RELATIVE INTERGENERATIONAL MOBILITY, A GLOBAL REVIEW

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

Co-Guide: Claudio Montenegro[‡]

May 17, 2023



Facultad de Economía y Negocios, Universidad de Chile

Abstract

Keywords: Social Movility, Elasticity, Intergenerational Inequality **JEL**: Código1 Código1 Código3 Código4 Código5

^{*}asanhuezac@fen.uchile.cl

 $^{^{\}dagger} jnunez@fen.uchile.cl$

 $^{^{\}ddagger} cmontene gro@worldbank.org$



1 Motivation & Literature Review

We can summarize the reasons for estimating relative intergenerational mobility in two types:

- 1) Normative reasons: There are economic/social differences due to factors that individuals can control, such as their effort, responsibility, choices, etc. But there are other exogenous factors that individuals do not control and that limit their opportunities to develop their life projects, such as, their parents' income or the money needed for education and health of quality. There is broad consensus that public policy must generate the necessary conditions for the existence of 'equality of opportunities'. This concept can be thought as the possibility that individuals have to realize their life projects. A society in which there are equality of opportunities is one in which circumstances do not determine the results. Therefore, the results however unequal they may be, would come from a fair process.
- 2) Economic Reasons: If there is equality of opportunity, and talents are spread throughout society, we should expect the relative position of individuals in the income/education distribution to change across generations. In reality, since there is not full equality of opportunity, much talent and ability is wasted as they are unable to realize their full potential, leading to an inefficient allocation of resources in the economy. If people feel that there is not a strong degree of equality of opportunity, this can impact on the effort given by people in the lower tail of the distribution as they feel that their efforts will be wasted and that their children's generation will end up relatively close to where they ended up. The same logic can be applied to the upper tail of the distribution.

There are many papers that have studied this issue, in different time windows, with different methodologies and data. To date, there are only two publications, conducted by the World Bank, that have studied intergenerational mobility with countries of all income levels, and from all regions of the world.

The paper seeks to contrast with the two works developed by Van der Weite et.al (2018 & 2021), on social mobility in the world. The difference is that we have more data available and the results will be presented using the same methodology without the need to mix our own results with those of other authors. Additionally we will perform an analysis using rank-rank measures, which has been shown by Chetty et.al (2014) to be a more robust estimator. We will also perform a cross section analysis using the year in which the survey was conducted, together with a cohort analysis. This is summarized and detailed in the following table.



	Van der waide et.al		This Work	
	2021 (Education) 2018 (Education & Income)		Education	Income
Surveys	< 500	<650	978	379
Methodology	IV & Coresidents	IV & Coresidents	Coresidents	Coresidents
Approach	Cohort	Cohort	Cohort + Cross Section	Cross Section
Estimations	Level	Level	Level & Range	Level & Range
Mix of results	Yes	Yes	No	No

*It is worth mentioning that for the income dimension we will concentrate primarily on the Latin American and Caribbean region, given the low number of surveys that are useful in the other regions of the world. We are currently working to add more surveys in this dimension.

Surveys: This would be the first study to have such a large coverage of surveys for countries around the world. We expect the results to be in line with existing literature for regional level work (LAC), and for the benchmark work we have from Van der Waide et.al. It should be noted that these surveys were chosen after applying a series of filters, which are detailed in later sections of this paper.

Methodology: The works of Van der Waide et al. first estimates relative intergenerational mobility using "retrospective questions" to the son to obtain information from the father. For example, it is possible to estimate mobility in education if the son reports how many years of education his father has, as this is fairly straightforward to achieve. On the other hand, income is difficult to report accurately, so if the son does not have this information from his father, it would be possible to estimate how much it would be, using instrumental variables based on the information provided by the son. In the case of not having this type of data, the authors use the coresidents methodology, which allows obtaining information from parents and children living in the same household. In our work we use only the latter methodology, which allows us to obtain more comparable results than if different ways of estimating the same thing were used.

Approach: The work of Van der Waide et al uses a cohort approach. This allows us to know how relative mobility has evolved over different generations (using the year of birth of individuals). For the education dimension we do the same, but in addition we add a cross section approach, which allows us to see how relative mobility has evolved over different moments in time (each moment takes into account different generations). We do the same for the income dimension, which allows us to generate a richer analysis of the phenomenon under study.

Estimates: Chetty et.al 2014 studying relative social mobility in the United States used rank measures noting that they are a more robust way of estimating relationships between variables. In this paper we estimate relative mobility in the traditional way by level, but



we also add rank estimates using spearman correlation. This would allow us to generate robustness analyses of the results found.

Mixing of results: The work of Van der Waide et.al shows results from different sources, the main one being Global Database on Intergenerational Mobility (GDIM) and the "Equalchances" database. The problem is that they do not document in detail the functional form of the regressions, or the filters of observations they used. It is possible that when presenting the results, they do so by mixing these methodologies, which is not ideal. In this paper we will estimate the results ourselves, using the same functional form and filters in each and every survey, without mixing results from other studies.

2 Research Questions

This paper seeks to answer two questions

1) How has intergenerational persistence in education and income behaved historically by cohort and survey years?

With this we seek to get a better idea of whether persistence is concentrated in countries with a certain level of income or in certain regions of the world, analyzing the trend over time.

2) How does intergenerational persistence correlate globally, with key variables from different models that attempt to explain it?

With this question we do not seek to present causal evidence as this requires a deeper analysis that is beyond the scope of this study. What we seek is to shed some light on whether the data has any relationship with variables that attempt to explain intergenerational persistence with countries around the world. The hypothesized relationship between intergenerational persistence and the variables is summarized in the following table.

Variables	Model/Study	Hypothesis
Inequality (Gini Coefficient)	Corak (2013)	Positive
Public Spent	Mayer & Lopoo (2007)	Negative
Economic Liberalization	Xie & Chang (2022)	Negative
Economic Growth	Yishay & Moav (1999)	Negative
Violence	Manduca & Sampson (2019)	Positive
Schooling Return	Solon (2004)	Positive



3 Data

The ideal data to generate this type of work are long panel data that track parents and children in their income history and years of schooling. These types of data are extremely expensive to collect, and if they exist at all, they are typically concentrated in developed world countries.

We have three data sources:

- 1) International Income Distribution Database (I2D2): In an effort to generate international evidence on different matters, the World Bank generated standardized databases from household and labor market surveys in countries around the world for over a decade. The databases covers economies from developed and developing regions. There is no censoring of any kind in the survey selection. Not all of economy/year/survey points are included in our analysis because some surveys lack key variables. If a survey was available and has all the needed variables, then it was included. A fundamental requirement for a survey to be part of this source is that it must be nationally representative. It should be noted that for a given economy/year, there could be more than one estimate. This is because for that economy/year, there is more than one survey available, or that the same survey is available more frequently than annually. It should be mentioned that many of these databases are not open to the public. This is why in a joint work with Claudio Montenegro we generated the necessary estimates for this research, maintaining the confidentiality of the information at all times.
- 2) Luxemburg Income Study (LIS): Surveys from this source are nationally representative. More up-to-date income data is available, especially for rich countries, with more developing countries recently included. This data source will be used as a secondary source if any country-year combination is not available in I2D2. LIS surveys will replace I2D2 surveys if income data are better. These data are more public, can be requested for transparency, but access is very limited. The opening is given by entering the server and putting the code to get the results that arrive via email.
- 3) Quality of Governance (Gothenburg University): In order to find correlational evidence at the international level between our estimates and macroeconomic variables of the countries, we will use this database. The observations are composed of a country-year combination.
- *) World Bank Country and Lending Groups: To classify countries by income level and geographic region, we use a World Bank classification updated to the year 2022. The income classification has 4 categories; High income, Low income, Upper middle and Lower middle (the only country that cannot be classified was Venezuela, since the data provided



is not reliable for the World Bank). The region classification has 7 categories; East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, North America, South Asia and Sub-Saharan Africa.

4 Empirical Approach

To measure social mobility across countries, we will use a cross-section approach (many cohorts in each survey) and a cohort approach (many surveys in each generation). The second way is the most used and allows us to see the situation of the older generations with respect to the newer ones.

To calculate the social mobility of countries, the following functional form is typically used

$$y_s = \alpha + \beta y_f + \gamma X + \epsilon_i$$

Where:

 $y_s = \ln(\text{Children's income})$ or Children's Education

 $y_f = \ln(\text{Father's income})$ or Father's Education

X=Control variables (following the existing literature, we used age and age squared as control variables to control for the life cycle of the children and the fathers)

The closer β is to 0, the greater the social mobility

4.1 Observation and survey filters

	Education	Income
Coresidents Son's Age	23-30	23-30
Semanal Working Hours	-	≥ 22
Father's Age	-	< 65
Exhibition top 0.5% of distribution	-	Yes
Control Variables	-	Son and father age
Surveys ≥ year 2000	Cross-Section Approach	Cross-Section Approach
N° Min Surveys per Cohort	2	2
Surveys: N° Obs Min & Max Median Error	200 y 0.2	200 y 0.2

1) Coresidents Son's Age: We believe that 23 years is a reasonable cut-off age for the coresident approach, given that around this age people finish their tertiary education cycle and start entering the labor market. On the other hand, since the surveys are conducted



3 to 5 years per country, we decided to set the upper cutoff age at 30 years in order to see the effect that exists in a cohort at different ages and to limit the possible coresidence bias. If we analyze the coresidency and labor participation rates of people in this age group, it is confirmed that this is an appropriate bracket.

- 2) Semanal Working Hours: In the income dimension, we seek to capture what is closest to people's permanent income, so we filter out observations that have 22 or more hours of work per week.
- 3) Father's Age: In the income dimension we seek to avoid a problem of selectivity of people leaving the labor market to retire. This was initially intended for men in Chile, but analyzing the retirement age for men in other countries, we consider it to be a good cut-off age.
- 4) Exhusion top 0.5% of distribution: For every survey, the top 0.5% of the income sample was eliminated. This is a World Bank practice to avoid possible biases due to wage outliers.
- 5) Control Variables: In the literature on intergenerational mobility there is the concept of "life-cycle bias". This bias occurs when calculating the intergenerational persistence of income of the father at a different age than that of the son. The solution given to solve this problem is to include as a control variable the age of the son and the age of the son squared. Some papers additionally include the father's age. In this paper we will use the son and father age as control variables. In the educational dimension, this bias does not occur because once people obtain an educational level they do not drop out of it, so estimating the intergenerational persistence of parents and children at different ages should not present a problem.
- 6) Year 2000: We stayed with these surveys from the year 2000 for 3 reasons. First because there is a greater coverage of countries during those years. Second because after the fall of the Berlin Wall and the Soviet Union many countries were created and destroyed, in order to give them time to organize themselves and to collect quality data, we think that the year 2000 is a good cut-off year. Finally, it is evident that the coresidency rate has been increasing over time for the relevant age range, so taking the latest years available is something that is useful in our research to reduce coresidency bias.
- 7) Approach: We only take into account those cohorts that have a minimum of 2 surveys as observations in order to distinguish the age effect from the cohort effect.
- 8) Surveys: We leave out surveys with relative mobility estimates less than 0.1 and greater than 0.8 since these extreme magnitudes have not been documented in the literature. Next, we excluded from the survey sample those surveys with a number of observations lower than 200 for the calculation of the statistics of interest. We also excluded those with a



deviation from the median of their country greater than 0.2 percentage points. A graph illustrating this can be found in the appendix.

To correctly capture the importance of the father's income in the children's income, we annualize the salaries to isolate the effect that the different frequencies with which each one receives it could have.

4.2 Cohort and Cross-Section approach

To understand the approaches, one can rely on the sample table in the appendix section. 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.

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 cohorts marked in red were eliminated for each economy, since in those cohorts we have less than two survey observations, which makes it impossible to distinguish the cohort effect from the age effect.

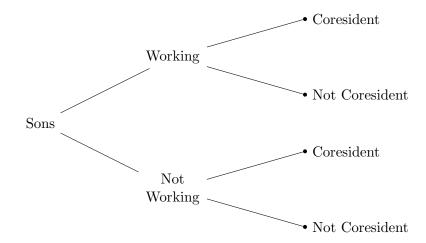
4.3 Selection Bias

There are two sources of selection bias in this work. The first is whether the person participates in the labor market, the second is whether the person coresides with his or her parents. 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 female labor participation rate is lower than males.

In this research we will only concentrate on those father-son combinations in which both works (have incomes). Since these father-son combinations are only observed with sons who live with their father, it is possible that there is a selection bias with respect to those sons who do not live with their fathers. This can be summarized in the following two



figures.



		Son Wage	Father Wage
Working	Coresident	✓	✓
WOLKING	Not Coresident	✓	×
Not Working	Coresident	×	✓
NOT WOLKING	Not Coresident	×	×

To take care of the second selection bias we will estimate by heckman maximum likelihood only for the income dimension since it has been documented that the coresident bias for education is marginal (Emran et.al, 2016 & Emran et.al, 2018).

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 $v_s^* > 0$, so we have the following

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

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

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

The estimates of the participation equation are obtained from the following regression model:

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



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

Age: As people get older, the rate of coresidence decreases (see appendix), so the probability of coresidence decreases as well.

Marital status: We define this variable as a dummy that is 1 if the person is married, with a partner, divorced or widowed, and 0 otherwise. It is expected that the probability of coresidence with parents should decrease if the person is in any of these categories.

Urban: We define this variable as a dummy that is 1 if the person lives in an urban area and 0 if not. The probability of coresidence is expected to decrease if the person lives in a city.

Crowd: This is a variable that captures the number of children living in the person's household. It is expected that in a household with more children, the probability of coresidency will decrease.

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.

The estimates of the income dimension are obtained from the following regression model:

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



5 Descriptive Statistics

Table 1: Education Economies

Economies	High income	Low income	Lower middle income	Upper middle income	Total
East Asia & Pacific	1	0	11	3	15
Europe & Central Asia	26	0	3	14	43
Latin America & Caribbean	5	0	5	11	21
Middle East & North Africa	2	2	7	2	13
North America	1	0	0	0	1
South Asia	0	1	6	1	8
Sub-Saharan Africa	0	22	15	5	42
Total	35	25	47	36	143

Self Made: World Bank

Table 2: Mean Statistics by Income

Income	Coresidence Rate	Labor Participation Rate	Father Schooling	Son Schooling
High income	.37	.7	11	12
Low income	.28	.74	5.1	7.5
Lower middle inc	.38	.78	6.2	9.3
Upper middle inc	.42	.78	8.4	11
Total	.39	.75	8.6	11

Self Made: World Bank

Table 3: Mean Statistics by Region

Region	Coresidence Rate	Labor Participation Rate	Father Schooling	Son Schooling
East Asia & Pacific	.38	.72	7.8	11
Europe & Central Asia	.43	.71	11	12
Latin America & Caribbean	.33	.83	7.1	10
Middle East & North Africa	.46	.76	7.3	11
North America	.23	.76	13	13
South Asia	.46	.85	5.2	8.5
Sub-Saharan Africa	.29	.72	5.4	8.1
Total	.39	.75	8.5	11

Self Made: World Bank



6 Results

6.1 Education

Table 4: Mean Estimations by Income

Income	Beta	Pearson	Spearman
High income	.31	.34	.35
Low income	.45	.44	.45
Lower middle inc	.45	.46	.47
Upper middle inc	.41	.45	.47
Total	.39	.41	.43

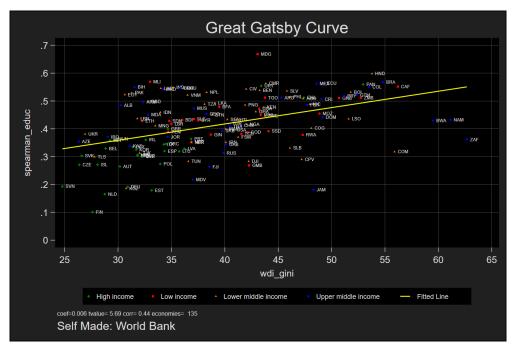
Self Made: World Bank

Table 5: Mean Estimations by Region

Region	\mathbf{Beta}	Pearson	Spearman
East Asia & Pacific	.43	.38	.45
Europe & Central Asia	.33	.35	.35
Latin America & Caribbean	.46	.53	.53
Middle East & North Africa	.31	.37	.37
North America	.23	.35	.4
South Asia	.46	.47	.5
Sub-Saharan Africa	.43	.44	.44
Total	.4	.41	.43

 $Self\ Made:\ World\ Bank$





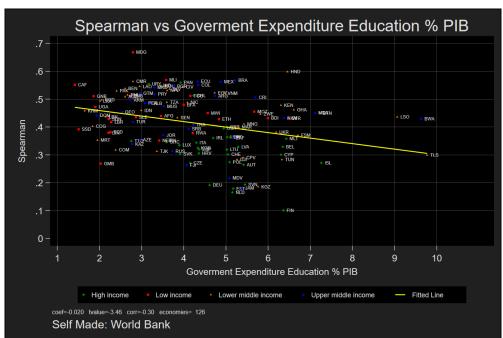




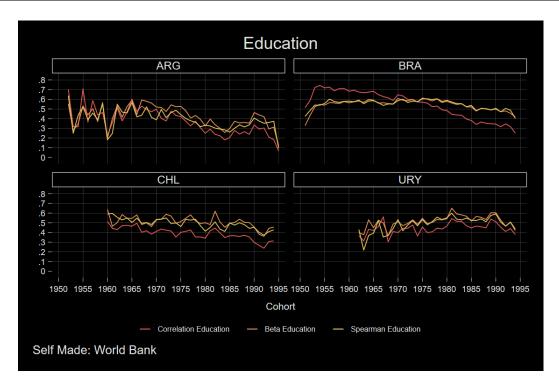
Table 6: Education Relative Mobility

	(1)	(2)	(3)
	Correlation	Beta Regression	Spearman
ln(Per Capita PPP)	-0.0265**	-0.0478***	-0.0146*
	(-2.69)	(-5.51)	(-2.16)
ln(Gini)	0.270***	0.168***	0.277***
	(10.60)	(6.26)	(12.77)
ln(Govt Education Expenditure % PIB)	-0.0302	-0.0551**	-0.0553***
,	(-1.79)	(-3.22)	(-4.20)
ln(Economy Liberalization)	-0.00754	-0.0119	-0.0799***
,	(-0.22)	(-0.56)	(-3.69)
Constant	-0.253	0.354^{*}	-0.0470
	(-1.53)	(2.12)	(-0.34)
R2_ajusted	0.301	0.361	0.499
Akaike	-936.1	-940.9	-1107.4
Bayes	-913.3	-919.3	-1085.7
Observations	702	555	568

t statistics in parentheses

^{*} p < 0.05, ** p < 0.01, *** p < 0.001







6.2 Income

Table 7: Mean Income Estimations

Economy	Heckman	Regresion	Elasticity	Pearson	Spearman
ARG	.3	.32	.31	.39	.42
BOL	.22	.24	.24	.31	.3
BRA	.37	.38	.38	.53	.54
CHL	.38	.39	.38	.52	.49
COL	.37	.39	.38	.46	.43
CRI	.29	.3	.29	.39	.4
DOM	.33	.31	.32	.41	.32
ECU	.24	.27	.26	.37	.4
GTM	.21	.24	.24	.35	.32
HND	.32	.33	.33	.46	.5
MEX	.23	.23	.24	.34	.35
NIC	.23	.24	.25	.36	.33
PAN	.35	.36	.35	.48	.45
PER	.21	.22	.21	.28	.24
PRY	.23	.23	.23	.37	.38
SLV	.18	.19	.2	.32	.32
URY	.28	.29	.29	.39	.41
Total	.28	.29	.29	.39	.39

 $Self\ Made:\ World\ Bank$

^{&#}x27;Regresion' corresponds to elasticity without age controls.



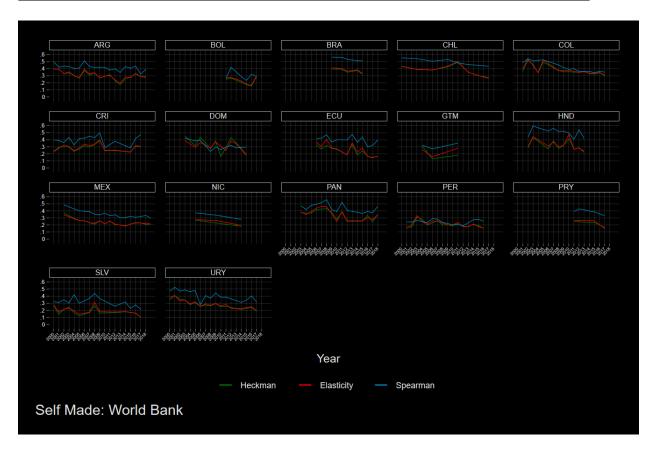




Table 8: LAC Income Relative Mobility

	(1)	(2)	(3)
	Heckman	Elasticity	Spearman
ln(Per Capita PPP)	0.134***	0.129***	0.132***
	(8.34)	(8.05)	(6.50)
$\ln(\mathrm{Gini})$	0.602***	0.592***	0.462***
	(11.09)	(11.14)	(7.37)
ln(Govt Education Expenditure % PIB)	-0.0836***	-0.0755***	-0.0193
m(co.v Zadodiož Ziponatvažo /(1 12)	(-3.59)	(-3.43)	(-0.53)
ln(Economic Globalization)	-0.0477	-0.0582	-0.0630
(,	(-1.56)	(-1.88)	(-1.91)
Constant	-3.008***	-2.885***	-2.379***
	(-10.48)	(-9.88)	(-6.91)
R2_ajusted	0.504	0.494	0.377
Akaike	-331.9	-330.1	-282.6
Bayes	-318.0	-316.2	-268.6
Observations	120	120	121

t statistics in parentheses

7 Conclusions

8 Bibliography

- Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. Journal of Economic Perspectives, 27(3), 79–102.
- Emran, M. Shahe; Greene, William; Shilpi, Forhad. 2016. When Measure Matters: Coresidency, Truncation Bias, and Intergenerational Mobility in Developing Countries. Policy Research Working Paper; No. 7608. © World Bank, Washington, DC
- Emran, M. Shahe; Shilpi, Forhad. 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

^{*} p < 0.05, ** p < 0.01, *** p < 0.001

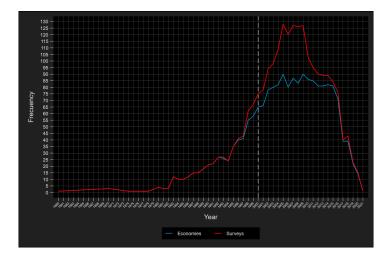


- Nuñez J & Risco C "Movilidad intergeneracional del ingreso en un país en desarrollo. El caso de Chile", Departamento de Economía Uchile, 2004
- Nuñez Sanhueza (2015) The expansion of education and the evolution of Intergenerational Mobility, Chile
- Mayer, S. E., Lopoo, L. M. (2008). Government spending and intergenerational mobility. Journal of Public Economics, 92(1-2), 139–158

Raj Chetty, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States, The Quarterly Journal of Economics, Volume 129, Issue 4, November 2014, Pages 1553–1623

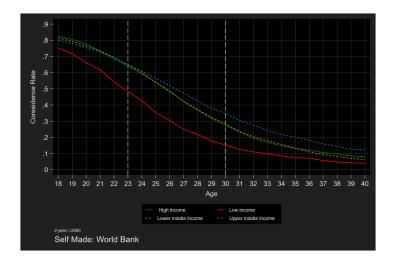
- Van der Weide, Roy; Narayan, Ambar; Cojocaru, Alexandru; Lakner, Christoph; Redaelli, Silvia; Mahler, Daniel Gerszon; Ramasubbaiah, Rakesh Gupta N.; Thewissen, Stefan. 2018. Fair Progress?: Economic Mobility Across Generations Around the World. Equity and Development. © Washington, DC: World Bank
- -Van der Weide, Roy; Lakner, Christoph; Mahler, Daniel Gerszon; Narayan, Ambar; Ramasubbaiah, Rakesh. 2021. Intergenerational Mobility around the World. Policy Research Working Paper; No. 9707. © World Bank, Washington, DC
- Xie, G. E., Zhang, X. M. (2022). The Effect of Trade Liberalization on Intergenerational Income Mobility. Open Journal of Business and Management, 10, 1784-1810.

9 Appendix





 $Coresidence Rate = \frac{Son's living with their father}{Total of Son's}$



 $\label{eq:Labor Force Participation Rate} \textbf{Employed+Unemployed} \\ \frac{\textbf{Employed+Unemployed}}{\textbf{Working age population}}$

