Relative Intergenerational Mobility, a global review

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

Co-Guide: Claudio Montenegro[‡]

April 14, 2023

Facultad de Economía y Negocios, Universidad de Chile

Abstract

---

[*]asanhuezac@fen.uchile.cl

[†]jnunez@fen.uchile.cl

[‡]cmontenegro@worldbank.org

1 Motivation & Literature Review

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

1) **Normative reasons**: There is broad consensus that public policy must generate the necessary conditions for the existence of a concept called *'equality of opportunities'*. This concept can be thought of as the possibility that individuals have to realize their life projects. There are economic/social differences due to factors that individuals choose, such as effort, responsibility, their 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 the money needed for education or violence in the neighborhood where they live. A society in which there is equality of opportunities is one in which opportunities (public policies) mean that circumstances do not determine the results. Thus, the results would come from a "fair" process.

2) **Economic Reasons**: Under the reasonable assumption that talents are spread across different strata of society and that there is equality of opportunity, the relative position of individuals in the income/education distribution should change over 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.

There are many papers that have studied this issue, in different time windows, with different methodologies and data. To date, there are only two papers, 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 complement the two studies by Van der Weite et.al (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 percentile correlation coefficients, 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 can be summarized in the following table.

	Currently	**This work**
Surveys	500 - 650	2308 + 728
Methodology	IV & Coresidents	Coresidents Only
Approach	Cohort	Cohort + Cross-Section
Estimations	Level	Level + Range
Mix of result	Yes	No

2   Research Question

1) How has intergenerational persistence in education and income fared globally by cohort and survey years?

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

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.

The data we will use are household surveys that contain information on the income and years of schooling of the parents and children residing with them.

We have four data sources:

1) International Income Distribution Database (I2D2): In an effort to generate international evidence on different matters, the World Bank generated databases from household and labor market surveys in countries around the world. The database covers economies from developed and developing regions. There is no censoring of any kind in the survey selection. If a survey was available and has all the needed variables, then it was included. For all the economies in this study, we have at least one point in time and in many several cases. An enormous effort was undertaken in standardizing the variable definitions across economies and time periods.

Not all of economy/year/survey points are included in our analysis because some surveys lack key variables. 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.

2) Luxemburg Income Study: All surveys are nationally representative. Has more up-to-date data on income, especially for rich countries. Has recently included more developing countries. Will be used as second order data if not available in I2D2. New economies Japan, Israel, Germany, etc. 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 mail.

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.

4) World Bank Country Classification: To classify countries by income level and geographic region, we use another World Bank database 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   Descriptive Statistics

5   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(Children's income) or Children's Education
y_f=ln(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)

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

5.1 Observation and survey filters

	Education	Income
Coresidents Son's Age	23-30	23-30
Semanal Working Hours	-	≥ 22
Father's Age	-	< 65
Exlusion top 0.5% of distribution	-	Yes
Control Variables	-	Son Age
Surveys $\geq$ year 2000	Cross-Section Approach	Cross-Section Approach
N° Min Surveys per Cohort	2	2
Surveys: N° Obs Min & Max Median Error	100 y 0.2	100 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) Exlusion 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 this for the robustness analysis. 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 intergen-

erational persistence of parents and children at different ages should not present a problem.

6) Surveys ≥ year 2000: We stayed with these surveys from the year 2000 for 3 reasons. Because there is a greater coverage of countries during those years. 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 excluded from the survey sample those surveys with a number of observations lower than 100 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.

5.2   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.

5.3   Selection Bias
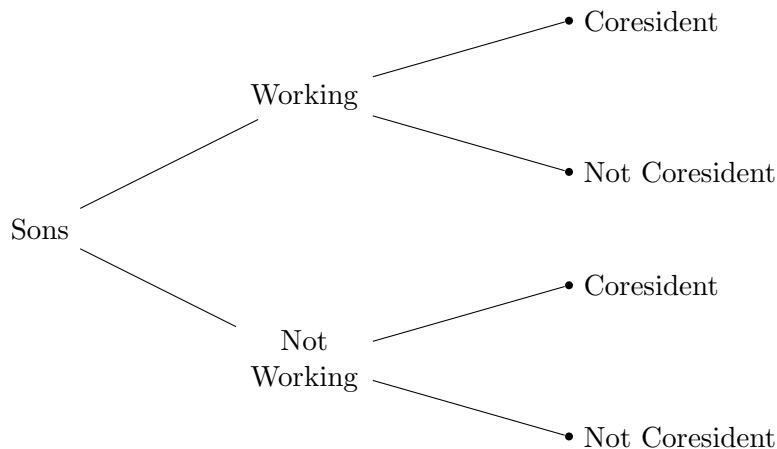
		Son wage	Father wage	Instruments
Working	Coresident	✓	✓	✓
	Not Coresident	✓	✕	✓
Not	Coresident	✕	✓	✓
Working	Not Coresident	✕	✕	✓

Variables	Controls	Instruments
Father wage	✓	✕
Son age	✓	✓
Son age^2	✓	✓
Civil status	✕	✓
Urban	✕	✓

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

$$y_s = \alpha + \beta y_f + \gamma age + \delta age^2 + \epsilon_s$$

$$Cores = \alpha + \beta ecivil + \gamma urban + \delta age + \rho age^2 + v$$

in this research we will only concentrate on those father-son combinations in which the son works. 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. To address this problem, we will use the Heckman Maximum Likelihood strategy for the income estimates only. In the educational dimension it has been shown that the coresident bias is insignificant.

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

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

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

Where Z_s corresponds to a vector of characteristics of the son that influence the decision to live or not with the father, including the following variables: marital status, urban or rural area and age

Coresidence rate

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

Labor participation and Employment rate

$$\boxed{\text{Labor Force Participation Rate} = \frac{\text{Employed+Unemployed}}{\text{Working age population}}}$$

$$\boxed{\text{Ocupation Rate} = \frac{\text{Employed}}{\text{Working age population}}}$$

6   Results

Coresidence rate

Labor participation and Employment rate

Social Movility

in line with the existing literature, it is observed that countries with higher income have higher income and education mobility than those with lower incomes

The literature has shown a negative relationship between income inequality, as measured by the gini index, and social mobility. This relationship is called the Great Gatsby Curve.

7   Conclusions

8   Bibliography

- Van der Weide R, Narayan A, et.al (2021). Fair Progress? : Economic Mobility Across Generations Around the World
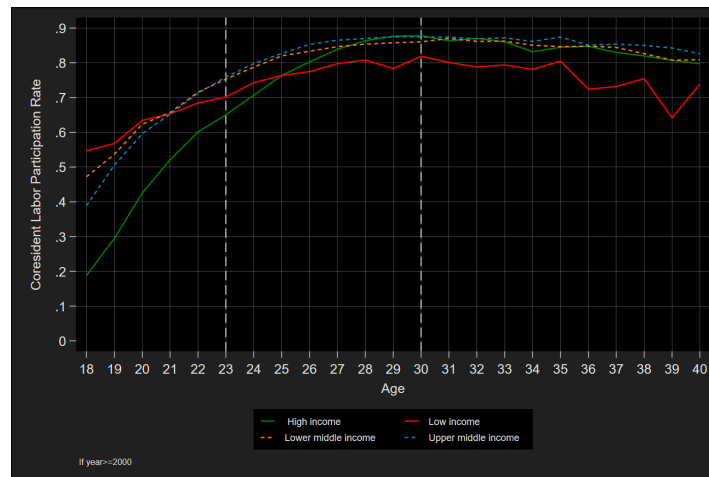
-Van der Weide R, et.al (2021). Intergenerational Mobility around the World

- 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

9   Appendix
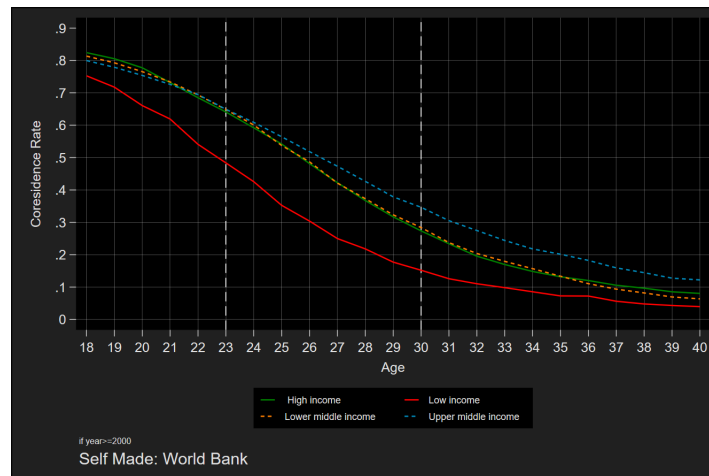
	2017	2015	2013	2011	2009	2006	2003	2000	1998	1996	1994	1992	1990	1987
1994	23	21	19	17	15	12	9	6	4	2	0	-2	-4	-7
1993	24	22	20	18	16	13	10	7	5	3	1	-1	-3	-6
1992	25	23	21	19	17	14	11	8	6	4	2	0	-2	-5
1991	26	24	22	20	18	15	12	9	7	5	3	1	-1	-4
1990	27	25	23	21	19	16	13	10	8	6	4	2	0	-3
1989	28	26	24	22	20	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

To see some results dynamically, click on the following link