

CAF - WORKING PAPER #2021/17

First version: December 15, 2021(current)

# Surnames and Social Rank: Long-term Traits of Social Mobility in Colombia and Chile

Juliana Jaramillo-Echeverri<sup>1</sup> | Andrés Álvarez<sup>2</sup> | Naim Bro<sup>3</sup>

<sup>1</sup>PhD Candidate, Department of Economic History, London School of Economics.

[j.jaramillo1@lse.ac.uk](mailto:j.jaramillo1@lse.ac.uk)

<sup>2</sup>Assistant Professor, Department of Economics, Universidad de Los Andes.

[ca.alvarez967@uniandes.edu.co](mailto:ca.alvarez967@uniandes.edu.co)

<sup>3</sup>Post-doctoral Fellow, Instituto Milenio Fundamentos de los Datos (IMFD - Chile). [naim.bro@imfd.cl](mailto:naim.bro@imfd.cl)

In the last two years, Colombia and Chile have witnessed strong social protests, characterized by slogans against inequality and the lack of social mobility. In this study we propose a comparative study on social mobility and the persistence of structural social inequalities in both countries. We collect evidence on the level of social immobility and test if it is rooted in historical forms of social segregation in both countries. We base our analysis in surname based methods. We conclude that there are clear indications of a significant persistence of upward immobility of the groups that were originally segregated during the colonial period: Afro-descendants (Colombia) and indigenous people (in both). Furthermore, we find that the downward social immobility of the elites shows an important persistence in both countries. However, in Chile the colonial elites (encomenderos and landowners) present greater persistence in their privileged status, while in Colombia those early elites seem to have converged more quickly to the mean. In both countries, there is a clear persistence of the elites of the second half of the 19<sup>th</sup> century in today's highest position of the social ladder.

## KEYWORDS

Social Status, Social Mobility, Historical Elites, Social Segregation, Surnames, Chile, Colombia

---

Small sections of text that are less than two paragraphs may be quoted without explicit permission as long as this document is acknowledged. Findings, interpretations and conclusions expressed in this publication are the sole responsibility of its author(s) and cannot be, in any way, attributed to CAF, its Executive Directors or the countries they represent. CAF does not guarantee the accuracy of the data included in this publication and is not, in any way, responsible for any consequences resulting from its use.

CAF - DOCUMENTO DE TRABAJO #2021/17

Primera versión: Diciembre 15, 2021 (actual)

# Apellidos y rango social: Rasgos de movilidad social a largo plazo en Colombia y Chile

Juliana Jaramillo-Echeverri<sup>1</sup> | Andrés Álvarez<sup>2</sup> | Naim Bro<sup>3</sup>

<sup>1</sup>Estudiante de Doctorado,  
Departamento de Historia  
Económica, London School of  
Economics.

[j.jaramillo1@lse.ac.uk](mailto:j.jaramillo1@lse.ac.uk)

<sup>2</sup>Profesor Asistente, Departamento  
de Economía, Universidad de Los  
Andes.

[ca.alvarez967@uniandes.edu.co](mailto:ca.alvarez967@uniandes.edu.co)

<sup>3</sup>Becario Postdoctoral, Instituto  
Milenio Fundamentos de los Datos  
(IMFD - Chile). [naim.bro@imfd.cl](mailto:naim.bro@imfd.cl)

En los últimos dos años, Colombia y Chile han sido testigos de fuertes protestas sociales, caracterizadas por consignas contra la desigualdad y la falta de movilidad social. En este trabajo proponemos un estudio comparativo sobre la movilidad social y la persistencia de las desigualdades sociales estructurales en ambos países. Recogemos evidencia sobre el nivel de inmovilidad social y comprobamos si está arraigado en formas históricas de segregación social en ambos países. Basamos nuestro análisis en métodos basados en los apellidos. Concluimos que existen claros indicios de una importante persistencia de la inmovilidad ascendente de los grupos que fueron originalmente segregados durante el período colonial: Afro-descendientes (Colombia) e indígenas (en ambos). Además, encontramos que la inmovilidad social descendente de las élites muestra una importante persistencia en ambos países. Sin embargo, en Chile las élites coloniales (encomenderos y terratenientes) presentan una mayor persistencia en su estatus privilegiado, mientras que en Colombia esas primeras élites parecen haber convergido más rápidamente a la media. En ambos países se observa una clara persistencia de las élites de la segunda mitad del siglo XIX en la actual posición más alta de la escala social.

## KEYWORDS

Estatus Social, Movilidad Social, Elites Históricas, Segregación Social, Apellidos, Chile, Colombia

---

Pequeñas secciones del texto, menores a dos párrafos, pueden ser citadas sin autorización explícita siempre que se cite el presente documento. Los resultados, interpretaciones y conclusiones expresados en esta publicación son de exclusiva responsabilidad de su(s) autor(es), y de ninguna manera pueden ser atribuidos a CAF, a los miembros de su Directorio Ejecutivo o a los países que ellos representan. CAF no garantiza la exactitud de los datos incluidos en esta publicación y no se hace responsable en ningún aspecto de las consecuencias que resulten de su utilización.

## 1 | INTRODUCTION

The most recent OECD report on social mobility places Latin America among the most unequal and least mobile regions in the world. Scholars have suggested that this pattern has roots in the region's deep past (Prados De La Escosura, 2005; Bértola and Ocampo, 2012; Bértola and Williamson, 2017; Williamson, 2010, 2015) but the evidence is not conclusive. Limited data, the lack of longitudinal sources and the complexities of defining a stable measure of social status are some obstacles to tackle this issue.

During the first decade of the 21<sup>st</sup> Century inequality decreased and social mobility increased in general in Latin America, including Colombia and Chile<sup>1</sup>. Even though both countries faced intense waves of popular demonstrations against economic inequality during 2019. Those social expressions reflect a common concern regarding social exclusion for some groups and the persistence of rigid social structures contributing to perpetuating the traditional elites. In this paper we would like to contribute to a better understanding of this perception, exploring empirical evidence about long-term existence of structural segregation against indigenous and black population and the persistence of historical elites.

In the past decade, researchers have used indirect methods to circumvent these problems, with promising results. New methods take advantage of surnames, an abundant form of data that links past and present generations over long stretches of time.<sup>2</sup> Surnames have been used to identify the linguistic (Novotný and Cheshire, 2012), ethnic (Mateos et al., 2011), genetic (King and Jobling, 2009), and socioeconomic (Bro and Mendoza, 2021) composition of populations. In intergenerational social mobility research, Clark (2014) uses rare surnames to estimate the social standing of families over the long run. He finds that it changes little from generation to generation and varies little across different countries. Güell et al. (2018) use surnames to study regional patterns in social mobility in Italy and find considerable regional variation across provinces.

There is evidence that a substantial segment of the Chilean upper-class today has been elite since the colonial period. For instance, Bro and Mendoza (2021) show that the ten most elite surnames in Santiago today are Errázuriz, Schmidt, García-Huidobro, Undurraga, Tagle, Vial, Larraín, Irarrázaval, Izquierdo, and Aldunate. Most of these surnames represent what Chilean historians call the "Castillian-Basque aristocracy". For instance, Irarrázaval was the first Chilean family to buy a nobility title, in 1684. García-Huidobro, Larraín, and Aldunate were also part of the colonial nobility. The Errázuriz produced three presidents in the 19th century, and the Vial produced two.

Clark (2014) uses surnames to estimate the rate of intergenerational mobility in Chile. He finds that the status correlation between two consecutive generations is about 0.8, and that many elite families in the past remain in the upper echelons of society today. They project that it will take another couple of centuries before the colonial elite regresses to the Chilean average. The proposed project updates the study of Clark and Díaz Vidal with novel data.

We use three surname-based methods for measuring social mobility. The "informational content of surnames" (ICS) approach, introduced by Güell et al. (2018), measures how much of the socioeconomic variation is captured by surnames. A second method, introduced by Clark et al. (2015), measures the correlation of social status between two generations holding

<sup>1</sup>According to a seminal research on social mobility in Latin America, Behrman et al. (2001) and the related paper published in Angulo et al. (2012), a comparison on intergenerational mobility in Chile, Colombia and Mexico shows a positive and strong correlation between the schooling of parents and children. Ferreira et al. (2013) describe contemporary patterns of equality of opportunity, inequality and social mobility for middle-class Latin American households. They find that Latin America experienced a considerable increase in social mobility during the first decade of the 21st century, and that the upward mobility levels of the middle class increased.

<sup>2</sup>Clark et al. (2015) and Santavirta and Stuhler (2020) survey this growing literature.

the same rare surnames. The third method, introduced by [Álvarez and Jaramillo-Echeverri \(2021\)](#), identifies rare surnames that were associated with the elites or the underclasses in the past, and measures the extent to which contemporary surname-holders are present in elite or underclass spaces.

As for the ICS, we find that surnames explain around 10% of the socioeconomic variation in both countries, a higher value than what [Güell et al. \(2018\)](#) found for Spain. As for the other two methods, we find that both countries have a high degree of elite and underclass persistence. We find that the elites in Chile have lower levels of intergenerational mobility than Colombia, although the indigenous groups appear to be more mobile in Chile. We update the literature by looking at elite and underclass groups in the long term.

[Torche \(2005\)](#) shows that Chile's level of social mobility is higher than expected given the country's high inequality. However, mobility occurs only in the middle of the income distribution: Chileans can make progress, but not enter the elites, which speaks of a mobile society but a strong elite. Indeed, Torche finds that Chile is among the most unequal countries in Latin America, but when the top decile is removed from the countries in the sample, it becomes one of the most equal.

In 2014, a comprehensive study on social mobility and inequality [Montenegro and Meléndez \(2014\)](#) showed that Colombia had a higher level of inequality and one of the lowest levels of intergenerational social mobility in Latin America. Using comparable longitudinal datasets for Colombia, Chile and Mexico [Angulo et al. \(2012\)](#) (p: 55) showed Colombia had the lowest social mobility among the three countries. Even though the three increased their social mobility during the first decade of the 21st century, Colombia is far from the levels of social intergenerational mobility of Mexico and Chile. In the three countries, the increase in intergenerational mobility is related with an increase in educational access. However, the effect of the fast enlargement of access to basic educational levels in Colombia has had a lower impact on accelerating social mobility.

[Angulo et al. \(2012\)](#) also show that the Colombian middle classes are more likely to descend than to ascend the social ladder. While the probability of rising from the lower spheres to the middle class is high, the likelihood of high-earning groups to downgrade is low. Like in Chile, social mobility occurs among the poor and the middle class, but less so the rich.

[Cárdenas Campo et al. \(2012\)](#) explore social discrimination and life quality segregation racial social groups in Colombia. They show substantial gaps separating native- and afro-Colombians versus the dominant mestizo population. Furthermore, they show that discrimination and social stigma reinforce the underrepresentation of indigenous and afrocolombians in politics and higher education.

[Álvarez and Jaramillo-Echeverri \(2021\)](#) assess the relative representation of historical groups in the Colombian educational system. Preliminary findings show that afro-descendant and indigenous surnames are underrepresented, or even fully absent, from high-quality high schools and elite higher education, suggesting a long term persistence of exclusion and low social mobility. Surnames associated with the 18<sup>th</sup> and 19<sup>th</sup> century elites, are overrepresented. Elite surnames in the 19<sup>th</sup> appear to be elite surnames today, when looking at access to schooling, while this is not the case of the elite surnames of the Colonial time.

All in all, this past decade has produced substantial progress in revealing patterns of intergenerational social mobility in Latin America. We build on this research program, and extend its time frame to assess how groups rise or fall over the centuries. This project tackles the long-term persistence of social status using surnames as the primary source of analysis.

Surnames can contain valuable information about social differences. On the one hand they are social markers, both about high and low status. But, on the other hand, they

also allow multiple generations to be linked, even very distant in time. Finally, the very way in which surnames are distributed within a population may reveal the persistence of hierarchical forms of organization and structural inequalities that do not easily appear in the standard analysis that uses contemporary information to estimate social mobility.

For the above mentioned reasons, this paper seeks to approach the problem posed in successive stages, aiming at extracting the most of the information surnames contain. Initially we evaluate the informative potential that surnames contain. We analyze whether they reveal patterns of high or low social mobility and then be able to constitute a path for unveiling the presence of rigid social structures. We then use historical and ethnographical sources to classify surnames according to historical social status. This allows us to evaluate to what extent to explore how far the long shadow of the past is still projected today through the family ties of elites and marginalized people in the history of both countries.

## 2 | WHAT'S IN A SURNAME? CONTEMPORARY PATTERNS OF SOCIAL STATUS AND THE DISTRIBUTION OF SURNAMES

Surnames are an abundant form of data that links past and present generations over long stretches of time.<sup>3</sup> Surnames have been used to identify the linguistic (Novotný and Cheshire, 2012), ethnic (Mateos et al., 2011), genetic (King and Jobling, 2009), and socioeconomic (Bro and Mendoza, 2021) composition of populations. Surnames have also been extensively used in intergenerational social mobility research, Clark (2014) uses rare surnames to estimate the social standing of families over the long run. He finds that it changes little from generation to generation and varies little across different countries. Güell et al. (2018) use surnames to study regional patterns in social mobility in Italy and find considerable regional variation across provinces.

In most societies surnames link parents and children in a probabilistic sense: the rarer the surname, the higher the probability that the surname is linking family members. In Western countries the distribution of surnames is skewed to the right. This characteristic presents an advantage for the method because the number of rare surnames is large, therefore allowing us to link several groups of families.

Figure 1 depicts the distributions of surnames for a comprehensive dataset. We use the most recent population census in Colombia (2018) and the civil identity register of Chile (1885-2020) as our main source for a comprehensive list of surnames. We are able to collect two surnames for almost every individual in the registries, as Colombia and Chile officially adopted the Spaniard surname rule: usually first surname from the father and second name from the first surname of the mother.

These datasets show that the surname distribution in both countries follows a distinctive pattern. There are many surnames with a low frequency of holders, while few common surnames with a very high number of holders. The long tails of the distribution tell us that rare surnames can contain an important amount of information about people connected through long-term familiar links.

Although Chile has fewer distinct names than Colombia, which is in line with the smaller population size (50 millions and 19 millions, respectively), the percentage of people holding a very common surname is similar in both countries, as well as the share of the population holding rare surnames (See table 1).

---

<sup>3</sup>Clark et al. (2015) and Santavirta and Stuhler (2020) survey this growing literature.

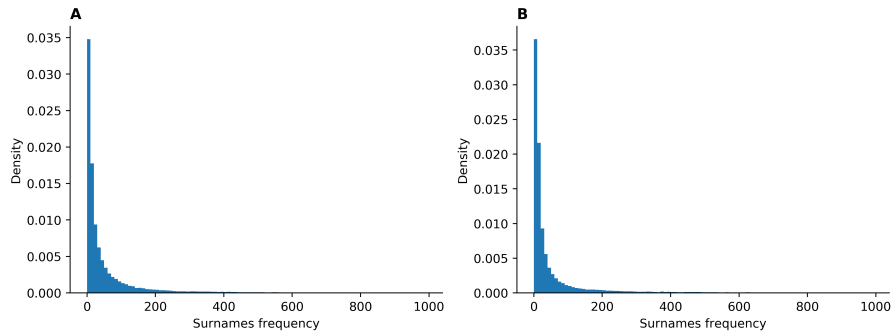


FIGURE 1 Surnames frequency distribution.

Notes: Figure A shows the distribution for Chile and Figure B shows the distribution for Colombia. Source: For Chile: Civil identity register, 1885-2020, for Colombia: Population and Housing Census, 2018.

TABLE 1 Surnames distribution: summary statistics

|                                     | Chile  | Colombia |
|-------------------------------------|--------|----------|
| Number of distinct names            | 11,527 | 56,329   |
| Mean frequency per name             | 86.42  | 316.89   |
| % of people holding top-20 surnames | 18     | 21       |
| % of people holding top-50 surnames | 32     | 33       |
| % of people holding rare surnames   | 24     | 23       |

Notes: Rare surnames are defined as those held by more than 3 holders to avoid misspellings and less than 2% of the most common surname in the sample.

Source: For Chile: Civil identity register, 1885-2020, for Colombia: Population and Housing Census, 2018.

Recent research has shown that socioeconomic status varies systematically across surnames: in particular suggesting that rare surnames have frequently higher status and that using rare surnames could lead to estimations in a selective and non-representative population (Santavirta and Stuhler, 2020). In this section we evaluate the socioeconomic gradient of the surnames in both countries and later we explore a technique of estimation of social mobility using the whole distribution of surnames, exploring patterns of persistence of social status beyond rarity.

## 2.1 | Surnames and social status.

A first approximation to the relation between contemporary social status and surnames is to correlate rarity (number of holders) with the average level of a comprehensive metric of socioeconomic position at the level of each surname. Figure 2 shows this relation using the average score in index of socioeconomic performance. In Chile, the rarity of surnames correlates positively with social status. This is not the case in Colombia, where rare family names are concentrated in the middle.

To perform our analysis we collected data from multiple sources, a detailed list of the primary sources is presented in Appendix (A.1). In order to define the rarity of surnames,

a source that is as representative as possible of the totality of surnames present in the population is needed today. In the case of Colombia, we have information on the frequency of each of the surnames registered in the 2018 population census. For Chile, we use the electoral register that contains the entire adult population of the country registered, obligatorily, to exercise the right to vote. We also use information from the civil identity register of Chile (1885-2020) to capture the evolution of the distribution of surnames since the late 19<sup>th</sup> century.

To evaluate socioeconomic performance of contemporary populations we impute income and social status at the level of a block in Santiago (Chile), for each person in the electoral register. In Colombia, we do not have information on monetary income but we have a source of information representative at the national level: the national system of social information (SISBEN - Sistema de información social para asignación de beneficios económicos). This source allows us to obtain a socioeconomic score and to identify the distribution of socioeconomic performance for at least 75% of the households.

Finally, to define the belonging of each surname to a group with historical marker we collect information from primary historical sources, as the Genealogies of the Spaniards inhabitants from the 17<sup>th</sup> century in both countries, agricultural census for Chile and a list of slaveowners for Colombia from the 19<sup>th</sup> century, among other sources. We describe each group in section 3.1 (see Table 3 below).

The difference between the two countries is likely to come from Chile's more sizable past migrant waves, the descendants of which hold rare last names and belong in the upper echelons of society. The first of such migrant waves was that of the Basques in the late colonial period (Thayer Ojeda 1904), which imported last names such as Larraín, Errázuriz, and Eyzaguirre - relatively unusual names associated with the upper classes. English (Edwards, Cox, etc) and German (Kunstmann, Niklitschek, etc) migrants arrived in the nineteenth century, and Palestianians (Jadue, Said, etc), Jews (Ergas, Cohen, etc), and Italians (Carozzi, Lucchetti, etc) in the twentieth century. Altogether, Chile's past migration is much smaller than that of Argentina, Uruguay, and Brazil, but their concentration in the upper strata produces the observed association between surname rarity and socioeconomic status. The positive association between last name rarity and social status is not uncommon. Collado et al. (2008) show that the Spanish elites tend to have rare last names. They trace this pattern to the nineteenth-century elite's practice of double-barreling their last names, and - like in Chile - to the overrepresentation of some migrant communities in the upper class.

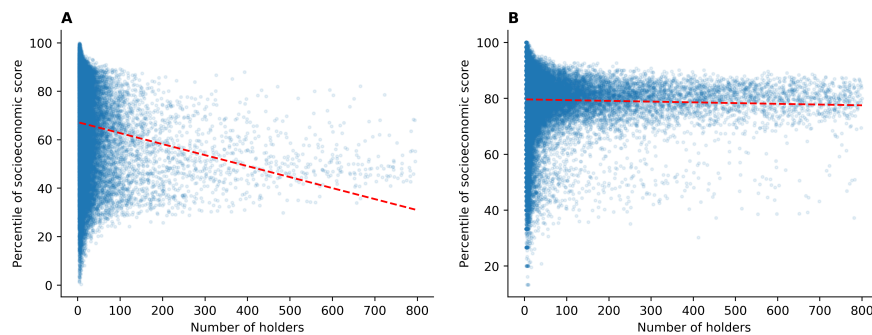


FIGURE 2 Socioeconomic status of surnames by number of holders.

Notes: Figure A shows the results for Chile and Figure B shows the results for Colombia.

Source: For Chile: Electoral register, 2020, for Colombia: Population and Housing Census, 2018.

In contrast with Chile, Colombia has not received significant migratory waves since the arrival of the Spanish conquerors and the slaves from Africa. For this reason, in Colombia, typical rare last names are indigenous or Hispanic, and do not tend to be concentrated at the top of the socioeconomic distribution. Interestingly, there are several surnames at the top of the distribution that are fairly common surnames, such as Restrepo, Quiroga and Betancur.

## 2.2 | The informational content of surnames (ICS).

As mentioned above, rarity itself is probably a signal of status in highly segregated societies. However, focusing on the analysis of sufficiently rare surnames makes it difficult to identify the information that surnames, and therefore family membership, may contain throughout the entire social status distribution. This is why we adopt a second methodology to explore how surnames can unveil traits of social mobility, or immobility.

Following Güell et al. (2018) we perform a first approximation to a measure of social mobility using surnames exploring the joint distribution of surnames and a particular economic outcome. This method is known as the Informational Content of Surnames (ICS) or R-squared method. It is based on cross-sectional data (i.e. census). The intuition is that if inheritance is important in a society, then surnames may contain a considerable amount of information about the distribution of socioeconomic variables within the population.

The ICS is defined as the difference in the aggregate performance of two linear regressions. The first one measures the actual relationship between the true surnames of the individuals, using as the regressors the matrix of the whole list of surnames in the form of a dummy variable taking the value of 1 when the individual is a holder of the surname. We name  $L$  the matrix of dummies for the actual distribution. On the left side of the regression we use a particular socioeconomic outcome ( $y_{i,s}$ ) for each individual  $i$  holder of a surname  $s$  (equation 1). The second regression uses a randomised ("fake")  $F$  assignment of surnames instead of the original matrix  $L$ , and we run a similar regression (equation 2):

$$y_{i,s} = b'L + \text{residual} \quad (1)$$

$$y_{i,s} = f'F + \text{residual}. \quad (2)$$

The difference between the R-squared of the first and second regression is the ICS and it reflects how much information surnames contain because of family linkages:

$$\text{ICS} = (R_L^2 - R_F^2) * 100. \quad (3)$$

To estimate the ICS for each country, we use two comparable sources of information. For Chile we use the electoral register to identify individuals, including both surnames that we match with a measure of social status. For Colombia we use the anonymized register at the individual level, including both surnames and a multidimensional measure of socioeconomic level from the 2018 census. The randomization process was performed more than 15 times, the variations in the values of the estimated R square were statistically insignificant.<sup>4</sup>

<sup>4</sup>The seminal paper, Güell et al. (2018), only reports one randomization step, but they also mention the results are fully robust to different methods and several rounds of the randomized distribution of surnames.



The result of these estimations gives us a comparable indicator of the importance of inheritance in Chile and Colombia. In the context of our research the ICS method can provide a first look at the potentiality of surnames to capture information about social status.

Table 2 shows the values of the R-square and the ICS for each country. First, it is worth noticing the R-square of the “fake” regression is practically nil, while the regression using the actual distribution of surnames has a value of 10.2% for Chile and 9.9% for Colombia. These values are high when compared with similar estimations in the literature. For instance, in Güell et al. (2018), estimations for Catalonia of the ICS using data on educational attainment are around 2% to 3%. They calibrated a model to derive intergenerational correlation (inheritance parameter) from the ICS. They obtain a measure of intergenerational correlation of 0.6, in line with previous results of intergenerational mobility oscillating around 0.4 to 0.6. The results are also in line with what Santavirta, and Stuhler, found using data for Finish veterans, and those in Güell et al. (2018).

From the ICS results it is clear that in Chile and Colombia surnames are a good predictor of socioeconomic status. Furthermore, this methodology provides significant evidence of the way family ties, and thus inherited social status, are determinant of the socioeconomic performance of contemporary generations.

TABLE 2 Informational Content of Surnames

|            | Chile     | Colombia |
|------------|-----------|----------|
| $R_L^2$    | 0.1018    | 0.100179 |
| $R_F^2$    | 5.882e-06 | 0.000793 |
| <b>ICS</b> | 10.2%     | 9.9%     |

### 2.3 | Linking generations through surnames: exploring social mobility in contemporary Chile and Colombia.

Given that rare surnames increase the predictive power of surnames as the probability of linking parents to children increases, we focus only on rare surnames to estimate coefficients of intergenerational mobility using the Chilean electoral register and the SISBEN database for Colombia. To do so we divide our samples in 2 generations: one from 1930-1959 and the other one from 1960-1989 and we follow a conventional approach of social mobility by regressing the social status of generation 2 against generation 1 using groups of surnames as the unit of analysis, we estimate independent regressions for each group, following Clark (2014). In this case, the main assumption is that we are comparing the social status of pseudo parents to pseudo children.<sup>5</sup>

$$y_{t,s} = \alpha + \beta_S y_{t-1,s} \quad (4)$$

We are interested then in the value of  $\beta$ , which is the coefficient of intergenerational mobility between the 1930-1950 cohort and the 1960-1989 cohort. Given that for the Colombian data we only observe around 75% of the population, we limit the Chilean data and we focus

<sup>5</sup>Appendix A.1 shows the results when we change this threshold to see how the coefficient of correlation is sensitive to surname rarity. The results are robust to different definitions of rarity in both countries.

only on the bottom 75% of the sample to make the samples comparable. Figure 3 shows the results of the linear model estimation.

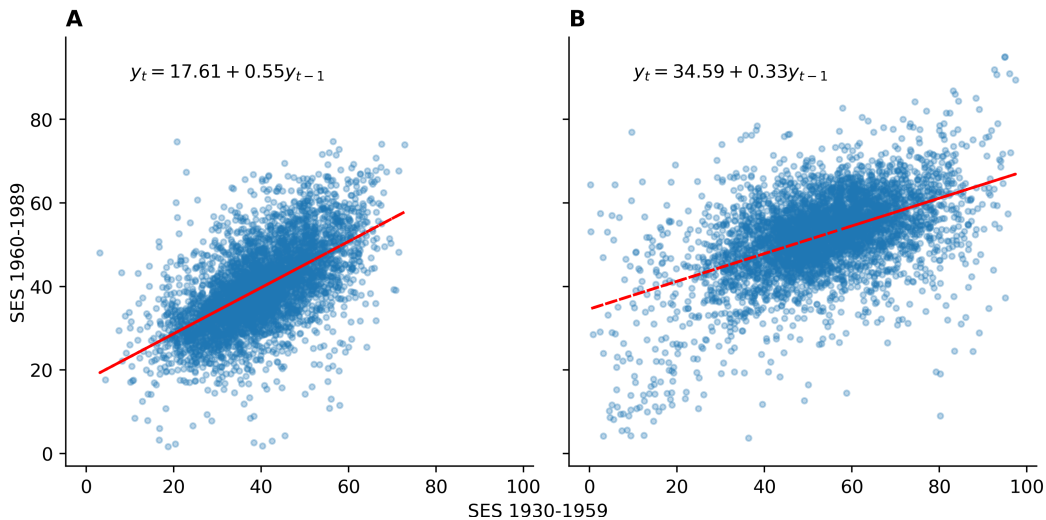


FIGURE 3 Coefficient of Intergenerational mobility.

*Notes:* The sample is based on rare surnames that have between 5 to 100 holders. Figure A shows the results for Chile and Figure B shows the results for Colombia.

*Source:* For Chile: Electoral register, 2020, for Colombia: SISBEN, 2013.

The results suggest that, when looking at the bottom 75% of the socioeconomic distribution, Chile has a lower intergenerational mobility than Colombia. The relative social status at the level of rare surnames show an important degree of stickiness in both countries with a  $\beta$  of 0.5 and 0.3 for Chile and Colombia, respectively. Excluding the highest 25% of the distribution of social status increases the degree of estimated social mobility in Chile, compared to the ICS estimation. This is worth noting as it is possibly showing a gradient of lesser social mobility among highest levels of socioeconomic status in Chile and given the results from the ICS estimation and the findings in section 3, this is possibly the case also in Colombia. In appendix B we show the equivalent of 3 using the whole distribution of income for Chile.

All the above results are obtained under in absence of any assumptions to group surnames. In this sense, hetherto we cannot infer anything from the historical point of view. In the following section we make use of ethnic, historical information and migration patterns to observe social status in the long run. We explore an additional method of analyzing the evolution of social status using groups of surnames instead of considering them as independent entities.

### 3 | THE STICKY SOCIAL LADDER: AN ANALYSIS OF LONG TERM PATTERNS OF MOBILITY AND PERSISTENCE.

In this section we use a historical taxonomy of the surnames to observe the long-term dimension of social status and to see to what extent social status has persisted in Colombia and Chile giving their low levels of social mobility. As we shall see, the conclusions about a higher level of social mobility in Colombia changes when grouping surnames using historical lists of elites and underclass groups from the late colonial to the early 20<sup>th</sup> century.

### 3.1 | Identifying historical elites and underclass groups along history.

As a way to disentangle which social groups have lower or higher social mobility we try to see if these patterns of social mobility can be traced back in the past, when the first Spaniards settled in Latin America around the 16<sup>th</sup> century. The Colony established the institution of the *encomienda* in which the notable Spaniards living in the colony were allowed to collect tribute such as gold and labour in exchange of protection of the Crown. Soon after, the indigenous were also forced to work in gold mines, haciendas and craft shops.

Colonial life differed substantially in Colombia and Chile. In Colombia, during the 16<sup>th</sup> century the striking demographic collapse of the indigenous population resulted in the importation of slaves from Africa to work in the gold mines. By the 18<sup>th</sup> century the economy of Nueva Granada was based on mining, agriculture and commerce, and in turn, these activities were established around slave labour. After Independence, the country started an institutional transformation. The importance of the slave economy started decreasing as new industries emerged, while mining lost importance in comparison with new agricultural products. Finally, by 1851 slavery was abolished but it was only until the late 19<sup>th</sup> that Colombia saw the beginning of the modern economic era with the emergence of early forms of industrial development and the banking sector in 1870.

In contrast, colonial Chile did not experience substantial slave trade, due to the country's lack of purchasing power compared to other colonies. The initial Spanish settlements grew around river gold mining, which decayed shortly after due to the exhaustion of the metal and the demographic collapse of the native workforce. The economic base of the colony became the export of tallow to Lima, and subsequently wheat (Góngora, 1970). Mahoney (2010) classifies Chile as one of the most marginal Spanish possessions throughout the colonial period, and few Spaniards arrived in the country. However, in the 19<sup>th</sup> century, a small cohort of Basque immigrants settled in the country, and eventually became the foundation of the country's traditional upper class (Villalobos 2006).

Chile and Colombia continue diverging in the second half of the 19<sup>th</sup> century. While Chile received a moderate flow of immigrants from Europe, very few arrived in Colombia.

The next section measures the contemporary social status of persons holding surnames from the historical groups listed in Table 3, and estimates how persistent status has been for the different historical groups.

TABLE 3 Historical groups

| Period             | Chile        | Colombia        |
|--------------------|--------------|-----------------|
| 16th century       | Mapuches     | Indigenous      |
| 17th century       |              | Afro-Colombians |
| 17th century       | Encomenderos | Encomenderos    |
| 19th century       | Landowners   | Slaveowners     |
| Early 20th century | Italians     | Jockey Club     |
| Mid 20th century   | Jews         |                 |

### 3.2 | Estimating social mobility using groups of surnames.

Similar to the traditional measures of intergenerational mobility presented in the previous section, following Clark (2014), we calculate the correlation of intergenerational mobility for each of our historical groups. Again, we focus on rare surnames that match with the

historical groups and divide the samples into two generations to observe the relationship between pseudo parents and pseudo offsprings. We calculate the correlation between the two cohorts at the level of a group of surnames,  $S$ , instead of the individual surname. We are interested in the coefficient  $\delta$ .

$$y_{t,S} = \delta y_{t-1,S} \quad (5)$$

The results are shown in table 4.

TABLE 4 Intergenerational mobility by historical groups

| Country  | Group        | Num observations | Correlation |
|----------|--------------|------------------|-------------|
| Chile    | Mapuche      | 835              | 0.425       |
|          | Encomenderos | 28               | 0.738       |
|          | Landowners   | 34               | 0.567       |
|          | Italians     | 122              | 0.506       |
|          | Jews         | 61               | 0.492       |
| Colombia | Indigenous   | 76               | 0.589       |
|          | Afro         | 41               | 0.624       |
|          | Encomenderos | 25               | 0.354       |
|          | Slaveowners  | 311              | 0.468       |
|          | Jockey Club  | 55               | 0.366       |

The results show that, in Chile, elite groups from the colonial period and the 19<sup>th</sup> century have higher levels of persistence in comparison to Colombia. The results for Chile are quantitatively similar to those found by [Clark \(2014\)](#). The results for Colombia on the other side show that underclass groups are the less mobile groups, while the Colonial Spaniard officers appear as the most mobile group of all the groups studied. Surprisingly, slave owners present low levels of persistence, which differs from what [Ager et al. \(2019\)](#) found for the USA case. In both countries, ethnic surnames like Afro Colombians and Jews have low mobile rates.<sup>6</sup> The historical upper class in Chile seems to be more mobile, while the indigenous and Afrocolombians seem to have very sticky floors.

#### 4 | EXPLORING LONG-TERM PATTERNS OF SOCIAL STATUS.

To go further exploring the influence of historical groups we search for a more robust estimation using the relative representation of each surname in a particular social category as a way to approach a comparison of social status of each group relative to the most common surnames. This approach allows us to define a benchmark to measure the historical performance of a group. A very common surname (i. e. Rodriguez in Colombia and González in Chile) has very similar frequency of holders in the dataset of reference (population census for Colombia and electoral census for Chile) as in any other dataset of surnames. This means that for a common surname when considering for example the frequency of holders in a

<sup>6</sup>The lower mobility of ethnic groups has been discussed also in [Chetty et al. \(2014\)](#)

very exclusive, high status, school, is not over or underrepresented relative to its frequency of holders in the whole population. This means a common surname represents the average performance in a social situation. In simple words, the probability to find a Colombian holding the surname Rodriguez in the elite is the same as to find it in a lower status situation.

We look at social status using the relative representation of surnames in different contemporary social categories as a way to measure persistence. In other words, we look at the share of a surname in a contemporary social situation relative to the share of that surname within the total population to determine whether it is over or under-represented in that social situation. In a completely mobile society, elite groups from the past should not appear overrepresented in high status categories today while underclass groups should not appear underrepresented in these high status categories. If a society has a strong persistence of social status, then we will observe the opposite. To construct the measure of relative representation we use the following equation.

$$RR_j = \frac{\text{Share of } j \text{ in contemporary socio economic outcome}}{\text{Share of } j \text{ in population}} \tag{6}$$

We standardize relative representation using a Z standardization to make our results comparable between the countries. We estimate the delta coefficients that capture how each historical social group explains the over representation of a surname in a particular social situation. The latent group of reference is the 20 most common surnames. Coefficients are thus expressed in terms of standard deviations to the average. We estimate the following equation, using a set of dummies variables taking the value of 1 when a surname belongs to a group  $g$ .  $D_g$  is the matrix of those dummy variables, from which we exclude a group as the latent standard of comparison:

$$Z[RR_j] = \alpha + \delta_g D_g + \epsilon_j. \tag{7}$$

The results are shown in Figure 4 and in the appendix A.2 and A.3.

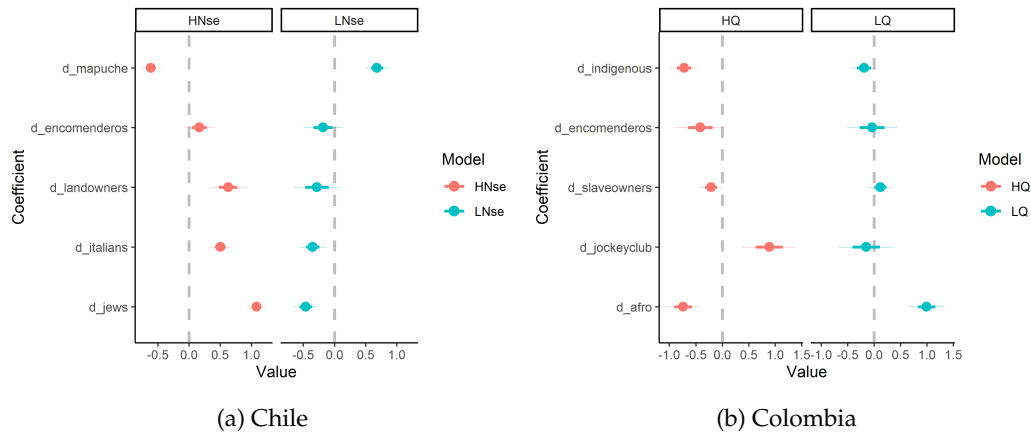


FIGURE 4 Persistence of social status by historical group.

Notes: The figure plots the coefficients from equation 7. The results can be found in Appendix

Source: For Chile: Civil identity register, 1885-2020, for Colombia: SABER 11

The results confirm in most cases what we found in the previous section looking at the coefficient of intergenerational mobility by historical groups. Perhaps not surprisingly but still concerning, indigenous and Mapuches are still at the bottom of the social-ladder, excluded from elite spaces and over-represented at low status categories. Thus, as indigenous groups are among the most immobile groups in both countries they have been trapped at the bottom for centuries. On the opposite side of the social ladder, elite groups of the 19th and early 20th century in both countries (landowners, Italians and members of the Jockey Club) appear as over-represented in elite spaces and under-represented in low status categories confirming the persistence of their social status.

But not all elite groups show immobility, as suggested previously. The Spaniard colonial officers in both countries appear as if they were slowly losing their elite status. This group shows the lowest persistence of all elite groups in both countries when measured by the relative representation in high status categories. In Chile, they behave almost as the commons, only slightly over-represented in high income groups. In Colombia, this regression to the mean is more evident as they appear slightly under-represented in high quality schooling. Finally, when comparing the persistence of Jews and AfroColombians we observe that, as with their similar levels of social mobility, both groups show high persistence of social status. As a result, they occupy very different places in the social ladder.

## 5 | CONCLUDING REMARKS

This paper makes three contributions to the literature.

It is the first to apply the Informational Content of Surnames method to Chile and Colombia. We find that surnames explain around 10% of the socioeconomic variance in both countries, a higher value than what Güell et al. (2018) find for Spain. This difference suggests lower mobility levels in the South American countries.

The paper is the first to apply the method proposed by Clark to Colombia, and it updates the data that Clark used for Chile. For Chile, it finds a slightly higher intergenerational correlation than the one that Clark found. For Colombia, it finds surprisingly high levels of elasticity, but we think this probably due to the nature of the Colombian data.

Using the method of relative representation proposed by Álvarez and Jaramillo-Echeverri (2021), we find that elites are more persistence in Chile than in Colombia. This result is surprising, because prior research has found lower mobility rates for Colombia. Further research will explore potential issues with the comparability of the data in both countries.

A key objective of the our project is to develop analytical tools to compare Latin American countries systematically. While the study begins with Colombia and Chile, a parallel goal is to build a network of researchers from other countries, and lay the foundations for an all-Latin American social mobility project. The results will reveal the determinants of social mobility in the long run, and as such they will inform social policy.

## ACKNOWLEDGEMENTS

We want to thank the Development Bank of Latin America (CAF) for its initiative to promote research projects on intergenerational mobility in Latin America. This project has received financial support and the possibility to participate in the *RED 2022: Intergenerational mobility in Latin America* held online on September 29<sup>th</sup> 2021. We are grateful to all the participants in the workshop, especially to Santiago Pérez for his comments on a very preliminary version of this work. We also acknowledge the fantastic help from our research assistants Josué Tapia and Andrés Felipe Rodríguez. All remaining errors, problems or omissions are our sole responsibility.

## REFERENCES

- Ager, P., Boustan, L. P. and Eriksson, K. (2019) The intergenerational effects of a large wealth shock: White southerners after the civil war. *Tech. rep.*, National Bureau of Economic Research.
- Angulo, R., Gaviria, A., Páez, G. N., Azevedo, J. P. et al. (2012) Movilidad social en Colombia. *Tech. rep.*, Universidad de los Andes-CEDE.
- Behrman, J. R., Gaviria, A., Székely, M., Birdsall, N. and Galiani, S. (2001) Intergenerational mobility in Latin America [with comments]. *Economía*, **2**, 1–44.
- Bértola, L. and Ocampo, J. A. (2012) *The economic development of Latin America since independence*. OUP Oxford.
- Bértola, L. and Williamson, J. (2017) *Has Latin American Inequality Changed Direction?: Looking Over the Long Run*. Springer Nature.
- Bro, N. and Mendoza, M. (2021) Surname affinity in Santiago, Chile: A network-based approach that uncovers urban segregation. *PLoS one*, **16**, e0244372.
- Cárdenas Campo, J. C., Ñopo, H., Castañeda Núñez, J. L. et al. (2012) *Equidad en la diferencia: políticas para la movilidad social de grupos de identidad* Misión de Movilidad Social y Equidad. Universidad de los Andes, Facultad de Economía, CEDE.
- 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**, 1553–1623.
- Clark, G. (2014) *The son also rises*. Princeton University Press.
- Clark, G., Cummins, N., Hao, Y. and Vidal, D. D. (2015) Surnames: A new source for the history of social mobility. *Explorations in Economic History*, **55**, 3–24.
- Collado, M. D., Ortín, I. O. and Romeu, A. (2008) Surnames and social status in Spain. *Investigaciones Económicas*, **32**, 259–287.
- Ferreira, F. G., Messina, J., Rigolini, J., López-Calva, L. F. and Vakis, R. (2013) *La movilidad económica y el crecimiento de la clase media en América Latina*. Banco Mundial.
- Góngora, M. (1970) *Encomenderos y estancieros: estudios acerca de la constitución social aristocrática de Chile después de la conquista, 1580-1660*. Universidad de Chile, Sede de Valparaíso, Area de Humanidades, Departamento . . .
- Güell, M., Pellizzari, M., Pica, G. and Rodríguez Mora, J. V. (2018) Correlating social mobility and economic outcomes. *The Economic Journal*, **128**, F353–F403.
- King, T. E. and Jobling, M. A. (2009) What's in a name? y chromosomes, surnames and the genetic genealogy revolution. *Trends in Genetics*, **25**, 351–360.
- Mahoney, J. (2010) *Colonialism and postcolonial development: Spanish America in comparative perspective*. Cambridge University Press.

- Mateos, P., Longley, P. A. and O'Sullivan, D. (2011) Ethnicity and population structure in personal naming networks. *PloS one*, **6**, e22943.
- Montenegro, A. and Meléndez, M. (2014) *Equidad y movilidad social: Diagnósticos y propuestas para la transformación de la sociedad colombiana*. Ediciones Uniandes-Universidad de los Andes.
- Novotný, J. and Cheshire, J. A. (2012) The surname space of the czech republic: examining population structure by network analysis of spatial co-occurrence of surnames. *PloS one*, **7**, e48568.
- Prados De La Escosura, L. (2005) Growth, inequality, and poverty in latin america: historical evidence, controlled conjectures. *UC3M Working Papers, Economic History and Institutions*.
- Santavirta, T. and Stuhler, J. (2020) Name-based estimators of intergenerational mobility. *Tech. rep.*
- Torche, F. (2005) Unequal but fluid: social mobility in chile in comparative perspective. *American Sociological Review*, **70**, 422–450.
- Williamson, J. G. (2010) Five centuries of latin american income inequality. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History*, **28**, 227–252.
- (2015) Latin american inequality: Colonial origins, commodity booms or a missed twentieth-century leveling? *Journal of Human Development and Capabilities*, **16**, 324–341.
- Álvarez, A. and Jaramillo-Echeverri, J. (2021) Surnames, status and schools: a long-term view of social status in colombia. *Tech. rep.*



## A | APPENDIX A

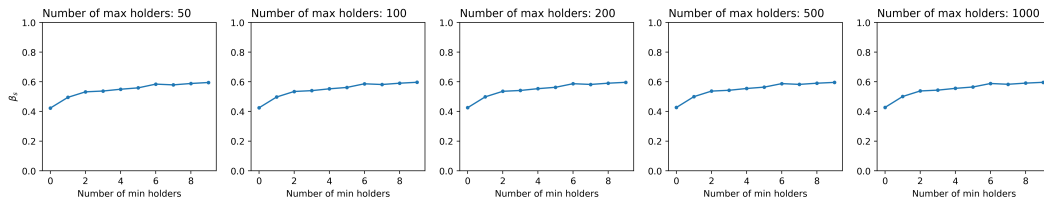
### A.1 | Primary sources

TABLE A.1 List of primary sources collected

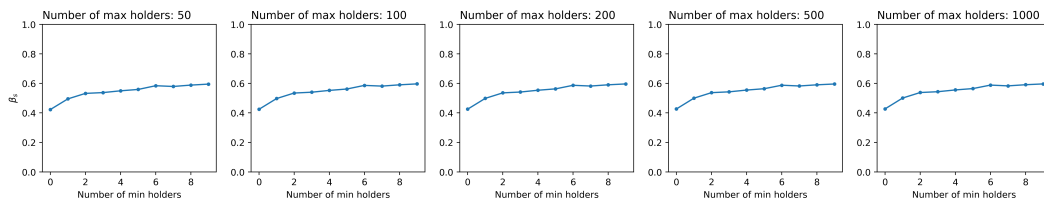
| Country   | Database  |
|---|---|
| Chile   | Electoral register 2020   |
|   | Civil identity register, 1885-2020  |
|   | Agricultural census 1874, 1908, 2007  |
|   | Roster of private companies' director appointments                              |
|   | List of Mapuche surnames  |
|   | List of Jewish surnames   |
| Colombia  | Population Census 2018.   |
|   | Genealogies of the "Nuevo Reino de Granada" 17th century                        |
|   | List of Afro-Colombian surnames   |
|   | List of Pre-Columbian peoples' surnames   |
|   | List of slaveowners by 1850   |
|   | List of bank shareholders from the free-banking period, 1867-1910               |
|   | List of the Jockey Club members, 1874-1920s                                     |
| SABER 11 - Register of the students attending higher education in Colombia. |   |
|   | SISBEN, Sistema de información social para asignación de beneficios económicos. |

### A.2 | Testing different threshold rarity

The following figure depicts the estimated coefficients of intergenerational mobility using different thresholds to define the minimum number of holders for a surname to be considered as rare.



(a) Chile



(b) Colombia

FIGURE A.1 Coefficient of Intergenerational mobility - Different rarity thresholds.

Source: For Chile: Civil identity register, 1885-2020, for Colombia: Population and Housing Census, 2018.

TABLE A.2 Chile

|                         | <i>Dependent variable:</i>           |                      |
|-------------------------|--------------------------------------|----------------------|
|                         | Relative representation (normalised) |                      |
|                         | High SES                             | Low SES              |
|                         | (1)                                  | (2)                  |
| Mapuche                 | −0.609***<br>(0.077)                 | 0.684***<br>(0.099)  |
| Encomenderos            | 0.165<br>(0.121)                     | −0.181<br>(0.155)    |
| Landowners              | 0.629***<br>(0.147)                  | −0.279<br>(0.190)    |
| Italians                | 0.505***<br>(0.084)                  | −0.346***<br>(0.109) |
| Jews                    | 1.084***<br>(0.079)                  | −0.458***<br>(0.101) |
| Constant                | 0.011<br>(0.076)                     | −0.244**<br>(0.098)  |
| Observations            | 3,494                                | 3,494                |
| Adjusted R <sup>2</sup> | 0.569                                | 0.286                |

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

TABLE A.3 Colombia

|                         | <i>Dependent variable:</i>           |                       |
|-------------------------|--------------------------------------|-----------------------|
|                         | Relative representation (normalised) |                       |
|                         | High Quality Schooling               | Low Quality Schooling |
|                         | (1)                                  | (2)                   |
| AfroColombians          | −0.738***<br>(0.166)                 | 0.996***<br>(0.168)   |
| Indigenous              | −0.720***<br>(0.134)                 | −0.187<br>(0.135)     |
| Encomendero             | −0.415*<br>(0.234)                   | −0.033<br>(0.237)     |
| Slaveowners             | −0.211*<br>(0.114)                   | 0.122<br>(0.115)      |
| JockeyClub              | 0.891***<br>(0.257)                  | −0.144<br>(0.260)     |
| Constant                | 0.298***<br>(0.099)                  | −0.108<br>(0.100)     |
| Observations            | 579                                  | 579                   |
| Adjusted R <sup>2</sup> | 0.101                                | 0.083                 |

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

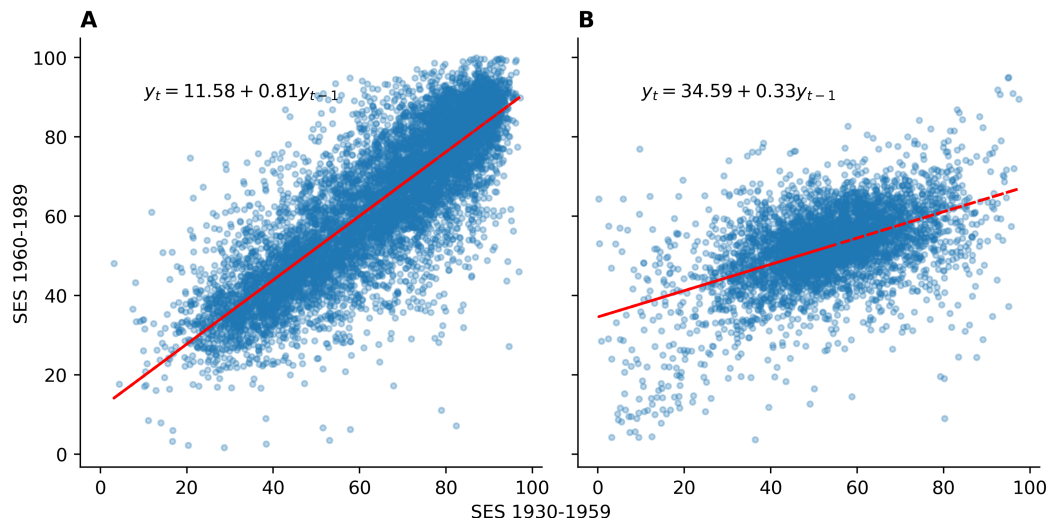
**B | APPENDIX B****B.1 | Estimations using the whole income distribution for Chile**

FIGURE B.1 Coefficient of Intergenerational mobility.

*Notes:* The sample is based on rare surnames that have between 5 to 100 holders. Figure A shows the results for Chile and Figure B shows the results for Colombia. This is the equivalent of 3 but it contains the whole income distribution for Chile

*Source:* For Chile: Electoral register, 2020, for Colombia: SISBEN, 2013.