IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16465

Economic Mobility and Fairness in a Developing Country: Evidence from Peru

Juan F. Castro
Gustavo Yamada
Santiago Medina
Joaquin Armas

SEPTEMBER 2023

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16465

Economic Mobility and Fairness in a Developing Country: Evidence from Peru

Juan F. Castro
Universidad del Pacifico

Gustavo Yamada
Universidad del Pacifico and IZA

Santiago Medina
Harvard University

Joaquin Armas
Stanford University

SEPTEMBER 2023

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Economic Mobility and Fairness in a Developing Country: Evidence from Peru

Periods of rapid economic growth in developing countries have been well studied in terms of poverty and income inequality reduction, but much less is known about the performance of these countries in terms of economic mobility. We study intragenerational mobility in Peru using an asset-based measure of wealth and longitudinal data from the Young Lives project (2002 - 2016). We find that Peruvian households enjoyed a moderately large degree of mobility in this period. Averages, however, mask significant differences between Spanish-speaking households and those that speak an indigenous language. We estimate a positive mobility gap in favor of Spanish-speaking households of 12.7 percentiles, and find that half of this gap persists after controlling for a comprehensive set of household characteristics that impact their ability to accumulate wealth. We propose a new measure of individual mobility and use it to assess the degree of inequality of opportunity for mobility, that is, to what extent is mobility caused by circumstances outside of households' control. We find that this fraction is at least 17.4% for the most disadvantaged half of the population, but only 1.9% for the more advantaged half.

JEL Classification: D63, D31, J60

Keywords: economic mobility, inequality of opportunity, development, wealth

Corresponding author:
Gustavo Yamada
Department of Economics
Universidad del Pacífico
Av. Salaverry 2020
Lima 11
Peru
E-mail: yamada_ga@up.edu.pe

1 Introduction

Several developing countries experienced a period of fast economic growth after the turn of the century. This process involved large and well documented increases in household incomes and reductions in poverty rates and income inequality. However, much less is known about the performance of developing countries in terms of economic mobility. This concept refers to the degree in which households can change their positions in the socioeconomic distribution relative to one another. The lack of studies for the developing world has had its cause, mainly, in data limitations (Iversen, Krishna, and Sen, 2019).

The analysis of mobility is important for several reasons. First, one can argue that mobility is necessary for a fair society, as individual efforts, decisions and hard work should lead to changes in one's position in the socioeconomic distribution. Second, people care for their welfare relative to their peers and not only about their welfare in absolute terms (Layard, 2006). Hence, one can infer that people value their ability to improve their situation relative to others. In addition, people's beliefs about their mobility impact their preferences for redistribution (Alesina and Giuliano, 2011). For example, misfortune may make an individual pessimistic about their upward mobility prospects and lead them to favor government policies that foster a more equitable distribution of income (Giuliano and Spilimbergo, 2008). Therefore, the degree of mobility in a society can be important in shaping individuals' perceptions and, ultimately, public policy.

The relation between economic mobility and social fairness has also become relevant in recent years. Latin America oversees the potential for social turmoil as several countries have witnessed calls for redistribution with demands for rapid and dramatic change. For example, in Chile, massive protests led to a referendum where citizens decided to write a new Constitution in 2020. Likewise, in 2021, Peru elected a political outsider who ran on the promise of calling on a Constitutional Assembly to change Peru's economic regime.

In this paper, we study economic mobility and its fairness in the context of a developing country. In particular, we focus on Peru and analyze its intragenerational mobility during a fourteen year period (2002 - 2016). We use the longitudinal information provided by the Young Lives project[1] to build a wealth index and track households' positions in the wealth distribution across time. We use the framework presented in Chetty et al. (2019, 2014) to document mobility and analyze whether there are significant differences between Spanish-speaking households and those that speak an indigenous language. In addition, we propose a new measure of individual mobility based on the residuals of the rank-rank equation used in Chetty et al. (2014). We use this measure to provide a lower bound for the proportion of mobility that can be related to circumstances outside households' control (as opposed to individual effort), following the framework developed in Roemer (1993, 1998) and the strategy proposed in Ferreira and Gignoux (2011). This allows us to measure the share of inequality of opportunity for mobility and to assess mobility from a normative perspective.

[1]The Young Lives project follows around 12,000 children of two age cohorts in four developing countries: Ethiopia, India, Peru, and Vietnam. The younger cohort was aged one in the first round, while the older was eight. Between 2002 and 2016, the project conducted five rounds of surveys.

Our main findings can be summarized as follows. We find that a 10 percentile increase in the initial wealth distribution is associated with a 7 percentile increase in the wealth distribution fourteen years later. By exploring additional mobility statistics, we observe that, on average, Peruvian households enjoyed a moderately large degree of mobility between 2002 and 2016. In particular, households who were at the very bottom of the wealth distribution in 2002 experienced an average increase of 14.6 percentiles after 14 years. Moreover, when taking the absolute value of all position changes, we find that the average household experienced a significant shift of 17 percentiles in the wealth distribution and that a third of all households experienced a shift of at least 20 percentiles between 2002 and 2016. Nevertheless, we observe that large mobility is mostly driven by movements in the middle of the distribution and that shifts from the bottom of the wealth distribution to the top (or vice-versa) are extremely rare.

The previous mobility statistics, however, mask important differences between Spanish-speaking households and those that speak an indigenous language (linguistic minorities). In fact, we estimate a positive mobility gap in favor of Spanish-speaking households of 12.7 percentiles, and find that half of this gap persists after controlling for a comprehensive set of household characteristics that impact their ability to accumulate wealth. Finally, we estimate that circumstances households cannot control cause at least 8% of the variations in mobility. Importantly, their explanatory power rises to 17.35% for households who started in the bottom half of the wealth distribution and is only of 1.87% for those who started in the top half. This suggests that equality of opportunity for mobility is significantly larger for Peruvians in the upper part of the wealth distribution.

This paper makes two important contributions to the study of economic mobility in developing countries. First, we overcome two important limitations to the study of intragenerational mobility in developing countries.[2] The first is the lack of reliable longitudinal information. We address this by exploiting the panel structure of the Young Lives sample, which suffers from little attrition. The second limitation is related to the presence of noise in income data, which we skirt by using detailed records of physical asset holdings and dwelling characteristics to build a wealth index less prone to measurement error than income measures found in national household surveys.

Second, to the best of our knowledge, we are the first study to provide a quantified normative assessment of mobility. We propose the rank-rank residuals as a new measure of household level mobility, and use it to provide an estimate of inequality of opportunity for mobility. In doing so, we combine the predominantly positive mobility framework of Chetty et al. (2014), with insights from literature on equality of opportunity, which asserts that social justice does not require equality of individuals' outcomes, but equality of the means or opportunities to reach them (see Rawls (1971), Sen (1980), Dworkin (1981a, 1981b), and Roemer (1993, 1998), among others).

[2] Empirical research on mobility has increased in recent years and after the work of Chetty et al. (2014), with a focus on developed countries (Acciari, Polo, and Violante, 2022; Braun and Stuhler, 2017; Chetty et al., 2019; Gregg, Macmillan, and Vittori, 2016; Güell et al., 2018; Landersø and Heckman, 2016; Modalsli, 2016). The study of mobility with the newest techniques and data has been particularly challenging in developing countries (Iversen, Krishna, and Sen, 2019) and most papers have focused on intergenerational mobility (see Torche 2015, Neidhöfer, Serrano, and Gasparini 2018, and Britto et al. 2022, for example). The lack of reliable longitudinal data makes the study of intragenerational mobility harder and has led researchers to use a variety of alternative methods (see Dang and Lanjouw 2018, Winkelried and Torres 2018).

The rest of the document is structured as follows. Section 2 elaborates on the concept of economic mobility. Section 3 describes the Young Lives data and our sample. Section 4 shows our main mobility estimates. Section 5 presents our analysis of mobility by native language. In Section 6, we discuss the measure of individual mobility proposed in this study and present the estimates for the share of mobility explained by circumstances outside households' controls. Finally, in Section 7 we present some concluding remarks.

2 Conceptual framework

2.1 Types of mobility

A broad definition of mobility would refer to it as the evolution of economic outcomes of households or individuals. This definition can be refined in several ways. A first division is concerned with whose economic outcomes one studies. If we compare outcomes between generations, we are talking about intergenerational mobility. For example, studying the relation between the incomes of a cohort and the income their parents had thirty years ago would fall into the category of intergenerational mobility. If we compare the economic outcomes from two different points in time, but for the same households or individuals, we are studying intragenerational mobility.

A second division in the study of economic mobility distinguishes between absolute and relative mobility. Measures of absolute mobility quantify changes in the economic outcomes of a specific group, whereas relative mobility ones measure how the outcomes of different households change relative to one another. Throughout this paper we will focus predominantly on relative mobility, as has been the case in the most recent literature. Moreover, we will most often only say "mobility" when referring to what has traditionally been called "relative mobility".

2.2 The rank-rank specification

The rank-rank specification has been typically used in intergenerational mobility studies (see, for example, Dahl and DeLeire (2008) and Chetty et al. (2014)) and consists in regressing the percentile ranks in the income distribution of a given cohort (generation) on the percentile ranks of their parents (the previous generation). Given that our focus is on intragenerational mobility, we will adjust this specification as follows:

$$y_{i,t} = \alpha + \beta y_{i,t_0} + \varepsilon_{i,t} \tag{1}$$

where $y_{i,t}$ is the percentile rank of household i in the wealth distribution of period t, and y_{i,t_0} is that same household's percentile rank in the wealth distribution of the baseline period (t_0). The percentile ranks take values in the $[0, 100]$ range and a rank of zero corresponds to the household in the very bottom of the wealth distribution.

Following Chetty et al. (2014), the coefficient α is a measure of absolute mobility. More precisely, it measures the expected final wealth percentile rank of a household that was in the lowest percentile of the initial wealth distribution. The coefficient β is a measure of relative mobility and captures the persistence of the initial positions. More precisely, it measures the predicted difference in the percentile ranks of period t between two households that were separated by one percentile rank in the initial period t_0.

Due to the nature of percentile ranks, β can only be in the $[-1, 1]$ range. If $\beta = 1$, each household's position in the final period (t) is the same as in the initial period (t_0). This is a scenario of perfect immobility. If $\beta = -1$, the positions in the initial period (t_0) are entirely reversed in the final period (t). An interesting case arises when $\beta = 0$, which means that households' initial positions have no predictive power on their final positions. These are all extreme cases, but they are useful to interpret less extreme values of β.

In practice, β will only takes values in the $[0, 1]$ range. Higher values of β represent lower relative mobility (higher persistence of initial positions), and lower values of β mean higher relative mobility (lower persistence). Finally, an important property of this regression is the relation between its slope $\hat{\beta}$ and its R-squared (R^2). In particular, one can show that $\hat{\beta}^2 = \mathrm{R}^2$.

3 Data

3.1 Description of the Young Lives sample

For our analysis of household-level intragenerational mobility, we use data from Young Lives, an international longitudinal study about childhood poverty. This project followed about 12,000 children and their families in four developing countries (Ethiopia, India, Peru, and Vietnam) and interviewed them in five different rounds between 2002 and 2016. Households with children in one of two age cohorts were surveyed. When the project began, the younger cohort was aged one, and the older cohort eight. In the case of Peru, the initial sample size of the younger cohort was 2,052 and the older cohort had 714 observations.

The sampling design was meant to oversample poor areas. To do so, population-weighted Peruvian districts where randomly selected, from the national pool of districts minus the wealthiest 5%. Within districts, population centres and census tracts were randomly selected and street blocks were visited following random number tables. All dwellings in each block were visited until households with children from the right ages were found. Although the survey is not intended to be nationally representative, the Peruvian sample has been found to closely resemble other nationally representative data and to properly reflect the diversity of children and families in Peru (Escobal and Flores, 2008).

Attrition rates by Round 5 are 9.36% and 14.85% for the younger and older cohorts, respectively. These are equivalent to attrition rates of 1.95% and 3.16% per round. Such attrition rates are relatively low compared to other longitudinal studies (Young Lives, 2018). Coupled with the richness of the questionnaire, these traits of the Young Lives survey offer the possibility to study intragenerational mobility reliably with panel data. As Iversen, Krishna, and Sen (2019) note, such studies have been scarce for the developing world context, but allow for much more robust and transparent measures.

3.2 Main Sample

We are interested in exploring intragenerational mobility at the household level. However, the main focus of the Young Lives study are children, as opposed to households or adults. Therefore, the study followed the child who was chosen in the first round and surveyed the household where the child was living in each subsequent round. Typically, a household in the sample would consist of the Young Lives child, their parents, and other relatives. Household heads tend to be the parents of the Young Lives child. However, there is no guarantee that a household where the Young Lives child lives in one round remains the same in the next one. In fact, there are several reasons why a child might move out of a household, like a parental split or the child reaching independence age.

To mitigate the risk of considering different households across time, we assume that, for every child, two observations in different points in time correspond to the *same* household if the child lived with at least one of their parents in both observations. Under this definition, for estimations that consider mobility between Round 1 and Round $t \in \{2, 3, 4, 5\}$, we will only include households that are the *same* in Rounds 1 and t.[3]

To ensure the comparability of our results across time horizons (for example, between Rounds 1-to-2 and Rounds 1-to-4), we additionally required that each household remained the same in the longest time horizon, which is between Rounds 1 and 5. Therefore, when estimating mobility between Rounds 1 and $t \in \{2, 3, 4, 5\}$, we will only inlcude households that are the same in Rounds 1 and 5, in addition to being the same in Rounds 1 and t.

In Table 1, we show the number of observations that we include in our sample for each time horizon. We also report the percentage that these observations represent out of the total number of households surveyed in each round. We also show these frequencies separately for the younger and older cohorts. For example, our sample up to the Round 5 includes 86.6% of all possible observations: 94% of the younger cohort and 64.1% of the older cohort.

As expected, our samples for Rounds 2, 3, and 4 have a smaller number of observations than the sample for Round 5, since belonging to the latter sample is required to belong to the former ones. The % out of the total number observations is increasing, as the total number of observation decreases with each round due to attrition and the observation count of our samples remains rather constant between Rounds 2 and 4, and increases in Round 5.

[3]While we do not require that it is the same parent that lived with Young Lives' child in Round 1 and Round t, we do not have a single case in which a child who lived only with their mother in Round 1 lived only with their dad in a subsequent round (or viceversa).

Table 1: Sample by rounds and cohorts

	Entire Sample				Only Young Cohort				Only Old Cohort			
	R2	R3	R4	R5	R2	R3	R4	R5	R2	R3	R4	R5
Obs.	2,084	2,080	2,053	2,138	1,703	1,702	1,697	1,748	381	378	356	390
% Total	0.787	0.794	0.809	0.866	0.868	0.876	0.892	0.940	0.556	0.558	0.561	0.641

Notes: The table reports the following statistics. **Obs**: The number of households in the sample of the round reported in the column. **% Total**: Obs. divided by the total number of households that were surveyed in the corresponding round.

In all the households of our sample, the Young Lives child lived with at least one of their parents. It is important to determine which parent it was and the role they had in the household, because we will later use parental characteristics as household characteristics in heterogeneity analysis and when assessing the degree of inequality of opportunity for mobility.

Table 2 provides summary statistics about the presence of parents in our sample. In Round 1 the father lived in 85.4% of the households in our sample, while the mother did so in 99.9% of them. These two fractions decrease by Round 5: the father lived in the household in 70.7% of the observations, while the mother did in 96.5% of them. If we focus only on the households where the parent was present in Round 1, the permanence rate of fathers in Round 5 goes up to 82.8%. Regarding the household roles played by parents, the father of the Young Lives child was the household head in 87.67% of the households in Round 1, and the mother was in 8.67% of them.

Last, we note that the average ages of the parents do not differ too much between the younger and older cohorts. Thus, we pool both samples in our estimations. In fact, the average ages of the fathers in the first round were 30 and 38 years old for the younger and older cohorts, respectively. These figures were 26 and 34 years for the mothers. In Appendix A, we show that these average age differences do not bias our estimations in the pooled sample.

Table 2: Presence of parents in the sample

	Fathers		Mothers	
	Round 1	Round 5	Round 1	Round 5
Obs. with father/mother	1,825	1,511	2,136	2,063
% Over sample	0.854	0.707	0.999	0.965
% Over present in first round		0.828		0.966

Notes: The table reports the following statistics. **Obs with father/mother**: The number of households in the sample of the respective round where the indicated parent lived in the household. **% Over sample**: Obs. with father/mother divided by the number of households in the sample of the respective round. **% Over present in first round**: Obs. with father/mother divided by the number of households in the sample of the fifth round where the indicated parent lived in the household in the first round.

3.3 The wealth index

To study economic mobility, we need to identify a variable that represents economic well-being and allows us to rank households. Income would be a natural candidate, but it is not a good measure of the real economic position of a household due to its transitory component. To address this problem, studies need to average income over several years (Chetty et al., 2014), but this might not always be feasible. Even if such strategy could be done, income is likely to be measured with error in developing countries, due to the dominance of agrarian sectors and informality (Iversen, Krishna, and Sen, 2019). In the same line of thought, Behrman et al. (2017) utilize consumption to measure the intergenerational transmission of poverty and inequality, as the authors state that it is a better indicator of the long-run resource constraint of a household. However, such strategy is not without its pitfalls, as detailed consumption measures often require a high degree of numerical recall capacity from the respondent.

To improve on previous efforts and considering the limitations that are particularly relevant in developing countries such as lack of panel data, or income data, we propose a measure of well-being that is based on physical assets. Such measure is advantageous for two reasons. First, asset possession is much less likely to be measured with error, as interviewers typically ask the household members if they have certain assets and this can be verified directly. Second, asset holdings are likely to have a smaller (if any) transitory component, due to their stock nature, contrasted to income and consumption, which are flows. Thus, we rank households using a variable based on assets: a wealth index. This idea is not entirely novel, but its use has been small in developing countries (see Krishna (2011) or Torche (2015) for further discussion on its advantages). According to Iversen, Krishna, and Sen (2019), such asset-based methods might be among the best currently available for measuring mobility in developing countries, if detailed information is available.

We will rely on the wealth index contained in the Young Lives survey. According to a technical note from Young Lives, the wealth index was designed to be the primary measure of socioeconomic status of households in the sample (Briones, 2017). The set of questions used to create the index have been asked in each round of the Young Lives survey and have negligible rates of non-response. Thus, this measure of well-being lends itself for valid comparisons between rounds.

The Young Lives wealth index is calculated as the simple average of three sub-indices, one for each of the following dimensions: (i) housing quality, (ii) access to services and (iii) consumer durables. Each sub-index is obtained as the average of other components. The housing quality index is composed of three dummy variables that take the value of one when a good quality wall, roof and floor, respectively, are observed in the household; and one continuous variable that measures how many rooms per person are there in the household (household density). The access to services index is the simple average of four dummy variables that take the value of one if the household has access to electricity, a safe drinking water source, managed sanitation (toilet), and uses a good quality cooking fuel, respectively. Finally, the components of the consumer durables index are twelve dummy variables that are equal to one if the household owns the following assets in sellable condition: radio, television, bicycle, motorbike, automobile, landline phone, mobile phone, refrigerator, stove, blender, iron, and record player.

Giving equal weight to each subindex and, similarly, to each component within subindex, is equivalent to assuming that each one has the same importance as the other. Moreover, the decisions on what type of wall was made of "good-quality" material or what source of drinking water was "safe" were made separately by each of the Young Lives country teams. Thus, they follow country-specific criteria (Briones, 2017). Finally, although the index is not a fully continuous variable in a strict sense, the amount of indicators that compose it renders a large quantity of possible combinations that can be ranked. Thus, for practical purposes we will treat the wealth index as a continuous variable that represents the household's material well-being in a range of $[0, 1]$.

We propose some modifications to improve the wealth index of Young Lives for the case of Peru. Concretely, we analyzed the evolution of the share of households that owned each asset. As households get richer, they should own more normal goods and fewer inferior goods. So we expect that the ownership of all the components of the consumer durables index is increasing in time. However, this is not the case in two components of the consumer durables index, as radio ownership is decreasing in time and landline phone ownership is rather constant.

These components were replaced with dummies that indicate the ownership of a computer or laptop, and of a microwave, since both assets satisfy the desired property of its ownership being increasing in time. We also censored the household density indicator, forcing it to take the value of one whenever more rooms than people were found.

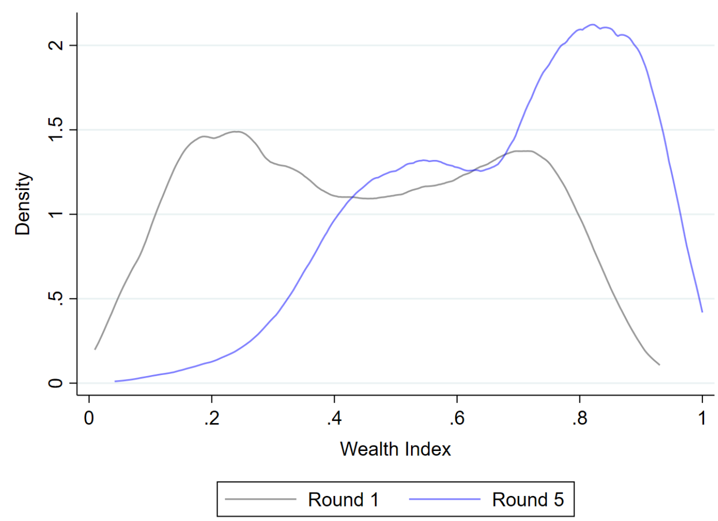
In Table 3 we show the evolution of the average of the wealth index, the sub-indices and their components. All are increasing in time. The housing quality index increases by 0.15 between Rounds 1-5; the services access index, by 0.16; and the consumer durables index experiences the largest increase, 0.33. In Figure 1 we show the wealth index distribution in both the first and fifth round. In Round 1, the distribution is bimodal and rather uniform between 0.1 and 0.7. In Round 5, the wealth index distribution becomes left-skewed as a large share of households get wealthier but a minority of them doesn't.

Table 3: Average wealth index and components, by round

Variable	Round 1	Round 2	Round 3	Round 4	Round 5
Wealth Index	0.44	0.50	0.58	0.64	0.68
Housing Quality Index	0.50	0.51	0.56	0.61	0.65
Good wall	0.33	0.36	0.38	0.44	0.50
Good roof	0.83	0.79	0.84	0.86	0.88
Good Floor	0.38	0.41	0.46	0.51	0.58
Modified density	0.46	0.49	0.54	0.62	0.64
Services Access Index	0.61	0.70	0.80	0.85	0.87
Good electricity	0.67	0.79	0.88	0.95	0.97
Good toilet	0.78	0.87	0.92	0.95	0.96
Good drinking water	0.52	0.64	0.82	0.82	0.83
Good cooking fuel	0.48	0.53	0.57	0.69	0.74
Consumer Durables Index	0.19	0.28	0.37	0.46	0.52
Owns T.V.	0.61	0.73	0.82	0.88	0.88
Owns motorbike	0.03	0.05	0.09	0.17	0.25
Owns car	0.04	0.04	0.05	0.10	0.13
Owns mobile phone	0.06	0.36	0.72	0.92	0.97
Owns fridge	0.17	0.24	0.30	0.45	0.52
Owns stove	0.45	0.59	0.65	0.77	0.86
Owns blender	0.33	0.47	0.58	0.65	0.70
Owns iron	0.39	0.50	0.56	0.61	0.63
Owns record player	0.18	0.26	0.32	0.29	0.35
Owns washing machine	0.02	0.06	0.11	0.20	0.28
Owns computer	0.03	0.07	0.16	0.34	0.42
Owns microwave	0.01	0.04	0.08	0.15	0.20

Notes: The table reports the average of the wealth index, its four sub-indices, and their components, by round. The wealth index and its four sub-indices are continuous variables with values between zero and one. All the components are dichotomous indicators of quality or ownership. Higher values of the wealth index represent strictly greater wealth.

Figure 1: Evolution of the wealth index distribution



4 Mobility estimations

Figure 1 reveals that, on average, households in the Young Lives sample are becoming richer in the 14-year period comprised between Rounds 1 and 5. However, to what extent can we say that the poorest households remain among the poorest and the richest households remain among the richest after 14 years? More precisely, to what extent is a household's position in the wealth distribution by Round 5 conditioned by the position it had in Round 1? This is a question related to the concept of absolute mobility [4] and in this section we will provide an answer based on three complementary ways of looking at the persistence of the wealth index percentile ranks observed in Round 1.

The first one is provided by the rank-rank equation. We estimate the intercept and slope of this equation for four different time horizons, comparing the wealth ranks observed in Rounds 2 to 5 vs. those observed in Round 1. As discussed above, the intercept of this equation provides an estimate of absolute mobility, while the slope provides an estimate of relative mobility. The second approach consists in analyzing the distribution of wealth rank changes that occurred between Rounds 1 and 5. Finally, we calculate the probability of being in each wealth quintile by Round 5, conditional on the household's wealth quintile in Round 1.

Table 4 shows the results of estimating the rank-rank equation. In each column, we present the result of regressing the wealth rank of the corresponding round on the wealth rank observed in Round 1. Estimates for the intercept of the rank-rank equation show that households at the bottom of the wealth distribution in Round 1 experienced an increase of 8.13 percentile points (pp) in their wealth rank after 4 years and a 14.59pp raise after 14 years. Absolute mobility is, therefore, increasing but this appears to be happening at a decreasing rate. In fact, there is a significant increase of 41% in

[4] A similar question in terms of relative mobility would be: to what extent are the differences in households' positions in the wealth distribution by Round 5 conditioned by the differences observed in Round 1?

the point estimate of the intercept between Rounds 2 and 3, but no difference between Rounds 4 and 5.

Estimates for the slope of the rank-rank equation reveal that a 10pp difference in Round 1 wealth rank is associated to a 8.37pp difference in wealth rank by Round 2 (after 4 years) and a 7.06 difference by Round 5 (after 14 years). One can also use the R-squared of the rank-rank equation to account for relative mobility (recall that $\hat{\beta}^2 = \text{R}^2$). According to the results reported in Table 4, variations in wealth ranks observed in Round 1 predict 73.8% of the variations observed in Round 2 and 49.2% of the variations observed in Round 5. As with absolute mobility, relative mobility increases with the length of the time horizon considered, but at a decreasing rate. In fact, mobility appears to stagnate after Round 4, so it only takes about 11 years for intragenerational mobility to stop growing.

What does this mean in practice? The average ages of parents in the first round are 31 for fathers and 27 for mothers. If the observed stagnation of mobility were to persist in a longer time horizon, we would have that households with fathers and mothers who were aged 31 and 27 years in 2002 can only look for changes in their expected wealth ranks (conditional on today's position) until they are 42 and 38 years old, respectively. In other words, conditional on their position today, their expected wealth rank later in life would not be different than the expected wealth rank for when they become about 40 years old.

In Appendix A we show that the rank-rank estimates presented in Table 4 are robust to a stricter definition of when a household remains the same between two periods. We also rule out that we are ignoring relevant age-based differences in mobility by pooling the younger and older cohorts together in our main estimations.

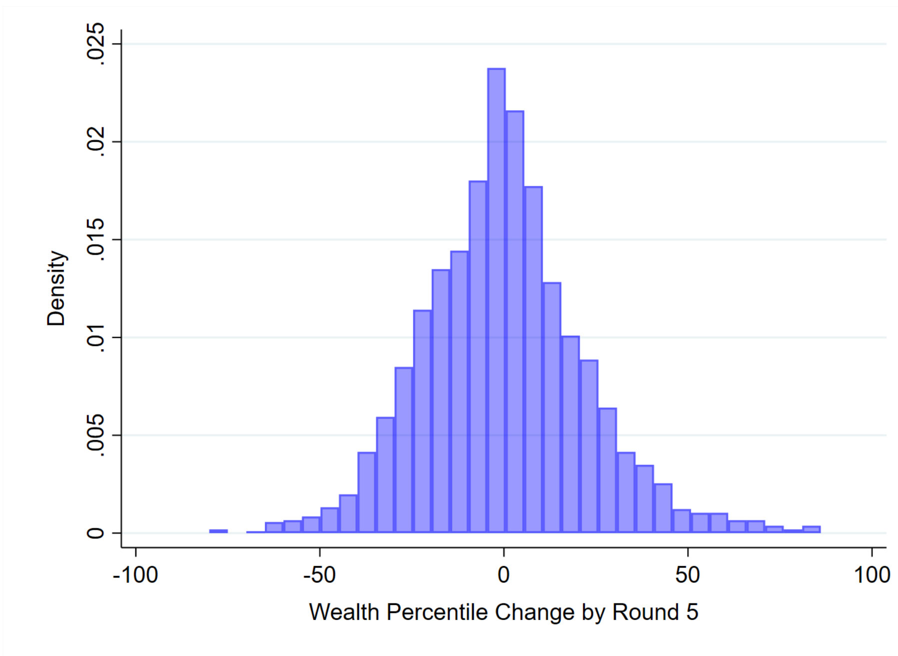
Table 4: Rank-rank equation estimates

	Round 2	Round 3	Round 4	Round 5
Relative Mobility (β)	0.837	0.769	0.718	0.706
	(0.019)	(0.028)	(0.036)	(0.038)
Absolute Mobility (α)	8.13	11.48	14.05	14.59
	(1.53)	(2.15)	(3.16)	(3.28)
R-squared	.738	.615	.492	.492
Observations	2074	2068	2043	2106
Years Since R1	4	7	11	14

Notes: The table presents the estimated rank-rank coefficients for four different time intervals. The dependent variable is the household wealth index percentile at the round indicated in the column, and the only independent variable is always the household wealth index percentile in the first round. Standard errors clustered at the sentinel site level are shown in parentheses. All estimated coefficients are statistically significant ($p < 0.01$).

In Figure 2, we present the distribution of wealth percentile changes between Rounds 1 and 5. As expected, this is a symmetric distribution where half of the households experienced an increase in wealth ranks and the other half experienced a decline. More importantly, when taking the absolute value of all position changes, we observe that the average household experienced a significant shift of 17 percentiles in the wealth distribution and a third of all households experienced a shift of at least 20 percentiles during this fourteen-year period.

Figure 2: Distribution of wealth index percentile changes by Round 5



Finally, we assess the transitions between wealth quintiles in Rounds 1 and 5 using a transition matrix shown in Table 5. The columns indicate final wealth quintiles, while the rows the initial wealth quintiles. We report the probability of finishing on each wealth quintile by Round 5, conditional on starting at a particular wealth quintile in Round 1. Under perfect immobility, this would be an identity matrix. If mobility was perfect, all probabilities would be the same and equal to 20%. Furthermore, we also report the median percentile change experienced by the households for every initial-final wealth quintile pair.

Three results are worth highlighting from Table 5. If one refers to transition probabilities, it reveals that, on average, the probability of remaining in the same wealth quintile between Rounds 1 and 5 (40.55%) is significantly lower than the probability of leaving it (59.45%). Second, we learn that these probabilities are explained by large median transitions and not by marginal shifts from one quintile to the one immediately above or below. For example, the median transition of households who started in the first quintile and finished in the second is 17.91 percentiles, which is almost an entire quintile of upward mobility. This median transition is similar for households moving from the second, third and fourth quintiles to the one immediately above or below.

Third, we observe that most of the quintile changes are concentrated in the middle of the socioeconomic distribution and not in its top nor bottom quintiles. For instance, the only quintiles that are harder to leave from than to stay in are the top and bottom ones. This is consistent with what is typically known as sticky floors and sticky ceilings in the literature. In fact, the probabilities of the largest changes occurring in a 14 years horizon are fairly small: a household that starts at the bottom quintile of the distribution only has a 2.62% chance of reaching the top one, and a household that starts at the top quintile only has a 0.49% chance of going all the way to the bottom one.

Table 5: Wealth transition matrix up to the fifth round

		Round 5				
Round 1	**Stat**	Q1	Q2	Q3	Q4	Q5
Q1	Proportion	53.33	27.62	11.90	4.52	2.62
	Median Movement	-1.21	17.91	39.03	60.79	72.27
	Observations	224	116	50	19	11
Q2	Proportion	34.68	34.68	19.24	8.08	3.33
	Median Movement	-18.24	0.04	19.39	37.51	57.61
	Observations	146	146	81	34	14
Q3	Proportion	9.72	25.83	30.81	23.46	10.19
	Median Movement	-33.80	-17.17	1.49	17.46	35.87
	Observations	41	109	130	99	43
Q4	Proportion	2.32	9.51	29.23	30.16	28.77
	Median Movement	-57.15	-35.80	-19.77	0.24	17.96
	Observations	10	41	126	130	124
Q5	Proportion	0.49	2.18	15.78	27.18	54.37
	Median Movement	-78.13	-59.51	-35.27	-17.06	0.06
	Observations	2	9	65	112	224

Notes: The table represents the transition matrix for the main sample. For every pair of initial wealth quintile and final wealth quintile, two statistics are presented. **Proportion**: The proportion of households that started in the row quintile and finished in the column one. In other words, the probability of finishing in the column wealth quintile conditional on starting on the row one. **Median Movement**: The median wealth percentiles movement of the households that started in the row wealth quintile and finished in the column one. **Obs**: The number of households that started in the row quintile and finished in the column one.

Overall, the results presented in this section indicate that, on average, the households in the Young Lives sample enjoyed a moderately large degree of (absolute) mobility between Rounds 1 and 5. Despite the fact that the rank-rank estimates suggest that mobility stops growing after 14 years, we do not find evidence to support the claim that the positions in the wealth distribution in 2002 strongly conditioned the positions observed in 2016. Even though there are sticky floors and ceilings, there is still some movement out of these quintiles to the neighboring ones and, importantly, there exist sizeable movements across the middle three ones.

5 Differences in mobility by native language

It is a well documented fact that there are significant differences in economic opportunities between individuals of different ethnic and linguistic backgrounds in Peru (Thorp and Paredes, 2010). In this section, we analyze whether these differences are also present in terms of mobility. For this, we will divide our sample using the native language of the household head. We rely on the first language because it is less prone to biases than other indicators such as the self-reported ethnic identity.

In particular, we compare households whose native language was only Spanish and households whose native language was at least one indigenous language. This comparison is relevant considering that native language is correlated with different geographic, economic, and social factors that strongly condition economic opportunity in Peru and, more generally, in countries with a large share of indigenous population. A large minority of the Peruvian population speaks Quechua, the former language of the Inca Empire and the most spoken indigenous language in the Americas. It is estimated that there are 8 million Quechua speakers throughout the Andes and across Colombia, Ecuador, Peru, Bolivia, and Argentina.

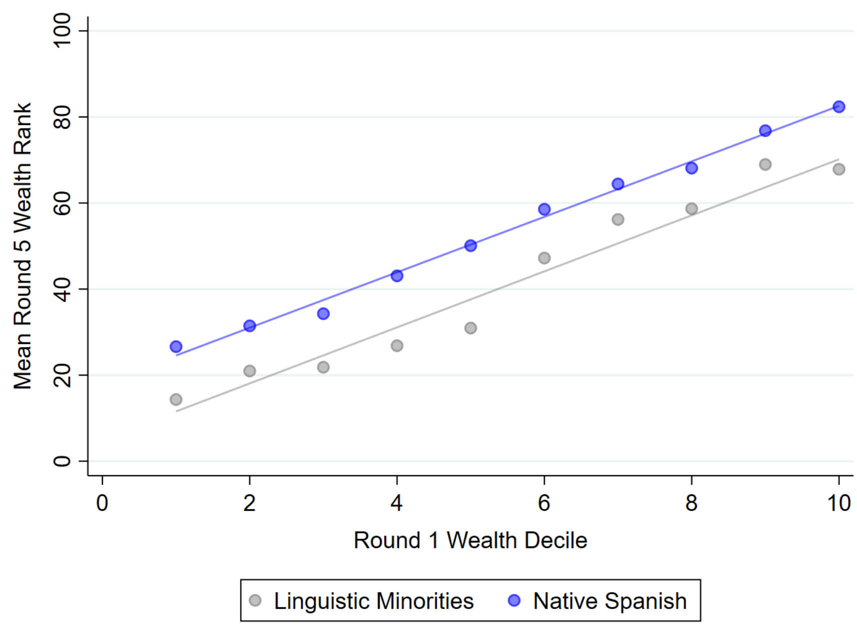
We assign a first (native) language to each household in the following way. First, we assign the household head's first language. If that information is not available, we assign the first language of a parent of the Young Lives child provided they lived with them in the first round. Priority is given to the father's first language if that information is available. Based on this, we identify two mutually exclusive groups: i) Spanish-speaking households where Spanish is their only first language; and ii) linguistic minorities households where indigenous languages such as Quechua and Aymara are their first language. Households that have two first languages (most commonly Quechua and Spanish) are considered as part of the second group. Spanish-speaking households represent 72.33% of the main sample, while linguistic minorities households represent 27.67%.

All between-groups comparisons in this section consider our longest available time horizon of fourteen years (Rounds 1 to 5). We document the differences in mobility measures between native Spanish-speaking households and linguistic minorities households in three different ways. First, we compute differences in the intercept and slope of the rank-rank equation of each group, following Chetty et al. (2019). Second, we analyze differences in the distributions of wealth index percentile changes. Finally, we compare the transition matrices of each group.

5.1 Rank-rank estimates

We present a graphical representation of the rank-rank regressions of Spanish-speaking and linguistic minorities households in Figure 3. Each dot represents the average final wealth index percentile rank of all the households that started in a given wealth index decile in the complete sample. Two important insights can be drawn from Figure 3. First, Spanish-speaking households appear to enjoy larger absolute mobility than their linguistic minorities counterparts, as the red line has a higher intercept than the blue line. Second, there are no visible differences in terms of relative mobility between these groups (the slopes of the two lines are very similar).

Figure 3: Rank-Rank estimates by native language



These two results suggest that, for *any* starting position in the wealth distribution, Spanish-speaking households are predicted to end higher up in the wealth distribution, after 14 years, than households belonging to a linguistic minority. Note that this doesn't mean that linguistic minorities are worse off in 2016 than in 2002 in absolute terms: Peru's rapid growth did lead to large reductions in poverty and increased standards of living. Nevertheless, growth appears to have had a pro-Spanish-speaking bias relative to equally rich households that belong to a linguistic minority.

To confirm the size and statistical significance of this mobility gap, we can estimate the following regression:

$$y_{i,R5} = \alpha_0 + \beta_0 y_{i,R1} + \alpha_1 Spanish_i + \beta_1 y_{i,R1} \times Spanish_i + \varepsilon_{i,R5} \tag{2}$$

Where α_0 corresponds to the intercept of the rank-rank equation for households belonging to a linguistic minority and β_0 is the rank-rank slope of this group. $Spanish_i$ is an indicator for being a Spanish-speaking household. Therefore, α_1 captures the difference in absolute mobility between a Spanish-speaking household and a household that belongs to a linguistic minority, and β_1 captures the difference in relative mobility between the two groups.

Table 6: Rank-Rank estimates by native language

(α_0) Constant	8.318**
	(3.473)
(β_0) R1 Wealth Percentile	0.651***
	(0.0492)
(α_1) Spanish	12.69**
	(4.663)
(β_1) Spanish x R1 Wealth Percentile	-0.00516
	(0.0647)
R-Squared	0.533
Observations	2084

Notes: The table reports the estimated coefficients from Equation 2. Standard errors clustered at the sentinel site level in parentheses. Stars denote statistical significance at the 10% (*), 5% (**) and 1% (***) level.

Table 6 presents the results of estimating Equation 2. Absolute and relative mobility estimates for households in the group of linguistic minorities group are 8.32 and 0.65, respectively. Estimates for coefficients α_1 and β_1 indicate that Spanish-speaking households enjoy a large and significant absolute mobility advantage of 12.6 percentiles and that there are no significant differences in terms of relative mobility between the two groups of households. Put together, these two results confirm that, for any pair comprised of a Spanish-speaking and a linguistic minority household that had the exact same position in the 2002 wealth distribution, the Spanish-speaking household will end up 12.69 percentiles above the linguistic minority household by year 2016.

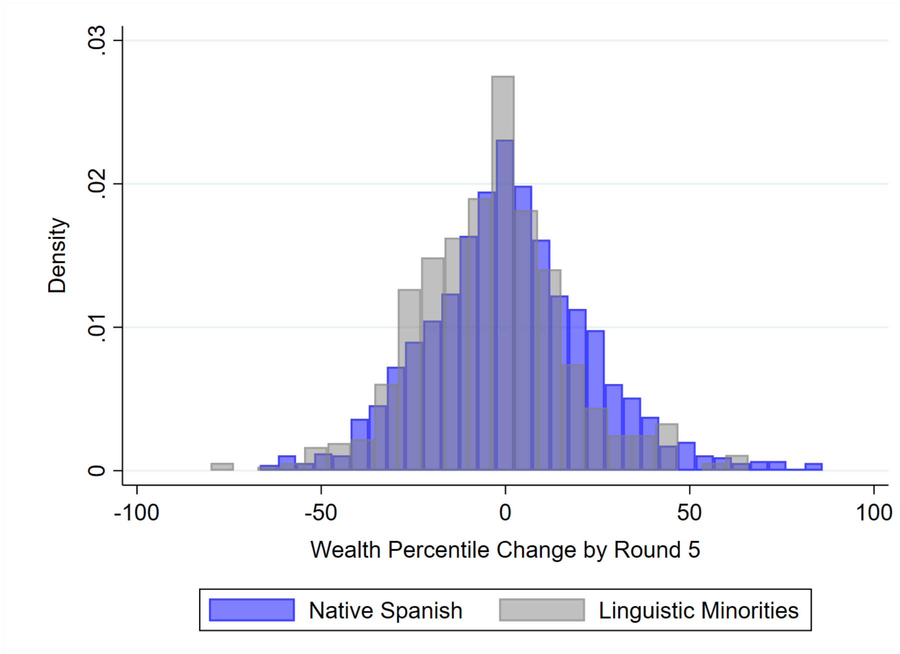
5.2 Wealth percentile changes by native language

In Figure 4 we present the distribution of wealth index percentile changes for Spanish-speaking households and households that belong to a linguistic minority. Two results are consistent with Spanish-speaking households having a significant mobility advantage.

First, notice that the share of Spanish-speaking households that experienced a positive shift or stayed in the same position (51.42%) is larger than that of the households that belong to a linguistic minority (40.98%). In fact, the distribution is fairly well balanced between positive and negative shifts for the Spanish-speaking group, but the distribution for households that belong to a linguistic minority is skewed towards negative changes.

Second, the two groups of households have a similar likelihood of experiencing a shift of at least 20 percentiles between 2002 and 2016 (34.83% and 31.52%, respectively). However, Spanish-speaking households are more likely to experience a positive 20-percentiles shift than households that belong to a linguistic minority (18.57% vs 9.81%), while the latter exhibit a higher probability of experiencing a negative 20-percentiles shift than Spanish-speaking households (21.72% vs 16.26%).

Figure 4: Wealth index percentile changes by native language



5.3 Transition matrices by native language

Table 7 presents transition matrices for Spanish-speaking households and households belonging to a linguistic minority. Every entry for each group indicates the probability of being in the column quintile of the entire wealth distribution by Round 5 (2016) conditional on starting in the row quintile of the entire wealth distribution in Round 1 (2002).

Overall, the probabilities of experiencing a quintile change between Rounds 1 and 5 are fairly similar between the two groups of households and are around 40%. However, we observe large differences in the likelihoods of experiencing positive and negative changes, consistent with the evidence presented above. Three results are worth highlighting.

First, if one averages the probabilities of positive and negative quintile changes, Spanish-speaking households exhibit a greater probability of experiencing an upward shift (30.80% vs. 20.32%) while households that belong to a linguistic minority have a larger probability of a downward shift (37.48% vs. 29.21%). Second, Spanish-speaking households starting in the bottom quintile are more likely to leave this quintile than they are to stay in it. The opposite is true for linguistic minorities. And, third, households that belong to a linguistic minority and start in the top quintile are more likely to leave it than they are to stay in it, while the opposite is true for Spanish-speaking households.

Table 7: Transition matrices by native language groups

		Native Spanish (R5)					Linguistic Minorities (R5)				
R1	**Stat**	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
Q1	Pr.	43.69	28.83	15.77	7.21	4.50	65.26	26.84	6.32	1.05	0.53
	Obs.	97	64	35	16	10	124	51	12	2	1
Q2	Pr.	22.27	35.71	26.47	11.34	4.20	50.83	33.70	9.94	3.87	1.66
	Obs.	53	85	63	27	10	92	61	18	7	3
Q3	Pr.	5.57	22.91	32.82	26.32	12.38	23.16	35.79	24.21	13.68	3.16
	Obs.	18	74	106	85	40	22	34	23	13	3
Q4	Pr.	1.88	8.85	28.69	29.49	31.10	5.88	13.73	35.29	33.33	11.76
	Obs.	7	33	107	110	116	3	7	18	17	6
Q5	Pr.	0.00	1.96	14.57	25.49	57.98	3.70	3.70	24.07	38.89	29.63
	Obs.	0	7	52	91	207	2	2	13	21	16

Notes: The table represents two transition matrices for the main sample, which has been divided by household head first language. For every pair of initial wealth quintile and final wealth quintile, two statistics are presented. **Pr**: The proportion of households that started in the row quintile and finished in the column one. In other words, the probability of finishing in the column wealth quintile conditional on starting on the row one. **Obs**: The number of households that started in the row quintile and finished in the column one.

5.4 Conditional wealth gaps by native language

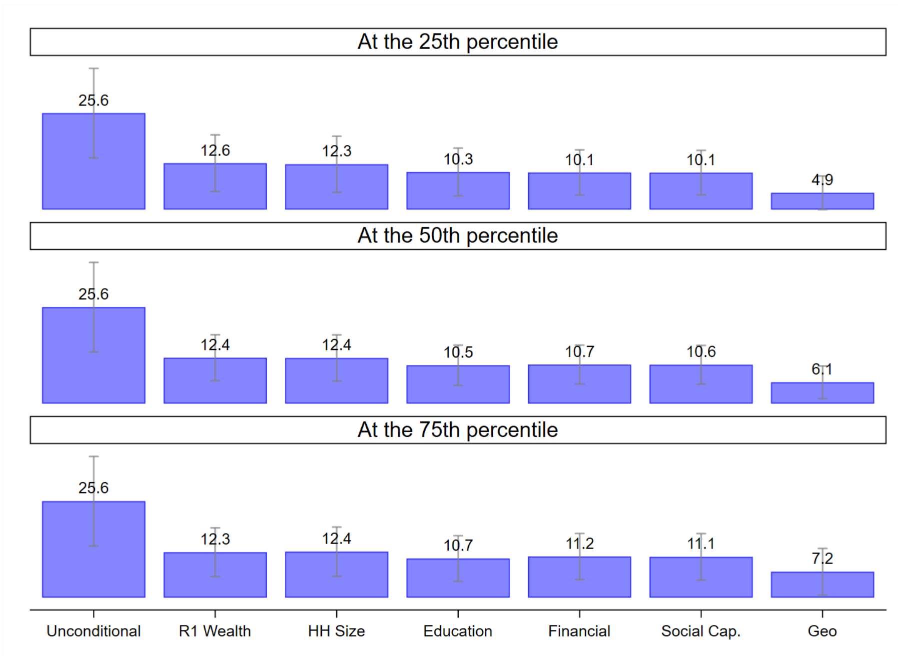
In this subsection we aim to analyze how much of the mobility gap between Spanish-speaking households and those that belong to a linguistic minority can be related to observable characteristics that can impact these household's ability to generate and accumulate wealth. Thus, we will follow Chetty et al. (2019) and estimate how much of the mobility gap persists after controlling for these characteristics.

For this, we estimate the following regression:

$$y_{i,R5} = \alpha_0 + \beta_0 y_{i,R1} + \alpha_1 Spanish_i + \beta_1 y_{i,R1} \times Spanish_i + \gamma' X_{i,R1} + \varepsilon_{i,R5} \quad (3)$$

where $X_{i,R1}$ is a vector of household characteristics measured in Round 1. In this specification, the wealth gap at a given initial wealth index percentile (y_{R1}) is given by: $\Delta_{y_{R1|X}} = \alpha_1 + \beta_1 y_{R1}$. We will estimate this gap at different values of y_{R1} and controlling for an increasing number of household characteristics. We consider five subsets of household characteristics: household size, education, financial, social capital and geographical characteristics. In Figure 5 we present our estimates of $\Delta_{y_{R1|X}}$ at the 25th, 50th and 75th percentiles of Round 1 wealth, and after progressively adding the five subsets of controls mentioned above.

Figure 5: Conditional wealth gaps by native language



Notes: The figure reports the conditional wealth gaps as estimated with Equation 3 at three different initial wealth percentiles (panels). First, we report the unconditional gap in the first bar of each panel. Subsequent bars sequentially add controls to this gap. **R1 Wealth**: The initial wealth percentile of the household. **HH Size**: The number of people who lived in the house. **Education**: The household head's years of education (measured in Round 2), and an indicator of whether the household head is disabled. **Financial**: An indicator of house ownership, an indicator of serious debt outstanding, the number of remittances sources and whether the household receives food donations. **Social Capital**: Dichotomous variables of different groups that support the household. **Geo**: An urban-rural indicator and community sites fixed effects. Unless otherwise indicated, all controls are measured in Round 1. We report 95% confidence intervals for each bar.

The first estimate corresponds to the unconditional Round 5 wealth gap between Spanish-speaking households and those that belong to a linguistic minority. This amounts to 25.6 percentiles. If one conditions this gap to households that start in the 50th percentile of the wealth distribution, the difference between Spanish-speaking and linguistic minorities is halved down to 12.4 percentiles. Remarkably, no significant differences are produced by introducing household size, education, financial and social capital controls. Urban-rural and community fixed effects, however, further reduce the Spanish-linguistic minority wealth gap down to 6.1 percentiles. Notice that these results are robust to conditioning the initial wealth at the 25th and 75th percentiles of the Round 1 wealth distribution.

Overall, the results presented in this section reveal a significant mobility advantage in favor of Spanish-speaking households. Around 50% of this gap can be related to their geographical location but a significant difference of almost 6 percentiles between Spanish-speaking and linguistic minorities households persists after controlling for a comprehensive set of characteristics that can influence their ability to accumulate wealth.

6 Good or bad mobility?

In previous sections, we analyzed different measures of mobility for the Peruvian case between 2002 and 2016. We concluded that there has been a moderately large degree of mobility during this period of high economic growth, although mobility stagnates once household heads reach their late 30s/early 40s (conditional on their position in 2002). Importantly, there exists a statistically significant mobility gap between native Spanish-speaking households and linguistic minorities, which was estimated at 12.7 percentiles.

This native language mobility gap persists even after controlling for various household characteristics that impact their ability to accumulate wealth. This suggests that native language itself (and not only the variables with which it correlates) matters for a households' ability to improve its position in the wealth distribution of Peruvian society. Just like native language, multiple other factors outside of a household's control can also impact their ability of moving up the economic ladder. This is clearly undesirable in terms of equality of opportunity.

Building on that observation, in this section we are interested in determining to what extent do characteristics outside of a households' control influence mobility. In the previous section, the estimation of the native language mobility gap was an exercise of statistical inference: its main purpose was to determine whether a mobility gap explained by native language existed in the first place and to quantify it. In the current section, we switch our focus to a predictive exercise, where we will analyze the explanatory power on mobility of variables that are outside of a household's control. The question now is: How much do variations in such characteristics predict changes in mobility?

6.1 Mobility from an equality of opportunity perspective

We propose that mobility is an outcome whose societal convenience depends on what causes it and not only on its occurrence. Our justification for this claim (and the upcoming empirical analysis) is based on the literature on equality of opportunity (see, for example, Rawls (1971), Sen (1980), Dworkin (1981a, 1981b), Arneson (1989), and Cohen (1989)). In particular, it follows the work of Roemer (1993; 1998) most closely.

In Roemer's framework, individuals are characterized by a set of circumstances, effort, and advantages. Advantages represent something desirable for the individual, like income or consumption, for example. Circumstances are factors beyond the individual's control, like sex, race, the education of their parents and their place of birth, for example. In this framework, differences in advantages due to circumstances are morally unjustified and warrant redistributive policies, as individuals cannot be held responsible for them since they are merely the result of the "birth lottery". This is the *compensation principle*. Effort is usually assumed to be unobservable. In contrast to circumstances, differences in advantages due to different degrees of effort are morally justified and should be preserved. This constitutes the *reward principle*. Finally, the acquisition of advantages is typically assumed to be completely determined by effort and circumstances.[5]

Under this framework, the population can be divided into different groups of Roemerian types, which are defined as individuals that share the exact same set of circumstances. If there exists equality of opportunity we should observe similar distributions of advantages across Roemerian types, because i) circumstances would be playing play no role in shaping outcomes and ii) effort would have a similar distribution between Roemerian types.[6]

Based on these premises, one of the key conclusions of this literature is that income inequality is not morally wrong in itself, but only insofar as it is attributable to inequality of opportunity. We consider that this same conclusion also applies to economic mobility. This follows directly from considering individual mobility to be an advantage and applying the previous framework, just like one would do when thinking of income or consumption as advantages. Hence, differences in mobility between individuals in a society would only be desirable as long as they are not caused by factors beyond their control (circumstances).

How is this linked with overall mobility in a society? A society is typically considered to be highly mobile if there is a high degree of exchange in the positions of individuals in the income or wealth distribution through time. A high degree of society-wise positional change is the natural consequence of there being large differences in individual mobility across people. Consequently, high levels of overall mobility in a society can be undesirable if what fuels it are not people's choices or exerted effort but discrimination against linguistic minorities, for example, and circumstances, in general. We consider that this approach is important because it provides us with a normative analysis of mobility.

In the next subsections, we apply a simple and intuitive methodology in order to estimate a lower bound for the fraction of mobility that is *caused* by circumstances. This would constitute a lower bound for the fraction of *bad mobility*, as the share of mobility caused by circumstances would be undesirable from an ethical point of view under the theoretical framework discussed above. Intuitively, the larger this fraction, the less meritocratic is observed mobility.

6.2 Empirical strategy

To estimate the share of bad mobility, we apply the strategy proposed by Ferreira and Gignoux (2011). In their paper, the authors estimate the share of inequality in income and consumption that is caused by inequality of opportunity. In what follows, we will use a measure of individual-level mobility to determine what percentage of its variability is caused by circumstances. Consider the

[5] This is a deterministic framework that ignores the role of luck. However, as discussed in Roemer and Trannoy (2016), luck can be incorporated to the framework. For example, Lefranc, Pistolesi, and Trannoy (2009) consider that "residual luck" is a necessary addition to effort and circumstances in order to account for all possible factors that determine advantages. They do not recommend complete compensation of luck, but only that it should not be correlated with circumstances.

[6] If effort does not have a similar distribution between Roemerian types, one should consider that the difference is caused by circumstances, which invalidates the assumption of equality of opportunity.

following linear model:

$$m_i = \theta + \delta' x_i + \gamma u_i \tag{4}$$

$$u_i = \tau + \phi' x_i + v_i \tag{5}$$

where m_i is an individual's measure of mobility, x_i is a vector of all the circumstances of the individual, and u_i is a scalar that summarizes the level of effort of the individual. Equation (4) provides a model for mobility (m_i), which is completely determined by circumstances (x_i) and effort (u_i), as in the Roemerian framework. Equation (5) provides a model for effort (u_i), which is also a function of circumstances (x_i) and of an idiosyncratic level of effort (v_i). Hence, this model allows for circumstances to have a direct impact on mobility (δ), but also an indirect one that is mediated by effort ($\gamma \cdot \phi$).

We will estimate a reduced form of the previous model by regressing individual mobility on a set of (observable) circumstances.

$$m_i = \lambda + \pi' x_i + \epsilon_i \tag{6}$$

Under equality of opportunity, circumstances should have no impact at all on mobility, neither direct nor indirect. Hence, the explanatory power of circumstances on mobility should be zero.

How can we reach a conclusion about circumstances in general if we observe only a subset of them? We are certainly unable to observe all relevant circumstances. Moreover, it is likely that some unobserved circumstances are correlated with the observed ones. This complicates identification of the *marginal* effects of each specific observed circumstance, but not necessarily the estimation of the *overall* effect of all circumstances. In other words, omitted variable bias would be a problem if our goal were to provide consistent estimates for the coefficients of each circumstance. However, our goal is different, as we aim to estimate the fraction of mobility that is explained by *all* circumstances, both observed and unobserved. More specifically, we will estimate a *lower bound* for this fraction.

As Ferreira and Gignoux (2011) argue, observing a subset of circumstances is enough for this. They prove formally that by estimating the explanatory (predictive) power of a strict subset of circumstances on a Roemerian advantage, one actually obtains a lower-bound estimate of the degree of inequality of opportunity. The logic of this result rests on the observation that if all circumstances were observable, one could estimate their joint explanatory power and that would be equal to the true degree of inequality of opportunity. As the joint explanatory power of circumstances increases as more circumstances are added, any estimate of a subset of all circumstances is a lower bound for inequality of opportunity. For our purposes, we will use the R-squared of the reduced form regression as a measure of the explanatory power of observed circumstances. Following Ferreira and Gignoux (2011), this would be a lower-bound estimate for the fraction of society-wide mobility that is caused by circumstances and is, consequently, morally unjustifiable.

A potential problem with this approach is that the R-squared of a regression estimated with OLS increases monotonically as more covariates are included in the regression. Hence, adding many circumstances could lead to overfit the data and to overstate their real explanatory power. This is a potentially relevant concern, as the literature on opportunity inequality recommends to consider as many circumstances as possible in order to tighten the lower bound. To the best of our knowledge, this concern has not been heavily considered in the past, mainly due to a lack of data on relevant circumstances, which reduces the likelihood of overfit.

To address this problem we use five-fold cross validation. Our main sample is divided in circumstance-balanced subsets. Then we estimate the reduced form five times, leaving a different one of the five subsets unused each time. Each of these five estimations constitutes a model training, and we test each of them in the respective left out subset of the sample by predicting mobility in this subset and computing the corresponding out of sample pseudo R-squared: the square of the correlation between the observed and predicted mobility. Finally, we report the average out of sample R-squared.

6.2.1 Measuring individual mobility ($m_{i,t}$)

Choosing a good measure of individual mobility is not a trivial decision. In fact, to the best of our knowledge no paper has proposed a measure of individual mobility that is based on the rank-rank framework. There exist several measures that characterize mobility at the population-level, with the rank-rank slope (β) being one of the most prevalent, but there is not a commonly accepted analog to characterize mobility at the individual level. To guide our choice of a good measure of mobility, we consider that it must satisfy the following three conditions: i) it must vary at the individual level, ii) it must draw a comparison between an individual's final and initial positions, and iii) it must be comparable across individuals. In what follows, we will explain why we think that those three conditions are necessary and why the residual of the rank-rank equation is a natural candidate that satisfies all of them. Finally, we will provide some further arguments as to why we consider it to be an adequate measure of mobility.

Recall that the rank-rank equation allows us to predict the final wealth position for each household using their initial position as the only regressor. Conditional on a household's initial wealth position, the rank-rank residual measures the vertical distance between the predicted final wealth position and the one actually obtained. Therefore, the rank-rank residual quantifies how much better or worse each household did relative to (the final position of) households that started in the same initial position.

Our first condition for an individual measure of mobility is that it must vary at the individual level. In other words, the mobility of each *individual* household must be quantified. This is necessary for this exercise given that we want to estimate a relationship between an individual's circumstances and that same individual's mobility. The residuals satisfy this because each household has a different residual. Notice that this is unlike the rank-rank coefficients (α and β), which are properties of the sample. The rank-rank coefficients represent structural or economy-wide factors by which the initial positions determine the expected value of the final ones, whereas the residuals account for the idiosyncratic household level factors that are unrelated to the initial positions.

Our second condition for an individual measure of mobility is that it must draw a comparison between an individual's final and initial positions. This is also a natural condition for a variable that aims at measuring mobility, as almost by definition measuring mobility involves quantifying the evolution of one outcome between two periods of time: one is not interested in $y_{i,t-s}$ or $y_{i,t}$ in isolation, but in a quantity that reflects a difference between them. Mathematically, one can think of mobility as necessarily being a function f whose arguments are $y_{i,t-s}$ and $y_{i,t}$, Mathematically, residuals also satisfy this second criterion, as they are defined as: $e_{i,t} = y_{i,t} - \hat{\alpha} - \hat{\beta} y_{i,t-s}$. Therefore, they are indeed a function of $y_{i,t-s}$ and $y_{i,t}$, and are an adjusted difference between them.

Our third condition is that the mobility measure must be comparable across households. In particular, we consider that it must allow for meaningful comparisons across households with different initial positions in the wealth distribution. To motivate this condition, consider the case of $\Delta y_{i,t}$, a seemingly natural candidate for an individual measure of mobility. $\Delta y_{i,t}$ does not allow for meaningful comparisons across households with different initial positions because the range of values it can take is mechanically determined by the value of $y_{i,t-s}$. For example, a household with $y_{i,t-s} = 10$, can only have $\Delta y_{i,t} \in [-10, 90]$. Similarly, a household with $y_{i,t-s} = 100$, can only have $\Delta y_{i,t} \in [-100, 0]$. Such a variable is not useful in terms of drawing conclusions about what predicts mobility within a cross-section of individuals. Why? Because basically any circumstance that predicts a low initial position will end up predicting a positive value for $\Delta y_{i,t}$, but not because that circumstance confers an advantage in terms of what is commonly understood as moving up the economic ladder, but simply because it predicts an initial position from which one can almost only move up.

It is important to stress that we are not imposing this third condition for convenience or due to the requirements of our goal in this section. We think that it is sensible to require some degree of comparability across households in order to talk meaningfully about mobility. However, we are not the first to impose such a requirement. The previously explained reasoning and general way of thinking about mobility has been (with varying degrees of explicitness) common in the more recent literature. For example, when Chetty et al. (2019) estimate the impact of race on mobility and conclude that White Americans have a mobility advantage over Black Americans, the baseline comparison is not an unconditional comparison of income ranks. An unconditional comparison would simply reflect the average White-Black gap in the income ranks distribution. What makes the income rank gap a mobility gap is precisely the fact that it is calculated by estimating separate rank-rank equations for Black and White Americans: the mobility gap is defined as the vertical difference between the expected final positions of Black and White Americans that started at the exact same position in the economic distribution. Implicitly, this reflects the notion that a necessary condition to make meaningful mobility comparisons is to account for the role of initial positions.

Residuals satisfy the third condition of comparability across households with different initial positions, because the values a residual can take are not mechanically related to the values of the initial positions. Residuals can be either positive or negative, and larger or smaller in absolute value regardless of where a household starts in the wealth distribution. Critically, they allow for comparisons across households with different initial positions given that they are the result of controlling for them. Conceptually, focusing on residuals as mobility is equivalent to a theoretical experiment where one equalizes initial positions across households, allows them to move, and then compares how well each household did relative to others. This is what allows for a meaningful comparison between households.

Appendix B contains descriptive statistics of our estimated rank-rank residuals and discusses their links with previous statistics of mobility presented in this paper.

6.2.2 Circumstances (x_i)

Circumstances are traditionally defined as factors that are completely beyond an individual's control, like the education of their parents or their place of birth. Given that our analysis is at the household-level, we will consider factors that the household head cannot control as the circumstances of the household. As we have done before, for each circumstance we consider the household head's information if available. Otherwise, we use the available information of one parent, provided that they were part of the household in the first round (prioritizing the father's information if available).

For our empirical analysis, we will use four different circumstances that can be grouped into two sets. The first one is the set of *strict circumstances* and includes three circumstances of the household head: i) native language, ii) self-reported ethnicity, and iii) region of birth. Peru has a total of 24 regions and 1 constitutional province. They are the first-level administrative subdivision of the country. These three circumstances are indisputably outside the control of any person as they are determined at birth and are immutable by nature. These circumstances are categorical variables, so we add them to our regression in the form of group dummies.

Our second set of circumstances follows a strand of more recent research on opportunity inequality, which proposes that the classical definition of circumstances is too narrow and that it should be widened to consider *before age of consent circumstances*. The key argument is that an individual should not be held responsible for their characteristics or lack of achievement before an *age of consent* (Roemer and Trannoy 2016, Hufe et al. 2017). According to this view, differences in opportunities due to choices an individual made as a kid are not morally justifiable and should be corrected by the compensation principle. Our fourth circumstance variable falls within this broader definition and is years of schooling censored at eleven years. Hence, this variable is only measuring how many years of primary and secondary education someone received, which are typically finished just before someone reaches adulthood.

We classify this variable as part of the second set of circumstances to be conservative, even though it could be argued that this is indeed a more strict circumstance. The decision of going to primary or secondary school is most likely not made by a child, but by its parents. Moreover, it is likely that several cases of children that drop out of school in Peru are caused by their parents requiring them to work since an early age.

6.3 Results

Table (8) presents our estimates for the explanatory power of circumstances on mobility, which constitute lower-bound estimates of the degree of inequality of opportunity in mobility (Ferreira and Gignoux (2011). As in previous instances, we estimate Equation (6) for four different time horizons:

up to Round 2 and onwards. For each time horizon, we estimate Equation (6) using the set of strict circumstances first and include the before-age-of-consent circumstance next. Since results barely change when expanding the set of circumstances, we will mention the estimates corresponding to the extended set of of circumstances when discussing the results. These estimates appear in the bottom three rows of Table (8).

For each time horizon and set of circumstances, we additionally estimate Equation (6) in two separate subsamples: one that contains households in the bottom 50% of the Round 1 wealth distribution and another one that contains the remaining 50%. This sample partition will allow us to say if opportunity inequality is different along the economic ladder.

Finally, in addition to the typical OLS R-squared, Table (8) also reports the five-fold cross validated (CV) R-squared of each regression. It measures the out-of-sample explanatory power of the model and, therefore, eliminates the risk of overestimating the true predictive power of circumstances. As one can see in Table (8), the CV R-squared of each regression is lower than their OLS counterpart, which confirms the risk that we may be overfitting the data. To be conservative, we will discuss the results in terms of the CV R-squared, which appears in the last four columns of Table (8).

Table 8: Mobility decomposition by circumstances

		Ordinary Least Squares				Cross Validated			
		Round 2	Round 3	Round 4	Round 5	Round 2	Round 3	Round 4	Round 5
Strict	All	6.86	8.55	10.49	10.90	3.50	4.91	6.63	7.31
Circumstances	R1 Bottom 50%	13.81	15.27	20.75	22.14	7.20	9.36	14.39	16.48
	R1 Top 50%	6.30	7.28	6.98	5.89	1.54	1.13	0.83	0.91
+ Household	All	7.30	8.99	11.06	11.73	4.00	5.22	7.08	8.03
Education	R1 Bottom 50%	14.15	15.83	21.61	23.33	7.55	9.65	15.09	17.35
	R1 Top 50%	8.51	8.26	8.04	7.33	2.85	1.60	1.26	1.87

Notes: The table reports the explanatory power of circumstances on mobility measured as the R-Squared of estimating Equation 6. The dependent variable is mobility by the column round as measured by the residuals of the rank-rank equation estimated by that round as well. The independent variables are circumstances: (i) strict circumstances include the household head's native language, ethnicity, and region of birth, (ii) household education is the household's head years of education censored to eleven (end of highschool). We limit the main sample to the households where all of these variables are observed to ensure comparability. Estimations are done first using the all observations, and then dividing the main in two: Round 1 Bottom 50% and Round 1 Top 50%. Two methods of estimation are used: Ordinary Least Squares and 5-Fold Cross Validation.

For our longest time horizon of 14 years (Round 5), we obtain that circumstances explain at least 8% of variations in mobility in the entire sample. In terms of the Roemerian framework, this means that at least 8% of total observed mobility is caused by opportunity inequality and is thus morally unjustifiable. This is what we call bad mobility. We consider that this result is modest in size, although we want to remark that it is a lower bound for the real share of mobility that is caused by factors outside of a household's control.

Holding the time horizon fixed at 14 years (Round 5), we can compare the share of bad mobility between the bottom and top halves of the wealth distribution. This comparison reveals a large difference between the two subgroups. Standing at 17.35%, the predictive power of circumstances on mobility is a lot higher for the bottom 50% of the population than for the top 50%. In fact, the circumstances that we observe (first language, ethnicity, region of birth, and education) have an explanatory power on mobility of almost zero for the top 50% of our sample. In other words, we do not find evidence of bad mobility for this group. This suggests that almost all of the observed mobility for the top 50% of the population is due to circumstances that are uncorrelated with the ones we observe, or due to effort. This is consistent with equality of opportunity in mobility. On the other hand, almost a fifth of mobility for the bottom 50% due to inequality of opportunity, and therefore bad mobility.

Focusing on our estimates as the time horizon lengthens, we observe that the predictive power of circumstances on mobility increases with time. The CV R-squared for the complete sample increases from 4% after 4 years to 8% after 14 years. Note that this increase is not a mechanical consequence from the documented increase in mobility over this same time horizon. Hence, we find that as mobility increases with a longer time horizon, the fraction of it that is caused by opportunity inequality increases as well.

Importantly, the increase in the share of bad mobility in the complete sample is exclusively driven by the bottom 50%. For the most disadvantaged half of the population, inequality of opportunity in mobility increases monotonically as the time horizon lengthens, going from at least 7.55% up to at least 17.35%. On the other hand, the share of mobility caused by inequality of opportunity remains almost constant and low in each of the four time horizons for the top 50%. A corollary from these two observations is that the gap in the share of bad mobility between the bottom and top 50% is increasing in the time horizon.

7 Conclusion

In this article we have employed longitudinal data collected by the Young Lives project to study intragenerational economic mobility in Peru over a period of fourteen years (2002-2016). We were able to characterize the evolution of intragenerational mobility because we observe the asset holdings of households in five different years. We found that both absolute and relative mobility were increasing in the time horizon during those years. However, mobility stagnates after 11 years and initial positions end up explaining about 50% of the variation in the final ones.

After assessing different measures of mobility, we concluded that mobility was moderately large during this period. On average, households experienced a significant shift (in either direction) of 17 percentiles in the wealth distribution. On top of that, about a third of households experienced a transition of over 20 percentile ranks in absolute value. However, this happens in a setting where there is some degree of sticky floors and ceilings and where most large changes occurred in the middle of the socioeconomic distribution.

Moreover, mobility was not equally distributed. We find that positive transitions are skewed in favour of households where the only native language is Spanish. We document significant differences in absolute mobility between native Spanish-speaking households and linguistic minorities. In fact, native Spanish-speaking households are expected to end 12.69 percentile ranks higher in the wealth distribution relative to a linguistic minority household with exactly the same initial position 14 years ago. This mobility gap shrinks but persists even after controlling for several variables that influence the ability of households to accumulate wealth.

Finally, we estimated the share of inequality of opportunity in mobility. To do so, we proposed a simple yet intuitive way of quantifying individual level mobility by using the residuals of the rank-rank equation. We then estimated that circumstances cause at least 8.03% of the variations in mobility for the whole sample. Importantly, their explanatory power rises to 17.35% for households in the bottom 50% of the Round 1 wealth distribution and is a mere 1.87% for top 50%. Going by these findings, it seems like Peruvians in the top 50% enjoy more equality of opportunity than those in the bottom 50%.

In face of the resurgent interest in mobility, we consider that the literature on equality of opportunity provides a suitable normative framework with which to study mobility. Applying its principles to this context, not all economic mobility is necessarily good and it is important to determine to what extent it is morally justified or not by looking at its causes. We think that further work is required to fully integrate these set of normative principles to the study of intergenerational mobility (and mobility in general), but that this would constitute a valuable step forward for a literature about issues that have important social justice implications.

References

Acciari, Paolo, Alberto Polo, and Giovanni L Violante. 2022. “And yet it moves: Intergenerational mobility in Italy.” *American Economic Journal: Applied Economics* 14 (3):118–63.

Alesina, Alberto and Paola Giuliano. 2011. “Preferences for redistribution.” In *Handbook of social economics*, vol. 1. Elsevier, 93–131.

Arneson, Richard J. 1989. “Equality and equal opportunity for welfare.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 56 (1):77–93.

Behrman, Jere R., Whitney Schott, Subha Mani, Benjamin T. Crookston, Kirk Dearden, Le Thuc Duc, Lia C. H. Fernald, and Aryeh D. Stein. 2017. “Intergenerational Transmission of Poverty and Inequality: Parental Resources and Schooling Attainment and Children’s Human Capital in Ethiopia, India, Peru, and Vietnam.” *Economic Development and Cultural Change* 65 (4):657–697.

Braun, Sebastian Till and Jan Stuhler. 2017. “The Transmission of Inequality Across Multiple Generations: Testing Recent Theories With Evidence From Germany.” *The Economic Journal* 128 (609):576–611.

Briones, Kristine. 2017. “’How many rooms are there in your house?’ Constructing the Young Lives Wealth Index.” *Young Lives Technical Note 43* .

Britto, Diogo G.C., Alexandre de Andrade Fonseca, Paolo Pinotti, Breno Sampaio, and Lucas Warwar. 2022. “Intergenerational Mobility in the Land of Inequality.” .

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2019. “Race and Economic Opportunity in the United States: an Intergenerational Perspective.” *The Quarterly Journal of Economics* 135 (2):711–783.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. “Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States.” *The Quarterly Journal of Economics* 129 (4):1553–1623.

Cohen, Gerald A. 1989. “On the currency of egalitarian justice.” *Ethics* 99 (4):906–944.

Dahl, Molly W and Thomas DeLeire. 2008. *The association between children’s earnings and fathers’ lifetime earnings: estimates using administrative data*. University of Wisconsin-Madison, Institute for Research on Poverty Madison . . . .

Dang, Hai-Anh H. and Peter F. Lanjouw. 2018. “Poverty Dynamics in India between 2004 and 2012: Insights from Longitudinal Analysis Using Synthetic Panel Data.” *Economic Development and Cultural Change* 67 (1):131–170.

Dworkin, Ronald. 1981a. “What is Equality? Part 1: Equality of Welfare.” *Philosophy and Public Affairs* 10:185–246.

———. 1981b. "What is Equality? Part 2: Equality of Resources." *Philosophy and Public Affairs* 10:283–345.

Escobal, Javier and Eva Flores. 2008. "An assessment of the Young Lives sampling approach in Peru." *Young Lives Technical Note 3* .

Ferreira, Francisco H. G. and Jérémie Gignoux. 2011. "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America." *Review of Income and Wealth* 57 (4):622–657.

Giuliano, Paola and Antonio Spilimbergo. 2008. "Growing up in bad times: Macroeconomic Volatility and the Formation of Beliefs." Tech. rep., UCLA mimeo.

Gregg, Paul, Lindsey Macmillan, and Claudia Vittori. 2016. "Moving Towards Estimating Sons' Lifetime Intergenerational Economic Mobility in the UK." *Oxford Bulletin of Economics and Statistics* 79 (1):79–100.

Güell, Maia, Michele Pellizzari, Giovanni Pica, and José V. Rodríguez Mora. 2018. "Correlating Social Mobility and Economic Outcomes." *The Economic Journal* 128 (612):F353–F403.

Hufe, Paul, Andreas Peichl, John Roemer, and Martin Ungerer. 2017. "Inequality of income acquisition: the role of childhood circumstances." *Social Choice and Welfare* 49 (3-4):499–544.

Iversen, Vegard, Anirudh Krishna, and Kunal Sen. 2019. "Beyond Poverty Escapes—Social Mobility in Developing Countries: A Review Article." *The World Bank Research Observer* 34 (2):239–273.

Krishna, Anirudh. 2011. *One illness away: Why people become poor and how they escape poverty.* Oxford University Press.

Landersø, Rasmus and James J. Heckman. 2016. "The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US." *The Scandinavian Journal of Economics* 119 (1):178–230.

Layard, Richard. 2006. *Hapiness.* Penguin Books.

Lefranc, Arnaud, Nicolas Pistolesi, and Alain Trannoy. 2009. "Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France." *Journal of Public Economics* 93 (11-12):1189–1207.

Modalsli, Jørgen. 2016. "Intergenerational Mobility in Norway, 1865–2011." *The Scandinavian Journal of Economics* 119 (1):34–71.

Neidhöfer, Guido, Joaquín Serrano, and Leonardo Gasparini. 2018. "Educational inequality and intergenerational mobility in Latin America: A new database." *Journal of Development Economics* 134:329–349.

Rawls, John. 1971. *A Theory of Justice.* Harvard University Press, Cambridge.

Roemer, John E. 1993. "A pragmatic theory of responsibility for the egalitarian planner." *Philosophy & Public Affairs* :146–166.

Roemer, John E. 1998. *Equality of Opportunity*. Harvard University Press, Cambridge.

Roemer, John E. and Alain Trannoy. 2016. "Equality of Opportunity: Theory and Measurement." *Journal of Economic Literature* 54 (4):1288–1332. URL https://www.aeaweb.org/articles?id=10.1257/jel.20151206.

Sen, Amartya. 1980. "Equality of what?" *The Tanner lecture on human values* 1:197–220.

Thorp, Rosmery and Maritza Paredes. 2010. *Ethnicity and the Persistence of Inequality: The Case of Peru*. Palgrave Macmillan.

Torche, Florencia. 2015. "Intergenerational Mobility and Gender in Mexico." *Social Forces* 94 (2):563–587.

Winkelried, Diego and Javier Torres. 2018. "Economic mobility along the business cycle. The case of Peru." *Applied Economics* 51 (18):1894–1906.

Young Lives. 2018. "Survey Design and Sampling: Round 5 in Peru." https://www.younglives.org.uk/content/survey-design-and-sampling-round-5-peru. Last access: 2020-04-30.

Appendices

A Robustness checks of mobility estimations

As we explained, we have assumed that, for every child, two observations in different points in time correspond to the same household if the child lived with at least one of their parents in both observations. We additionally strengthened our sample selection criterion by requiring that each household also remained the same in the longest time horizon (between Rounds 1 and 5).

We conducted a robustness check on our sample selection procedure by imposing a stricter requirement. We additionally required that the child lived with at least one of their parents in each round from the first one to the round of interest. Naturally, under this definition, the number of observations we keep becomes smaller.

In Table A1, we show the number of observations and the results of estimating the rank-rank equation in our main sample and in the sample with the additional restriction. Our estimates are practically identical, which is expected given the small difference in the number of observations between the two samples. This suggests that we are not significantly altering the results by using a looser sample selection criterion that the one introduced in this appendix.

Table A1: Comparison of observation counts and rank-rank slope coefficients between the main and restricted samples

	Round 2		Round 3		Round 4		Round 5	
	N | (R^2)	β	N | (R^2)	β	N | (R^2)	β	N | (R^2)	β
Main Sample	2,074	0.837	2,068	0.769	2,043	0.718	2,106	0.706
	(0.701)	(0.019)	(0.592)	(0.028)	(0.516)	(0.036)	(0.497)	(0.038)
Further Restricted	1,954	0.838	1,953	0.770	1,953	0.721	1,936	0.714
	(0.702)	(0.020)	(0.593)	(0.029)	(0.521)	(0.038)	(0.509)	(0.040)
Years Since R1	4		7		11		14	

Notes: The table presents the number of observations (N), the R^2, and the estimated rank-rank coefficients (β) for four different time intervals. The dependent variable in the rank-rank regression is the household wealth index percentile at the round indicated in the column, and the only independent variable is the household wealth index percentile in the first round. In the columns with header N | (R^2), the number of observations appears first, and the R^2 is shown below, in parentheses. In the columns with header β, the point estimate appears first, and the standard errors (clustered at the sentinel site level) are shown in parentheses below. All estimated coefficients are statistically significant. The first row ("Main sample") shows results and observation counts that correspond to the main sample of the paper, whereas the second row ("Further restricted") does the same for the sample that imposes the additional restriction discussed in this appendix.

We also rule out that we are ignoring relevant age-based differences in mobility by pooling the younger and older cohorts together in our main estimations. For instance, estimating the rank-rank regression with a sample of parents that are 50 years old in the beginning of the period and 64 years old in the end would probably yield very different mobility measures than doing the same with a sample of parents that are 30 years old in the beginning and 44 in the end.[7] Our data has the benefit of containing households that were selected on the basis of having children of exactly the same ages. This limits the variance of the age of the household head within the younger and older cohorts, respectively. However, we have a group of households that had children aged one in 2002 and another with children aged 8 in the same year. Therefore, we do end up with two groups of households that have different average ages of the household head.

To address the previous concern, we divided our main sample in two subsamples: one contained households with children from the younger cohort and the other from the older cohort. We then estimated the rank-rank equation separately in each subsample. The results are shown in Table A2.

Table A2: Comparison of observation counts and rank-rank slope coefficients between the younger and older cohorts

	Round 2		Round 3		Round 4		Round 5	
	N | (R^2)	β	N | (R^2)	β	N | (R^2)	β	N | (R^2)	β
Main Sample	2,074	0.837	2,068	0.769	2,043	0.718	2,106	0.706
	(0.701)	(0.019)	(0.592)	(0.028)	(0.516)	(0.036)	(0.497)	(0.038)
Young Cohort Only	1,698	0.831	1,696	0.771	1,691	0.723	1,723	0.703
	(0.690)	(0.018)	(0.587)	(0.027)	(0.516)	(0.037)	(0.489)	(0.039)
Old Cohort Only	376	0.863	372	0.755	352	0.673	383	0.712
	(0.716)	(0.030)	(0.562)	(0.048)	(0.441)	(0.057)	(0.478)	(0.056)
Years Since R1	4		7		11		14	

Notes: The table presents the estimated rank-rank coefficients (β) for four different time intervals. The dependent variable in the rank-rank regression is the household wealth index percentile at the round indicated in the column, and the only independent variable is the household wealth index percentile in the first round. In the columns with header N | (R^2), the number of observations appears first, and the R^2 is shown below, in parentheses. In the columns with header β, the point estimate appears first, and the standard errors (clustered at the sentinel site level) are shown in parentheses below. All estimated coefficients are statistically significant. The first row ("Main sample") shows results and observation counts that correspond to the main sample of the paper. The second and third rows ("Younger cohort only" and "Older cohort only") do the same, but using only the observations that correspond to the younger and older cohorts, respectively.

As can be seen, the results are fairly similar between cohorts for each time horizon. The largest difference in the rank-rank coefficient between cohorts occurs when we analyze mobility up to Round 4. In this case, $\beta_{young} = 0.72$ while $\beta_{old} = 0.67$. Nevertheless, the gap rapidly closes and reaches just 0.011 when we analyze mobility up to Round 5. Moreover, the gap is never statistically significant. This evidence is consistent with the absence of relevant age-based differences in mobility between cohorts.

[7] This is similar to the concern for life cycle bias expressed in Chetty et al. (2014).

B Characterizing individual-level mobility

In Table B3, we present descriptive statistics of the rank-rank equation residuals. The average of the residuals is zero by construction, but other moments exhibit patterns that are consistent with what we have documented about mobility so far. For example, the variance of the residuals is increasing in the time horizon. This result is mechanically linked to the gradual reduction in the R^2 of the rank-rank regressions, which itself was linked to the decreasing rank-rank coefficient.

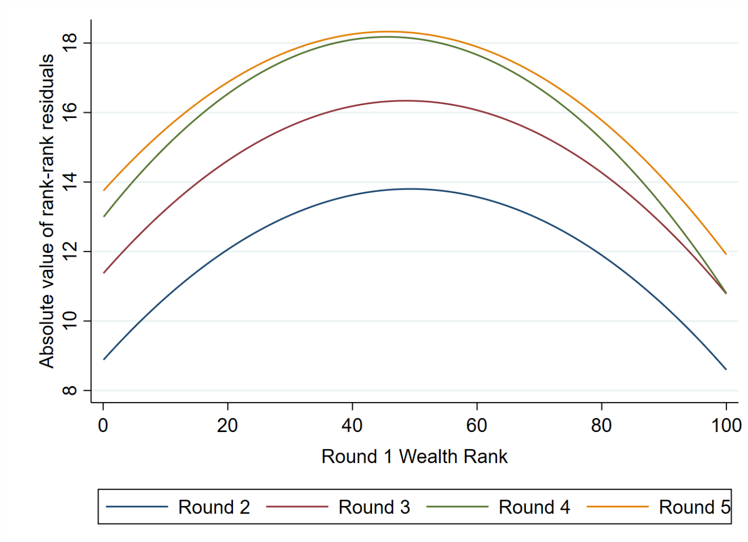
Table B3: Descriptive statistics of rank-rank residuals

	Obs.	Mean	SD	Min	P-25	Median	P-75	Max
A. Residuals								
Resids. from R2: $e\_i, 2$	2008	-0.07	15.70	-60.69	-10.36	-1.04	9.11	74.09
Resids. from R3: $e\_i, 3$	1974	-0.12	18.29	-65.01	-12.69	-0.43	12.13	69.67
Resids. from R4: $e\_i, 4$	1943	-0.13	20.02	-65.42	-14.10	-1.00	13.52	73.63
Resids. from R5: $e\_i, 5$	1988	-0.12	20.41	-65.80	-15.25	-0.51	13.77	74.42
B. Absolute value of residuals								
Abs. resids. from R2: $\|e\_i, 2\|$	2008	12.12	9.97	0.00	4.56	9.83	16.94	74.09
Abs. resids. from R3: $\|e\_i, 3\|$	1974	14.60	11.02	0.01	6.11	12.51	20.49	69.67
Abs. resids. from R4: $\|e\_i, 4\|$	1943	16.06	11.95	0.01	6.84	13.77	22.66	73.63
Abs. resids. from R5: $\|e\_i, 5\|$	1988	16.48	12.04	0.02	7.25	14.59	22.81	74.42

Notes: Rank-rank residuals are the residuals of the rank-rank regression: the dependent variable is the household wealth index percentile at the round indicated in the row and the only independent variable is the household wealth index percentile in the first round. All the residuals were estimated with our main sample. $e_{i,t}$ denotes a residual corresponding to the rank-rank equation of mobility between periods 1 and t. Panel A shows descriptive statistics of the rank-rank residuals and Panel B shows descriptive statistics of their absolute value.

The sign of each residual indicates whether the household experienced upward or downward mobility. Thus, the absolute value of a residual measures mobility in any direction. This allows us to tell whether the chances of economic mobility (in any direction) are the same along the wealth distribution. To do so, we analyze the relation between a household's residual in absolute value and their initial position. We graph this relation in Figure B1, where we have plotted quadratic regression curves for the four different time horizons.

Figure B1: Quadratic regression curves: Conditional expectation of the absolute value of rank-rank residuals on Round 1 wealth percentile rank



We observe that the curves have an inverted U shape. This suggests that mobility is higher for households in the middle of the distribution and that it is lower for households at the top and the bottom. This phenomenon is sometimes called sticky floors and sticky ceilings. Once again, this evidence is consistent with what we had found previously, but using transition matrices. The curves in Figure B1 also show an interesting asymmetry between mobility starting from the top and bottom quintiles. Even though mobility is smaller in them than in the center, the curve shows that the poorest households are a bit more mobile than the richest ones.