Exploring Worldwide Relative Social Mobility: Insights from a Homogeneous Methodology

Student: Agustín Sanhueza[*] Guide: Javier Nuñez[†]

Co-Guide: Claudio Montenegro[‡]

June 17, 2024

School of Business and Economics, Universidad de Chile

Abstract

This study analyzes relative social mobility through intergenerational persistence using a standardized methodology of coresidents, focusing on both educational and income dimensions. Utilizing a unique dataset of previously unavailable survey data and employing a cross-sectional approach combined with seven distinct measures of intergenerational persistence, this research enhances our capacity to compare global social mobility. The findings indicate that high-income countries tend to exhibit greater levels of social mobility, while regions such as South Asia and Latin America and the Caribbean experience comparatively lower rates. The conducted regressions offer valuable empirical insights, establishing correlations between social mobility and macroeconomic indicators such as income inequality and government investment in education. These insights serve as a foundation for further investigations, enabling researchers to incorporate additional parameters, filters, and key variables, such as returns on schooling, into their analyses.

Keywords: Social Mobility, Intergenerational Persistence, Inequality

JEL Codes: D3, I2, J3

[*]asanhuezac@fen.uchile.cl

[†]jnunez@fen.uchile.cl

[‡]cmontenegro@worldbank.org

Contents

List of Figures

1 Introduction

Individuals possess an inherent aspiration for economic advancement throughout their lifetimes, seeking to enhance their socioeconomic status either absolutely, by improving their own circumstances, or relatively, by surpassing the average societal position. Examining the channels facilitating such social mobility is imperative, as it paves the way for the formulation of public policies geared towards fostering expedited, more inclusive progress and heightened opportunities. This issue assumes particular significance within economically disadvantaged nations, where upward mobility remains a great challenge. Relative intergenerational mobility, herein construed as the capacity of individuals to transition from one relative position within the education/income distibution to another across successive generations, keep our interest. A society bereft of relative intergenerational mobility manifests a recurrent pattern wherein identical individuals consistently occupy fixed strata within the distribution. The most extreme scenario illustrates itself in a caste based society, where regardless of an individual's innate talent or potential, they remain confined to their birthplace on the social hierarchy, thus precluding vertical mobility. This principle operate equivalently in the opposite direction. In light of these considerations, we can distinguish the reasons for estimating relative intergenerational mobility into two categories:

Normative reasons. Disparities in economic and social outcomes arise from factors within individuals control, such as their effort, responsibility, and choices. However, there exist external factors beyond individuals control that constrain their opportunities to pursue their life goals, such as residing in a safe neighborhood, having a robust social network, access to quality education and healthcare, among others. These factors are often transmitted through parental education and income, as well as influenced by government policies. There is a consensus in favor of public policies creating the necessary conditions for fostering *equality of opportunities*. This concept refers to the capacity individuals have to pursue their life aspirations. In a society characterized by equality of opportunities, circumstances do not predetermine outcomes. Therefore, even if there are disparities in outcomes, they are regarded as just because they stem from a fair process. Sen (2000) explores this concept, emphasizing individuals capabilities to engage in various aspects of their lives.

Economic reasons. Assuming the existence of equality of opportunity and the equitable distribution of talents across society, we would anticipate shifts in individuals relative positions within the income and education distribution across generations. In practice, due to the absence of complete equality of opportunity, much talent and potential remain unrealized, resulting in an inefficient allocation of resources in the economy. In addition, when people perceive a lack of robust equality of opportunity, it can impact the efforts exerted by those at the lower end of the distribution. They may believe that their efforts will yield limited returns, and that their children will likely end up in a similar relative position, discouraging their pursuit of advancement. Conversely, when individuals at the upper levels of society believe that their status is guaranteed, they may not place a high priority on ensuring that their children receive a better education than they did. They rely on the inheritance they pass down and their network of connections to be sufficient in preserving their social standing.

The foundational framework for understanding the intergenerational transmission of income and human capital can be traced back to the pioneering work of Becker and Thomes in their seminal publications from 1979 and 1986. In their model, families are depicted as rational agents seeking to maximize their utility by making optimal choices between their own consumption and investments in their children's human capital. This model operates under the initial assumptions of a perfectly functioning credit market and the presence of public investment as a viable alternative to private investment. Subsequent scholars have built upon this foundation, either relaxing or refining these assumptions to provide a more nuanced and comprehensive analysis of intergenerational dynamics.

In recent years, there has been a significant research focus on understanding intergenerational socioeconomic mobility and its global variations. Corak, in his works from 2006 and 2013, provides estimates of intergenerational coefficients for 11 high-income economies from various studies. He also explores the mechanisms through which income inequality influences a country's relative social mobility, developing the well-known *Great Gatsby Curve*. Causa and Johansson (2010) investigate relative mobility across OECD countries using diverse metrics, concluding that Southern European nations and Luxembourg exhibit low mobility, while Nordic countries display high mobility. Brunori, Ferreira, and Peragine (2013) conduct a meta-analysis of indices measuring equality of opportunity, linking it with GDP per capita, inequality, and social mobility across 41 countries. Their findings suggest a positive association between income inequality and inequality of opportunities, emphasizing the role of limited opportunities in connecting high inequality to low social mobility.

However, international evidence on intergenerational mobility remains scarce. Many studies rely on surveys that gather panel data or cross-sectional data with information on both parents and children, but these surveys are relatively rare worldwide and are typically concentrated in developed countries. To address this issue, this investigation introduces an homogeneous methodology based on *coresidents* to produce comparable estimates of intergenerational educational and income mobility persistence for 87 and 150 economies, respectively, spanning the period 2000-2020. This approach leverages the increasing trend of sons residing with their fathers, as observed in socioeconomic surveys worldwide in recent decades, which allows for the observation of income and education levels for both fathers and sons. Similar methods have been employed in previous studies, including those by Sanhueza (2011), Nuñez and Sanhueza (2015), Shilpi et.al (2016), and Van der Weide et.al (2018, 2021).

One concern associated with coresiding data pertains to the potential presence of bias in the estimates of intergenerational coefficients. However, recent literature, particularly in the case of intergenerational education coefficients, suggests that this bias is generally low and often negligible, especially among younger sons, as demonstrated in studies by Emran et.al (2016, 2018) and Van der Weide et.al (2018). We contend that a similar pattern of low potential bias exists in the case of intergenerational income mobility among younger sons, where coresidence rates tend to be higher. To address this, we compute coefficients for sons aged 23-30, a range characterized by higher coresidence rates and therefore a lower potential for bias. Nevertheless, to further address potential biases in estimating intergenerational income coefficients, we employ a two-stage Heckman procedure. This approach models the coresiding stage based on determinants of coresidence, as identified in the literature, with references to studies by Sanhueza (2011) and Morales (2014). This additional step helps us account for potential selection biases in the estimation process and enhances the robustness of our results.

The results obtained from both simple OLS and Two-stage Heckman estimates exhibit remarkable similarity, aligning with the prevailing understanding of low bias in intergenerational coefficients, especially among younger sons, as documented in the literature. This consistency reinforces the robustness of our findings. Furthermore, our estimates demonstrate a noteworthy degree of alignment with existing international estimates derived from full sample father/son income data, particularly for similar age groups, as illustrated in the work by Van der Weide et.al (2018). Our results indicate that wealthier nations tend to exhibit higher levels of both educational and income intergenerational mobility. Conversely, regions characterized by lower mobility include South Asia, Latin America & the Caribbean, and South Africa. This empirical evidence underscores the role of macroeconomic factors, such as government expenditures on education and economic inequality, as mechanisms contributing to intergenerational mobility. These findings underscore the intricate interplay between social mobility and broader economic and policy variables.

In the following sections, this investigation is organized as follows: In section 2 we review the literature that has estimated relative intergenerational mobility in countries of all income levels and regions. Section 3 provides an in depth examination of our data sources and the methodology used to construct our empirical results. Section 4 elaborates on the empirical aspects of our research, filters applied and selection bias issues. Section 5 offers descriptive statistics, both at the aggregate level and segmented by income and global regions. Our research findings are presented in Section 6, addressing the two dimensions mentioned in Section 2. Finally, Section 7 concludes the investigation, summarizing our results and suggesting potential avenues for future research building on our findings.

2 Related literature

To date, only two publications conducted by the World Bank have comprehensively studied intergenerational mobility across countries of all income levels and regions. A comparative summary of these works with this investigation is provided in the following table.

Table 1: **Comparison of researchs**

	Van der Weide et.al		**This work**	
	2021 (Education)	2018 (Education & Income)	Education	Income
Surveys [1]	<500	<650	1.074	536
Methodology	IV & Coresidents	IV & Coresidents	Coresidents	Coresidents
Approach	Cohort	Cohort	Cross Section	Cross Section
Estimations	Level	Level	Level & Range	Level & Range
Mix of results	Yes	Yes	No	No

This study stands as a big effort by achieving an extensive coverage of surveys spanning various economies across the globe. It is worth highlighting that the selection of these surveys underwent a rigorous filtering process, as elaborated upon in subsequent sections. It is important to note that the number of surveys indicated represents the minimum count of surveys obtained using a specific estimator. Alternative estimators may yield a higher count of surveys than those presented in the table. For a comprehensive overview of the distribution of surveys and economies over time, please refer to the appendix section.

[1] The World Bank's work does not directly state the number of surveys that were used, so we infer a higher figure.

Van der Weide et.al's methodology employed a two-stage estimation approach, obtaining information from sons about their fathers through retrospective questions. For instance, they estimated educational mobility based on the son's report of his father's education level, a relatively straightforward measure to report. However, for income, they used instrumental variables derived from the son's information, or, when unavailable, turned to the coresidents method. Our study employs the coresident methodology consistently across all surveys, ensuring that we generate estimates in a uniform and directly comparable manner.

We adopt a cross-sectional approach, allowing us to explore how relative mobility has evolved at various time points, encompassing individuals from different birth cohorts. This differs from Van der Waide et al.'s cohort approach, which tracks changes in relative mobility across generations based on birth years. Additionally, following the approach of Chetty et.al (2014), we estimate relative mobility by both levels and ranks, which enhances the robustness of our findings and facilitates sensitivity analyses.

Van der Weide et.al utilizes household and labor surveys to compile their findings in the *Global Database on Intergenerational Mobility* (GDIM), and they integrate these results with estimates from the *Equalchances* database. However, they do not provide detail information regarding the specific functional forms of their regression models or the filters employed in the selected surveys. As a result, their approach may lead to a blend of results obtained through different methodologies, which may not be ideal. In contrast, our study maintains a distinct approach by not mix results from other sources. We generate estimates by thoroughly documenting the functional forms used and justifying the choice of variables. Furthermore, we offer detailed information about the filters applied at both the observation and survey levels. This transparency enables us to compare countries over time with confidence, knowing that any observed differences are not a result of methodological disparities.

3 Data

The ideal data for conducting research of this nature would consist of long panel datasets that meticulously track the income histories and years of schooling of both parents and their children. However, obtaining such data is often prohibitively expensive, and if available at all, it tends to be concentrated in developed countries. The next best option would be to have partial information about the parents, which would allow us to estimate key variables like income. This approach would enable us to generate parent-child combinations from a representative sample with a substantial number of observations. The third best option, and the one employed in this investigation, is to work with data where parent-child combinations are observed solely for those living in the same household. While not as comprehensive as long panel data, this type of data still provides valuable insights for the research at hand. The 3 sources of data available to us are as follows:

International Income Distribution Database (I2D2). To compile international evidence from various economies, the World Bank meticulously curated standardized databases through household and labor market surveys spanning both, developed and developing economies over a decade. The database extends from the work of Monenegro and Patrinos (2014). The survey selection process did not involve any form of censoring, but some surveys were excluded due to missing key vari-

ables. Inclusion criteria mandated that the surveys must be nationally representative. Multiple estimates may exist for a given economy and year due to the quarter or half-yearly frecuency of the survey. Additionally, many of these databases are not publicly accessible, which prompted us to collaborate with Claudio Montenegro to generate the necessary estimates while maintaining data confidentiality. Our analysis drew from a total of 2,308 surveys, reduced to 911 surveys for the education dimension and 477 for the income dimension following a series of filters detailed in the empirical approach section.
Luxemburg income study (LIS): Surveys sourced from the LIS are nationally representative and offer more contemporary income data, particularly for affluent nations, with a growing inclusion of developing countries. While these data are more accessible to the public, access is somewhat restricted, requiring users to enter a server and utilize a code to receive results via email. This source encompasses a total of 728 surveys, with 163 surveys utilized for the education dimension and 59 for the income dimension. The LIS data will serve as a secondary source if specific country-year combinations are unavailable in I2D2.
Quality of governance (Teorell et.al, 2023): To establish correlational evidence at the international level between our estimates and macroeconomic variables of countries, we rely on this database. It contains country-year observations gathered from international organizations such as the World Bank and the International Monetary Fund. This dataset is openly accessible to the public.

In addition to these data sources, we categorize countries by income level and geographic region using a World Bank classification updated to the year 2022. The income classification encompasses four categories[2]: High income, Low income, Upper middle, and Lower middle. The region classification consists of seven categories[3]: East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, North America, South Asia, and Sub-Saharan Africa. Detailed lists of countries in each category can be found in the appendix section.

4 Empirical approach

4.1 Intergenerational persistence coefficients

To assess relative social mobility across different economies, we employ four distinct measures: Pearson correlation, Spearman correlation, Ordinary Least Squares (OLS), and Heckman Maximum Likelihood. It's important to note that the Heckman Maximum Likelihood measure is specifically applied in the income dimension, and its detailed explanation can be found in the selection bias subsection.

Pearson and Spearman correlations are utilized as robustness measures, providing complementary insights into the relationship between variables. Mathematical representations of these measures are further detailed in the appendix section. These measures play a crucial role in evaluating the degree of association and robustness in our analysis of social mobility across economies.

[2]The only country that cannot be classified was Venezuela, since the data provided is not reliable for the World Bank

[3]The classification may not be geographic. For example, Mexico is geographically in North America, but in the database it is classified as Latin America and the Caribbean.

Ordinary least squares regression typically has the following functional form:

$$s_{i,t,k} = \alpha + \beta f_{i,t,k} + \gamma X + \epsilon_{i,t,k}$$

Where:
$s_{i,t,k}$= ln(income) / education of son i, in the year t, in the economy k
$f_{j,t,k}$= ln(income) / education of father i, in the year t, in the economy k.
X = Following the functional form of Solon (1992) we used age and age squared as control variables to control for the life cycle of the sons and the fathers (in the income dimension)

In our analysis we annualize the salaries. This annualization process is crucial for isolating the effect that different frequencies of income receipt may have on intergenerational income mobility.

There is a well-known relationship between the regression coefficient (β) and the pearson correlation (ρ). The derivation of this relationship can be found in the appendix section.

$$\beta = \rho \frac{\sigma_s}{\sigma_f}$$

In simple, this relationship tells us that the regression coefficient is sensitive to the inequality (dispersion) that exists in the income/schooling distribution of the parent and child generation. Therefore, correlation measures are appropriate robustness instruments. In case the inequality in both generations are equal $\frac{\sigma_s}{\sigma_f} = 1$, then $\beta = \rho$

Let's use an example to better understand the relationship between social mobility and the beta coefficient. Suppose we have a situation where a parent's income is 10% above the average income of their generation, and the beta coefficient in the regression is 0.5. This beta coefficient essentially represents the degree of intergenerational income mobility. In this scenario the parent's children, on average, are expected to have incomes that are 5% (10% * 0.5) above the mean income of their generation. This implies some degree of social mobility. Now, let's consider a different scenario where the relative position of the parent remains the same, but the value of the beta coefficient is 1. In this situation on average, the children will have incomes that are 10% (10% * 1) above the mean income of their generation. In this case, the relative position of the parent and their children in the income distribution remains constant from one generation to the next, indicating a complete lack of social mobility.

To put it another way, as Torche (2013) explains, a regression coefficient of 0.4 indicates that a 10% difference in parents' earnings leads, on average, to a 4% difference in their children's earnings. This means that if two fathers' earnings differ by 10%, their children's earnings will differ, on average, by 4%. This coefficient provides a quantitative measure of the extent to which parental income influences the income of the next generation and thus serves as a key indicator of social mobility within a society.

Therefore, the closer β (intergenerational persistence) is to 0, the greater the social mobility

It's important to note that estimating the income dimension of intergenerational mobility poses more challenges compared to the educational dimension. Income can exhibit greater variability over time, including transitory shocks and significant fluctuations within the study window. This variability introduces a higher risk of measurement error when compared to years of schooling, which are relatively stable.

4.2 Observations and surveys filters

To ensure the accuracy and relevance of our analysis in both, the education and income dimensions, we applied a set of filters to the observations and surveys. These filters are summarized in the following table and are elaborated upon below:

Table 2: **Observations and surveys filters**

Level	Filters	Education	Income
Observations	Coresidents son's age	23-30	23-30
	Semanal working hours	-	≥ 22
	Father's age	-	≤ 65
	Exclusion top 0.5% of distribution	No	Yes
	Control variables	-	Son and father age
Surveys	$\geq$ Year 2000	Yes	Yes
	Plausible estimates	[0.1 ; 0.8]	[0.1 ; 0.8]
	Min observation & Max median error	200 & 0.2	100 & 0.2

Coresidents son's age. The choice of a specific age range for coresident sons in our analysis is a critical decision. We have selected an age range of 23 to 30 years for two main reasons. *Transition to Labor Market*: At around 23 years of age, individuals typically complete their tertiary education and begin transitioning into the labor market. This age represents a pivotal point in their economic and educational trajectories, making it relevant for studying intergenerational mobility. *Coresidency and Labor Participation*: An analysis detailed in the appendix section, confirms that this age bracket is appropriate for our analysis. It aligns with the typical life stage during which individuals reside with their parents and actively participate in the labor market.

Semanal working hours. In the income dimension, our filter excludes observations with 22 or fewer hours of work per week. This approach is in line with the concept of permanent income and enhances the accuracy of our analysis by excluding individuals with very low work hours, which may not provide a consistent measure of income.

Father's age. We exclude observations where fathers are beyond 65 years old. This filter was initially designed for men in Chile, further analysis of retirement age patterns in other countries has led us to determine that the chosen cutoff age is appropriate and applicable more broadly. By setting this age limit, we aim to minimize the impact of retirement-related income changes on our analysis, focusing on individuals who are actively participating in the labor market and contributing to the measurement of income mobility.

Truncation of income distribution. In order to ensure the robustness and reliability of our income dimension analysis, we have applied a filter that excludes the top 0.5% of income earners from each survey. This filtering approach helps to produce more representative and accurate measures of income mobility by focusing on the majority of income earners while minimizing the impact of extreme outliers. This practice is consistent with a World Bank standard procedure.

Control variables. In the income dimension, we incorporate both, the age of the son and the age of the father as control variables. This approach follows the functional form proposed by Solon (1992) and is designed to account for potential age related variations in income persistence. Including these control variables allows us to better isolate the impact of intergenerational income mobility, taking into consideration the ages of both generations, mitigating the famous "*life-cycle bias*". In contrast, in the educational dimension, this concept is less relevant because educational levels typically remain stable once attained, and individuals do not drop out of their educational level. Therefore, estimating intergenerational educational persistence at different ages does not present the same problem, and the inclusion of control variables related to age is not necessary in this context.

Time window. Our decision to focus on surveys conducted from the year 2000 onwards is based on several compelling reasons. *Country coverage*: During this period, there is a greater coverage of countries included in the surveys. This broader representation allows us to capture a more comprehensive view of intergenerational mobility across diverse economies. *Post-Transition Period*: After the fall of the Berlin Wall and the dissolution of the Soviet Union, numerous countries experienced significant political, economic, and social transformations. It was essential to allow these nations time to stabilize, rebuild, and collect high-quality data before including them in our analysis. The year 2000 represents a reasonable cutoff point to ensure that these countries had sufficient time to organize themselves and generate reliable data. *Coresidency rate trends*: Another critical consideration is the trend in coresidency rates over time, particularly within the relevant age range. We observed that the coresidency rate has been increasing over time for individuals within the specified age group. Focusing on the latest available years is valuable for our research as it helps mitigate coresidency bias, allowing us to capture a more accurate picture of intergenerational mobility. The first and third reasons can be seen graphically in the appendix section.

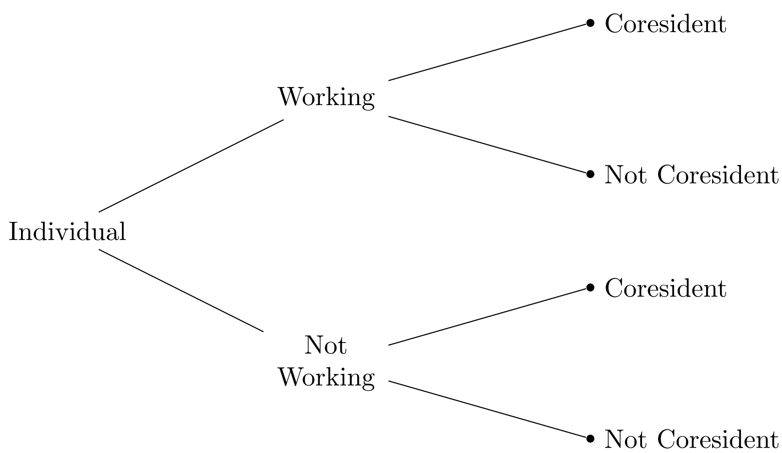
Plausible estimates. In our analysis, we have applied a filter to exclude surveys that yield relative mobility estimates that are either less than 0.1 or greater than 0.8. The idea behind this filter is to focus on estimates that fall within a reasonable range and align with empirical observations documented in the existing literature. By excluding such extreme estimates, we aim to maintain the credibility and relevance of our analysis, ensuring that our findings are consistent with established patterns of relative mobility and plausible real-world scenarios.

Surveys. To maintain the statistical reliability and robustness of our analysis, we implemented two type of filters related to the surveys. *Number of observations*: Surveys with fewer than 200 observations in the educational dimension or 100 observations in the income dimension were excluded from the sample used for calculating the coefficients of interest. This criterion ensures that the results are not unduly influenced by sampling errors that may arise in surveys with very small sample sizes. *Deviation from Median*: Surveys were also excluded if they exhibited a deviation from the median value for their respective country that exceeded 0.2 percentage points. This filter helps ensure that the results are not skewed by surveys that significantly deviate from the typical patterns within their respective countries.

4.3 Selection bias

We acknowledge the presence of two sources of selection bias that can impact our findings. *Labor market participation bias.* Individuals who are not actively engaged in the labor force may have different income outcomes compared to those who are employed. This bias can influence our estimates of intergenerational income mobility. *Coresidence bias.* Individuals who live with their parents may have different income experiences than those who have left the parental household.

Figure 1: **Selection Bias Sources**



To attenuate the first selection bias we will focus only on the father-son combination, leaving aside the mother-son, father-daughter or mother-daughter combinations since it has been documented female labor participation rate is lower than males. In the following table, the impossibility of observing the salary is marked with a $\times$ and the possibility of seeing it in the database is marked with a $\checkmark$.

Table 3: **Father-Son income combinations**

		Son Wage	**Father Wage**
Working	Coresident	$\checkmark$	$\checkmark$
	Not Coresident	$\checkmark$	$\times$
Not Working	Coresident	$\times$	$\times$
	Not Coresident	$\times$	$\times$

In our analysis, we encounter a data limitation related to the ability to observe the income of both fathers and sons. Here's how this limitation manifests. *Son's Income*: Regardless of whether the son lives with his father, we can observe the son's salary as long as he actively participates in the labor market. *Father's Income*: On the other hand, we can only observe the father's salary if he participates in the labor market and lives with his son. This limitation arises from the fact that our databases do not contain questions that provide information about the father when the son lives alone or independently. Therefore, we rely on the coresident methodology to estimate the father's income in these cases.

However, it's worth noting that if we had access to additional information about the father, such as his schooling, occupation, or other characteristics, we could employ a technique known as the Two-Sample Instrumental Variable (TSIV) methodology to estimate the father's income. This method, explained and applied in the work of Angrist and Krueger (1992), allows for the estimation of a variable (in this case, the father's income) when only limited information is available. Unfortunately, our current databases do not provide this level of detail on fathers, necessitating the use of the coresident approach in our analysis.

To address the potential selection bias associated with sons who do not live with their fathers, we will employ the Heckman maximum likelihood estimation method. This approach is specifically applied to the income dimension of our analysis, as it has been documented that the coresident bias for education is minimal, as indicated by previous research (Emran et al., 2016 & Emran et al., 2018). The Heckman maximum likelihood method allows us to account for any selection bias that may arise due to differences between sons who live with their fathers and those who do not. By statistically adjusting for this potential bias, we can obtain more accurate estimates of intergenerational income mobility, ensuring that our findings are not unduly influenced by the coresidence status of the individuals in our sample.

If we consider the latent variable v_s^* as the benefit that the son obtains by living with his father, we will observe the sample only when $v_s^* > 0$. We have:

$$v_s^* = \gamma Z_s + \epsilon_s$$

$$v_s = \begin{cases} 1 & \text{if } \gamma Z_s + \epsilon_s > 0 \\ 0 & \text{if } \gamma Z_s + \epsilon_s \leq 0 \end{cases}$$

Where Z_s corresponds to a vector of variables that influence the son's decision to live or not with his father.

Taking into consideration the variables used in the works of Sanhueza (2011) and Morales (2014), together with the variables we have available in our surveys, we decided that the participation equation will be as follows.

$$Cores = \alpha + \beta ecivil + \gamma urban + \delta_1 age + \delta_2 age^2 + \rho crowd + \eta \text{Liability Ratio} + v$$

Cores. This is a binary dummy variable that takes the value of 1 if the person lives with their father and 0 if not.

Civil status. This is another binary dummy variable that takes the value of 1 if the person is married, in a partnership, divorced, or widowed, and 0 otherwise. It is expected that individuals in these civil status categories are less likely to live with their parents, impacting their coresidence status.

Urban. This binary dummy variable is set to 1 if the person lives in an urban area and 0 if not. It accounts for the influence of urban versus rural living environments on coresidence patterns.

Age. Is a continuous variable representing the age of the individual. As people get older, the likelihood of coresidence with parents typically decreases, and this variable helps capture that effect.

Crowd. Quantifies the number of children living in the person's household. It is expected that households with more children will have a lower probability of coresidence, as there may be fewer available resources and space for adult children.

Liability ratio. This variable divides the number of "liabilities" with the number of "assets" per household. We define "active" persons as those between the ages of 23 and 65. We define "passive" persons as those who are 75 years of age or older.

Following the functional form employed by Solon (1992), the estimates of the income dimension (principal equation) are obtained from the following regression model:

$$y_s = \alpha + \beta y_f + \gamma_1 age + \gamma_2 age^2 + \delta_1 \text{father age} + \delta_2 \text{father } age^2 + \epsilon_s$$

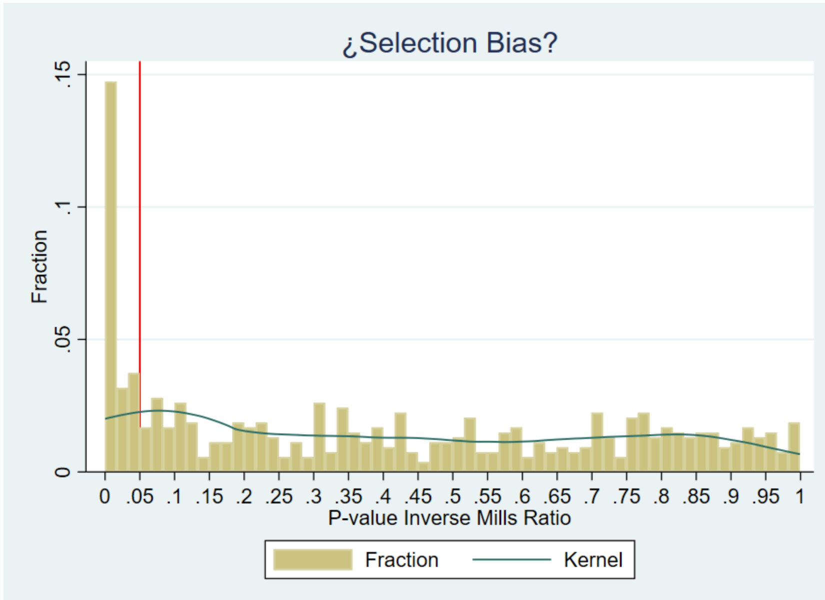
In summary, the following table shows the variables used as controls (main equation) and those used as instruments (participation equation).

Table 4: **Control and instrument variables**

Variables	**Controls**	**Instruments**
Father's wage	✓	×
Son's age & age^2	✓	✓
Father's age & age^2	✓	×
Civil status	×	✓
Urban	×	✓
Crowd	×	✓
Liability ratio	×	✓

We conclude the section by showing the distribution that the p-value of the inverse mills ratio has when estimated by Heckman maximum likelihood.

Figure 2: **Inverse mills ratio distribution**



Author's Calculations: Base on World Bank and Luxembourg Income Study Data

We have 536 heckman estimates. 116 (21.6%) of them have a p-value less than 5% in their inverse mills ratio. However, in the results section we can see that the difference between the Heckman and OLS estimates are not very different from each other.

5 Descriptive statistics

Presented here are frequency tables depicting the distribution of economies by region and income level. Furthermore, we will examine descriptive statistics on coresidence rates, labor force participation, and schooling years.

Table 5: **Education economies**

Economies	High income	Low income	Lower middle income	Upper middle income	Total
East Asia & Pacific	3	0	11	3	17
Europe & Central Asia	29	0	3	14	46
Latin America & Caribbean	5	0	5	11	21
Middle East & North Africa	3	2	7	2	14
North America	2	0	0	0	2
South Asia	0	1	6	1	8
Sub-Saharan Africa	0	22	15	5	42
Total	42	25	47	36	150

Author's Calculations: Based on World Bank and Luxembourg Income Study Data

Table 6: **Income economies**

Economies	High income	Low income	Lower middle income	Upper middle income	Total
East Asia & Pacific	3	0	5	2	10
Europe & Central Asia	21	0	3	10	34
Latin America & Caribbean	3	0	4	10	17
Middle East & North Africa	2	1	3	2	8
North America	2	0	0	0	2
South Asia	0	1	5	0	6
Sub-Saharan Africa	0	5	4	1	10
Total	31	7	24	25	87

Author's Calculations: Based on World Bank and Luxembourg Income Study Data

The frequency tables provided represent the distribution of economies by region and income level, using the lower bound of available countries by dimension. If we were to choose a different statistic and construct tables based on that, we could potentially include more countries in the analysis.

Table 7: **Descriptive statistics of coresiding sons (ages 23-30) by income level**

	Coresidence Rate	Labor Participation Rate	Father's Schooling (years)	Son's Schooling (years)
High Income				
Mean	.45	.80	11.0	12.0
Median	.46	.82	12.0	12.0
Max	.95	1.00	20.0	16.0
Minimum	.00	.00	0.0	0.0
Standard Deviation	.23	.14	1.7	1.2
Surveys	-	-	770	770
Low Income				
Mean	.30	.77	4.8	7.1
Median	.25	.82	4.4	7.2
Max	.97	1.00	12.0	12.0
Minimum	.00	.00	0.9	2.5
Standard Deviation	.19	.16	2.6	2.5
Surveys	-	-	99	99
Lower Middle Income				
Mean	.46	.83	6.2	9.2
Median	.45	.86	5.8	9.3
Max	.98	.99	13.0	15.0
Minimum	.01	.00	0.9	3.8
Standard Deviation	.20	.13	2.5	2.1
Surveys	-	-	273	273
Upper Middle Income				
Mean	.50	.84	8.4	11.0
Median	.47	.88	7.8	11.0
Max	1.00	1.00	18.0	15.0
Minimum	.03	0.00	2.3	5.1
Standard Deviation	.21	.13	2.5	1.5
Surveys	-	-	471	471

Author's calculations: Based on World Bank and Luxembourg Income Study data

Table 8: **Descriptive statistics of coresiding sons (ages 23-30) by region**

Region	Coresidence Rate	Labor Participation Rate	Father's Schooling (years)	Son's Schooling (years)
East Asia & Pacific	.52	.82	8.4	11.0
Europe & Central Asia	.51	.80	12.0	12.0
Latin America & Caribbean	.37	.87	7.1	11.0
Middle East & North Africa	.60	.84	8.0	12.0
North America	.21	.80	13.0	13.0
South Asia	.56	.89	5.2	8.6
Sub-Saharan Africa	.32	.76	5.3	7.9
Total	.46	.82	9.2	11.0

Author's Calculations: Based on World Bank and Luxembourg Income Study Data

It's notable that across all income levels and regions, there has been a consistent pattern of absolute positive social mobility in terms of years of schooling between the generation of parents and their children. This indicates an overall trend of increasing educational attainment in younger generations. For our study's purposes, having both a high coresidency rate and labor force participation rate is crucial to minimize potential selection bias. Interestingly, the economies with the highest rates in these categories are upper middle-income economies, along with South Asian economies.

6 Results

In this section, we will begin with a statistical summary of the three estimators of intergenerational persistence in the education dimension and the four estimators in the income dimension. We will then present these estimators disaggregated by income level of the economies and by region. We will show the relationships between the estimators and conclude the section by presenting correlational evidence between the intergenerational persistence estimators and macroeconomic variables.

Table 9: **Intergenerational persistence coefficients summary**

	Education			Income			
	Beta [4]	Pearson	Spearman	Elasticity [5]	Pearson	Spearman	Heckman
Economies	150	150	150	89	88	87	88
Surveys	1,074	1,092	1,101	544	549	618	536
Mean	.38	.42	.42	.28	.34	.33	.28
Median	.38	.43	.42	.27	.35	.34	.27
Maximum	.79	.70	.69	.79	.73	.74	.77
Minimum [6]	.10	.10	.10	.10	.10	.10	.10
Standard Deviation	.14	.12	.12	.11	.14	.14	.12

Author's Calculations: Based on World Bank and Luxembourg Income Study Data

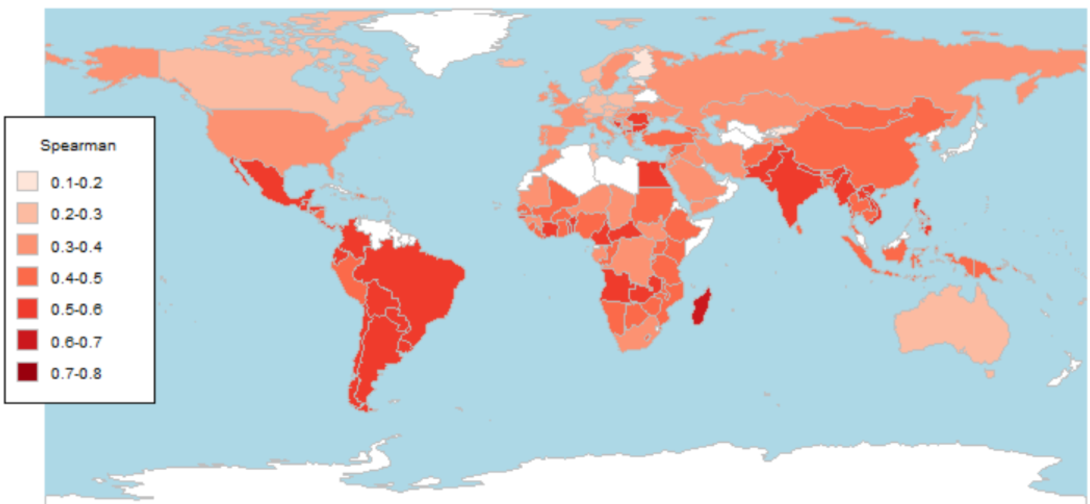
From the table, it can be observed that the robustness measures, Pearson and Spearman, have higher values than those of the regressions. There are no significant differences between the mean

[4] OLS regression coefficient
[5] OLS regression coefficient
[6] This values are explained due the surveys filter

and median estimates. Finally, it is noted that the Ordinary Least Squares estimates are not very different from those obtained using the Heckman method. It's evident from the data that we have a substantial number of economies and surveys available for conducting our study, with the education dimension having approximately 70% more countries and almost twice as many surveys compared to the income dimension.

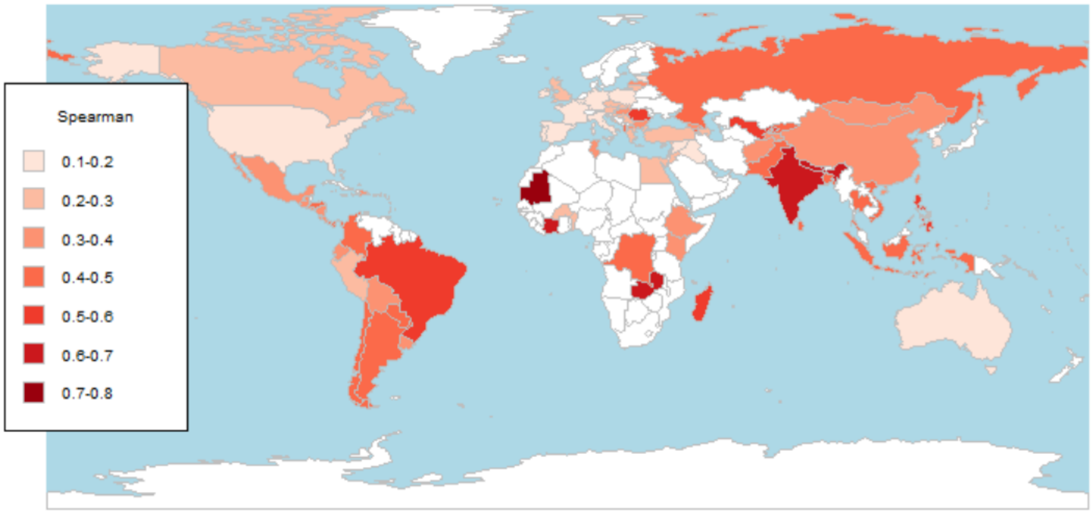
To provide an initial overview of the average persistence of economies worldwide from the year 2000 onwards, we have generated two figures that utilize Spearman's measure for both the education and income dimensions. These figures provide a visual representation of the relative intergenerational persistence across various economies. If an economy appears blank in the figures, it indicates that there is no available information for that particular region.

Figure 3: **Educational intergenerational persistence coefficient world map**



Author's Calculations: Base on World Bank and Luxembourg Income Study Data

Figure 4: **Income intergenerational persistence coefficient world map**



Author's Calculations: Based on World Bank and Luxembourg Income Study Data

These figures serve as a preliminary snapshot of intergenerational persistence, offering an initial glimpse into how different economies fare in terms of education and income mobility. Further analysis will delve into the specifics of these patterns and provide more in-depth insights into the factors influencing intergenerational mobility across the globe.

6.1 Intergenerational persistence coefficients by income level

We will now look at the results of relative social mobility by disaggregating the economies by income level. We will show this using different statistics and time series graphs.

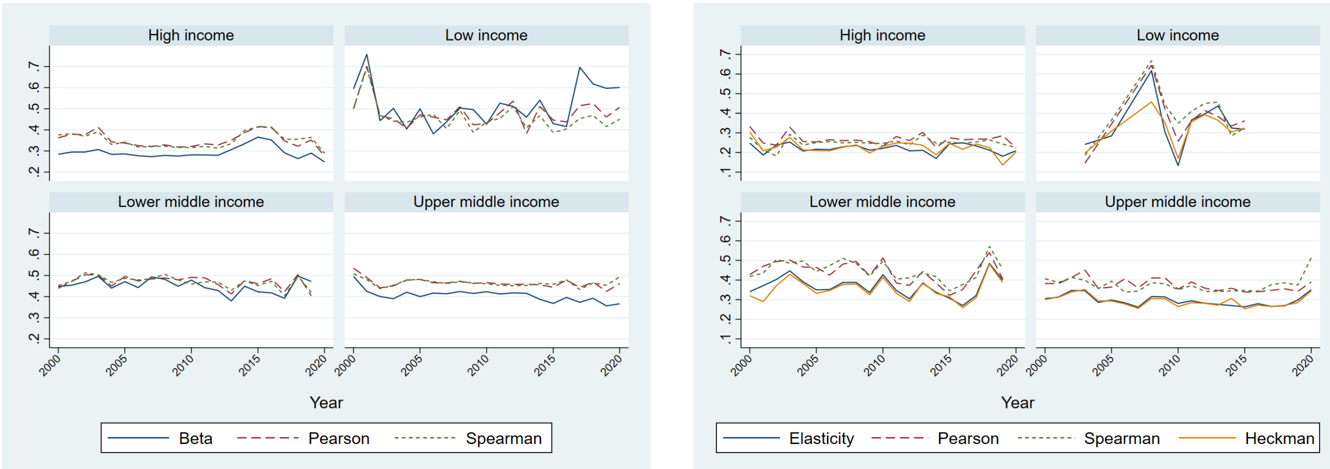
Table 10: **Intergenerational persistence by income level**

	Education			Income			
	Beta [7]	Pearson	Spearman	Elasticity [8]	Pearson	Spearman	Heckman
High Income							
Mean	.29	.35	.34	.22	.26	.25	.23
Median	.26	.34	.33	.21	.21	.21	.21
Max	.67	.64	.62	.53	.70	.65	.6
Minimum	.10	.10	.10	.10	.10	.10	.10
Standard Deviation	.11	.11	.11	.09	.14	.13	.10
Surveys	409	421	424	208	218	280	199
Economies	42	42	42	33	32	31	31
Low Income							
Mean	.48	.46	.45	.36	.37	.40	.33
Median	.47	.46	.44	.34	.38	.37	.34
Max	.79	.70	.69	.62	.65	.67	.52
Minimum	.22	.22	.24	.13	.15	.19	.17
Standard Deviation	.14	.09	.09	.13	.12	.13	.11
Surveys	76	79	80	14	13	14	14
Economies	25	25	25	7	7	7	7
Lower Middle Income							
Mean	.45	.47	.47	.36	.44	.45	.35
Median	.46	.49	.50	.35	.42	.43	.34
Max	.77	.68	.65	.79	.73	.74	.77
Minimum	.13	.14	.13	.11	.15	.21	.10
Standard Deviation	.12	.10	.10	.13	.12	.12	.13
Surveys	244	245	251	132	131	133	130
Economies	47	47	47	24	24	24	25
Upper Middle Income							
Mean	.41	.46	.46	.29	.37	.37	.29
Median	.41	.47	.48	.28	.38	.36	.28
Max	.78	.66	.63	.54	.60	.56	.58
Minimum	.12	.19	.18	.11	.11	.12	.13
Standard Deviation	.11	.09	.09	.08	.10	.09	.08
Surveys	345	347	346	190	187	191	193
Economies	36	36	36	25	25	25	25

Author's Calculations: Based on World Bank and Luxembourg Income Study Data

[7] OLS regression coefficient
[8] OLS regression coefficient

Figure 5: **Average intergenerational persistence coefficients by income level**



Author's Calculations: Based on World Bank and Luxembourg Income Study Data

The analysis reveals a noteworthy pattern. High income economies consistently exhibit lower intergenerational persistence compared to poorer countries. Interestingly, the disparities in intergenerational persistence are less marked between upper-middle and lower-middle income countries. Moreover, the estimates across all income levels exhibit a high degree of robustness. It's worth noting that both Pearson and Spearman correlations consistently show greater intergenerational persistence compared to regression coefficients.

6.2 Intergenerational persistence coefficients by region

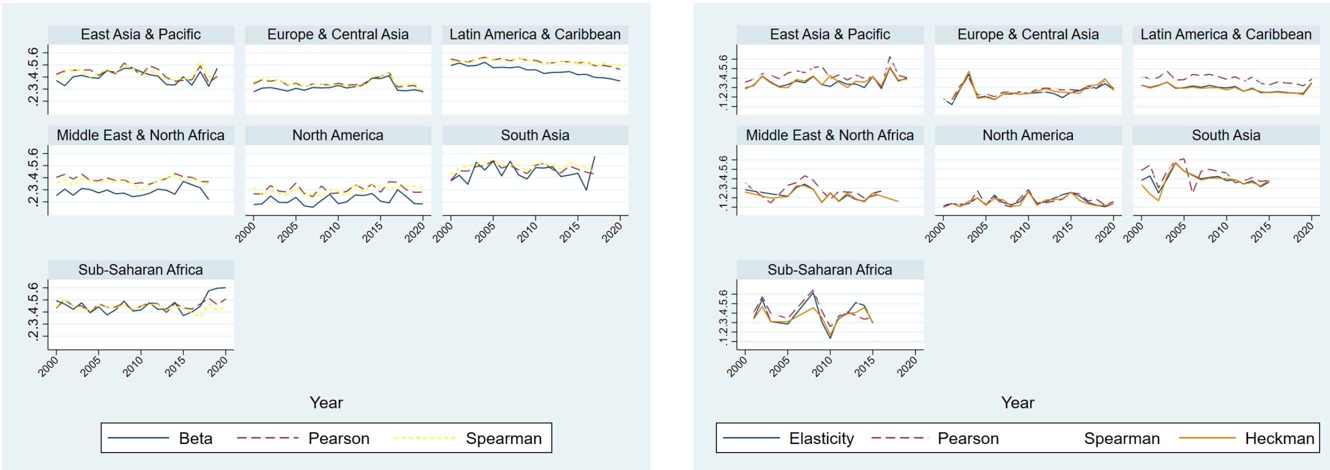
In this subsection we will do the same by disaggregating the economies by region. Since there are a large number of regions the table of statistics will only show the average.

Table 11: **Average intergenerational persistence coefficients by region**

	Education			Income			
	Beta [9]	Pearson	Spearman	Elasticity [10]	Pearson	Spearman	Heckman
East Asia & Pacific	.41	.43	.43	.35	.43	.43	.36
Europe & Central Asia	.32	.35	.34	.24	.27	.25	.24
Latin America & Caribbean	.46	.53	.53	.29	.39	.39	.28
Middle East & North Africa	.30	.39	.37	.24	.28	.25	.23
North America	.22	.30	.30	.17	.17	.18	.16
South Asia	.46	.48	.50	.40	.46	.47	.39
Sub-Saharan Africa	.44	.45	.44	.40	.41	.43	.38
Total	.38	.42	.42	.28	.34	.33	.28

Author's Calculations: Based on World Bank and Luxembourg Income Study Data

Figure 6: **Average intergenerational persistence coefficients by region**



Author's Calculations: Based on World Bank and Luxembourg Income Study Data

Consistently, the regression coefficient retains its position as the statistic with the lowest persistence at the regional level. The regions with the highest intergenerational persistence are South Asia and Sub-Saharan Africa. This suggests that, within these regions, there may be significant challenges in achieving upward social and economic mobility between generations. These findings shed light on the varying degrees of intergenerational persistence across different regions, emphasizing the need for tailored policy interventions to address mobility issues in specific geographic contexts.

6.3 Intergenerational persistence coefficients relationships

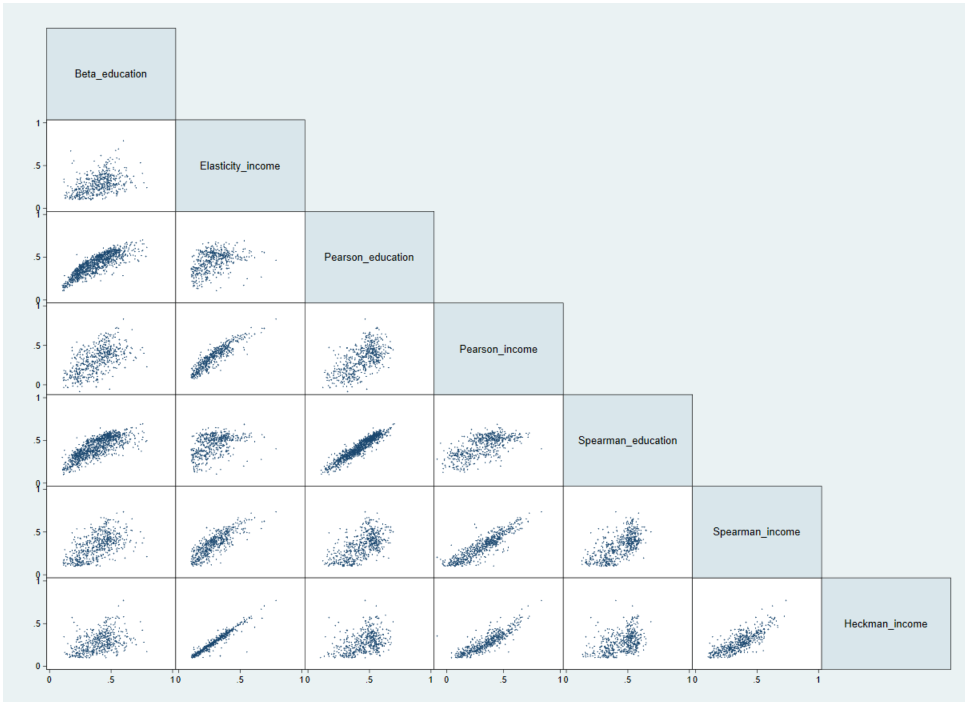
Next, we will examine the relationship between different measures of intergenerational persistence. We will start with a correlation matrix and then proceed to a graphical analysis.

Table 12: **Intergenerational persistence coefficients, correlation matrix**

Pairwise Correlation	Beta_education	Elasticity	Pearson_education	Pearson_income	Spearman_education	Spearman_income	Heckman_income
Beta_education	1.00						
Elasticity	0.43	1.00					
Pearson_education	0.84	0.41	1.00				
Pearson_income	0.59	0.90	0.64	1.00			
Spearman_education	0.78	0.43	0.94	0.65	1.00		
Spearman_income	0.59	0.82	0.62	0.90	0.65	1.00	
Heckman_income	0.41	0.96	0.39	0.86	0.40	0.80	1.00

[9] OLS regression coefficient
[10] OLS regression coefficient

Figure 7: **Intergenerational persistence coefficients, scatter matrix**



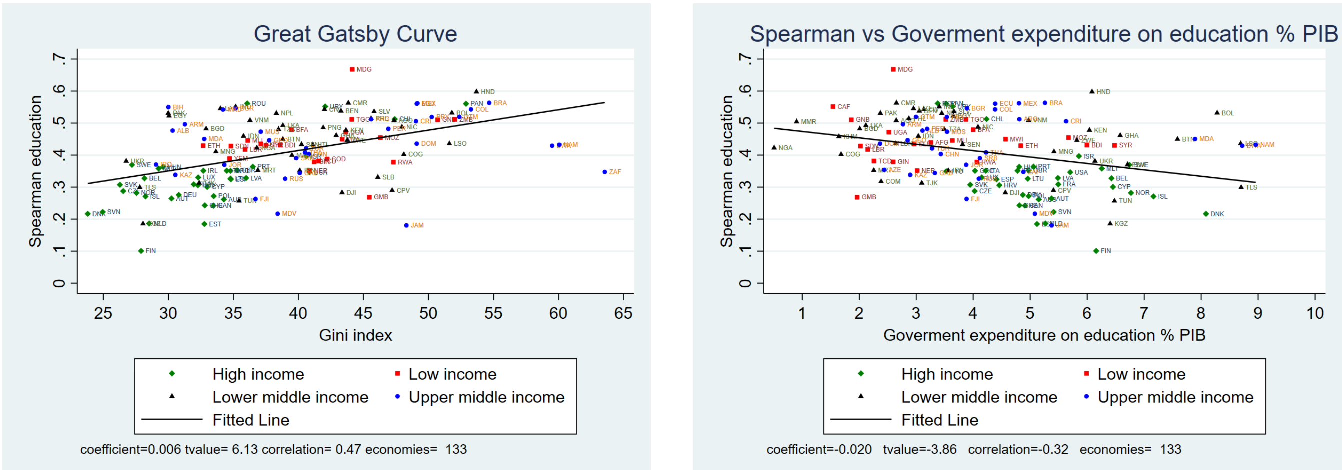
Author's Calculations: Based on World Bank and Luxembourg Income Study Data

It's notable that all combinations of intergenerational persistence exhibit a positive relationship and high correlation. This suggests that there is a strong and consistent connection between the intergenerational persistence of education and income.

6.4 Regresions

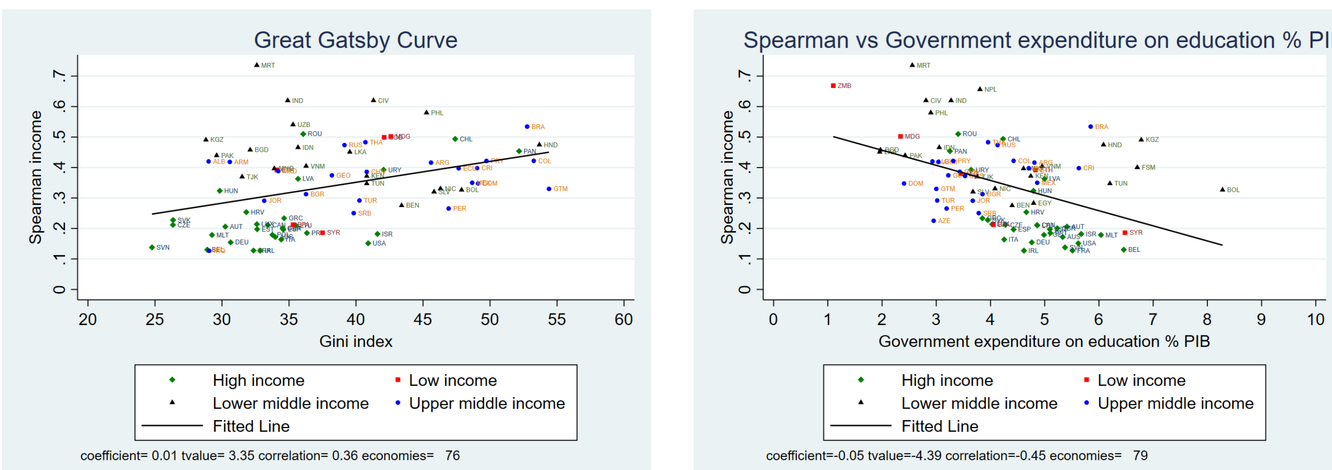
Next, we will graphically examine the correlation between some measures of relative intergenerational persistence and different macroeconomic variables. Subsequently, we will perform a regression analysis and discuss the resulting coefficients.

Figure 8: **Educational intergenerational persistence coefficients vs macroeconomic variables**



Author's Calculations: Based on World Bank and Luxembourg Income Study Data

Figure 9: **Income intergenerational persistence coefficients vs macroeconomic variables**



Author's Calculations: Based on World Bank and Luxembourg Income Study Data

As previously documented in the literature, there is a consistent observation that countries with lower levels of income inequality and higher public spending on education tend to have, on average, higher social mobility. This higher social mobility is reflected in lower intergenerational persistence, indicating that when societies invest in reducing income inequality and providing greater access to quality education, it tends to result in more equitable opportunities for individuals to improve their social and economic positions across generations. These findings highlight the important role of public policies and investments in shaping social mobility outcomes and suggest that addressing income inequality and promoting accessible, high quality education are key factors in fostering greater intergenerational mobility.

Table 13: **Intergenerational persistence coefficients**

	Education			**Income**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ln(Beta)	ln(Pearson)	ln(Spearman)	ln(Elasticity)	ln(Pearson)	ln(Spearman)	ln(Heckman)
ln(Per Capita PPP)	-0.168***	-0.0778*	-0.0732*	-0.129	-0.144*	-0.210**	-0.128
	(-3.82)	(-2.35)	(-2.25)	(-1.97)	(-2.29)	(-3.01)	(-1.85)
ln(Gini)	0.453*	0.766***	0.712***	0.281	0.721**	0.613**	0.249
	(2.57)	(6.42)	(6.24)	(1.28)	(3.43)	(2.86)	(1.14)
ln(Govt Education Expenditure)	-0.156*	-0.116*	-0.122*	-0.287*	-0.370**	-0.197	-0.333**
	(-1.99)	(-2.16)	(-2.19)	(-2.55)	(-2.79)	(-1.38)	(-2.69)
ln(Economy Liberalization)	-0.0742	-0.0853	-0.149	-0.230	-0.237	-0.208	-0.199
	(-0.57)	(-0.72)	(-1.19)	(-0.97)	(-1.14)	(-0.82)	(-0.82)
Constant	-0.567	-2.458***	-2.044**	0.190	-0.937	-0.259	0.231
	(-0.58)	(-3.67)	(-2.97)	(0.14)	(-0.72)	(-0.20)	(0.18)
R^2 Adjusted	0.370	0.421	0.431	0.262	0.412	0.419	0.230
Surveys	706	715	722	376	382	437	374

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In our analysis, we employ the reciprocal of the standard deviation of the dependent variable as the expansion factor for the regressions. This approach helps to account for variations in the dependent variable and ensure that the results are appropriately weighted. Furthermore, it's worth noting that all the regressors included in our models exhibit the expected signs, and the models demonstrate a high level of goodness of fit. For calculating standard errors, we utilize clustering by economy to account for potential heterogeneity within different economies. Additionally, the dependent variables are expressed in natural logarithms, allowing us to interpret the coefficients as elasticities.

To properly assess the economic significance of our findings, it's crucial to understand how the variables are constructed and measured. *Income variable*: This variable represents income and is measured in 2017 dollars adjusted for purchasing power parity (PPP). Adjusting for PPP allows for meaningful international comparisons of income levels, accounting for differences in the cost of living and currency exchange rates. *Inequality variable*: The inequality variable is measured using the Gini coefficient, which is a standard measure of income inequality. The Gini coefficient is expressed on a scale from 0 to 100, where 0 represents perfect equality (everyone has the same income), and 100 represents perfect inequality (one person has all the income). *Economic openness variable*: It is also measured on a scale from 0 to 100. This variable assesses the degree to which an economy is open to international trade and foreign investments. A higher value indicates a more open economy, while a lower value suggests a more closed or protected economy. *Government spending on education*: This variable measures the percentage of a country's gross domestic product (GDP) that is allocated to government spending on education. It provides insights into the level of investment made by the government in the education sector.

In the educational dimension, analyzing the significance using the regression coefficient (beta), we can observe the following: If inequality were to increase by 10%, intergenerational persistence, as measured by the regression coefficient, would increase on average by 45.3%. This suggests a strong positive relationship between income inequality and intergenerational persistence in education. Higher inequality is associated with lower educational mobility. Conversely, if a country's globalization (economic openness) were to increase, intergenerational persistence would decrease on average by -0.74%. This indicates that greater economic openness is associated with slightly lower educational persistence across generations. Increasing government spending on education by 10% would lead to a significant reduction in intergenerational persistence. On average, intergenerational persistence would decrease by 15.6%. This underscores the importance of public investment in education for promoting educational mobility. Similarly, if GDP per capita were to increase, intergenerational persistence would fall by an average of 16.8%. Higher per capita income is linked to lower educational persistence, suggesting that economic prosperity can contribute to greater educational mobility.

The income, inequality and education expenditure variables are statistically significant in the educational dimension. No statistical significance is seen for the globalization variable in any model. In our view, the variables that are economically significant are inequality and education spending, given the magnitude of the average response to the above mentioned changes. In the income dimension, the statistical significance of the variables is concentrated in the Pearson model. The only economically relevant variable in terms of magnitude appears to be inequality of income.

7 Conclusion

In conclusion this investigation provides comprehensive estimates of intergenerational mobility for a large number of economies, covering 150 economies for education and 87 economies for income over the period from 2000 to 2020. The use of a consistent coresidents methodology across surveys allows for meaningful cross country comparisons. The study confirms the well documented pattern that high-income countries generally exhibit lower intergenerational persistence, indicating higher social mobility. In contrast, regions such as South Asia and Latin America and the Caribbean (LAC) have higher intergenerational persistence, suggesting lower social mobility in these areas.

The findings suggest that governments have the potential to improve social mobility by addressing income inequality and increasing investment in education. These policy measures can help promote equality of opportunities and reduce the persistence of socioeconomic disparities across generations.

The rich dataset and methodology presented in this paper offer numerous opportunities for researchers to expand our understanding of intergenerational mobility, both within and across countries. Researchers could explore how changing the age bracket of children, the time window, or another filter, impacts the estimates of intergenerational mobility. Adding variables related to returns to education, as suggested by Solon (2004), could enhance the analysis. Examining how differences in returns to education influence intergenerational mobility could provide deeper insights into the education-income relationship. These future research directions can contribute to a more comprehensive and nuanced picture of social mobility dynamics globally.

8 References

Angrist, J & Krueger, A. (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, 87(418), 328

Becker, G & Tomes, N. (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. Journal of Political Economy, 87(6), 1153–1189

Becker, G. & Tomes, N. (1986). Human Capital and the Rise and Fall of Families. Journal of Labor Economics, 4(3, Part 2), S1–S39.

Brunori, P; Ferreira, F & Peragine, V, (2013). Inequality of opportunity, income inequality and economic mobility : some international comparisons. Policy Research Working Paper Series 6304, The World Bank

Causa, O & Johansson, A. (2010). Intergenerational Social Mobility in OECD Countries. OECD Journal: Economic Studies Volume 2010

Corak, M. (2006). Do Poor Children Become Poor Adults? Lessons from a Cross Country Comparison of Generational Earnings Mobility. Institute for the Study of Labor

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. Journal of Economic Perspectives, 27(3), 79–102.

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

Emran, S; Greene, W & Shilpi, F. (2016). When Measure Matters: Coresidency, Truncation Bias, and Intergenerational Mobility in Developing Countries. Policy Research Working Paper; No. 7608. World Bank, Washington, DC

Emran, S & Shilpi, F. (2018). Estimating Intergenerational Mobility with Incomplete Data: Coresidency and Truncation Bias in Rank-Based Relative and Absolute Mobility Measures. Policy Research Working Paper; No. 8414. World Bank, Washington, DC

Jantti, M; Bratsberg, B; Røed, K; Raaum, O; Naylor, R; Osterbacka, E; Bjorklund, A; & 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. IZA Discussion Paper No. 1938.

Nuñez, J & Sanhueza, J (2015) - The expansion of education and the evolution of Intergenerational Mobility, Department of Economics, Universidad de Chile

Morales, C. (2014). ¿Movilidad Intergeneracional de Ingresos: Una perspectiva de Largo Plazo?. Undergraduate thesis. Universidad de Chile

Sanhueza, J. (2011). Mecanismos de transmisión de la movilidad intergeneracional: Chile Cohortes 1962-1984. Postgraduate thesis. Universidad de Chile

Sen, A. (2000). El desarrollo como libertad. Gaceta Ecológica Núm. 55 Pág. 14-20

Solon, G. (1992). Intergenerational Income Mobility in the United States, The American Economic Review. Vol.82, n°3, pp.393-408

Solon, G. (2004). A model of intergenerational mobility variation over time and place. Generational Income Mobility in North America and Europe, 38–47

Torche, F. (2013). How do we characteristically measure and analyze intergenerational mobility?. The Stanford Center on Poverty and Inequality. New York University

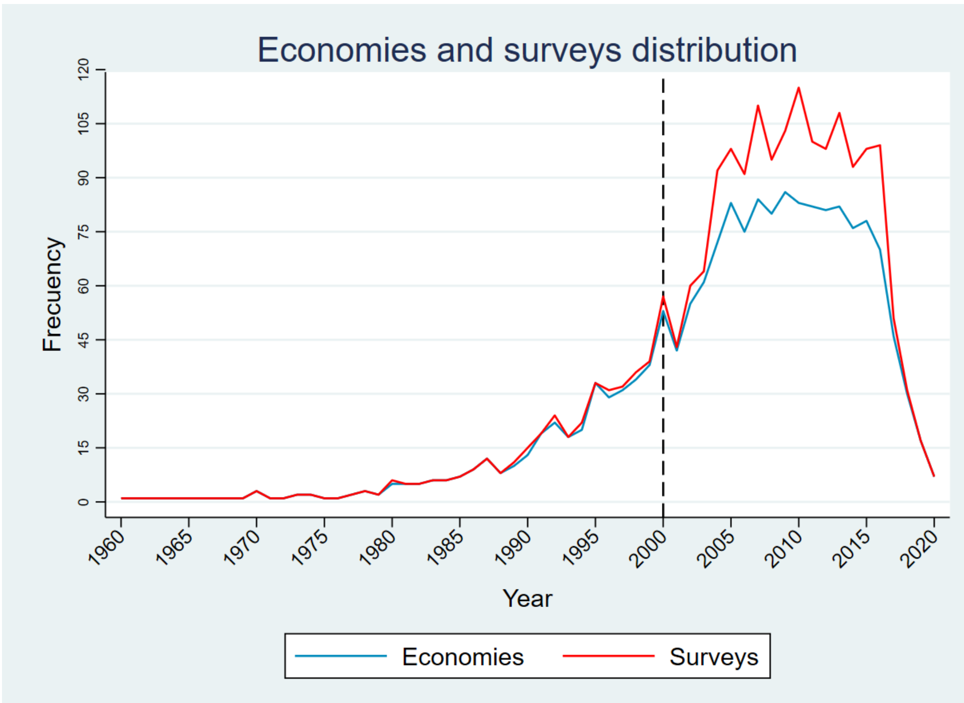
Teorell, J; Sundström A; Holmberg S; Rothstein, B, Alvarado N, Mert, C & Meijers, Y. (2023). The Quality of Government Standard Dataset. University of Gothenburg: The Quality of Government Institute

Van der Weide, R; Narayan, A; Cojocaru, A; Lakner, C; Mahler; S, Gerszon, D; Ramasubbaiah, R; Stefan, T. (2018). Fair Progress?: Economic Mobility Across Generations Around the World. Equity and Development. Washington, DC: World Bank

Van der Weide, R; Lakner, C; Mahler, D; Narayan, A; Ramasubbaiah, R. (2021). Intergenerational Mobility around the World. Policy Research Working Paper; No. 9707. World Bank, Washington, DC

9 Appendix

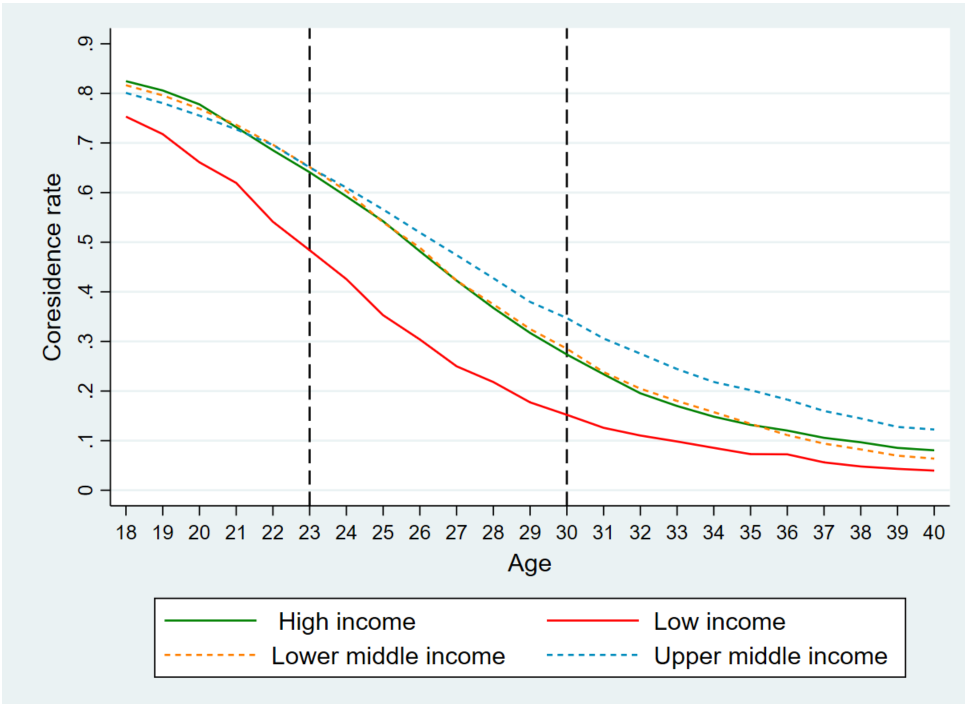
Figure 10: **I2D2 & LIS economy-survey distribution**



Author's Calculations: Based on World Bank and Luxembourg Income Study Data

$$\boxed{\text{Coresidence rate} = \frac{\text{Son's living with their father}}{\text{Total of son's}}}$$

Figure 11: **Coresidence rate over life cycle**



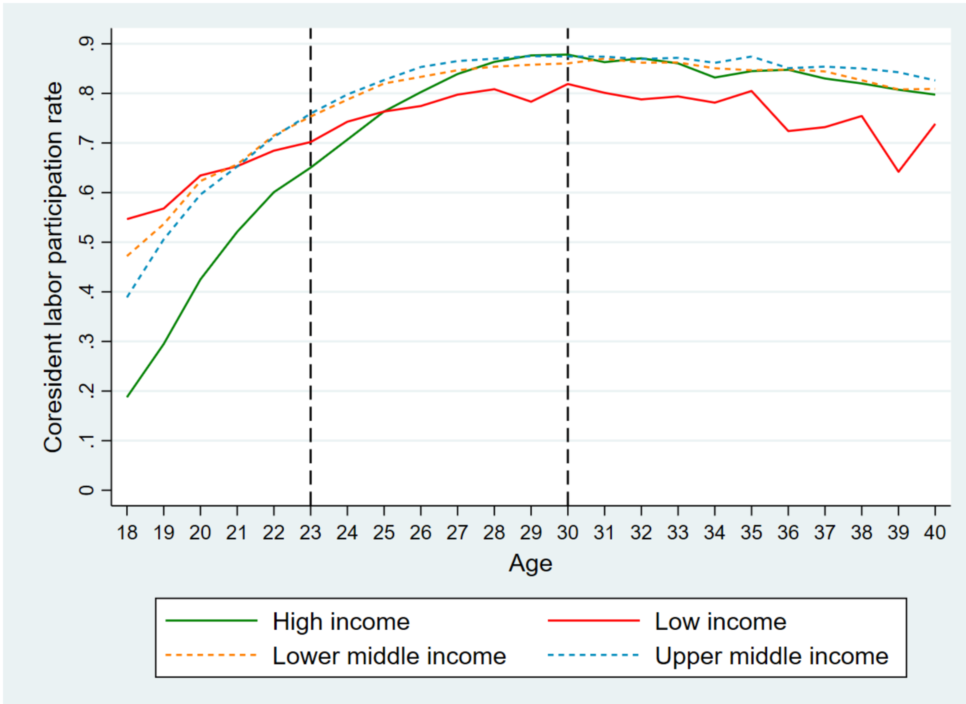
Author's Calculations: Based on World Bank and Luxembourg Income Study Data

Figure 12: **Coresidence rate over time**

Author's Calculations: Based on World Bank and Luxembourg Income Study Data

$$\boxed{\text{Labor participation rate} = \frac{\text{Employed} + \text{unemployed}}{\text{Working age population}}}$$

Figure 13: **Coresident labor participation rate over life cycle**



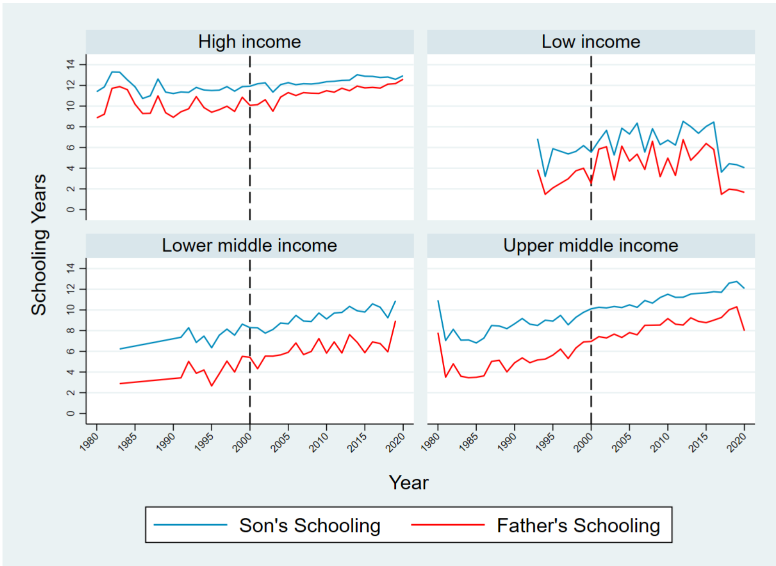
Author's Calculations: Based on World Bank data

Figure 14: **Coresident labor participation rate over time**

Author's Calculations: Based on World Bank data

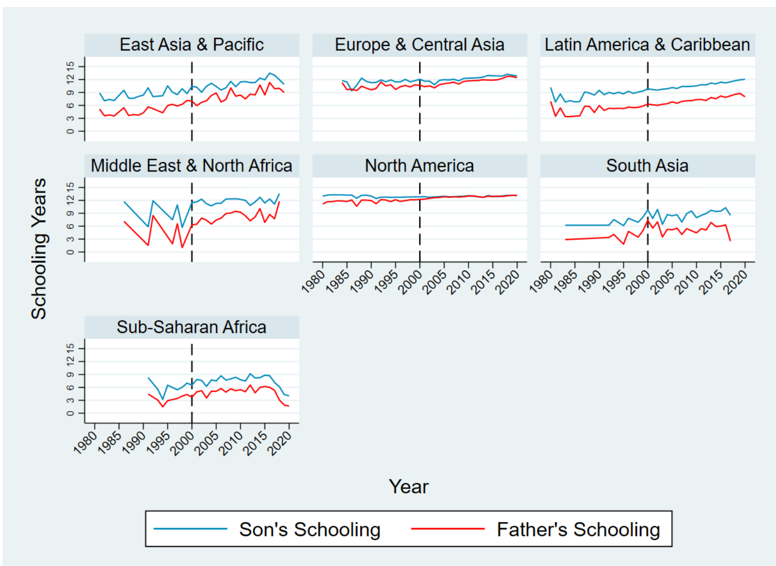
The average labor participation rate has fallen over time in the relavant age bracket, presumably due to the increased opportunities that have opened up in tertiary education.

Figure 15: **Schooling years of coresident sons by income level**



Author's Calculations: Based on World Bank and Luxembourg Income Study Data

Figure 16: **Schooling years of coresident sons by region**



Author's Calculations: Based on World Bank and Luxembourg Income Study Data

9.1 Cohort and cross-section approach, Chilean example

To understand the cross-section approach of this paper, it is useful to show the following matrix that also illustrates the cohort approach used in the World Bank's work.

Figure 17: **Cohort and cross-section approach, Chilean example**

	2017	2015	2013	2011	2009	2006	2003	2000	1998	1996	1994	1992	1990	1987
1994	23	21	19	17	15	12	9	6	4	2	0	-2	-4	-7
1993	24	22	20	18	16	13	10	7	5	3	1	-1	-3	-6
1992	25	23	21	19	17	14	11	8	6	4	2	0	-2	-5
1991	26	24	22	20	18	15	12	9	7	5	3	1	-1	-4
1990	27	25	23	21	19	16	13	10	8	6	4	2	0	-3
1989	28	26	24	22	20	17	14	11	9	7	5	3	1	-2
1988	29	27	25	23	21	18	15	12	10	8	6	4	2	-1
1987	30	28	26	24	22	19	16	13	11	9	7	5	3	0
1986	31	29	27	25	23	20	17	14	12	10	8	6	4	1
1985	32	30	28	26	24	21	18	15	13	11	9	7	5	2
1984	33	31	29	27	25	22	19	16	14	12	10	8	6	3
1983	34	32	30	28	26	23	20	17	15	13	11	9	7	4
1982	35	33	31	29	27	24	21	18	16	14	12	10	8	5
1981	36	34	32	30	28	25	22	19	17	15	13	11	9	6
1980	37	35	33	31	29	26	23	20	18	16	14	12	10	7
1979	38	36	34	32	30	27	24	21	19	17	15	13	11	8
1978	39	37	35	33	31	28	25	22	20	18	16	14	12	9
1977	40	38	36	34	32	29	26	23	21	19	17	15	13	10
1976	41	39	37	35	33	30	27	24	22	20	18	16	14	11
1975	42	40	38	36	34	31	28	25	23	21	19	17	15	12
1974	43	41	39	37	35	32	29	26	24	22	20	18	16	13
1973	44	42	40	38	36	33	30	27	25	23	21	19	17	14
1972	45	43	41	39	37	34	31	28	26	24	22	20	18	15
1971	46	44	42	40	38	35	32	29	27	25	23	21	19	16
1970	47	45	43	41	39	36	33	30	28	26	24	22	20	17
1969	48	46	44	42	40	37	34	31	29	27	25	23	21	18
1968	49	47	45	43	41	38	35	32	30	28	26	24	22	19
1967	50	48	46	44	42	39	36	33	31	29	27	25	23	20
1966	51	49	47	45	43	40	37	34	32	30	28	26	24	21
1965	52	50	48	46	44	41	38	35	33	31	29	27	25	22
1964	53	51	49	47	45	42	39	36	34	32	30	28	26	23
1963	54	52	50	48	46	43	40	37	35	33	31	29	27	24
1962	55	53	51	49	47	44	41	38	36	34	32	30	28	25
1961	56	54	52	50	48	45	42	39	37	35	33	31	29	26
1960	57	55	53	51	49	46	43	40	38	36	34	32	30	27
1959	58	56	54	52	50	47	44	41	39	37	35	33	31	28
1958	59	57	55	53	51	48	45	42	40	38	36	34	32	29
1957	60	58	56	54	52	49	46	43	41	39	37	35	33	30

The table contains information from Chile's Casen surveys between the years 1987 and 2017. The first column corresponds to the cohort of the person, i.e. the year of his birth. The first row corresponds to the year of the survey. The values in the matrix correspond to the age in each cohort-survey year combination.
For example, in the 2017 Casen survey, sons in the 1984 cohort were 23 years old. Since we work with the sons who are between 23 and 30 years old, we can see marked in green color the boxes that show the evolution of the relevant bracket in the different cohort-survey-year combinations. This same exercise can be extrapolated to the rest of the economies.

When we obtain results of relative social mobility with the cross section approach, we are doing a column-by-column analysis. In each column we obtain how much is the average estimator for the relevant age bracket over different cohorts. When we get results with the cohort approach, we are doing a row-by-row analysis. In each row we get how much is the average estimator for the relevant age bracket across different surveys.

The cohort approach is one of the most widely used in the literature, since it allows us to see the evolution of social mobility according to the year of birth of individuals. When the world bank use the coresidents methodology, this approach has the disadvantage of having few parent-child observations in case there are many years between one survey and another. For the Chilean case illustrated in the appendix, this is not so serious since the surveys are 2 or 3 years apart. For other economies with 5 or more years between one survey and another this would represent a problem, since there would be less statistical power in obtaining social mobility coefficients.

9.2 Mathematical

Pearson correlation quantifies the linear association between two variables at the level. It provides a value ranging between -1 and 1, with 1 indicating a perfect positive linear correlation, -1 indicating a perfect negative linear correlation, and 0 indicating no linear correlation.

$$\text{Pearson Correlation } (\rho) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

$$\rho = \frac{Cov(X,Y)}{Sd(X) \cdot Sd(Y)} = \frac{\sigma_{x,y}}{\sigma_x \cdot \sigma_y}$$

Spearman correlation measures the association between two variables in a non-linear manner, considering the rank order of the data rather than the raw values. It also yields a value between -1 and 1, offering a robust measure of association that doesn't rely on the specific distribution of the data.

$$\text{Spearman Correlation} = 1 - \frac{6\sum[R(x_i) - R(y_i)]}{n(n^2 - 1)}$$

The beta coefficient provides the average response of one variable in relation to a one-unit increase in another variable, while holding all other variables in the regression constant. In contrast to Pearson or Spearman correlations, the regression coefficient is not constrained within the range of -1 to 1

$$\text{Regression Coefficient } (\beta) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\beta = \frac{Cov(X,Y)}{Var(X)} = \frac{\sigma_{x,y}}{\sigma_x^2}$$

After dividing by the number of observations in the numerator and denominator in both parameters β and ρ, we get the relationship between the regression coefficient and pearson correlation:

$$\boxed{\beta = \rho\frac{\sigma_y}{\sigma_x}}$$

9.3 Economy Categories

Table 14: **High Income Economies**

Economy	**Code**	**Region**
Aruba	ABW	Latin America & Caribbean
Andorra	AND	Europe & Central Asia
United Arab Emirates	ARE	Middle East & North Africa
Antigua and Barbuda	ATG	Latin America & Caribbean
Australia	AUS	East Asia & Pacific
Austria	AUT	Europe & Central Asia
Belgium	BEL	Europe & Central Asia
Bahrain	BHR	Middle East & North Africa
Bahamas, The	BHS	Latin America & Caribbean
Bermuda	BMU	North America
Barbados	BRB	Latin America & Caribbean
Brunei Darussalam	BRN	East Asia & Pacific
Canada	CAN	North America
Switzerland	CHE	Europe & Central Asia
Channel Islands	CHI	Europe & Central Asia
Chile	CHL	Latin America & Caribbean
Curaçao	CUW	Latin America & Caribbean
Cayman Islands	CYM	Latin America & Caribbean
Cyprus	CYP	Europe & Central Asia
Czech Republic	CZE	Europe & Central Asia
Germany	DEU	Europe & Central Asia
Denmark	DNK	Europe & Central Asia
Spain	ESP	Europe & Central Asia
Estonia	EST	Europe & Central Asia
Finland	FIN	Europe & Central Asia
France	FRA	Europe & Central Asia
Faroe Islands	FRO	Europe & Central Asia
United Kingdom	GBR	Europe & Central Asia
Gibraltar	GIB	Europe & Central Asia
Greece	GRC	Europe & Central Asia
Greenland	GRL	Europe & Central Asia
Guam	GUM	East Asia & Pacific
Hong Kong SAR, China	HKG	East Asia & Pacific
Croatia	HRV	Europe & Central Asia
Hungary	HUN	Europe & Central Asia
Isle of Man	IMN	Europe & Central Asia
Ireland	IRL	Europe & Central Asia
Iceland	ISL	Europe & Central Asia
Israel	ISR	Middle East & North Africa
Italy	ITA	Europe & Central Asia
Japan	JPN	East Asia & Pacific
St. Kitts and Nevis	KNA	Latin America & Caribbean
Korea, Rep.	KOR	East Asia & Pacific
Kuwait	KWT	Middle East & North Africa
Liechtenstein	LIE	Europe & Central Asia
Lithuania	LTU	Europe & Central Asia
Luxembourg	LUX	Europe & Central Asia
Latvia	LVA	Europe & Central Asia
Macao SAR, China	MAC	East Asia & Pacific
St. Martin (French part)	MAF	Latin America & Caribbean
Monaco	MCO	Europe & Central Asia
Malta	MLT	Middle East & North Africa
Northern Mariana Islands	MNP	East Asia & Pacific
New Caledonia	NCL	East Asia & Pacific
Netherlands	NLD	Europe & Central Asia
Norway	NOR	Europe & Central Asia
Nauru	NRU	East Asia & Pacific
New Zealand	NZL	East Asia & Pacific
Oman	OMN	Middle East & North Africa
Panama	PAN	Latin America & Caribbean
Poland	POL	Europe & Central Asia
Puerto Rico	PRI	Latin America & Caribbean
Portugal	PRT	Europe & Central Asia
French Polynesia	PYF	East Asia & Pacific
Qatar	QAT	Middle East & North Africa
Romania	ROU	Europe & Central Asia
Saudi Arabia	SAU	Middle East & North Africa
Singapore	SGP	East Asia & Pacific
San Marino	SMR	Europe & Central Asia
Slovak Republic	SVK	Europe & Central Asia
Slovenia	SVN	Europe & Central Asia
Sweden	SWE	Europe & Central Asia
Sint Maarten (Dutch part)	SXM	Latin America & Caribbean
Seychelles	SYC	Sub-Saharan Africa
Turks and Caicos Islands	TCA	Latin America & Caribbean
Trinidad and Tobago	TTO	Latin America & Caribbean
Taiwan, China	TWN	East Asia & Pacific
Uruguay	URY	Latin America & Caribbean
United States	USA	North America
British Virgin Islands	VGB	Latin America & Caribbean
Virgin Islands (U.S.)	VIR	Latin America & Caribbean

Table 15: **Upper Middle Income Economies**

Economy	**Code**	**Region**
Albania	ALB	Europe & Central Asia
Argentina	ARG	Latin America & Caribbean
Armenia	ARM	Europe & Central Asia
American Samoa	ASM	East Asia & Pacific
Azerbaijan	AZE	Europe & Central Asia
Bulgaria	BGR	Europe & Central Asia
Bosnia and Herzegovina	BIH	Europe & Central Asia
Belarus	BLR	Europe & Central Asia
Belize	BLZ	Latin America & Caribbean
Brazil	BRA	Latin America & Caribbean
Botswana	BWA	Sub-Saharan Africa
China	CHN	East Asia & Pacific
Colombia	COL	Latin America & Caribbean
Costa Rica	CRI	Latin America & Caribbean
Cuba	CUB	Latin America & Caribbean
Dominica	DMA	Latin America & Caribbean
Dominican Republic	DOM	Latin America & Caribbean
Ecuador	ECU	Latin America & Caribbean
Fiji	FJI	East Asia & Pacific
Gabon	GAB	Sub-Saharan Africa
Georgia	GEO	Europe & Central Asia
Equatorial Guinea	GNQ	Sub-Saharan Africa
Grenada	GRD	Latin America & Caribbean
Guatemala	GTM	Latin America & Caribbean
Guyana	GUY	Latin America & Caribbean
Iraq	IRQ	Middle East & North Africa
Jamaica	JAM	Latin America & Caribbean
Jordan	JOR	Middle East & North Africa
Kazakhstan	KAZ	Europe & Central Asia
Libya	LBY	Middle East & North Africa
St. Lucia	LCA	Latin America & Caribbean
Moldova	MDA	Europe & Central Asia
Maldives	MDV	South Asia
Mexico	MEX	Latin America & Caribbean
Marshall Islands	MHL	East Asia & Pacific
North Macedonia	MKD	Europe & Central Asia
Montenegro	MNE	Europe & Central Asia
Mauritius	MUS	Sub-Saharan Africa
Malaysia	MYS	East Asia & Pacific
Namibia	NAM	Sub-Saharan Africa
Peru	PER	Latin America & Caribbean
Palau	PLW	East Asia & Pacific
Paraguay	PRY	Latin America & Caribbean
Russian Federation	RUS	Europe & Central Asia
Serbia	SRB	Europe & Central Asia
Suriname	SUR	Latin America & Caribbean
Thailand	THA	East Asia & Pacific
Turkmenistan	TKM	Europe & Central Asia
Tonga	TON	East Asia & Pacific
Türkiye	TUR	Europe & Central Asia
Tuvalu	TUV	East Asia & Pacific
St. Vincent and the Grenadines	VCT	Latin America & Caribbean
Kosovo	XKX	Europe & Central Asia
South Africa	ZAF	Sub-Saharan Africa

Table 16: **Lower Middle Income Economies**

Economy	**Code**	**Region**
Angola	AGO	Sub-Saharan Africa
Benin	BEN	Sub-Saharan Africa
Bangladesh	BGD	South Asia
Bolivia	BOL	Latin America & Caribbean
Bhutan	BTN	South Asia
Côte d'Ivoire	CIV	Sub-Saharan Africa
Cameroon	CMR	Sub-Saharan Africa
Congo, Rep.	COG	Sub-Saharan Africa
Comoros	COM	Sub-Saharan Africa
Cabo Verde	CPV	Sub-Saharan Africa
Djibouti	DJI	Middle East & North Africa
Algeria	DZA	Middle East & North Africa
Egypt, Arab Rep.	EGY	Middle East & North Africa
Micronesia, Fed. Sts.	FSM	East Asia & Pacific
Ghana	GHA	Sub-Saharan Africa
Honduras	HND	Latin America & Caribbean
Haiti	HTI	Latin America & Caribbean
Indonesia	IDN	East Asia & Pacific
India	IND	South Asia
Iran, Islamic Rep.	IRN	Middle East & North Africa
Kenya	KEN	Sub-Saharan Africa
Kyrgyz Republic	KGZ	Europe & Central Asia
Cambodia	KHM	East Asia & Pacific
Kiribati	KIR	East Asia & Pacific
Lao PDR	LAO	East Asia & Pacific
Lebanon	LBN	Middle East & North Africa
Sri Lanka	LKA	South Asia
Lesotho	LSO	Sub-Saharan Africa
Morocco	MAR	Middle East & North Africa
Myanmar	MMR	East Asia & Pacific
Mongolia	MNG	East Asia & Pacific
Mauritania	MRT	Sub-Saharan Africa
Nigeria	NGA	Sub-Saharan Africa
Nicaragua	NIC	Latin America & Caribbean
Nepal	NPL	South Asia
Pakistan	PAK	South Asia
Philippines	PHL	East Asia & Pacific
Papua New Guinea	PNG	East Asia & Pacific
West Bank and Gaza	PSE	Middle East & North Africa
Senegal	SEN	Sub-Saharan Africa
Solomon Islands	SLB	East Asia & Pacific
El Salvador	SLV	Latin America & Caribbean
São Tomé and Príncipe	STP	Sub-Saharan Africa
Eswatini	SWZ	Sub-Saharan Africa
Tajikistan	TJK	Europe & Central Asia
Timor-Leste	TLS	East Asia & Pacific
Tunisia	TUN	Middle East & North Africa
Tanzania	TZA	Sub-Saharan Africa
Ukraine	UKR	Europe & Central Asia
Uzbekistan	UZB	Europe & Central Asia
Vietnam	VNM	East Asia & Pacific
Vanuatu	VUT	East Asia & Pacific
Samoa	WSM	East Asia & Pacific
Zimbabwe	ZWE	Sub-Saharan Africa

Table 17: **Low Income Economies**

Economy	**Code**	**Region**
Afghanistan	AFG	South Asia
Burundi	BDI	Sub-Saharan Africa
Burkina Faso	BFA	Sub-Saharan Africa
Central African Republic	CAF	Sub-Saharan Africa
Congo, Dem. Rep.	COD	Sub-Saharan Africa
Eritrea	ERI	Sub-Saharan Africa
Ethiopia	ETH	Sub-Saharan Africa
Guinea	GIN	Sub-Saharan Africa
Gambia, The	GMB	Sub-Saharan Africa
Guinea-Bissau	GNB	Sub-Saharan Africa
Liberia	LBR	Sub-Saharan Africa
Madagascar	MDG	Sub-Saharan Africa
Mali	MLI	Sub-Saharan Africa
Mozambique	MOZ	Sub-Saharan Africa
Malawi	MWI	Sub-Saharan Africa
Niger	NER	Sub-Saharan Africa
Korea, Dem. People's Rep.	PRK	East Asia & Pacific
Rwanda	RWA	Sub-Saharan Africa
Sudan	SDN	Sub-Saharan Africa
Sierra Leone	SLE	Sub-Saharan Africa
Somalia	SOM	Sub-Saharan Africa
South Sudan	SSD	Sub-Saharan Africa
Syrian Arab Republic	SYR	Middle East & North Africa
Chad	TCD	Sub-Saharan Africa
Togo	TGO	Sub-Saharan Africa
Uganda	UGA	Sub-Saharan Africa
Yemen, Rep.	YEM	Middle East & North Africa
Zambia	ZMB	Sub-Saharan Africa