and third column add class and school
controls. In all cases, we highlight that the instruments that explain the probability

of tertiary educational attendances are the portfolio and external references. In the case of Portfolio, a 1 SD increase in a teacher's instrument Portfolio in a single grade increases the probability of attendance of graduate students who attended public schools by 1.25-1.52 percentage points depending on the specification relative to a mean tertiary educational attendance rate, which means an increase between 3.2% and 2.6% of mean attendance in the regression sample. For External References, a 1 SD increase in a teacher's instrument in a single grade increases the probability of attendance of graduate students who attended public schools by 1.06-1.54 percentage points depending on the specification relative to a mean attendance rate, which means an increase between 3.2% and 2.2% of mean attendance in the regression sample. Finally, for the case of the Self-evaluation score, the results, although positive, are marginal, while for the Peer Interview score in all specifications, the results are negative.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Tertiary Education Attendance			Vocational Education Attendance		
Portfolio Score	1.517	1.048	1.248	-0.387	-0.397	-0.269
	(0.000)	(0.000)	(0.000)	(0.001)	(0.006)	(0.021)
Self Evaluation Score	0.410	0.236	0.256	-0.019	-0.037	0.047
	(0.007)	(0.176)	(0.075)	(0.855)	(0.776)	(0.643)
Peer Interview Score	-0.441	-0.237	-0.375	0.472	0.402	0.450
	(0.013)	(0.204)	(0.020)	(0.000)	(0.007)	(0.000)
External References Score	1.536	1.498	1.061	-1.028	-0.959	-0.782
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions and Year t-2 Test Score		x			x	
Class and School Controls			x			x
Mean Dep. Var. (Public School)	47.27	47.92	47.27	22.97	22.97	22.97
Observations	1,805,012	1,165,653	1,805,012	1,805,012	1,165,653	1,805,012
Portfolio Score	3.2%	2.2%	2.6%	17.0%	14.7%	15.3%
Self Evaluation Score	0.9%	0.5%	0.5%	-0.7%	-1.1%	-0.9%
Peer Interview Score	-0.9%	-0.5%	-0.8%	-5.9%	-5.2%	-2.9%
External References Score	3.2%	3.1%	2.2%	21.8%	17.2%	19.3%

Table 3.8: Instruments of Teacher Evaluation on Outcomes - Linear Regression.

Each column reports coefficients from an OLS regression between the dependent variable and ED instruments (portfolio, self evaluation, peer interview and external references), with standard errors clustered by school-cohort and p-value in parentheses. Columns 1-3 use an indicator for tertiary education attendance of students graduated from public schools; columns 4-6 use an indicator for vocational education attendance. In the first column of each outcome-specification, we use the baseline control vector detailed in note of Table 3.3. In the second column of each outcome-specification, we use the baseline control vector adding the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. In the third column of each outcome-specification, we use the baseline control vector adding the twice-lagged test scores. The last row of each table corresponds to the ratio between the impact of the Teacher Evaluation on the average of the dependent variable for students graduated from public schools.

Variable	(7)	(8)	(9)	(10)	(11)	(12)
	University Attendance			Top-3 University Attendance		
Portfolio Score	1.904	1.444	1.518	0.207	0.230	0.186
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Self Evaluation Score	0.429	0.273	0.209	-0.008	-0.017	-0.011
	(0.009)	(0.143)	(0.167)	(0.794)	(0.640)	(0.685)
Peer Interview Score	-0.913	-0.639	-0.825	-0.072	-0.081	-0.035
	(0.000)	(0.008)	(0.000)	(0.046)	(0.106)	(0.282)
External References Score	2.564	2.457	1.843	0.266	0.270	0.235
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions and Year t-2 Test Score		x			x	
Class and School Controls			x			x
Mean Dep. Var. (Public School)	24.30	24.95	24.30	1.219	1.567	1.219
Observations	1,805,012	1,165,653	1,805,012	1,805,012	1,165,653	1,805,012
Portfolio Score	7.8%	5.8%	6.2%	13.0%	12.9%	12.4%
Self Evaluation Score	1.8%	1.1%	0.9%	0.3%	0.3%	-0.2%
Peer Interview Score	-3.8%	-2.6%	-3.4%	-5.6%	-5.4%	-5.7%
External References Score	10.6%	9.8%	7.6%	19.8%	19.8%	15.4%

Table 3.9: Instruments of Teacher Evaluation on Outcomes - Residual Regression.

Each column reports coefficients from an OLS regression between the dependent variable and ED instruments (portfolio, self evaluation, peer interview and external references), with standard errors clustered by school-cohort and p-value in parentheses. Columns 7-9 use an indicator for university attendance; columns 10-12 use an indicator for Top-3 university attendance. In the first column of each outcome-specification, we use the baseline control vector detailed in note of Table 3.3. In the second column of each outcome-specification, we use the baseline control vector adding the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. In the third column of each outcome-specification, we use the baseline control vector adding the twice-lagged test scores. The last row of each table corresponds to the ratio between the impact of the Teacher Evaluation on the average of the dependent variable for students graduated from public schools.

3.6.2 Second Approach for Teacher Evaluation

For the second approach, we residualize the dependent variable, Y_{it}^*, in this case, the tertiary educational attendance, including baseline controls and teacher fixed effect, obtained Y_{it}, where Y_{it} isolates the impact that a certain teacher has on tertiary educational attendance, as explained in the case of value-added estimates.

Then, we estimate the linear specification:

$$Y_{it} = \alpha + \rho ED_{jt} + \omega ijt \tag{3.14}$$

Where the variable ED_{jt} corresponds to normalized weighted total score of teacher j, ED_{jt}/σ_{ED}, and represents the reduced form of the impact of an increase of one standard deviation of teacher evaluation for a given year, or grade, on tertiary educational attendance.

It should be noted that ρ will correspond to the teacher evaluation impact, measured through the students' grades, on the students' future tertiary educational attendance. In other words, a teacher may affect students' attendance in ways other than those associated with their grades, such as their confidence or aptitude when applying for tertiary education.

Our second methodology for evaluating the impact of teacher evaluation on the rates of the tertiary, vocational, university, and top-3 university attendance estimates is presented in Table 3.10.

We observe stability in the results, which is slightly lower than those observed in our first methodology used in the case of teacher evaluation (Table 3.7). Specifically, we see a 1 SD increase in a teacher's evaluation in a single grade increases the probability of tertiary education attendance of graduate students who attended public schools by 1.63-1.91 percentage points (an increase between 4.0% and 3.4% of mean attendance in the regression sample), decreases the probability of vocational education attendance by 0.33-0.41 percentage points (a decrease between 1.4% and 1.8%

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Tertiary Education Attendance			Vocational Education Attendance		
Weighted Total Score	1.910	1.910	1.633	-0.327	-0.327	-0.407
	(0.000)	(0.000)	(0.000)	(0.007)	(0.007)	(0.007)
Mean Dep. Var.	47.27	47.27	47.92	22.97	22.97	22.97
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Observations	1,805,012	1,805,012	1,165,653	1,805,012	1,805,012	1,165,653
VA Over Mean Dep. Var	4,0%	4,0%	3,4%	-1,4%	-1,4%	-1,8%

Variable	(7)	(8)	(9)	(19)	(20)	(21)
	University Attendance			Top-3 University Attendance		
Weighted Total Score	2.237	2.237	2.040	0.183	0.183	0.194
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean Dep. Var.	24.30	24.30	24.95	1.22	1.22	1.57
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Observations	1,805,012	1,805,012	1,165,653	1,805,012	1,805,012	1,165,653
VA Over Mean Dep. Var	9,2%	9,2%	8,2%	15,0%	15,0%	12,4%

Table 3.10: Teachers' Evaluation Test Outcomes - Residual Regression.

Each column reports coefficients from an OLS regression between the residual of dependent variable and Teacher Evaluation for public schools, with standard errors clustered by school-cohort and p-value in parentheses. Columns 1-3 use an indicator for tertiary education attendance of students graduated from public schools; columns 4-6 use an indicator for vocational education attendance; columns 7-9 use an indicator for university attendance; columns 10-12 use an indicator for Top-3 university attendance. In the first column of each outcome-specification, we residualize each dependent variable using the baseline control vector detailed in note of Table 3.3. In the second column of each outcome-specification, we use the baseline control vector adding the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. In the third column of each outcome-specification, we use the baseline control vector adding the twice-lagged test scores. The regressions are run on the sample restricted to observations with a non-missing Teacher Evaluation and estimated for each subject and according to level. The Weighted Total Score of Teacher Evaluation is scaled in units of student test score standard deviations and is estimated using data from classes taught by the same teacher in other years. The last row of each table corresponds to the ratio between the impact of the Weighted Total Score on the average of the dependent variable for students graduated from public schools.

of mean attendance in the regression sample), increases the probability of university education by 2.24-2.04 percentage points (an increase between 9.2% and 8.2% of mean attendance in the regression sample), and increases the probability of a top-3 university education by 0.18-0.19 percentage points (an increase between 12.4% and 15.0% of mean attendance in the regression sample).

Figures 3.7a to 3.7d plot the residual of each educational attendance rate for students in school year t against ED_{jt}. To construct this binned scatter plot, we follow the same procedure detailed in section 3.5.2.

In the case of Teacher's Evaluation, our results suggest that, moving a student from a teacher in the fifth to the ninety-fifth percentile of ED score distribution would lead to increases the probability of tertiary education attendance of graduate students who attended public schools by 6.23 percentage points in a single grade[14]. Figure 3.6 illustrates the estimated effects of moving a student from a teacher in the fifth to the twenty-second (+1 SD or +0.25 points of ED score), fifty-eighth (+2 SD or +0.50 points of ED score) and ninety-fifth (+3.26 SD or +0.81 points of ED score) percentile of ED score distribution, respectively, on the probability of tertiary education attendance.

Finally, we replicate the study associated with each of the ED instruments. The results of Tables 4.11 and 4.12 in Appendix, remain the same as those obtained in section 3.6.1, being the most relevant instruments in explaining the probability of tertiary educational attendance, portfolio (1.30-1.71 percentage points), and external references (1.74-1.86 percentage points).

---

[14]A teacher who is at the ninety-fifth percentile is 3.26 standard deviation better than one at the fifth percentile. Therefore, the effect is 3.26 x ρ, i.e., 3.26 x 1.91 = 6.23
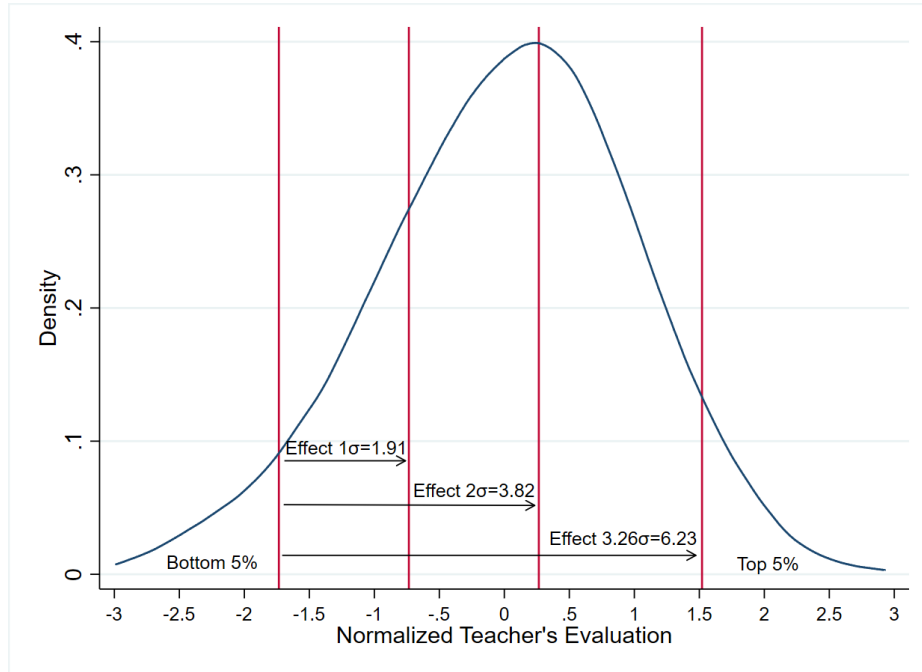
(a) Normalized ED score distribution.
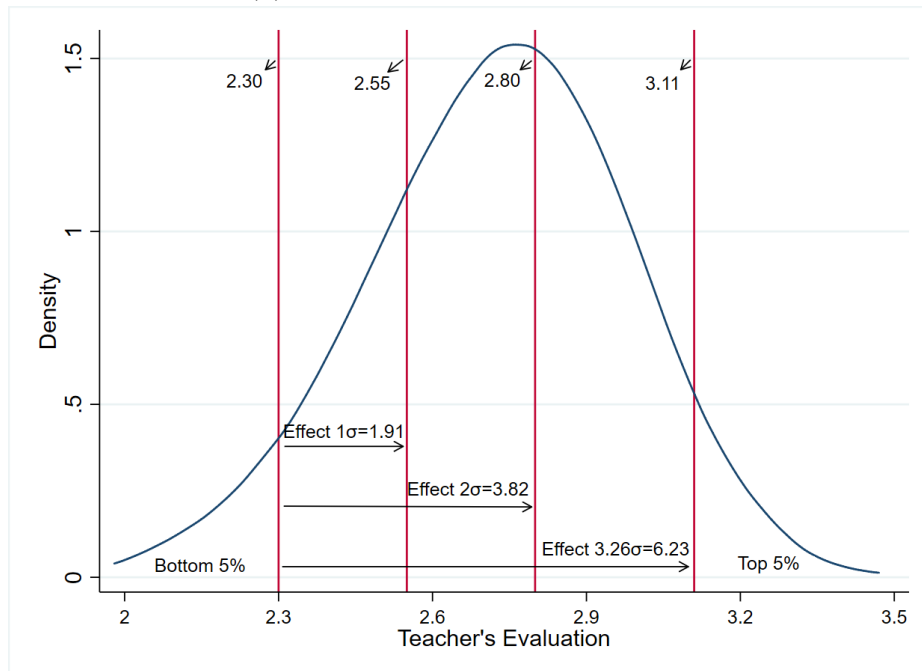
(b) ED score distribution.

Figure 3.6: Effects of moving a student from Teacher's Evaluation in the fifth percentile of the ED score distribution in public schools.

Distribution of ED score and normalized ED score for language in public schools. The arrows in the figure correspond to the effect of moving to a student from a teacher in the 5th to the 22th, 58th and 95th percentile of the ED score distribution. Its calculation corresponds to $1 \times \rho = 1.91$, $2 \times \rho = 3.82$, and $3.26 \times \rho = 6.23$, respectively.

148

(a) Tertiary Education Attendance

(b) Vocational Education Attendance

(c) University Attendance

(d) Top-3 Universities Attendance
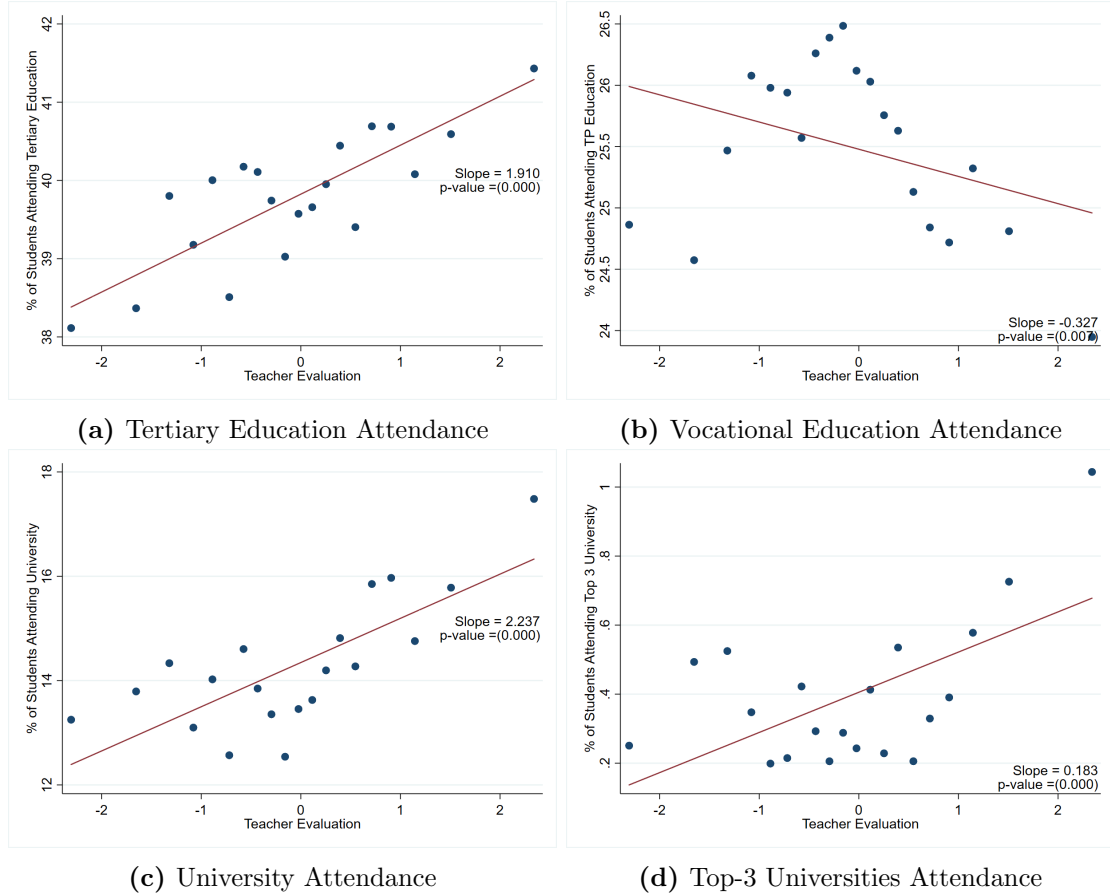
Figure 3.7: Conditional mean of Teachers' Evaluation on Educational Outcomes in Public Schools

Panels (a) to (d) are binned scatter plots of tertiary education attendance rates, vocational education attendance rates, university attendance rates and top-3 university attendance rates vs. Teacher's Evaluation ED_{jt}. These plots correspond to the regressions in the first column of each outcome-specification of Table 3.10 and use the same sample restrictions and variable definitions. To construct these binned scatter plots, we first residualize the dependent variable with respect to the baseline control vector separately within each subject, using within-teacher variation to estimate the coefficients. We then divide the VA estimates ED_{jt} into twenty equal-sized groups (vingtiles) and plot the means of the dependent variable residuals within each bin against the mean value of ED_{jt} within each bin. Finally, we add back the unconditional mean of the dependent variable in the estimation sample to facilitate interpretation of the scale. The solid line shows the best linear fit estimated on the underlying micro data using OLS.

3.7 Comparison of two mesuares of teacher's quality

In this section, we compare both measures of teacher quality and their impact on entry into tertiary education. In the same regression of the previous sections, we include both measurements simultaneously to corroborate whether the found results are maintained.

Considering both instruments' application to the same sample, we use the methodology used in sections 3.5.2 and 3.6.2 only considering students who have graduated from public schools. For the above, we estimate the following equation, including controls by level, subject, and their interaction, as shown in equations 3.11 and 3.14:

$$Y_{it} = \alpha + \tau \hat{m}_{jt} + \rho ED_{jt} + \omega ijt \tag{3.15}$$

Where Y_{it} is the residual of attendance, eliminating the effect of observable characteristics X_{it} from the regression (including the previous grade and teacher fixed effect), the same as we do for estimating equation 3.11 (Value-Added) and equation 3.14 (Teacher Evaluation). $\hat{m}_{jt} = \hat{\mu}_{jt}/\sigma_\mu$ is the normalized teacher value-added j, such that τ represents the reduced form of the impact of an increase of 1 SD on teacher value-added for a given year, or grade, on any variable of attendance. ED_{jt} is the normalized weighted total score of teacher j, ED_{jt}/σ_{ED}, such that ρ represents the reduced form of the impact of an increase of 1 SD on teacher evaluation for a given year, or grade, on any variable of attendance. Last, ωijt are the unobserved determinants of attendance, which are assumed to be orthogonal to teacher value-added estimates and teacher evaluation.

The Tables 3.11 and 3.12 present the results of the estimates for graduated students from public schools in the rates of tertiary, vocational, university, and top-3 university attendance and their respective robustness. The first column of each

educational outcome specification is the result of equation 3.15 using our baseline controls. The second and third columns replicate the specification with the baseline control vector and add the vulnerability condition of student and twice-lagged test scores, respectively.

For the case of tertiary education attendance, a 1 SD increase in a teacher's evaluation in a single grade increases the probability of tertiary education attendance by 1.67-1.93 percentage points, which means an increase of 3.5%-4.1% of mean tertiary education attendance in the regression sample. For teachers' true VA test scores in a single grade, the effect is an increase of 0.64-0.69 percentage points, which means a 1.3%-1.5% increase in the mean tertiary education attendance. In both cases, the null hypothesis that teachers' evaluation and value-added have no effect on tertiary education attendance is rejected with a $p-value < 0.001$.

For the case of vocational attendance, the results for VA are not significant. At the same time, for teacher evaluation, a 1 SD increase in ED decreases the probability between 0.32 and 0.44 percentage points (1.4%-1.9% decrease in the mean vocational education attendance in the regression sample).

For universities, the result is that a 1 SD increase in a teacher's evaluation in a single grade increases the probability of university attendance by 2.10-2.26 percentage points (an increase of 4.8%-4.4% of mean university education attendance in the regression sample). For the case of teachers' true VA test scores in a single grade, the effect is an increase of 0.92-1.07 percentage points (an increase of 1.9%-2.2% of mean university attendance). In both cases, the null hypothesis that teachers' evaluation and value-added have no effect on tertiary education attendance is rejected with a $p-value < 0.001$.

Finally, the VA results are not significant for top-3 university attendance. At the same time, for teacher evaluation, a 1 SD increase in ED increases the probability by 0.19-0.20 percentage points (0.9% increase in the mean university attendance in the regression sample).

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Tertiary Education Attendance			Vocational Education Attendance		
Weighted Total Score	1.936	1.936	1.669	-0.324	-0.324	-0.435
	(0.000)	(0.000)	(0.000)	(0.010)	(0.010)	(0.004)
Teachers' Value-Added	0.689	0.666	0.639	-0.230	-0.225	-0.435
	(0.004)	(0.005)	(0.014)	(0.173)	(0.180)	(0.024)
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Mean Dep. Var. (Public School)	47.27	47.27	47.95	22.96	22.96	22.95
Observations	1,726,710	1,726,710	1,152,901	1,726,710	1,726,710	1,152,901
Over Mean Dep. Ver.						
Weighted Total Score	4.1%	4.1%	3.5%	-1.4%	-1.4%	-1.9%
Teachers' Value-Added	1.5%	1.4%	1.3%	-1.0%	-1.0%	-1.9%

Table 3.11: Teachers' Value-Added and Teacher Evaluation on Outcomes.

Each column reports coefficients from an OLS regression between the residual of dependent variable and Teacher VA and Teacher Evaluation for public schools, with standard errors clustered by school-cohort and p-value in parentheses. Columns 1-3 use an indicator for tertiary education attendance of students graduated from public schools; columns 4-6 use an indicator for vocational education attendance. In the first column of each outcome-specification, we residualize each dependent variable using the baseline control vector. In the second column of each outcome-specification, we use the baseline control vector adding the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. In the third column of each outcome-specification, we use the baseline control vector adding the twice-lagged test scores. The regressions are run on the sample restricted to observations with a non-missing in VA and Teacher's Evaluation model, and estimated for each subject and according to level. The score for VA and Weighted Total Score of Teacher Evaluation is scaled in units of student test score standard deviations and is estimated using data from classes taught by the same teacher in other years. The last row of each table corresponds to the ratio between the impact of each measure on the average of the dependent variable for students graduated from public schools.

Variable	(7)	(8)	(9)	(10)	(11)	(12)
	University Attendance			Top-3 University Attendance		
Weighted Total Score	2.260	2.260	2.104	0.196	0.196	0.202
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Teachers' Value-Added	0.919	0.891	1.074	0.123	0.122	0.158
	(0.011)	(0.013)	(0.007)	(0.121)	(0.123)	(0.121)
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Mean Dep. Var. (Public School)	24.31	24.31	25.00	1.25	1.25	1.57
Observations	1,726,710	1,726,710	1,152,901	1,726,710	1,726,710	1,152,901

Over Mean Dep. Ver.						
Weighted Total Score	9.3%	9.3%	8.4%	15.7%	15.7%	12.8%
Teachers' Value-Added	3.8%	3.7%	4.3%	9.9%	9.8%	10.0%

Table 3.12: Teachers' Value-Added and Teacher Evaluation on Outcomes.

Each column reports coefficients from an OLS regression between the residual of dependent variable and Teacher VA and Teacher Evaluation for public schools, with standard errors clustered by school-cohort and p-value in parentheses. Columns 7-9 use an indicator for university attendance; columns 10-12 use an indicator for Top-3 university attendance. In the first column of each outcome-specification, we residualize each dependent variable using the baseline control vector. In the second column of each outcome-specification, we use the baseline control vector adding the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. In the third column of each outcome-specification, we use the baseline control vector adding the twice-lagged test scores. The regressions are run on the sample restricted to observations with a non-missing in VA and Teacher's Evaluation model, and estimated for each subject and according to level. The score for VA and Weighted Total Score of Teacher Evaluation is scaled in units of student test score standard deviations and is estimated using data from classes taught by the same teacher in other years. The last row of each table corresponds to the ratio between the impact of each measure on the average of the dependent variable for students graduated from public schools.

Our last exercise was to verify the contemporary correlation of both measures. Panels (a) and (b) of Figure 3.8 present the conditional means of the Normalized Weighted Total Score (ED_{jt}) and the Normalized Portfolio Score, respectively. These indicators are within vingtiles means constructed from the Normalized Teacher's Value-Added ($\hat{m}_{jt}$).

We find that the correlation between measurements of teachers' quality that come from the Teacher's Evaluation vs Teacher's Value-Added approach appears to be null. Specifically, the correlation between ED_{jt} and $\hat{m}_{jt}$ is -0.018, and the correlation between the Normalized Portfolio Score and $\hat{m}_{jt}$ is -0.011. The correlations between $\hat{m}_{jt}$ and the rest of the Teacher's Evaluation instruments are shown in Table 4.13 of the Appendix.

(a) ED_{jt} and $\hat{m}_{jt}$ **(b)** Normalized Portfolio Score vs. $\hat{m}_{jt}$
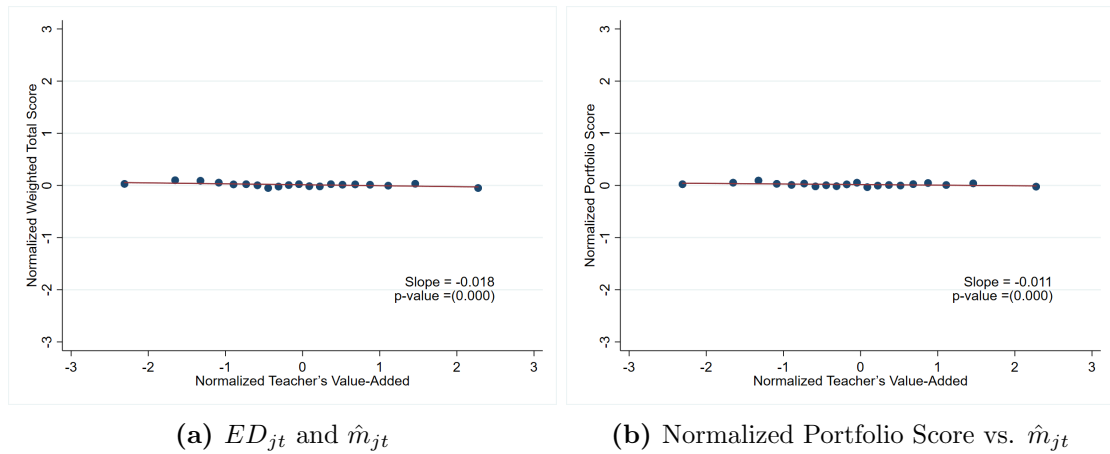
Figure 3.8: Binned scatter plots of Normalized Weighted Total Score vs. Portfolio Score vs. Normalized Teacher's Value-Added.

Panel (a) are binned scatter plots of Normalized Weighted Total Score vs. Normalized Teacher's Value-Added and Panel (b) are binned scatter plots of Normalized Portfolio Score vs. Normalized Teacher's Value-Added. To construct this binned scatterplot, we divide the $\hat{m}_{jt}$ into twenty equal-sized groups (vingtiles) and plot the means of the ED_{jt} for Panel (a) and Normalized Portfolio Score for Panel (b) within each bin against the mean value of $\hat{m}_{jt}$ within each bin. The line shows the best linear fit estimated on the underlying micro data using OLS of ED_{jt} and Normalized Portfolio Score on the $\hat{m}_{jt}$, respectively.

As shown by these results, we verify that i) both impacts remain in the same ranges of equations 3.11 and 3.14 for each of the variables of attendance; ii) a 1 SD

increase in a teacher's evaluation affects between two or three times the effect of a 1 SD increase in a teacher's value-added; and iii) both measurements turn out to be orthogonal, from which we can infer that both would capture different dimensions or abilities of teachers when studying their impact on tertiary education attendance.

3.8   Conclusions

This chapter contributes to the discussion of the impact teachers have on their students' academic results once they graduate from secondary education. To this end, we analyze two measurements used by most countries: the value-added methodology and the teacher's evaluation. We use the same sample of students and teachers to study each of them, including all students who graduated from public schools for the same period.

In the case of the value-added methodology, all evidence available in the literature is verified to ensure that the results present minimal bias that allows us to conclude and infer from them. In this sense, the results of both observables and quasi-experiment bias are in ranges similar to those found by Chetty et al. (2014a) for the US.

Similarly, in the case of teacher's evaluation, its effects are estimated from two methodologies, finding practically the same results in both cases. The first is commonly used by the literature (Kane et al., 2008, Kane et al., 2011, Briole 2019, among others), and the second allows us to estimate both measures simultaneously. Additionally, this chapter contributes to quantifying which of a series of instruments, used to a greater or lesser extent by all countries that apply this type of evaluation, has a greater relationship with tertiary, vocational and university attendance.

Our results suggest that the correlation between teacher's evaluation and teacher's value-added appears to be null in school outcomes. However, our analysis also reveals that both measures, ED and TVA, positively affect the probability of tertiary

education attendance, indicating that both measures are complementary in measuring teacher quality in the middle run. These results have relevance from the public policy point of view as unlike countries (e.g. USA) where VA is used for teacher's promotions and personnel decisions, in countries where VA is not used for teacher's personnel decisions (e.g. Chile), VA seems to be useful to measure teacher quality. Furthermore, our findings are consistent with the argument of the multidimensionality of teaching quality, because even though in the short run VA and ED seem to be orthogonal, in the medium run they seem to be complementary tools to measure teacher effectiveness.

Finally, suppose we weigh the potential costs of each, especially for a developing economy. In that case, these results give an account of certain instruments that could be applied to the extent that their economic resources allow. In our evaluation and the availability to access certain students' socioeconomic variables and their grades, the value-added measurement is highly cost-effective. Likewise, an external evaluation from the hierarchical superiors, based on a precise and structured guide based on some good teaching framework, which includes several questions about the evaluated teacher's performance, turns out to be a cost-effective tool. Additionally, an instrument that presents evidence of the pedagogical practice of the teacher in an objective way designed and adapted for each context by a centralized institution, in our case the Portfolio, is a third tool that would help to identify the teachers who could have a better impact on their students' academic results. However, as reviewed during this chapter, this tool may be time-consuming for teachers and costly from a fiscal perspective.

Chapter 4

Appendix

Year	Labor Force	Total Workers	Formal Employees	Private Formal Employees	% Private Formal Empl/Formal Empl.	Proj. Private Formal Empl.	% Formal Empl/Total Workers	% Proj. Private Formal empl/Total Workers	Total UIPD	% Total UIDP/Labor Force	% Total UIDP/Formal Empl.	% Total UIDP/Proj. Private Formal Empl.
2003	6,395.5	5,786.4	3,670.5			3,108.8	63.4%	53.7%	727.5	11.4%	19.8%	23.4%
2004	6,605.1	5,942.4	3,807.7			3,225.0	64.1%	54.3%	1,392.0	21.1%	36.6%	43.2%
2005	6,798.5	6,171.1	3,986.1			3,376.1	64.6%	54.7%	1,897.9	27.9%	47.6%	56.2%
2006	6,806.2	6,274.4	4,166.3			3,528.7	66.4%	56.2%	2,276.1	33.4%	54.6%	64.5%
2007	6,943.7	6,449.4	4,360.9			3,693.5	67.6%	57.3%	2,643.2	38.1%	60.6%	71.6%
2008	7,201.3	6,638.9	4,582.5			3,881.2	69.0%	58.5%	2,958.8	41.1%	64.6%	76.2%
2009	7,302.4	6,594.7	4,501.9			3,812.9	68.3%	57.8%	3,039.4	41.6%	67.5%	79.7%
2010	7,774.6	7,148.5	4,910.5	4,159.0	84.7%	4,159.0	68.7%	58.2%	3,313.7	42.6%	67.5%	79.7%
2011	8,054.2	7,478.8	5,143.0	4,392.9	85.4%	4,392.9	68.8%	58.7%	3,646.1	45.3%	70.9%	83.0%
2012	8,150.2	7,627.1	5,361.9	4,553.9	84.9%	4,553.9	70.3%	59.7%	3,950.8	48.5%	73.7%	86.8%
2013	8,280.4	7,785.2	5,481.4	4,669.4	85.2%	4,669.4	70.4%	60.0%	4,112.5	49.7%	75.0%	88.1%
2014	8,437.7	7,904.2	5,530.7	4,666.4	84.4%	4,666.4	70.0%	59.0%	4,190.7	49.7%	75.8%	89.8%
2015	8,557.0	8,022.8	5,651.1	4,752.5	84.1%	4,752.5	70.4%	59.2%	4,298.1	50.2%	76.1%	90.4%
2016	8,686.3	8,122.4	5,661.7	4,789.5	84.6%	4,789.5	69.7%	59.0%	4,389.7	50.5%	77.5%	91.7%
2017	8,867.1	8,276.1	5,713.0	4,766.7	83.4%	4,766.7	69.0%	57.6%	4,440.8	50.1%	77.7%	93.2%
2018	9,020.1	8,391.8	5,821.0	4,815.1	82.7%	4,815.1	69.4%	57.4%	4,618.6	51.2%	79.3%	95.9%

Table 4.1: Representativity of the unemployment insurance program dataset

This dataset is compared with the information of the ENE (Encuesta Nacional de Empleo) questionnaire administered by the government statistics agency in Chile (INE-Instituto Nacional de Estadísticas). All columns correspond to annual averages. Total Workers recorded by ENE refers to the total number of workers including: Private Formal Employees, Public Sector Employees, Informal Workers, Training Contracts and Domestic Workers. Projection Private Formal Employees is recorded by ENE since 2010. For years 2003 to 2009 we project the number of Private Formal Employees by multiplying the number Private Formal Employees by the proportion of Private Formal Employees recorded by 2010, 84.7%. Total UIPD is the annual averages of private formal employees in the database of the UIP.

Year—	N	Min	Mean	Max	p1	p5	Median	p95	p99
2003	1,349,584	2	384,908	12,379,695	18,804	67,062	267,458	1,105,923	2,475,924
2004	1,849,529	2	428,928	12,432,321	23,508	76,579	289,927	1,316,578	2,486,467
2005	2,337,830	2	456,767	11,231,467	24,410	86,533	306,313	1,429,162	2,477,661
2006	2,701,345	2	491,342	7,432,997	27,325	99,207	327,247	1,576,341	2,477,666
2007	3,103,157	1	526,906	12,848,755	30,519	101,730	350,557	1,663,517	2,495,274
2008	3,309,297	1	538,045	34,705,504	26,991	97,940	353,730	1,720,412	2,479,460
2009	3,419,851	1	581,561	21,906,338	28,923	108,215	382,093	1,890,408	2,472,360
2010	3,742,474	1	622,926	8,101,156	30,870	110,580	407,878	2,038,509	2,678,778
2011	4,052,453	1	655,923	7,992,452	33,467	112,763	434,478	2,158,657	2,734,946
2012	4,286,460	1	698,879	10,768,408	38,684	118,068	468,135	2,296,852	2,777,116
2013	4,404,045	1	745,498	14,414,099	40,927	124,889	507,729	2,426,510	2,913,642
2014	4,502,329	1	758,235	16,735,737	38,477	125,925	518,157	2,476,340	2,990,906
2015	4,611,049	1	772,466	16,963,680	34,989	129,608	532,841	2,497,008	3,026,780
2016	4,695,182	1	790,532	19,047,738	35,946	134,800	546,676	2,528,353	3,073,080
2017	4,861,557	1	813,370	44,150,312	40,455	138,969	565,972	2,574,900	3,131,013
2018	5,010,358	1	835,629	38,659,216	40,000	144,000	580,000	2,623,760	3,238,980

Table 4.2: Descriptive statistics of the UIP database

Region (Number)—	Size KM^2	Population	Poverty rate	% GDP Chile
Arica y Parinacota (15)	16,873	226,068	8.4%	0.8%
Tarapacá (1)	42,226	330,558	6.4%	2.5%
Atacama (3)	75,176	286,168	7.9%	2.6%
Coquimbo (4)	40,580	757,586	11.9%	3.1%
Valparaíso (5)	16,396	1,815,902	7.1%	9.1%
Metropolitana de Santiago (13)	15,403	7,112,808	5.4%	46.4%
Libertador Gral Bernardo O'Higgins (6)	16,387	914,555	10.1%	4.8%
Maule (7)	30,296	1,044,950	12.7%	3.4%
Ñuble (16)	13,179	480,609	16.1%	-
Biobío (8)	23,890	1,556,805	12.3%	7.9%
La Araucanía (9)	31,842	957,224	17.2%	2.8%
Los Ríos (14)	18,430	384,837	12.1%	1.4%
Los Lagos (10)	48,584	828,708	11.7%	3.3%
Aysén del Gral C. Ibáñez del Campo (11)	108,494	103,158	4.6%	0.6%
Magallanes y la Antártica Chilena (12)	132,291	166,533	2.1%	1.2%

Table 4.3: Information and indicators for different Chilean regions
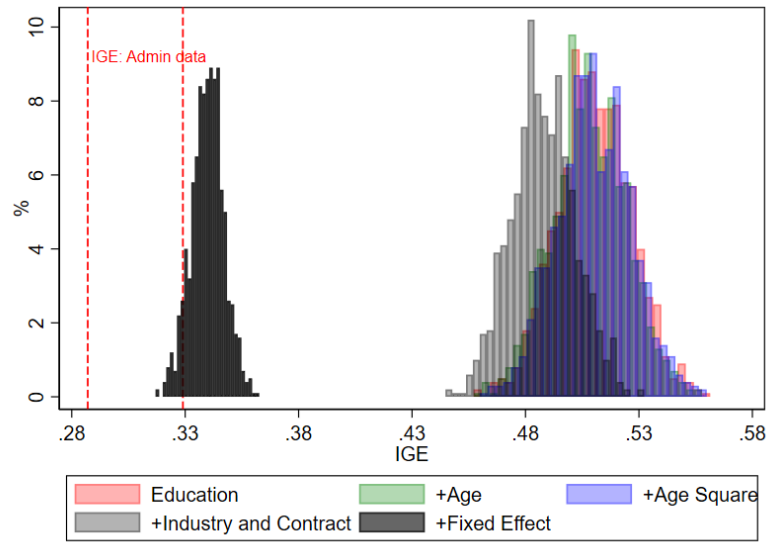
Figure 4.1: IGE Distribution: Alternative set of variables[†]

Note[†]: IGE distributions are shown when alternative set of controls are included in the Mincer's equation. Each distribution is generated by running the regression 1,000 times.

Region	N	β_r	α_r	p_{15}	p_{11}	p_{55}	r_{abs}	r_{per}
Santiago	4711	0.212	43.360	0.183	0.207	0.388	48.654	63.474
Cerrillos	2207	0.175	43.128	0.165	0.184	0.291	47.514	59.792
Cerro Navia	4743	0.140	42.220	0.096	0.216	0.240	45.720	55.520
Conchalí	3876	0.131	46.015	0.159	0.163	0.314	49.286	58.444
El Bosque	5855	0.171	40.689	0.132	0.248	0.258	44.954	56.893
Estación Central	3277	0.207	42.257	0.154	0.203	0.359	47.434	61.931
Huechuraba	2523	0.269	36.537	0.123	0.226	0.366	43.263	62.098
Independencia	1664	0.197	43.278	0.170	0.175	0.368	48.207	62.009
La Cisterna	2259	0.197	44.285	0.236	0.189	0.379	49.506	63.283
La Florida	11614	0.226	42.181	0.154	0.192	0.388	47.819	63.607
La Granja	4257	0.162	43.585	0.158	0.208	0.300	47.646	59.015
La Pintana	6475	0.141	38.496	0.090	0.264	0.173	42.021	51.889
La Reina	2460	0.313	43.944	0.268	0.158	0.582	51.778	73.713
Las Condes	5619	0.409	41.259	0.235	0.166	0.676	51.489	80.132
Lo Barnechea	2353	0.551	29.397	0.149	0.222	0.771	43.172	81.741
Lo Espejo	3789	0.144	41.046	0.091	0.237	0.211	44.641	54.709
Lo Prado	3061	0.180	41.495	0.135	0.188	0.277	45.991	58.578
Macul	3074	0.247	39.830	0.145	0.215	0.372	46.008	63.307
Maipú	17481	0.186	45.625	0.197	0.190	0.351	50.286	63.337
Ñuñoa	3655	0.264	44.424	0.234	0.227	0.467	51.024	69.502
Pedro Aguirre Cerda	3350	0.174	41.523	0.130	0.235	0.283	45.879	58.075
Peñalolén	7261	0.288	34.718	0.109	0.263	0.400	41.917	62.075
Providencia	1627	0.248	50.776	0.284	0.205	0.589	56.983	74.363
Pudahuel	7465	0.151	43.272	0.137	0.221	0.265	47.055	57.646
Quilicura	5469	0.184	43.138	0.114	0.170	0.293	47.728	60.580
Quinta Normal	3050	0.169	44.225	0.153	0.195	0.302	48.760	60.618
Recoleta	4729	0.162	43.364	0.126	0.214	0.267	47.416	58.763
Renca	4892	0.142	44.065	0.135	0.186	0.266	47.620	57.573
San Joaquín	2650	0.160	43.306	0.102	0.208	0.273	47.304	58.498
San Miguel	2111	0.247	42.157	0.184	0.197	0.420	48.328	65.607
San Ramón	2890	0.153	38.985	0.098	0.250	0.179	42.799	53.478
Vitacura	1570	0.271	56.960	NA	NA	0.721	63.740	82.724
Puente Alto	19612	0.189	41.116	0.145	0.220	0.292	45.852	59.113
Pirque	510	0.262	38.296	0.103	0.191	0.509	44.857	63.227
San José de Maipo	385	0.332	36.005	NA	NA	0.381	44.297	67.516
Colina	3551	0.223	38.426	0.098	0.227	0.350	43.990	59.571
Lampa	1593	0.186	39.806	0.081	0.244	0.343	44.466	57.513
Tiltil	630	0.138	46.137	0.195	0.195	0.326	49.599	59.292
San Bernardo	9491	0.189	39.324	0.102	0.250	0.286	44.049	57.277
Buin	2918	0.235	36.048	0.086	0.299	0.313	41.932	58.408
Calera de Tango	650	0.360	32.700	0.144	0.288	0.568	41.697	66.887
Paine	1970	0.184	38.631	0.092	0.235	0.311	43.235	56.125
Melipilla	3228	0.182	39.959	0.110	0.266	0.286	44.498	57.207
Alhué	172	0.244	46.174	NA	NA	NA	52.280	69.380
Curacaví	783	0.134	43.136	0.133	0.235	0.216	46.489	55.876
María Pinto	327	0.136	41.334	0.058	0.256	NA	44.740	54.277
San Pedro	196	0.150	43.680	NA	NA	NA	47.426	57.914
Talagante	2241	0.234	38.799	0.113	0.242	0.363	44.641	60.999
El Monte	827	0.213	38.600	0.072	0.258	0.268	43.930	58.854
Isla de Maipo	1028	0.154	39.982	0.125	0.253	0.330	43.832	54.612
Padre Hurtado	1494	0.126	43.783	0.105	0.145	0.270	46.939	55.775
Peñaflor	2381	0.178	41.758	0.144	0.187	0.319	46.214	58.689

Table 4.4: Intergenerational mobility indicators by municipality in the Metropolitan region. "Santiago" refers to the municipality, not the city.

Percentile	UIP	ENE
1%	152,889	170,614
5%	218,433	231,840
10%	263,508	250,902
25%	343,076	330,000
50%	490,707	451,624
75%	767,851	700,000
90%	1,173,052	1,003,609
95%	1,544,161	1,304,692
99%	2,371,979	2,500,000

**Table 4.5:** Comparison of earnings between our dataset and ENE for individuals between 28-33 years old for 2018.

This dataset is compared with the information of the ENE (Encuesta Nacional de Empleo) questionnaire administred by the goverment statistics agency in Chile (INE-Instituto Nacional de Estadisticas). W ENE refers to the earnings percentiles for all workers – formal, informal and self employed. Units are in 2018 Chilean pesos.

	(1)	(2)	(3)	(4)	(5)	(6)
y_p	0.399***	0.413***	0.344***	0.412***	0.437***	0.342***
	(0.018)	(0.025)	(0.029)	(0.021)	(0.028)	(0.033)
Constant	8.313***	8.245***	8.836***	8.397***	8.219***	9.055***
	(0.216)	(0.292)	(0.338)	(0.230)	(0.308)	(0.362)
Observations	12,498	6,861	5,637	12,498	6,861	5,637
R-squared	0.080	0.116	0.045	0.076	0.114	0.040

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4.6: OLS estimates of the intergenerational earnings elasticity using CASEN 2017 for main sample and CASEN 2000 for auxiliary sample of Pseudo parents

Earnings are measured as monthly earnings (labor income) reported in the survey. We impute $\hat{y}^p$ by Mincer equation using categorical variables of parents' educational type. Column (1) report results for male and female children on the father's earnings. Column (2) report results only for female children on the father's earnings. Column (3) report results only for male children on the father's earnings. Column (4) report results for male and female children on the mother's earnings. Column (5) report results only for female children on the mother's earnings. Column (6) report results only for male children on the mother's earnings.
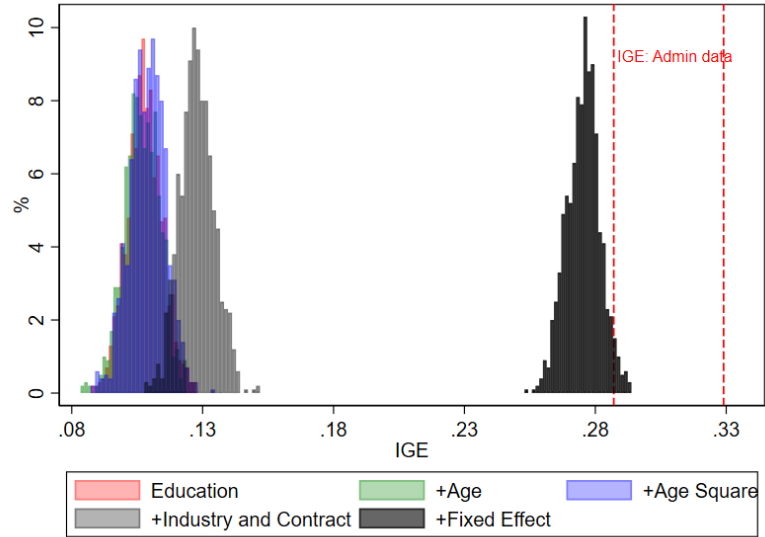
**Figure 4.2:** IGE Distribution: Alternative set of variables under variance correction [†]

Note[†]: IGE distributions are shown when alternative set of controls are included in the Mincer's equation. Each distribution is generated by running the regression 1,000 times under variance correction.
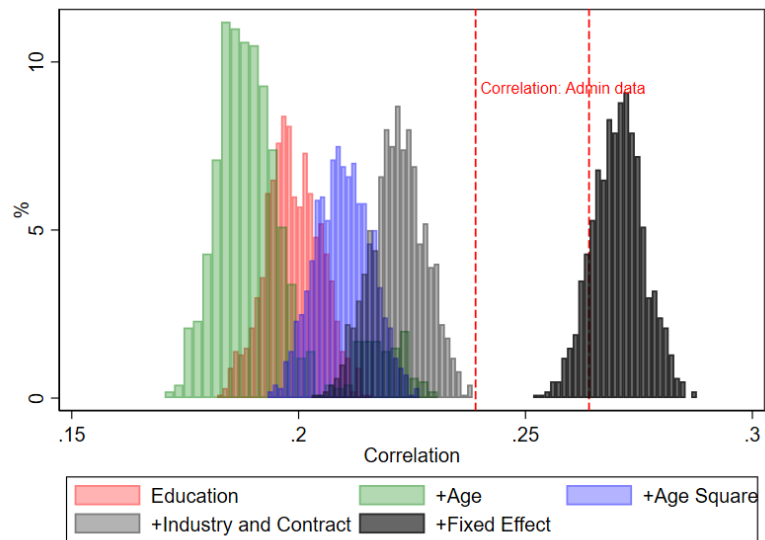
Figure 4.3: Rank-rank Correlation Distribution: Alternative set of variables[†]

Note[†]: Rank-rank correlation distributions are shown when alternative set of controls are included in the Mincer's equation. Each distribution is generated by running the regression 1,000 times.
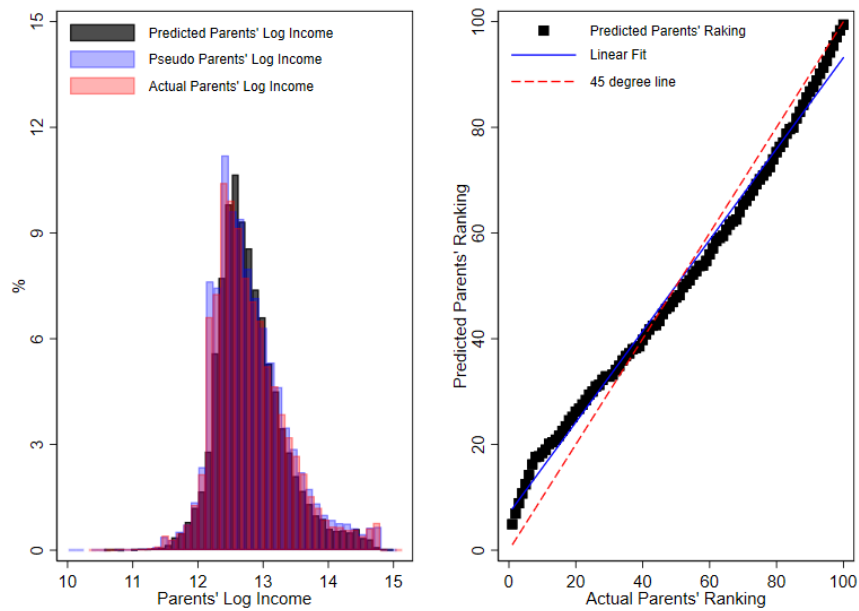
Figure 4.4: Parents' Log Income and Ranking, TSTSLS Method with fixed effects[†]

Note[†]: Left panel shows the distribution of the predicted parents' log income, pseudo parents' log income and actual parent' log income. Right panel depicts the predicted parents' ranking as a function of the actual ranking. A linear fit line and a 45 degree line are also included. Income prediction is constructed with education, age, age square, industry, type of contract and individual fixed effects as control variables.

|	(1) min	(2) p25	(3) p50	(4) mean	(5) p75	(6) max
R^2: with education	0.192	0.207	0.211	0.211	0.216	0.235
R^2: + Age	0.192	0.207	0.212	0.212	0.216	0.235
R^2: + Age Square	0.195	0.210	0.214	0.214	0.218	0.238
R^2: + Industry and Contract	0.243	0.259	0.263	0.263	0.267	0.286
R^2: + Fixed Effect	0.800	0.809	0.811	0.811	0.813	0.822

Table 4.7: R square, Mincer's equation, Alternative set of variables[†].

Note[†]: This table shows a descriptive statistics on R^2 under alternative set of variables.
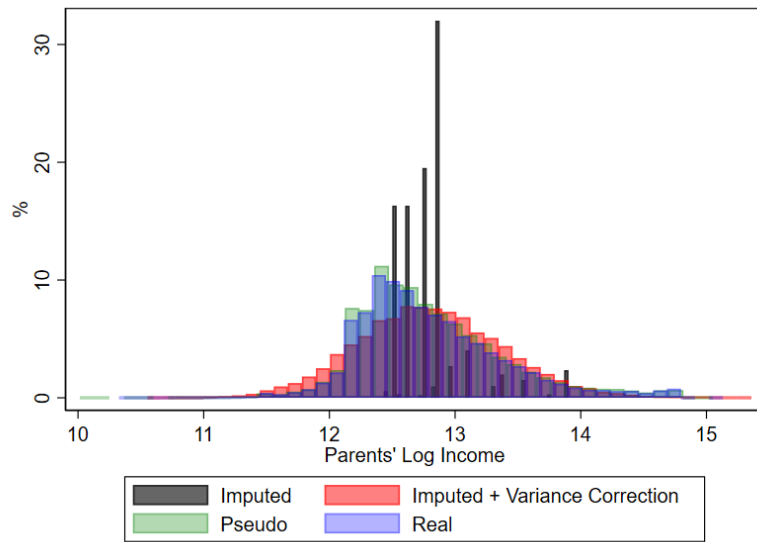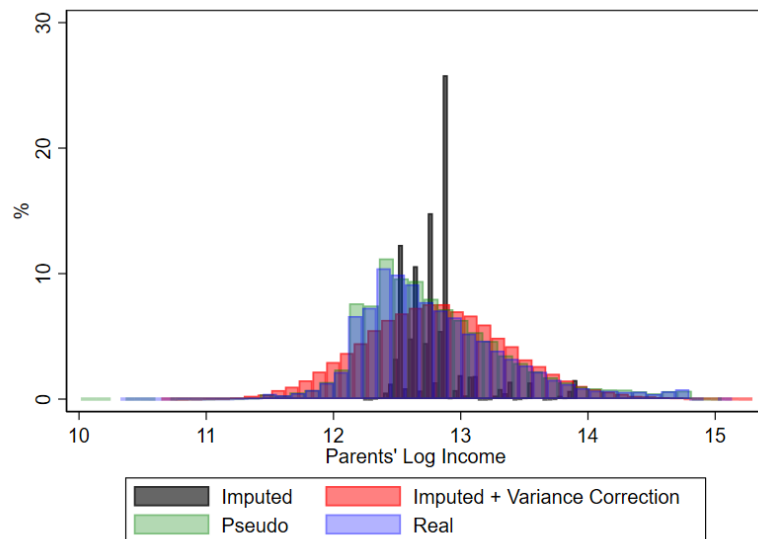

Figure 4.5: Parents' Log Income Distribution: education as control[†]

Note[†]: Income distributions are shown for actual, pseudo and predicted parents. Two predicted parents' incomes are depicted under deterministic and stochastic imputation. Mincer equation includes education as a control variable.
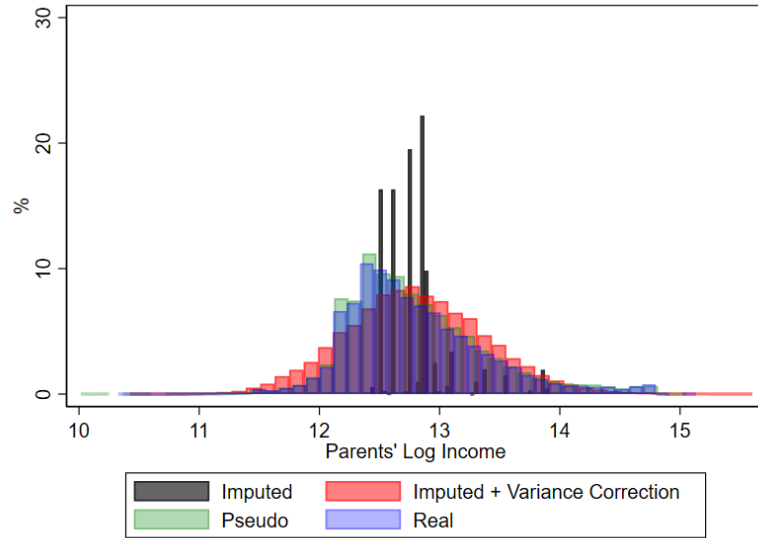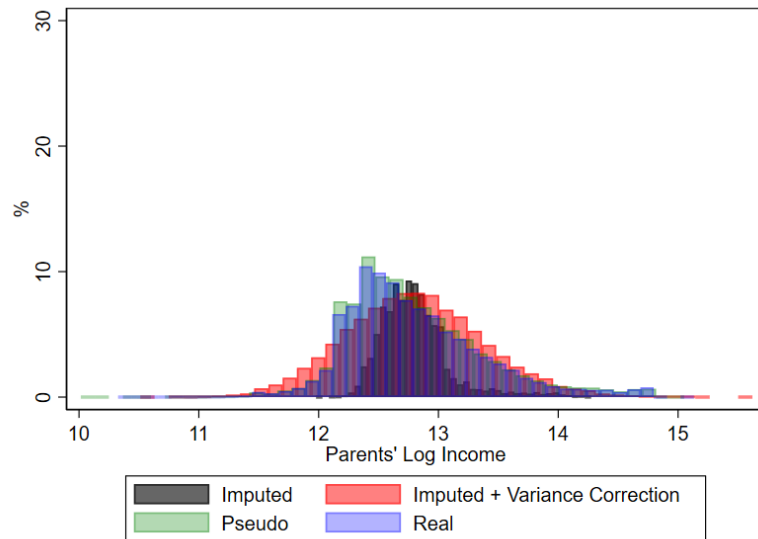
Figure 4.6: Parents' Log Income Distribution: education and age as controls[†]

Note[†]: Income distributions are shown for actual, pseudo and predicted parents. Two predicted parents' incomes are depicted under deterministic and stochastic imputation. Mincer equation includes education and age as control variables.
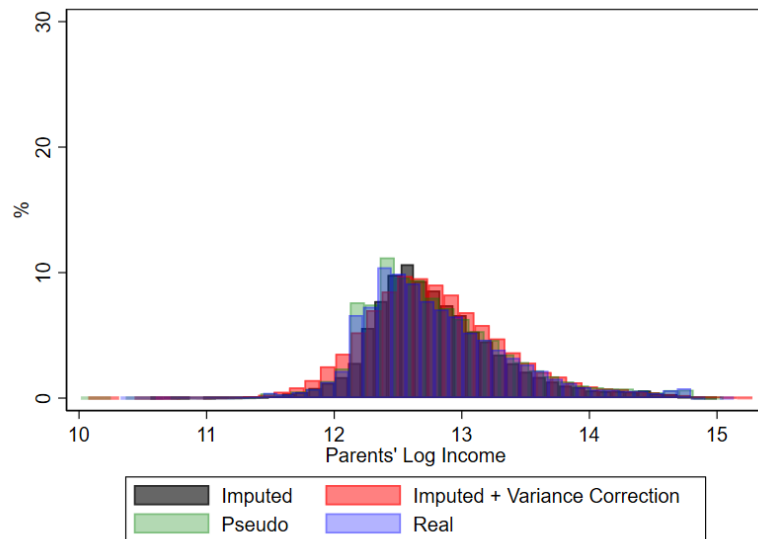

Figure 4.7: Parents' Log Income Distribution: education, age and age square as controls[†]

Note[†]: Income distributions are shown for actual, pseudo and predicted parents. Two predicted parents' incomes are depicted under deterministic and stochastic imputation. Mincer equation includes education, age and age square as control variables.

Figure 4.8: Parents' Log Income Distribution: education, age, age square, and industry, contract as controls[†]

Note[†]: Income distributions are shown for actual, pseudo and predicted parents. Two predicted parents' incomes are depicted under deterministic and stochastic imputation. Mincer equation includes education, age, age square, industry and type of contract as control variables.

Figure 4.9: Parents' Log Income Distribution: education, age, age square, industry, contract, and fixed effects as controls[†]

Note[†]: Income distributions are shown for actual, pseudo and predicted parents. Two predicted parents' incomes are depicted under deterministic and stochastic imputation. Mincer equation includes education, age, age square, industry, type of contract and individual fixed effects as control variables.

Domain A: Teaching, Planning, and Preparation	Domain B: Development of an Adequate Environment for Learning	Domain C: Instruction	Domain D: Professional Responsibilities
A1. Masters the content of the discipline s/he teaches and the national curriculum framework	B1. Establishes an environment of acceptance, equity, trust, solidarity and respect	C1. Communicates learning goals in a precise and clear way	D1. Systematically reflects on his/her performance
A2. Consciousness of students' characteristics, knowledge and experiences	B2. Expresses high expectations about the possibilities of learning and development of all students	C2. The teaching strategies are challenging and significant for the students	D2. Forms teams and professional relationships with his/her colleagues
A3. Masters the didactics of the disciplines s/he teaches	B3. Establishes and keeps consistent rules of behavior in the classroom	C3. The content of the class is rigorously covered and it is understandable for the students	D3. Takes responsibility for his/her students' guidance
A4. Organizes the goals and content in a way that it is coherent with the national curriculum framework and the distinctive feature of his/her students	B4. Organizes a structured environment and uses the available resources	C4. Optimizes the available teaching time	D4. Fosters collaborative and respectful relationships with students' parents or representatives
A5. The evaluation strategies are coherent with the learning goals, the discipline taught and the national curriculum framework and allow all students to show what they have learned		C5. Promotes the development of thinking	
		C6. Evaluates and monitors the process of understanding and appropriation of the contents by the students	

Table 4.8: Domains and Criteria of the Marco para la Buena Enseñanza (MBE). To save space, the descriptors are not included in the table.

The MBE has 4 domains, 20 criteria, and 71 descriptors.

Domain A: Teaching Planning, and Preparation					
Criteria A2: Consciousness of students' characteristics, knowledge, and experiences					
Descriptor	Portfolio (Module 1)	Portfolio (Module 2)	External References	Peer Interview	Self-Evaluation
A.2.1.Consciousness of the developmental characteristics of his/her students according to their age	*			*	*
A.2.2. Consciousness of the family and cultural heterogeneity of the students	*	*	*	*	*
A.2.3. Consciousness of the strengths and weaknesses of the students relative to the content	*	*		*	*
A.2.4. Consciousness of the different learning methods of the students	*	*		*	*

**Table 4.9:** Mapping between the Instruments and the MBE.

Taken from Manzi et al. (2011). In this example we see that descriptor A.2.2 is measured with all the instruments, while the other descriptors are measured with fewer instruments. *=measured by the instrument.

| Module	Dimension	Teaching Practice
Module 1	A. Organization of the unit	a.a. Formulation of goals a.b. Relationship between activities and goals a.c. Sequence of the unit
	B. Analysis of the class	b.a. Analysis based on students' characteristics b.b. Analysis of the carried-out unit b.c. Analysis of the class
	C. Quality of the evaluation	c.a. Evaluations and rubrics used c.b. Relationship between evaluations and goals
	D. Reflection on students' results	d.a. Responsibility for students' results d.b. Students' feedback
Module 2 (video)	F. Environment of the class	f.a. Work environment f.b. Promotion of students' participation f.c. Activity's support and guidance
	G. Structure of the class	g.a. Quality at the beginning of the class g.b. Quality at the end of the class g.c. Activity's contribution to the fulfillment of goals
	H. Pedagogical interaction	h.a. Developed explanations h.b. Quality of the questions asked of the students h.c. Feedback quality h.d. Curricular emphasis

Table 4.10: Dimensions and Teaching Practices Evaluated by the Portfolio.

Own elaboration based on the document "Evaluación Docente como herramienta de gestión escolar." CPEIP, Mineduc. Dimension E was not considered in the estimation as it was discontinued.
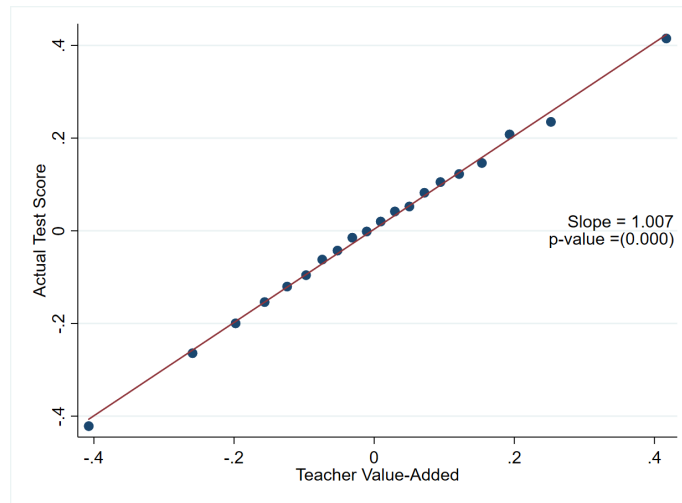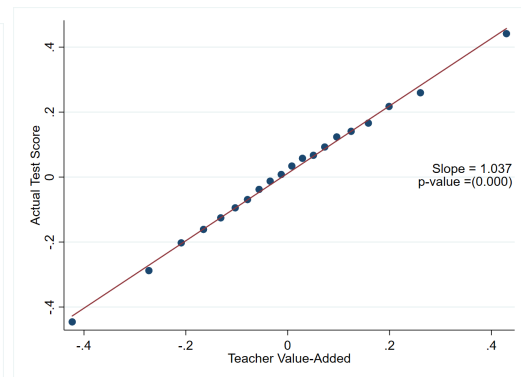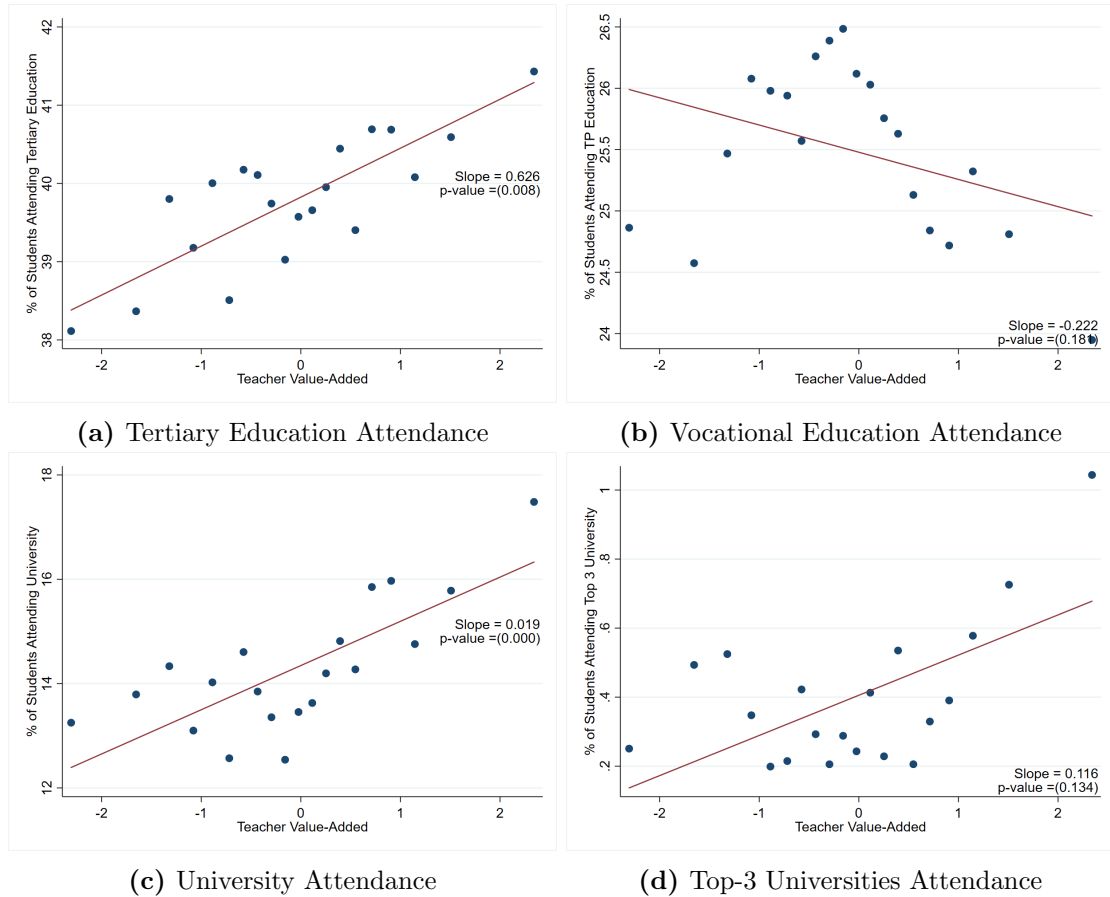
(a) Baseline
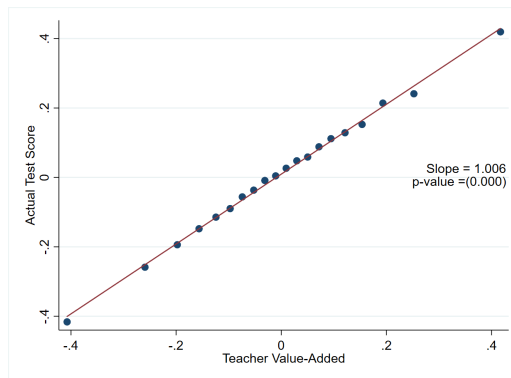
(b) Added includes Two Lags of Test Score

(c) Added includes Vulnerability Conditions

Figure 4.10: Effect of Teachers' Value-Added on Actual Scores in Public Schools

These Figures are constructed using the sample used to estimate VA model for public schools. To construct this binned scatter plot, we first residualisze the actual test score with respect to the baseline control vector (detailed in the note of Table 3.3) separately within each subject and using within-teacher variation to estimate the coefficients for public schools. Then divide the VA estimates $\hat{\mu}_{jt}$ into twenty equal-sized groups (vingtiles) and plot the means of the actual test score residuals within each bin against the mean value of $\hat{\mu}_{jt}$ within each bin. The line shows the best linear fit estimated on the underlying micro data using OLS

(a) Tertiary Education Attendance

(b) Vocational Education Attendance

(c) University Attendance

(d) Top-3 Universities Attendance

Figure 4.11: Conditional mean of Teachers' VA on Educational Outcomes in Public Schools

Panels (a) to (d) are binned scatter plots of tertiary education attendance rates, vocational education attendance rates, university attendance rates and top-3 university attendance rates vs. normalized teacher VA $\hat{m}_{jt}$ for public schools. These plots correspond to the regressions in the first column of each outcome-specification of Table 3.6 and use the same sample restrictions and variable definitions. To construct these binned scatter plots, we first residualize the dependent variable with respect to the baseline control vector separately within each subject, using within-teacher variation to estimate the coefficients. We then divide the VA estimates $\hat{m}_{jt}$ into twenty equal-sized groups (vingtiles) and plot the means of the dependent variable residuals within each bin against the mean value of $\hat{m}_{jt}$ within each bin. Finally, we add back the unconditional mean of the dependent variable in the estimation sample to facilitate interpretation of the scale. The solid line shows the best linear fit estimated on the underlying micro data using OLS.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Tertiary Education Attendance			Vocational Education Attendance		
Portfolio Score	1.708	1.678	1.297	-0.305	-0.299	-0.346
	(0.000)	(0.000)	(0.000)	(0.014)	(0.015)	(0.018)
Self Evaluation Score	0.461	0.454	0.332	-0.002	-0.001	-0.024
	(0.003)	(0.003)	(0.071)	(0.981)	(0.994)	(0.856)
Peer Interview Score	-0.470	-0.426	-0.429	0.475	0.466	0.410
	(0.010)	(0.018)	(0.033)	(0.000)	(0.000)	(0.008)
External References Score	1.741	1.730	1.863	-0.977	-0.974	-0.938
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Mean Dep. Var. (Public School)	47.27	47.27	47.92	22.97	22.97	22.97
Observations	1,805,012	1,805,012	1,165,653	1,805,012	1,805,012	1,165,653
Over Mean Dep. Ver.						
Portfolio Score	3.6%	3.5%	2.7%	-1.3%	-1.3%	-1.5%
Self Evaluation Score	1.0%	1.0%	0.7%	0.0%	0.0%	-0.1%
Peer Interview Score	-1.0%	-0.9%	-0.9%	2.1%	2.0%	1.8%
External References Score	3.7%	3.7%	3.9%	-4.3%	-4.2%	-4.1%

Table 4.11: Instruments of Teacher Evaluation on Outcomes - Residual Regression.

Each column reports coefficients from an OLS regression only for public schools, with standard errors clustered by school-cohort and p-value in parentheses. Dependent variable of first column of each outcome-specification correspond to residuals of educational outcome using the same control vector used to estimate baseline VA model detailed in Table 3. Dependent variable of second column of each outcome-specification, add in the estimate of residuals of educational outcome the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. Dependent variable of third column of each outcome-specification, add in the estimate of residuals of educational outcome the twice-lagged test scores.The regressions are run on the sample used to estimate the baseline Teacher Evaluation model, restricted to observations with a non-missing Teacher Evaluation. There is one observation for each student-subject-grade-school year in all regressions.The score for each instrument of Teacher Evaluation is scaled in units of student test score standard deviations and is estimated using data from classes taught by the same teacher in other years. Impact of each instrument of Teacher Evaluation is estimated for each subject (language and mathematics) and according to level (primary and high school). The last row of each table corresponds to the ratio between the impact of each instruments of Teacher's Value-Added on the average of the dependent variable.

Variable	(7)	(8)	(9)	(10)	(11)	(12)
	University Attendance			Top-3 University Attendance		
Portfolio Score	2.013	1.977	1.643	0.158	0.157	0.194
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Self Evaluation Score	0.464	0.454	0.356	0.004	0.004	-0.003
	(0.006)	(0.006)	(0.073)	(0.896)	(0.903)	(0.936)
Peer Interview Score	-0.945	-0.892	-0.839	-0.068	-0.066	-0.090
	(0.000)	(0.000)	(0.002)	(0.063)	(0.069)	(0.085)
External References Score	2.718	2.705	2.801	0.241	0.241	0.241
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Baseline Controls	x	x	x	x	x	x
Vulnerability Conditions		x			x	
Year t-2 Test Score			x			x
Mean Dep. Var. (Public School)	24.30	24.30	24.95	1.219	1.219	1.567
Observations	1,805,012	1,805,012	1,165,653	1,805,012	1,805,012	1,165,653

Over Mean Dep. Ver.						
Portfolio Score	8.3%	8.1%	6.6%	13.0%	12.9%	12.4%
Self Evaluation Score	1.9%	1.9%	1.4%	0.3%	0.3%	-0.2%
Peer Interview Score	-3.9%	-3.7%	-3.4%	-5.6%	-5.4%	-5.7%
External References Score	11.2%	11.1%	11.2%	19.8%	19.8%	15.4%

Table 4.12: Instruments of Teacher Evaluation on Outcomes - Residual Regression.

Each column reports coefficients from an OLS regression only for public schools, with standard errors clustered by school-cohort and p-value in parentheses. Dependent variable of first column of each outcome-specification correspond to residuals of educational outcome using the same control vector used to estimate baseline VA model detailed in Table 3. Dependent variable of second column of each outcome-specification, add in the estimate of residuals of educational outcome the condition of vulnerability of the students, a discrete variable equal to 1 if the student belongs to the most vulnerable 40% of the population. Dependent variable of third column of each outcome-specification, add in the estimate of residuals of educational outcome the twice-lagged test scores. The regressions are run on the sample used to estimate the baseline Teacher Evaluation model, restricted to observations with a non-missing Teacher Evaluation. There is one observation for each student-subject-grade-school year in all regressions. The score for each instrument of Teacher Evaluation is scaled in units of student test score standard deviations and is estimated using data from classes taught by the same teacher in other years. Impact of each instrument of Teacher Evaluation is estimated for each subject (language and mathematics) and according to level (primary and high school). The last row of each table corresponds to the ratio between the impact of each instruments of Teacher's Value-Added on the average of the dependent variable.

Correlations Matrix	Teachers' Value Added	Weighted Total Score	Portfolio Score	Self Evaluation Score	Peer Interview Score	External References Score
Teachers' Value Added	1.000					
Weighted Total Score	-0.018	1.000				
Portfolio Score	-0.011	0.793	1.000			
Self Evaluation Score	-0.001	0.249	0.051	1.000		
Peer Interview Score	-0.045	0.637	0.162	0.061	1.000	
External References Score	0.043	0.414	0.140	0.123	0.110	1.000

Table 4.13: Correlations between Normalized Teacher's Value-Added and Normalized Instruments of Teacher's Evaluation Test.

Corresponds to the correlations between Normalized Teacher's Value-Added and Normalized Instruments of Teacher's Evaluation Test for teachers teaching in public schools. We standardize according to year and grade for each instruments of Teacher's Evaluation Test, such that for each year-grade, it has a mean of zero and variance of one.

Bibliography

[1] An, Y., Wang, Le. and Xiao, R. (2020). "A Nonparametric Nonclassical Measurement Error Approach to Estimating Intergenerational Mobility Elasticities". Journal of Business & Economic Statistics, Pages 1-17.

[2] Aaronson, D., Barrow, L., and Sander, W. (2007). "Teachers and student achievement in Chicago Public High Schools". Journal of Labor Economics, 24(1), Pages 95-135.

[3] Aaronson, D. and Mazumder, B. (2008) "Intergenerational economic mobility in the united states, 1940 to 2000", Journal of Human Resources, 43(1), Pages 139-172.

[4] Acciari, P., Polo, A. and Violante, G., "'And Yet, it Moves': Intergenerational Mobility in Italy (April 2019)". CEPR Discussion Paper No. DP13646. Available at SSRN: https://ssrn.com/abstract=3368143.

[5] Alesina, A., Hohmann, S., Michalopoulos, S. and Papaioannou, E. (2019). "Intergenerational Mobility in Africa" (No. w25534). National Bureau of Economic Research.

[6] Angrist, Joshua D and Krueger, Alan B. (1992). "The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples". Journal of the American statistical Association, Volume 87, No 418, Pages 328-336

[7] Angrist, J., Hull, P., Pathak, P. and Walters, C. (2017). "Leveraging Lotteries for School Value-Added: Testing and Estimation". Quarterly Journal of Economics, Volume 132, Issue 2, Pages 871-919.

[8] Asher, S., Novosad, P., and Rafkin, C. (2018). "Intergenerational Mobility in India: Estimates from New Methods and Administrative Data". World Bank Working Paper. Available at: http://www. dartmouth. edu/ novosad/anr-india-mobility. pdf (accessed December 2018).

[9] Bacher-Hicks, A., Chin, M., Kane, T., and Staiger, D. (2017). "An evaluation of bias in three measures of teacher quality: Value-added, classroom observations, and student surveys". National Bureau of Economic Research.

[10] Barbieri, T., Bloise, F. and Raitano, M. (2020), "Intergenerational earnings inequality: new evidence from italy", Review of Income and Wealth, 66(2), Pages 418-443.

[11] Bau, N. and Das, J. (2020) "Teacher Value Added in a Low-Income Country". American Economic Journal: Economic Policy, 12(1), Pages 62-96.

[12] Bell, C., Gitomer, D., McCaffrey, B., Hamre, B., Pianta, R. and Qi, Y. (2012). "An argument approach to observation protocol validity". Educational Assessment, 17(2-3), 62-87.

[13] Björklund, A., and Jäntti, M. (1997). "Intergenerational income mobility in Sweden compared to the United States". The American Economic Review, 87(5), Pages 1009-1018

[14] Björklund, A., and Kjellström, C. (2002). "Estimating the return to investments in education: how useful is the standard Mincer equation?". Economics of Education Review, Volume 21, No 3, Pages 195-210.

[15] Blanden, J. (2013), "Cross-country rankings in intergenerational mobility: a comparison of approaches from economics and sociology", Journal of Economic Surveys 27(1), Pages 38-73.

[16] Black, S., and Devereux, J. (2011). "Recent Developments in Intergenerational Mobility". In Handbook of Labor Economics, edited by David Card and Orley Ashenfelter, 4, Part B:Pages 1487-1541. Elsevier.

[17] Bloise, F., Brunori, P., Piraino, P. (2020). "Estimating intergenerational income mobility on sub-optimal data: a machine learning approach", Technical report.

[18] Bollinger, C., Hirsch, B., Hokayem, Ch., and Ziliak, J. (2018). "Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch". The Journal of Political Economy. Forthcoming.

[19] Bratsberg, B., Roed, K., Raaum, O., Naylor, R., Jäntti, M., Eriksson, T. and Österbacka, E. (2007). "Nonlinearities in intergenerational earnings mobility: consequences for cross-country comparisons". The Economic Journal, 117(519), C72-C92.

[20] Briole, S. and E. Maurin (2019) "Does Evaluating Teachers Make a Difference?". IZA Discussion Paper Series No. 12307

[21] Canales, A. and Maldonado, L. (2018) "Teacher quality and student achievement in Chile: Linking teachers' contribution and observable characteristics". International Journal of Educational Development, Volume 60, Pages 33–50.

[22] Card, D. and Krueger, A. (1992) "Does school quality matter? Returns to education and the characteristics of public schools in the United States". Journal of political Economy, Volume 100, No 1, Pages 1-40.

[23] Carmichael, F., Darko, C., Ercolani, M., Ozgen, C. and Siebert, W. (2020). "Evidence on intergenerational income transmission using complete dutch population data", Economics Letters 189, 108996.

[24] Celhay, P., Sanhueza, C. and Zubizarreta, J. (2010). " "Intergenerational Mobility of Income and Schooling: Chile 1996-2006."

[25] Chetty, R., Hendren, N., Kline, P. and 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), Pages 1553-1623.

[26] Chetty, R., Hendren, N., Kline, P., Saez, E. and Turner, N. (2014). "Is the United States still a land of opportunity? Recent trends in intergenerational mobility". American Economic Review, Volume 104, No 5.

[27] Chetty, R., Friedman, J. and Rockoff, J. (2014a). "Measuring the impacts of teacher I: Evaluating the bias in teacher value-added estimates". American Economic Review, 104(9), Pages 2593-2632.

[28] Chetty, R., Friedman, J. and Rockoff, J. (2014b). "Measuring the impacts of teacher II: Teacher Value-Added and Student Outcomes in Adulthood". American Economic Review, 104(9), Pages 2633-2679.

[29] Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R. and Narang, J. (2017). "The fading American dream: Trends in absolute income mobility since 1940". Science, 356(6336), Pages 398-406.

[30] Chetty, R. and Hendren, N. (2018a). "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects". The Quarterly Journal of Economics, 133(3), Pages 1107-1162.

[31] Chetty, R., and Hendren N. (2018b). "The Impacts of Neighborhoods on Intergenerational Mobility II: County Level Estimates". Quarterly Journal of Economics, 133 (3), Pages 1163-1228.

[32] Chin, M. and Goldhaber, D. (2015) "Exploring Explanations for the 'Weak' Relationship Between Value Added and Observation-Based Measures of Teacher Performance". Mimeo, Center for Education Policy Research at Harvard University.

[33] Connolly, M., Corak, M. and Haeck, C. (2019), "Intergenerational mobility between and within canada and the united states", Journal of Labor Economics 37(S2), S595–S641.

[34] Contreras, D., Bustos, S. and Sepúlveda, P. (2010). "When schools are the ones that choose: The effect of screening in Chile". Social Science Quarterly, 91, Pages 1349-1368.

[35] Corak, M. and Heisz, A. (1999). "The intergenerational earnings and income mobility of Canadian men: Evidence from longitudinal income tax data". Journal of Human Resources, 504-533.

[36] Corak, M. (2006). "Do poor children become poor adults? lessons from a cross-country comparison of generational earnings mobility", Research on economic inequality, 13(1), Pages 143-188.

[37] Corak, M. (2013). "Income inequality, equality of opportunity, and intergenerational mobility". Journal of Economic Perspectives, 27(3), Pages 79-102.

[38] Corak M. (2019). "The Canadian Geography of Intergenerational Income Mobility", The Economic Journal, uez019, https://doi.org/10.1093/ej/uez019

[39] Darling-Hammond, L. (2015) "Can Value Added Add Value to Teacher Evaluation?" Educational Researcher, Volume 44, issue 2, Pages 132-137.

[40] Daude, C. and Robano, V. (2015). "On intergenerational (im) mobility in Latin America". Latin American Economic Review, 24(1), 9.

[41] Deutscher, N, and Mazumder, B. (2020). "Intergenerational mobility across Australia and the stability of regional estimates". Labour Economics, 66, 101861.

[42] Dunn, Christopher E. (2007). "The intergenerational transmission of lifetime earnings: Evidence from Brazil". The BE Journal of Economic Analysis & Policy, Volume 7, No 2.

[43] Durlauf, S., Kourtellos, A. and Tan, Ch. (2017). "Status traps". Journal of Business & Economic Statistics, Volume 35, No 2, Pages 265-287.

[44] Emran, M., Ferreira, F., Jiang, Y. and Sun, Y. (2019). "Intergenerational Educational Mobility in Rural Economy: Evidence from China and India". Available at SSRN 3393904.

[45] Emran, M. S. and Shilpi, F. J. (2019), "Economic approach to intergenerational mobility: Measures, methods, and challenges in developing countries", Technical report, WIDER Working Paper.

[46] Ferreira, S. and Veloso, F. (2006). "Intergenerational mobility of wages in Brazil". Brazilian Review of Econometrics, Volume 26, No 2, Pages 181-211.

[47] Flores, I., Atria, J., Sanhueza, C. and Mayer, R. (2019). "Top Incomes in Chile: A Historical Perspective of Income Inequality (1964-2017)". Review of Income and Wealth (forthcoming).

[48] Fortin, N. and Lefebvre, S. (1998). "Intergenerational income mobility in Canada. Labour markets, social institutions, and the future of Canada's children", (89-553).

[49] Glazerman, S., Potik, A., Teh, B., Bruch, J., Max, J. and Warner, E. (2013). "Transfer incentives for high-performing teachers: Final results from a multi-site randomized experiment". United States, Department of Education.

[50] Goldhaber, D. and Chaplin, D. (2015). "Assessing the Rothstein falsification test. Does it really show teacher value added models are biased?" Journal of Research on Educational Effectiveness, 8(1), Pages 8-34.

[51] Goldhaber, D. and Hansen, M. (2013). "Is it just a bad class? Assessing the stability of measured teacher performance". Economica, 80(319), Pages 589-612.

[52] Gong, H., Leigh, A. and Meng, X. (2012). "Intergenerational income mobility in urban china", Review of Income and Wealth 58(3), Pages 481-503.

[53] Grawe, N. (2001). "Intergenerational mobility in the U.S. and abroad : quantile and mean regression measures", Doctoral Dissertation/University of Chicago.

[54] Grawe, N. (2004). "Intergenerational mobility for whom? The experience of high-and low-earning sons in international perspective". Generational income mobility in North America and Europe, Pages 58-89.

[55] Grossman, P., Loeb, S., Cohen, J. and Wyckoff, J. (2013). "Measure for measure: The relationship between measures of instructional practice in middle school English arts and teachers' value-added scores". American Journal of Education, 119, Pages 445-470.

[56] Guarino, C., Reckase, M., Stacey, B. and Wooldridge, J. (2015). "Can value-added measures of teacher performance be trusted?" Education, Finance and Policy, 10(1), Pages 117-156.

[57] Güell, M., Rodríguez Mora, J. and Telmer, C. (2015). "The informational content of surnames, the evolution of intergenerational mobility, and assortative mating". The Review of Economic Studies, 82(2), Pages 693-735.

[58] Güell, M., Pellizzari, M., Pica, G. and Rodríguez Mora, J. (2018). "Correlating social mobility and economic outcomes". The Economic Journal, 128(612), pp.F353-F403. (2018): Correlating Social Mobility and Economic Outcomes, The Economic Journal, 128(July), F353-F403.

[59] Haider, S. and Solon, G. (2006). "Life-cycle variation in the association between current and lifetime earnings", American Economic Review, 96(4), Pages 1308-1320.

[60] Hanushek, E. and Rivkin, S. (2010). "Generalizations about using value added measures of teacher quality". American Economic Review, 100(2), Pages 26-271.

[61] Harris, D. (2012). "How do Value-Added indicators compare to other, measures of teacher effectiveness?". Draft. Carnegie Knowledge Network What we know series: Value-added methods and application. Stanford, CA. Retrieved from http://www.carnegieknowledgenetwork.org/wp-content/uploads/2012/10/CKN_201210_Harris.pdf

[62] Heidrich, S. (2017). "Intergenerational mobility in Sweden: a regional perspective". Journal of Population Economics, 30(4), Pages 1241-1280.

[63] Heckman, J., Lochner, L. and Todd, P. (2003). "Fifty years of Mincer earnings regressions". National Bureau of Economic Research.

[64] Hill, H. (2005). "Content across communities: Validating measures of elementary mathematics instruction". Educational Policy, 19(3), Pages447-475.

[65] Hill, H., Kapitula, L. and Umland, K. (2011). "A validity argument approach to evaluating teacher value-added scores". American Educational Research Journal, 48(3), Pages 794-831.

[66] Inoue, A. and Solon, G. (2010). "Two-sample instrumental variables estimators", The Review of Economics and Statistics 92(3), 557-561.

[67] Isoré, M. (2009). "Teacher evaluation: Current practices in OECD countries and a literature review". OECD Education Working Papers, No. 23, OECD Publishing, Paris.

[68] Jackson, K. (2014). "Teacher quality at the high-school level: The importance of accounting for tracks". Journal of Labor Economics, 23(4), Pages 645-684.

[69] Jacob, B. and Lefgren, L. (2008). "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education". Journal of Labor Economics, 26(1), Pages 101-136.

[70] Jäntti, M., Bratsberg, B., Roed, K., Raaum, O., Naylor, R., Osterbacka, E., Bjorklund, A. and Eriksson, T. (2006). "American exceptionalism in a new light: a comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States."

[71] Jäntti, M. and Jenkins, S. (2015). "Income mobility". In Handbook of income distribution (Volume 2, Pages 807-935). Elsevier.

[72] Jerrim, J., Choi, Á. and Simancas, R. (2016). "Two-sample two-stage least squares (tstsls) estimates of earnings mobility: how consistent are they?", in 'Survey Research Methods', Vol. 10, Pages 85-101.

[73] Jiménez, M. (2011). 'Un análisis empírico de las no linealidades en la movilidad intergeneracional del ingreso. El caso de la Argentina". Centro de Estudios Distributivos, Laborales y Sociales.

[74] Kane, T., Rockoff, J. and Staiger, D. (2008). "Estimating teacher impacts on student achievement: An experimental evaluation". NBER Working Paper 14607.

[75] Kane, T., Taylor, E., Tyler, J. and Wooten, A. (2011). "Identifying effective classroom practices using student achievement data". Journal of Human Resources, 46(3), Pages587-613.

[76] Kane, T., McCaffrey, D., Miller, T. and Staiger, D. (2013). "Have we identified effective teachers? Validating measures of effective teaching using random assignment". Seattle, WA: Bill and Melinda Gates Foundation.

[77] Klevmarken, A. (1982). "Missing variables and two-stage least-squares estimation from more than one data set", Technical report, IUI Working Paper.

[78] Koedel, C., Mihaly, K. and Rockoff, J. (2015). "Value added modeling: A review". Working Paper 1501, University of Missouri.

[79] Lambert, S., Ravallion, M. and Van de Walle, D. (2014). "Intergenerational mobility and interpersonal inequality in an African economy". Journal of Development Economics, 110, Pages 327-344.

[80] Lara, B., Mizala, A. and Repetto, A. (2010). "Una mirada a la efectividad de los profesores en Chile". Estudios Públicos, 120, Pages 147–182.

[81] Leites, M., Sena, E. and Vilá, J. (2014). "Movilidad intergeneracional de ingresos en Uruguay. Una mirada basada en registros administrativos". El Futuro en Foco. Cuadernos sobre Desarrollo Humano.

[82] Lindahl, L. (2008). "Do birth order and family size matter for intergenerational income mobility? evidence from sweden", Applied economics 40(17), 2239-2257. Little, R. J. and Rubin, D. B. (2019), Statistical analysis with missing data, Volume 793, John Wiley & Sons.

[83] Lombardi, M. (2019). "Is the remedy worse than the disease? The impact of teacher remediation on teacher and student performance in Chile". Economics of Education Review, Volume 73.

[84] Lynch, K., Chin, M. and Blazar, D. (2017). "Relationships between observations of elementary teacher mathematics instruction and student achievement: Exploring variability across districts". American Journal of Education, Volume 123, No 4.

[85] Manzi, J., González, R. and Sun, Y. (2011). "La Evaluación Docente en Chile". Mide UC, Facultad de Ciencias Sociales, Escuela de Psicología, Universidad Católica de Chile.

[86] Mazumder, B. (2005). "Fortunate sons: New estimates of intergenerational mobility in the united states using social security earnings data", Review of Economics and Statistics, 87(2), Pages 235-255.

[87] McCaffrey, D., Sass, T., Lockwood, J. and Mihaly, K. (2009). "The intertemporal variability of teacher effect estimates". Education, Finance and Policy, 4(4), Pages 572-606.

[88] Meckes, L. and Bascopé, M. (2012). "Uneven distribution of novice teachers in the chilean primary school system". Education Policy Analysis Archives, 20, (2012)

[89] Meneses, F. (2021). "Intergenerational Mobility After Expanding Educational Opportunities: A Quasi Experiment". Job Market Paper, Duke University.

[90] Mincer, J. (1974). "Schooling, Experience, and Earnings". Human Behavior & Social Institutions No. 2.

[91] Mincer, J. (1996). "Changes in wage inequality, 1970-1990". Human Behavior & Social Institutions No. 2.

[92] Mizala, A. and Torche, F. (2012). "Bringing the schools back in: the stratification of educational achievement in the Chilean voucher system". International Journal of Educational Development, Volume 32 (1), Pages 132–144.

[93] Mocetti, S. (2007). "Intergenerational earnings mobility in italy", The BE Journal of Economic Analysis & Policy 7(2).

[94] Murphy, K. and Topel, R. (2002). "Estimation and inference in two-step econometric models", Journal of Business & Economic Statistics, 20(1), Pages 88-97.

[95] Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Mahler, D., Ramasubbaiah, R. and Thewissen, S. (2018). "Fair Progress?: Economic Mobility Across Generations Around the World", World Bank Publications.

[96] Neidhöfer, G. (2019). "Intergenerational mobility and the rise and fall of inequality: Lessons from Latin America". Available at SSRN 2740395 .

[97] Neidhöfer, G., Serrano, J. and Gasparini, L. (2018). "Educational inequality and intergenerational mobility in Latin America: A new database". Journal of Development Economics, 134, Pages 329-349.

[98] Nicoletti, C. and Ermisch, J. F. (2008). "Intergenerational earnings mobility: changes across cohorts in britain", The BE Journal of Economic Analysis & Policy 7(2).

[99] Núñez, J. and Risco, C. (2004). "Intergenerational income mobility in a developing country: the case of chile: The case of chile", Documento de Trabajo, 210.

[100] Núñez, J. and Miranda, L. (2010). "Intergenerational income mobility in a less-developed, high-inequality context: The case of Chile". The BE Journal of Economic Analysis and Policy, 10(1).

[101] Núñez, J., and Miranda, L. (2011). "Movilidad intergeneracional del ingreso y la educación en zonas urbanas de Chile". Estudios de economía, 38(1), Pages 195-221.

[102] OECD. (2013a). "Synergies for Better Learning: An International Perspective on Evaluation and Assessment". OECD Reviews of Evaluation and Assessment in Education, Editions OCDE, Paris.

[103] OECD. (2013b). "Teachers for the 21st Century: Using Evaluation to Improve Teaching". OECD Publishing.

[104] OECD. (2018). "A broken social elevator? how to promote social mobility", COPE Policy Brief.

[105] Olivetti, C. and Paserman, M. D. (2015). "In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1940", American Economic Review 105(8), Pages 2695-2724.

[106] Olivetti, C., Paserman, M. D. and Salisbury, L. (2018). "Three-generation mobility in the United States, 1850-1940: The role of maternal and paternal grandparents". Explorations in Economic History, Volume 70, pages 73-90.

[107] Pacini, D. and Windmeijer, F. (2016). "Robust inference for the two-sample 2sls estimator", Economics letters, 146, Pages 50-54.

[108] Pianta, R., Hamre, B., Haynes, N., Mintz, S. and La Paro, K. (2006). "Classroom Assessment Score System (CLASS) manual: Middle/secondary version pilot". Charlottesville: Curry School of Education, University of Virginia.

[109] Piraino, P. (2007). "Comparable estimates of intergenerational income mobility in italy", The BE Journal of Economic Analysis & Policy 7(2).

[110] Piraino, P. (2015). "Intergenerational earnings mobility and equality of opportunity in south africa", World Development, 67, Pages 396-405.

[111] Rivkin, S., Hanushek, E. and Kain, J. (2005). "Teachers, schools, and academic achievement". Econometrica, 73(2): Pages 417-458.

[112] Rothstein, J. (2010). "Teacher quality in educational production: Tracking, decay, and student achievement". Quarterly Journal of Economics, 125(1), Pages 175-214.

[113] Rubin, D. (1978). "Multiple imputations in sample surveys-a phenomenological bayesian approach to nonresponse", in 'Proceedings of the survey research methods section of the American Statistical Association', Vol. 1, American Statistical Association, Pages 20-34.

[114] Rubin, D. (1996). "Multiple imputation after 18+ years", Journal of the American statistical Association 91(434), Pages 473-489.

[115] Rubin, D. (2004). "Multiple imputation for nonresponse in surveys", Vol. 81, John Wiley & Sons.

[116] Sapelli, C. (2013). "Movilidad intrageneracional del ingreso en chile", Estudios Públicos, 131, Pages 1-35.

[117] Schnelle, K. (2015). "Intergenerational Mobility in Norway: Transition Probabilities and Directional Rank Mobility", Mimeo, University of Bergen.

[118] Sehnbruch, K. and Carranza, R. (2015). "The Chilean system of unemployment insurance savings accounts". Universidad de Chile, Departmento de Economía.

[119] Simard-Duplain, G. and St-Denis, X. (2020). "Exploration of the Role of Education in Intergenerational Income Mobility in Canada: Evidence from the Longitudinal and International Study of Adults". Canadian Public Policy, (aop), e2019072.

[120] Solon, G. (1992). "Intergenerational income mobility in the united states", The American Economic Review, Pages 393-408.

[121] Solon, G. (2002). "Cross-country differences in intergenerational earnings mobility". Journal of Economic Perspectives 16.3 (2002), Pages59-66.

[122] Taylor, E. and Tyler, J. (2012). "The effect of evaluation on teacher performance". American Economic Review, 102(7), Pages 3628-51.

[123] Toledo, G. and Valenzuela, J.P. (2015). "Over-estimating the effects of teacher attributes on school performance in the Chilean education system". Estudios de Economía, Volume 42, No. 1, June 2015.

[124] Torche, F. (2005). "Unequal but fluid: social mobility in Chile in comparative perspective". American Sociological Review, 70(3), Pages 422-450.

[125] Torche, F. (2014). "Intergenerational mobility and inequality: The Latin American case". Annual Review of Sociology, 40, Pages 619-642.

[126] Valenzuela, J., Bellei, C. and De los Ríos D. (2013). "Socioeconomic Segregation in a market-oriented Educational System. The Case of Chile". Journal of Educational Policy, 29 (2), Pages 217-241

[127] Wooldridge, J. (2010). "Econometric analysis of cross section and panel data", MIT press.

[128] Wyness, G., Murphy, R. and Weinhardt, F. (2018). "Who Teaches the Teachers? a RCT of Peer-to-Peer Observation and Feedback in 181 Schools". Discussion Paper No.116, Rationality & Competition.

[129] Zimmerman, D. (1992). "Regression toward mediocrity in economic stature", The American Economic Review, Pages 409-429.

[130] Zimmerman, S. (2019). "Elite colleges and upward mobility to top jobs and top incomes". American Economic Review, 109(1), Pages 1-47.