IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18254

Inherited Inequality in Latin America

Francisco H. G. Ferreira
Paolo Brunori
Guido Neidhöfer
Pedro Salas-Rojo
Louis Sirugue

NOVEMBER 2025

DISCUSSION PAPER SERIES

IZA DP No. 18254

Inherited Inequality in Latin America

Francisco H. G. Ferreira
LSE and IZA

Paolo Brunori
LSE and the University of Florence

Guido Neidhöfer
Turkish-German University and ZEW Mannheim

Pedro Salas-Rojo
CUNEF

Louis Sirugue
LSE

NOVEMBER 2025

ISSN: 2365-9793

ABSTRACT

Inherited Inequality in Latin America*

This paper argues that relative measures of intergenerational mobility and inequality of opportunity are closely related ways of quantifying the inheritability of inequality. We review both literatures for Latin America, looking both at income and educational persistence. We document very high levels of intergenerational persistence and inequality of opportunity for education, with inherited characteristics predicting 29% to 52% of the current-generation variance in years of schooling. Inherited circumstances are somewhat less predictive of educational achievement, measured through standardized test scores, accounting for 20% to 30% of their variance. Our estimates of inequality of opportunity for income acquisition suggest that between 46% to 66% of contemporary income Gini coefficients can be predicted by a relatively narrow set of inherited circumstances, making Latin America a region of high inequality inheritability by international standards. Our review also finds a very wide range of intergenerational income elasticity estimates, with substantial uncertainty driven by data challenges and methodological differences.

JEL Classification: D31, I39, J62, O15

Keywords: inherited inequality, intergenerational mobility, inequality of opportunity, Latin America

Corresponding author:
Francisco H. G. Ferreira
London School of Economics
Houghton Street
London WC2A 2AE
United Kingdom

E-mail: F.D.Ferreira@lse.ac.uk

* We are grateful to the Editors, Steven Durlauf and Bhash Mazumder, as well as to conference and seminar participants at the University of Chicago, the European University Institute, University College London, the Institute for Fiscal Studies, Monash University, the University of Stockholm (SOFI), and the Catholic University of Chile for helpful comments on earlier drafts. All errors are ours alone.

1. Introduction

Despite considerable intra-regional heterogeneity, as well as much uncertainty about the exact levels of income inequality, Latin America is widely believed to be one of the two most unequal regions in the world (Alvaredo et al., 2025).[2] The other candidate to this unfortunate title is Africa, but a reliable comparison between the two regions is difficult, because inequality is mostly measured for incomes in Latin America, and for consumption expenditures in Africa. In most Latin American countries, there is also considerable inequality in educational attainment and achievement (see Fernández et al., 2025), agricultural land (Gáfaro et al., 2025) and, to the extent that information is available, wealth more generally (Carranza et al., 2025).

While not immutable, these inequalities have generally proved broadly persistent over time, owing to various mechanisms of inequality reproduction (Attanasio et al., 2025). It has been argued that the region's high levels of inequality have deep historical roots, dating back at least to colonial times, when factor endowments and unequal wealth distributions gave rise to institutions that acted to entrench inequality over time (see, e.g. Sokoloff and Engerman, 2000, and Eslava and Valencia, 2025, for a recent review).

But is this persistence of cross-sectional inequality characterized by lots of 'churning' across generations – so that the poor today may just as easily be the children of rich parents as of poor parents – or is it instead characterized by persistence in socioeconomic status within families and lineages across generations? To what extent can important outcomes – such as education or income levels – be predicted by the achievements and characteristics of one's parents and grandparents, or other inherited attributes?

Despite a challenging data landscape, researchers have attempted to answer these questions for many Latin American countries for at least half a century, typically taking one of two approaches. The first approach consists of a literature on intergenerational mobility (IGM), which essentially consists of measuring the association between a particular outcome – say, household income or years of schooling – for parents and their children. The stronger that association, the more predictive parental outcomes are of their children's outcomes, and the lower is intergenerational mobility.

The second approach is associated with the literature on inequality of opportunity (IOp) and draws on a larger set of variables. This literature was not originally developed for, or interpreted as,

[2] In a recent review of the measurement of income inequality in Latin America and the Caribbean, Alvaredo et al. (2025) compile over 5600 estimates of Gini coefficients (plus many other measures) for 34 countries in the region over the last 75 years. They report wide variation in estimates for the same country-year combinations depending, most importantly, on whether household surveys, administrative tax data, distributed National Accounts estimates, or combinations of these data sources, are used. They argue that no single data source or adjustment method is clearly superior to all others, implying that there is genuine uncertainty about inequality levels in most countries in the region, although there is much more agreement on trends.

measuring intergenerational dependence. The seminal references in economics, such as Roemer (1993, 1998), van de Gaer (1993), and Fleurbaey (1994), drew on earlier work by moral philosophers and were intended as contributions to social choice theory. They adopted or developed normative principles suggesting that differences in outcomes due to the exercise of choices that people could be held responsible for were fair, whereas differences due to factors beyond people's control or responsibility were unfair and should be compensated by society. These factors that individuals should not be held responsible for became known as "circumstances", and an empirical literature developed suggesting that the extent to which circumstances could predict outcomes such as income or education could be interpreted as a measure of unfair inequality, or inequality of opportunity for that outcome (see e.g., Bourguignon et al., 2007).

We argue that, despite their different origins, these two literatures are closely related and usefully complement each other. This chapter sets out to review, summarize, and compare findings within and across the two approaches, focusing on two classes of variables: incomes and educational outcomes. The remainder of the chapter is organized as follows. Section 2 briefly describes the concept of inherited inequality as an umbrella term that can encompass both intergenerational persistence (the converse of mobility) and inequality of opportunity. The basic idea is that an important class of measures of intergenerational mobility – sometimes known as "relative" or "origin independence" measures – and measures of inequality of opportunity share the same basic structure and can be interpreted as shares of observed inequality that can be predicted by inherited characteristics.

Section 3 then reviews the literature on intergenerational mobility in the region, looking first at studies of intergenerational *educational* mobility, and then at intergenerational *income* mobility. For education, we review studies of persistence in educational attainment, as well as achievement. For income mobility, we pay close attention to the evolving data landscape and to the resulting changes in methods, concluding that – perhaps similarly to the cross-sectional inequality literature – we may be in a transitional phase where there is considerable uncertainty about the IGM estimates. Section 4 briefly reviews the early literature on inequality of opportunity in the region, also covering studies on income and education. Section 5 then focuses on the more recent approaches to inequality of opportunity, a 'second generation' of studies characterized by the use of data-driven methods to select appropriate prediction functions. This section provides some new estimates that update and extend the work of Brunori et al. (2025). Section 6 concludes.

2. Inherited Inequality: A simple framework

Relative, or origin-independent, approaches to intergenerational mobility basically consist of measuring the degree of association in a particular outcome – say, per capita household income – across generations. If we observe, for a given population, the incomes of children (y_c) and the income of their parents (y_p), then we observe the joint distribution $F(y_c, y_p)$. Like any joint distribution, this one is characterized by the marginal distributions $F_c(y_c)$ and $F_p(y_p)$, and by the copula $C(F_c, F_p)$, which describes the association between the two margins. Relative approaches to intergenerational mobility seek to summarize features of this association in various different ways. Transition matrices, which underpin many of the earlier measures of mobility, are essentially 'discretized' copula densities, with each cell giving the frequency of children in a certain interval of $F_c(y_c)$ whose parents are in some other given interval of $F_p(y_p)$. Statistics such as the Pearson correlation coefficient $\rho_{y_c y_p}$, the rank-rank correlation coefficient, or the regression coefficient of y_c on y_p (β^{IGE}, when the variables are in logs), are alternative scalar summary measures of that same association.[3]

Just as relative intergenerational mobility has been measured in a number of different ways, so too has inequality of opportunity. Ferreira and Peragine (2016) and Roemer and Trannoy (2016) provide reviews of this literature.[4] But the different approaches to measuring IOp share a common essence: a vector of "circumstances" C is identified and used to predict an outcome of interest, say income or a measure of educational achievement, through a prediction function $y_c = f(C, \varepsilon)$, generating a 'counterfactual distribution' denoted either by the vector $\hat{\boldsymbol{y}}_{\boldsymbol{c}}$ or, continuously, by $F_{\hat{Y}}(\hat{y}_c)$, where, in either case, $\hat{y}_c = \hat{f}(C)$. Because income differences due to circumstances were judged unfair, inequality of opportunity was simply inequality measured in this counterfactual distribution: $I(\hat{y}_c)$. Or, if expressed in relative terms, $I(\hat{y}_c)/I(y_c)$, where $I(y_c)$ denotes inequality in the marginal distribution of income in the children's generation, $F_c(y_c)$. In what follows, for notational simplicity, we will denote $F_c(y_c)$ interchangeably as $F_Y(y)$, and similarly drop the *c* subscript when referring to income vectors or distributions in the IOp literature.

It has recently been suggested that if one restricts the vector of circumstances to a subset $H \subseteq C$ of attributes that people inherit at birth[5] – such as their biological sex, race or ethnicity, place of birth, and family background variables such as parental income, education, occupation, etc. – then $I(\hat{y})$ or $I(\hat{y})/I(y)$, where $\hat{y} = \hat{f}(H)$, could be interpreted as *inherited inequality*: the amount or share of inequality in outcome *y* that can be predicted by a set of inherited circumstances.[6] It was also noted that some measures of relative mobility, such as the Pearson correlation coefficient, are closely related to $I(\hat{y})/I(y)$. Indeed $\rho^2_{y_c y_p} = Var(\hat{y})/Var(y)$, when $\hat{y} = \hat{\alpha} + \hat{\beta} y_p$, the predicted

[3] Although, as we shall see, the regression coefficient is not, in fact, a pure measure of association.

[4] One of the key distinctions in this literature is between 'ex-ante' and 'ex-post' measures of IOp, a distinction that originates from (slightly) different formulations of how unfair inequality should be defined. In the normative literature, these formulations are known as versions of the Principle of Compensation. For our present purposes, these distinctions would simply imply different functional forms for the prediction function $\hat{f}(\quad)$. Most of the literature we review, as well as our own novel estimates in Section 5, are of the 'ex-ante' variety. See Brunori, Ferreira and Salas-Rojo (2023) for an 'ex-post' alternative.

[5] Or up to some early age "of responsibility".

[6] See Brunori et al. (2023a) and Ferreira and Brunori (2024).

child income from a standard Galtonian regression, $y_c = \alpha + \beta y_p + u$. In the log-log version of this, equation $\rho^2_{y_c y_p}$ is the ratio of the variance of logarithms of predicted to actual child incomes.

The concept of inherited inequality is therefore one that theoretically encompasses both relative intergenerational mobility and inequality of opportunity and, empirically, lies somewhere between them. As we take a narrower and narrower view of the set of inherited characteristics and, in the limit, consider only parental income, then $H \to y_p$, then inherited inequality converges towards relative intergenerational mobility. Conversely, if we take a progressively broader view of inherited circumstances, allowing this set to include a larger and larger group of variables outside people's control or beyond their responsibility, then $H \to C$, and inherited inequality converges towards inequality of opportunity.

In practice, of course, the exact measure of inherited inequality depends on three things: the set of predictors, *H*; the specific prediction function $\hat{f}(\quad)$; and the inequality index $I(\quad)$. As we will see below, each of those choices can matter a great deal. It is nonetheless useful to bear in mind that, when we restrict the set of circumstances to include only inherited characteristics, intergenerational mobility and inequality of opportunity are closely related approaches to quantifying the extent to which today's inequality can be predicted by inherited factors. This perspective will inform our reading of both of these literatures as applied to Latin American countries in what follows.

3. Intergenerational Mobility in Latin America: A review

Latin America has attracted considerable attention in the study of the transmission of inequality from one generation to the next. In the 1960s and 1970s, research on intergenerational transmission mechanisms in the region was predominantly carried out by sociologists, who used small-scale, urban-focused surveys to study occupational mobility. Torche (2014) provides a detailed overview of these early studies. It wasn't until the 1990s that economists also began to examine this topic, using more comprehensive household surveys. This allowed a new body of literature to emerge, relying on increasingly better – if still far from ideal – data to describe intergenerational persistence processes in a more granular way. While the earlier literature, which we do not review here, had focused mostly on occupational mobility, the economics literature from the 1990s onward concentrated on educational and income mobility, and we look at these two bodies of work in turn.

Intergenerational Mobility in Education

A large share of the research on intergenerational mobility in Latin America uses education as an indicator for the socio-economic status of parents and their children. Indeed, education is a crucial factor in both current and future well-being, and it is arguably less influenced by personal preferences than income or occupation. Moreover, education has practical benefits as a measurement tool; it is less subject to lifetime fluctuations than income or earnings and it is usually completed early in adulthood (typically between ages 18 and 30). Therefore, it serves as a stable and consistent indicator of socio-economic status that can be tracked across generations in many datasets.

The early research on intergenerational educational mobility in Latin America relied mostly on cross-sectional data. To overcome the lack of longitudinal panels necessary to link parents' outcomes to those of their children, two different approaches were used. The first approach examined children still living with their parents, using so-called co-resident samples. Behrman, Birdsall, and Székely (1999), as well as Dahan and Gaviria (2001), who analysed intergenerational mobility across 16 Latin American countries, are examples of this first procedure. The second used survey data that contained answers (by a current generation of respondents) to retrospective questions about the education and occupation of their parents. Behrman, Gaviria, and Székely (2001), who studied intergenerational mobility in Brazil, Colombia, Mexico, and Peru, were an example of this latter approach.

Beyond this main distinction, the methodological approaches followed by early studies varied along many other dimensions, from the choice of summary index to the definition of the variables used to capture education and socio-economic status. For instance, Behrman, Birdsall, and Székely (1999) measured mobility by the degree of association between family background and the 'schooling gap', a variable which they defined as the difference in years of schooling between the grade attained by the child and the grade corresponding to their age. Dahan and Gaviria (2001), on the other hand, used a method based on sibling correlations in years of schooling. Most other studies from this period focused on single countries, examining intergenerational mobility either directly or indirectly, namely through the lens of parental socio-economic status and its impact on children's education or labour market outcomes. Examples include Behrman and Wolfe (1987) for Nicaragua; Binder and Woodruff (2002) for Mexico; Heckman and Hotz (1986) for Panama; and Lam and Schoeni (1993) for Brazil.

The main conclusion from these early studies is that family background played a significant role in determining educational success in Latin America, indicating rather low intergenerational mobility compared, for instance, to the United States. Indeed, Behrman, Gaviria, and Székely (2001) reported intergenerational regression coefficients (IGRC) for years of schooling around 0.7 for Brazil and Colombia, around 0.5 for Mexico and Peru, and 0.35 for the United States. In that study, these estimates were obtained by regressing the child's years of schooling (E_c) on the parents' year of schooling (E_p) as follows:

$$E_c = \alpha + \beta E_p + u \qquad (1)$$

More recent research on intergenerational mobility in Latin America often spans multiple countries. Furthermore, newer studies usually exploit nationally representative household surveys incorporating retrospective questions on parental education. Studies focusing on adult children still living with their parents have become much less common because the exclusion of children who left their parents' home was found to generate a "co-residency" selection bias (Emran, Green and Shilpi, 2018). This newer line of research, which provided more representative and comparable estimates across countries, reveals significant variations in intergenerational mobility of education across Latin America (Daude and Robano, 2015; Ferreira et al., 2013; Neidhöfer, Serrano, and Gasparini, 2018). For instance, Neidhöfer, Serrano, and Gasparini (2018) found that for the 1964-1967 cohort, the average regression coefficient of child years of schooling on parents' years of schooling across eighteen Latin American countries was approximately 0.50, but this average concealed substantial variation, with coefficients ranging from around 0.35 in Venezuela and Costa Rica to around 0.6 or higher in Guatemala and El Salvador.

Global comparative studies on intergenerational mobility of education confirm these patterns, consistently placing Latin America among the regions with the lowest levels of intergenerational mobility worldwide (Ahsan et al., 2023; Hertz et al., 2008; Narayan et al., 2018; Van der Weide et al., 2024). Further findings indicate that intergenerational mobility in Latin America is negatively correlated with income inequality and periods of economic crises, and positively correlated with economic growth, the quality of education, and public educational expenditures, among other factors (Daude and Robano, 2015; Ferreira et al., 2013; Marteleto et al., 2012; Neidhöfer, 2019).

Table 1 provides an overview of results. It presents intergenerational regression coefficients for Galtonian regressions in years of schooling, averaged across birth cohorts, drawing on five studies that offer reasonably comparable estimates across multiple Latin American and Caribbean countries. While the cohorts studied are rather similar – with years of birth for the children's generation roughly between 1940 and 1990, except for the Hertz et al. (2008) study – the datasets used and the method of incorporating parental education in the regression model differ, generating variation in the estimates from one study to another. Methodological differences include options such as accounting for the education of one or both parents and using the maximum years of education between the two parents or their average, as indicated in the last row.

Table 1 – Educational mobility in Latin America: Average estimates of intergenerational regression coefficients.

	Van der Weide et al. (2024)	*Neidhöfer et al. (2018)*	*Hertz et al. (2008)*	*Ciaschi et al. (2025)*	*Celhay and Gallegos (2025)*
ARG	0.484	0.437			
BOL	0.679	0.540			

BRA	0.548	0.578	0.950	0.763	
CHL	0.476	0.444	0.640	0.489	0.453
COL	0.692	0.572	0.800		0.521
CRI	0.386	0.408			
DOM	0.477	0.438			
ECU	0.651	0.574	0.720	0.768	
GTM	0.815	0.696			
HND	0.585	0.538			
HTI	0.585				
MEX	0.510	0.492		0.648	0.672
NIC	0.511	0.525	0.820		
PAN	0.598	0.521	0.730	0.728	
PER	0.603	0.532	0.880		
PRY	0.548	0.549			0.459
SLV	0.577	0.620			0.553
URY	0.473	0.480			0.351
VEN	0.378	0.392			
Cohorts	1940-1989	1940-1987	1916-1983*	1940-1989	1940-1990**
Parental education	Maximum	Maximum	Average	Both (Lubotsky-Wittenberg estimate)	Either father or mother***

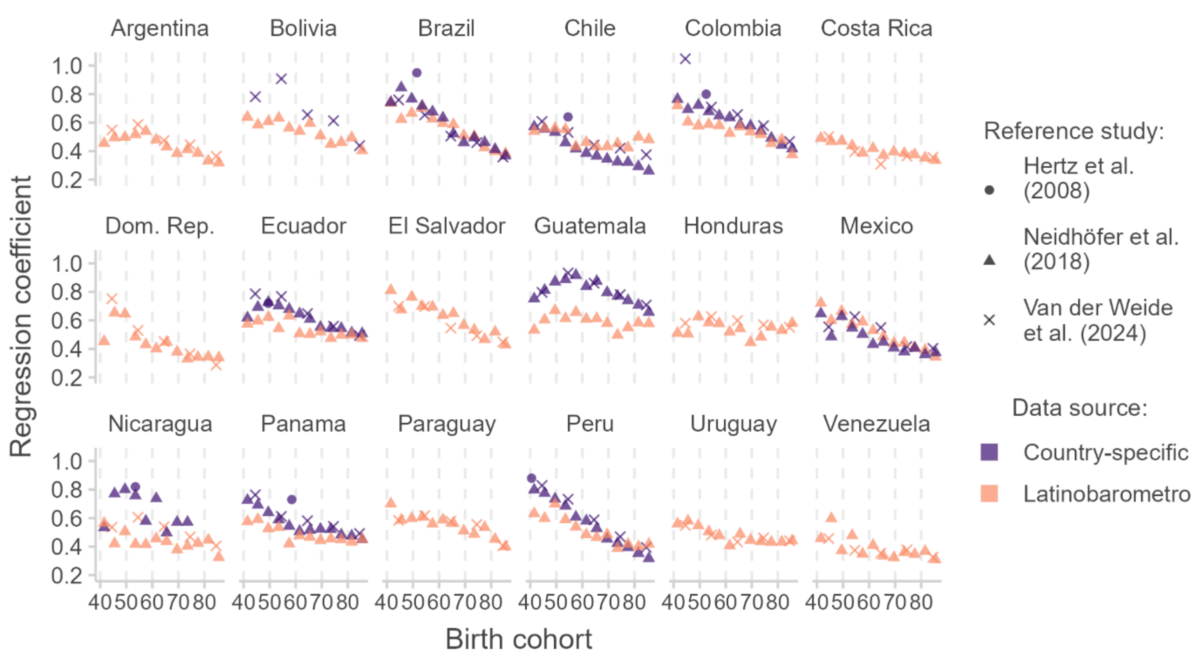
*Notes: *BRA 1927-76, CHL 1930-79, COL 1928-77, ECU 1925-74, NIC 1929-78, PAN 1934-83, PER 1916-65; ** While the youngest children in data are indicated to be born in 1990, we could not find any information on the oldest ones. 1940 is an approximation based on the parents' cohorts, which are 1920-1970; ***Estimates are obtained from census data where respondents are asked about their children's education. Depending on whether the respondent is male or female the estimate refers to father's or mother's education.*

Despite these methodological differences, some common findings do arise from this set of studies. First the IGRCs are typically quite high, implying limited educational mobility in the region. Second, differences across countries are large, and the co-variance of country-level estimates across studies is high: There is generally more educational persistence in countries such as Guatemala, El Salvador, Colombia and Ecuador, whereas Venezuela, Costa Rica, Argentina and Uruguay tend to display higher intergenerational mobility.

This more recent literature also examines trends – rather than just levels – in intergenerational mobility in education in Latin America. Neidhöfer et al. (2018), for example, find that the intergenerational regression coefficient of years of schooling decreased significantly between older and younger cohorts. For individuals born in the 1940s, an additional year of parental education translated to 0.6 more years of schooling for their children, while this advantage reduced to 0.4 years for those born in the 1980s. Over the same period, the probability of completing secondary education for individuals with low-educated parents more than doubled, reaching over 50% in many countries However, not all countries in the region follow this pattern. In countries such as Guatemala, Honduras, and Nicaragua, upward educational mobility remains exceptionally low and has shown little change over time. For the 1980s cohort in these countries, only about one in ten children with less educated parents completes secondary education (Neidhöfer et al., 2018).

Figure 1 illustrates this phenomenon by showing intergenerational regression coefficients for years of schooling from three studies that contain comparable estimates across multiple countries. The three studies featured in this graph were selected based on two criteria. First, they cover a comprehensive set of Latin American countries in a harmonized framework. Second, they rely on retrospective questions on parental background, and hence, are not restricted to individuals still living with their parents after they finished their education. That is: they do not suffer from co-residency bias.

Figure 1 – Intergenerational education mobility trends across cohorts in eighteen Latin American countries.



Source: Own elaboration, based on IGRC estimates from Hertz et al. (2008), Neidhöfer et al. (2018) and van der Weide et al. (2024).

The Figure consists of 18 panels, each corresponding to one country. In every panel, each point represents the value of a regression coefficient estimated for a given birth cohort. The higher the regression coefficient, the lower mobility. The shape of the points denotes the source study, and their colour indicates whether they were estimated using Latinobarometro data or country-specific data.[7] Although there is considerable variation across countries in both current and historical levels of mobility, as well as in the rate of increase in mobility over time, a general positive trend – a negative trend in persistence – is observed in most countries, with the main exceptions being Honduras (where there is no trend) and Guatemala (which displays an inverted U pattern).[8]

This dynamic picture is, at least at first glance, more encouraging than the static one. While older cohorts exhibit very low levels educational mobility, younger cohorts show mobility levels comparable to those observed in more developed countries. Intergenerational regression coefficients based on years of schooling for the 1980s cohorts in countries like Argentina, Brazil, Costa Rica, and Venezuela, ranging between 0.33 and 0.35, are similar to those observed in Italy (0.33), Spain (0.31), and the US (0.33) (Narayan et al., 2018).

Note, however, that the regression coefficient from (1) can be expressed as the correlation coefficient multiplied by the ratio of the standard deviation (σ) of children's years of schooling to that of their parents:

$$\beta = \rho_{E_c E_p} \cdot \frac{\sigma_{E_c}}{\sigma_{E_p}} \tag{2}$$

Declining IGRCs such as we see in Figure 1 may therefore arise from a decrease in the association between the two margins, as measured by $\rho_{E_c E_p}$, or by a falling variance in years of schooling of children relative to their parents'. In other words, the intergenerational regression coefficient combines information on degree of pure association between the two variables – which is sometimes referred to as 'relative mobility' – with information on changes in cross-sectional inequality, expressed by the ratio of standard deviations of the two marginal distributions. It is not a pure measure of association and cannot be derived solely from the copula of the joint distribution; it is sensitive to the marginal distributions.

So, what was happening in Latin America during the time-period covered in Figure 1? It turns out that the declining regression coefficients were primarily driven by changes in cross-section educational inequality (by falling $\frac{\sigma_{E_c}}{\sigma_{E_p}}$) across generations, rather than by changes in pure association. Figure 2 shows this by plotting the 3rd-order polynomial fits obtained from the country-specific trends in the regression coefficients depicted in Figure 1 on the left panel, alongside the fits obtained from the same process applied to the correlation coefficients on the

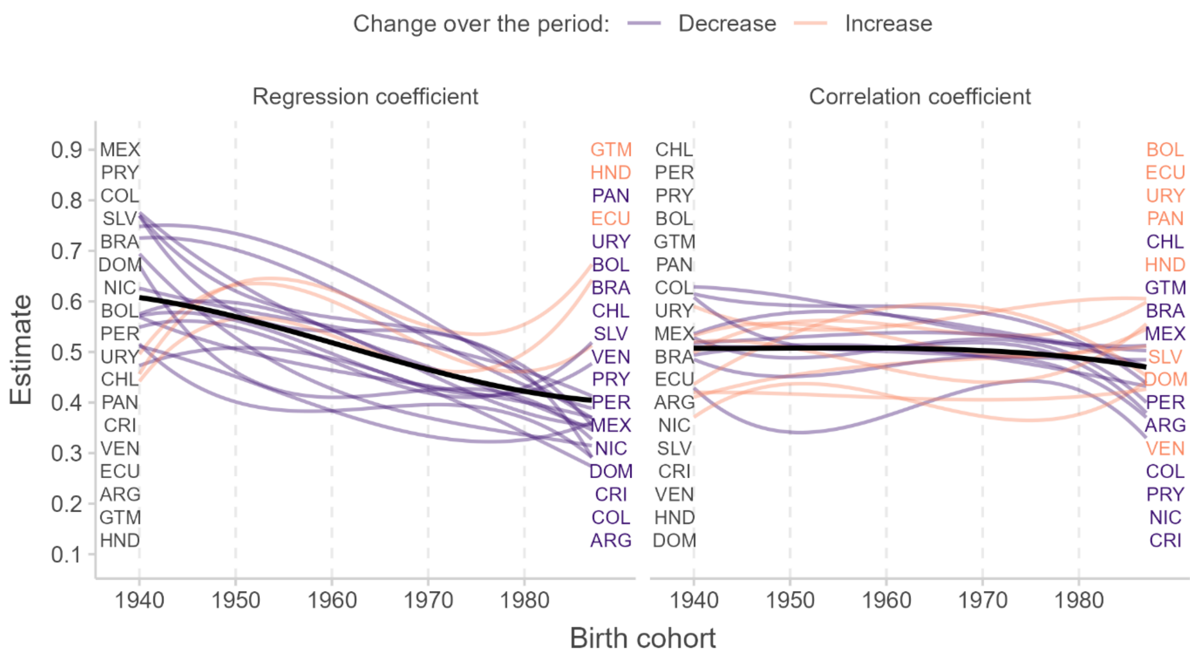
[7] Latinobarometro is a Chilean organization that conducts harmonized and nationally representative annual public opinion surveys across these eighteen Latin American countries. Survey waves since 1998 usually include retrospective questions on parental education.

[8] This pattern also helps explain why the IGRCs from the Hertz et al. (2008) study reported in Table 1 tended to be higher than those for the other four studies – with very few exceptions: The cohorts covered in that paper are older than those in the other four.

right panel. Lines are lightly coloured when values rise over the period, and darker when they fall. The thick black line denotes unweighted cross-country averages.

To maximize the comparability between the two statistics, the underlying values are restricted to the estimates based on Latinobarometro data from Neidhöfer et al. (2018). While regression coefficients follow a clear decreasing trend in most cases, the correlation coefficients appear much more stable. This suggests that the decline in the regression coefficients is driven primarily by changes in cross-sectional educational inequality at the margins, while relative mobility remained mostly stable.

Figure 2 – Trends in the regression and correlation coefficient between children's and parents' education for eighteen Latin American countries.



Note: Ordering of country names shows the ranking in the first and last cohort. Estimates are based on Latinobarometro data from Neidhöfer et al. (2018).

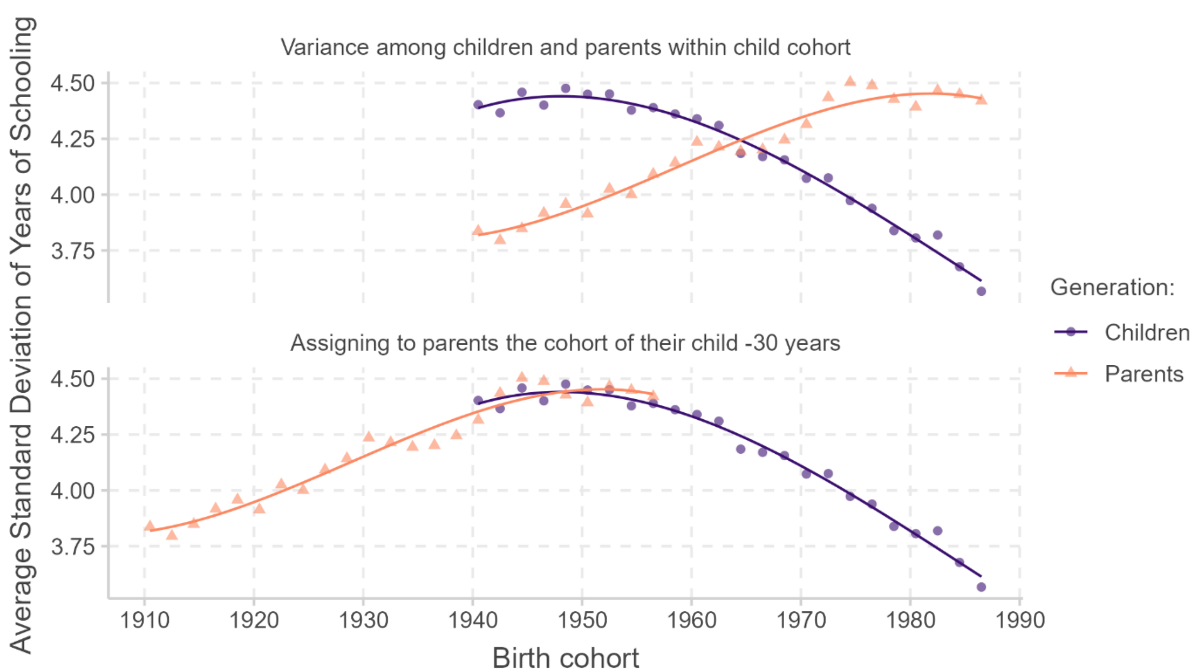
Of course, if the βs are falling while the ρs are largely stable, either σ_{E_c} is falling, or σ_{E_p} is rising – or both. Figure 3 shows the trend in these two variables, which are driving the decrease in the intergenerational regression coefficients of education. The upper panel shows the unweighted cross-country average of children's (purple) and parents' (orange) standard deviations in years of education for each cohort based on the children's year of birth.[9] We observe both a clear reduction

[9] The estimates for the standard deviations reported here are based on the same sample used by Neidhöfer et al. (2018).

in children's education inequality and an increase in their parent's education inequality over the period.

These opposing trends suggest that if we were to observe inequality in children's education further into the past, we would likely obtain an inverted U-shaped curve. This is illustrated in the lower panel of Figure 3, where the standard deviations of the parents' education are shifted by 30 years to the left, so they are roughly placed at their own birth cohorts instead of their children's. This clear inverted U-shaped pattern might be explained by the increase in the average level of education over time along a bounded domain, shifting from a low-mean and right-skewed distribution of education to a more symmetric one within the parents' generation, and then continuing to a more high-mean and left-skewed distribution within the children's generation. Since the standard deviations of children's and parents' education are approximately the same for children born in 1963 and 1964, the regression and correlation coefficient are mechanically equal for those cohorts. The further apart the inequality level of the two generations, the greater the difference between the two mobility measures.

Figure 3 – Evolution of the standard deviations for years of education in the children's and the parents' generation.



The bottom panel of Figure 3 depicts the approximate trajectory of the second moment of the distribution of years of schooling in Latin America during most of the 20th Century. Alongside it, but not shown, there was a pronounced rise in the first moment: average years of schooling increased very considerably over this period, in a massive educational expansion which was accompanied by a significant increase in upward absolute educational mobility (Neidhöfer et al., 2018). While the probability of children from less educated families completing secondary education improved steadily, relative mobility in education, as measured by the correlation coefficient (whether in levels or ranks) has so far remained largely unchanged.

As this brief summary suggests, the literature on intergenerational mobility in this region that uses education as the key outcome of interest has generated rather rich and interesting results. Nonetheless, it can clearly only provide a partial picture of socioeconomic mobility. Education matters a great deal to people, both as an end in itself and as a means to other desirable ends, such as higher earnings or greater political agency. But people evidently also care about their incomes, and the mapping between education and income depends on a myriad features of the labour market, household formation, etc. Despite dauting data challenges, researchers have therefore also attempted to estimate intergenerational mobility in incomes in Latin America.

Intergenerational Mobility in Income

The obligatory starting point to any review of this literature in the region is that limitations to data availability pose significant challenges to the study of intergenerational income mobility in Latin America. As is well known, to avoid life-cycle bias, accurate measures of income mobility ideally require longitudinal data that tracks several income spells over time for each generation (see, e.g. Jäntti and Jenkins, 2015). Such estimates of intergenerational income mobility, based on directly observed links between parents' and children's lifetime incomes, or ranks on their respective distribution of income, are available for only a few countries worldwide, such as Australia (Deutscher and Mazumder, 2020), Canada (Corak, 2020), Norway (Bratberg et al., 2017), Sweden (Heidrich, 2017), and the United States (Chetty et al., 2014). These estimates rely on panel data, usually from administrative sources, that include multiple income observations for both parents and children.

Such data is typically not available for Latin American countries. Although some studies have provided consistent estimates of *intra*generational income mobility – that is, concerning income changes within a single generation – for several Latin American countries (e.g., Fields et al., 2007; Cuesta, Ñopo, and Pizzolitto; 2011; Beccaria et al.; 2022), the lack of longitudinal data that includes several income spells for both parents and children remains a significant barrier to the analysis of intergenerational income mobility in most countries.[10]

To circumvent this limitation, researchers have often employed the two-sample-two-stage least squares (TSTSLS) estimator, as proposed by Björklund and Jäntti (1997). This method uses two

[10] This limitation has contributed to the prevalence of studies focusing on educational mobility (reviewed above), as well as to the emergence of the literature of inequality of opportunity (which we discuss below) in the region. The very first empirical paper seeking to measure inequality of opportunity for income acquisition – anywhere – was written for the case of Brazil (Bourguignon et al., 2007).

samples: one comprising the children's generation and the other the parents' generation, with no observed family links between them; basically, just two standard, independent cross-sectional surveys. In essence, the approach requires identifying parental characteristics that are present in the children's generation sample (typically from retrospective questions about parental education and occupation when the child was younger) as well as in the parents' generation sample. Then, in the first stage, a statistical prediction model of income as a function of those common observed characteristics is estimated in the parents' generation sample. Finally, in the second stage, the coefficients of that model are used to predict the parents' incomes in the children's generation sample. Most studies of intergenerational income mobility in Latin America rely on this method, including Jiménez (2016) for Argentina, Ferreira and Veloso (2006) and Dunn (2007) for Brazil, Nuñez and Miranda (2010) for Chile, Grawe (2004) for Peru and Ecuador, and Daza Báez (2025) for Mexico. More recently, Muñoz and van der Weide (2025) have also used the TSTSLS method to estimate intergenerational elasticities around the world, including nine Latin American countries.

Only very recently has progress in matching parents and children in large administrative datasets – such as tax and social security records – led to studies able to avoid the TSTSLS method and to estimate OLS regressions directly on data that contain observed incomes for children and their parents at comparable stages in their lifetimes. To the best of our knowledge, there are four such studies for the region: Britto et al. (2022) for Brazil, Cortes Orihuela et al. (2024) for Chile, Del Pozo and Moreno (2024) for Ecuador, and Leites et al. (2022) for Uruguay.[11] Making use of the large sample size of administrative data, these studies provide novel insights on regional disparities, non-linearities in intergenerational persistence along the income distribution, and gender differences.

This new wave of studies has a greater focus on rank-based mobility estimators. The typical approach consists in dividing each generation into 100 percentiles, called ranks, from the bottom 1% to the top 1% of the income distribution. Since these percentile income ranks follow a uniform distribution from 1 to 100 among both generations, the intergenerational regression coefficient equals the rank-rank correlation coefficient. This approach allows researchers to conveniently investigate non-linearities in the rank-rank relationship by plotting the expected income rank of children for each parental income rank. In Brazil, the rank-rank association is largely linear, except at the very top, where persistence increases sharply. In Chile and Uruguay, however, the relationship exhibits marked non-linearities, with high persistence at both extremes of the income distribution. In Chile, mobility is relatively high across the lower 80% of the parental income distribution but declines steeply among the richest households. In Uruguay, persistence is also much higher at the top, but non-linear regression models reveal strong persistence at the bottom as well, suggesting the presence of an intergenerational poverty trap.

[11] But relatedly, see also Doruk et al. (2022) for Brazil and Panama, using linked census data.

Absolute mobility estimators are also given increasing attention. The most common ones are absolute upward mobility, and cells of transition matrices. Absolute upward mobility refers to the expected income rank for children whose parents are located at a given percentile of their own income distribution, typically the 25th. It provides further perspective by illustrating the average advancement of children from low-income backgrounds. Transition matrices document the probability for children to end up in each income quantile conditional on parents' income quantile, typically quintiles. In Brazil, individuals born to below-median-income parents reach on average the 36th percentile of the income distribution in adulthood. Only 2.5% of those born in the bottom quintile make it to the top quintile, and nearly half of all individuals born in the bottom and top quintiles remain there as adults. In Ecuador, children of parents in the 25th percentile advance on average to the 44th percentile, while roughly one third remain in the same economic position as their parents. The probability of moving from the lowest to the highest quintile is around 11%, higher than Brazil's but slightly below Chile's estimate of 12%. These probabilities, particularly for Chile and Ecuador, are rather high by international standards. They compare with 7.5% in the United States (Chetty et al., 2014), 9.7% in France (Kenedi and Sirugue, 2023) and 12.3% in Australia (Deutscher and Mazumder, 2020).

A consistent result across all these studies on administrative data is the presence of pronounced gender heterogeneity in intergenerational persistence. Mobility tends to be lower for daughters than for sons, particularly among families in the lower and middle parts of the income distribution, while gender gaps tend to narrow, vanish, or even reverse at the top.

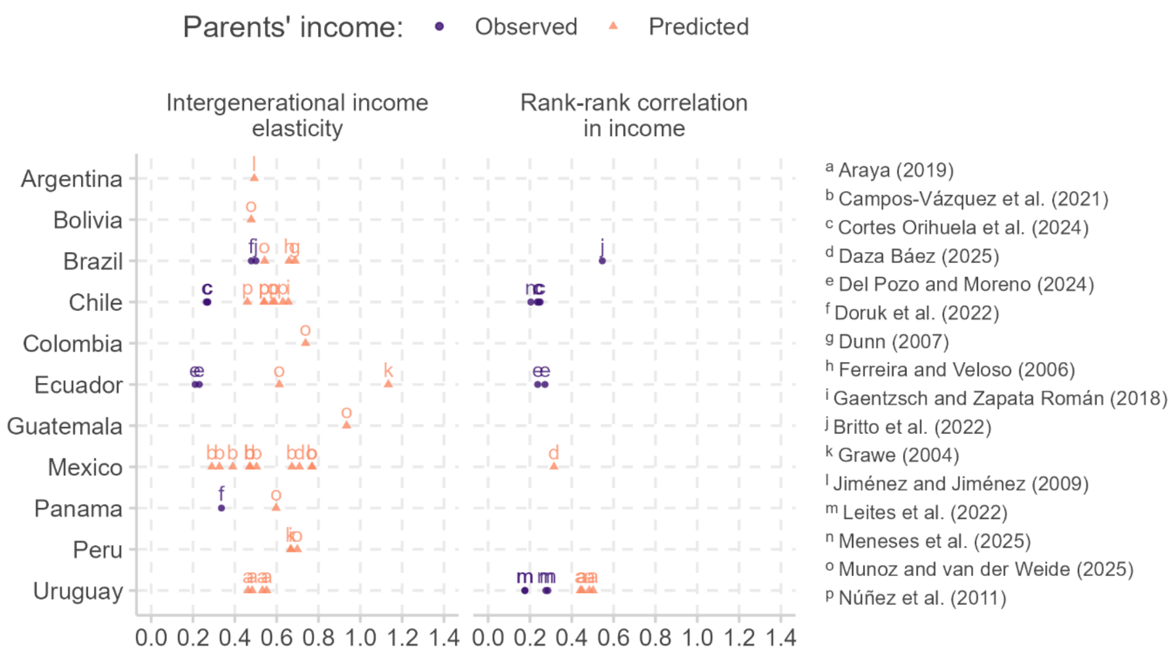
The large sample size of administrative datasets has also allowed the most recent studies to document geographical variations in intergenerational income mobility within countries. In Brazil, there is a stark divide between the more mobile Centre-South and the more persistent North. Regions where slavery was historically less prevalent, or those recently boosted by soy-driven economic growth, exhibit the highest upward mobility. In Ecuador, similar spatial inequalities are evident: vulnerable areas such as the Andean Highlands display low mobility, while provinces such as Galápagos show much higher rates of upward mobility.

While certainly commendable, these efforts to use administrative records for measuring mobility have run up against a different data challenge, often negligible in developed countries but critical in developing countries: the size and significance of the informal sector. Workers (or spells by workers) in that sector – which accounts for sizable proportions of the labour force in most Latin American countries – are, by definition, missing from official statistics. So, even when administrative data is finally exploited in Latin America, informality poses a considerable new challenge.

The challenge is not insurmountable: Leites et al. (2022) implemented various strategies to mitigate the bias caused by the absence of informal sector income in administrative records. Their findings indicate that intergenerational persistence is significantly higher for families less attached to the formal labour market. Similarly, Britto et al. (2022) accounted for informal income by imputing it based on survey data, and Del Pozo and Moreno (2024) combine machine-learning methods with social-security records to estimate total income, both reaching similar conclusions. In contrast, Cortes Orihuela et al. (2024) rely exclusively on formal-sector earnings, thus focusing on the middle-to-upper end of the distribution. Nonetheless, this does mean that imputation – a feature of TSTSLS one wished to move away from – is still present in nationally representative IGM estimates for Latin America.

We attempt to summarize this complex landscape in Figure 4, which shows the available estimates of intergenerational income mobility for countries in the region, distinguishing between methods which impute parental income (e.g., using the TSTSLS method), in orange triangles, and others in which parental income is actually observed in the (formal sector) data, in purple circles. While educational mobility has been estimated for many countries in the region, as shown earlier in Table 1 and Figure 1, intergenerational income elasticities have been estimated for eleven countries (Panel 1), and rank-rank correlations in income for five countries only (Panel 2).

Figure 4 – Intergenerational income mobility estimates for Latin American countries.



Sources: IGE and rank-rank correlation estimates come from the sixteen studies indicated on the right-hand side of the Figure.

The first striking finding from this figure is that intergenerational persistence estimates vary widely both across and within countries, especially for intergenerational income elasticities. For instance, estimates from Chile range from 0.26 (Cortes Orihuela et al., 2024) to 0.66 (Gaentzsch and Zapata

Román, 2018), and ranges are even wider for Ecuador and Mexico. These variations are due in part to various methodological differences, such as the age at which income is observed or imputed, the number of years income is averaged over, the birth cohorts represented in the sample, etc. But an important driver of these large within-country variations is clearly the fact that persistence estimates based on linked administrative data are systematically lower – and sometimes much lower – than survey-based TSTSLS estimates.

Several factors may contribute to this discrepancy. First, a systematic upward bias in TSTSLS estimates is likely at play. Indeed, even though the potential bias from the two-stage approach could theoretically go either way, the bias is generally observed to be upward in practice (Cortes Orihuela et al., 2025). Second, the discrepancy may also arise from a downward bias in the studies using administrative data, due to imperfect corrections for the absence of informal sector earnings. Although estimates from linked data are expected to be lower than TSTSLS estimates, it is surprising that those obtained for Ecuador and Chile are as low as what is typically observed for northern European countries. Rank-rank correlations obtained for Uruguay are also quite low by international standards. This casts doubt on how the informal sector is accounted for in these countries. In Chile, as noted, it is missing altogether. Some of the difference may also reflect genuine variation across the populations and cohorts covered in the different studies.

To get a clearer view on the cross-country variations in intergenerational income mobility, we select from Figure 4 our 'preferred' estimates for each country and present them in Table 2. Attempting to enhance comparability between selected estimates, we consider only those computed on samples of individuals aged between 25 and 45 and born after 1965 on average. When countries have more than one estimate meeting these conditions, we favour estimates based on *observed* parents' income (from administrative data), as opposed to *predicted* parents' income (from TSTSLS). We then select estimates obtained from individual incomes measured at older ages. Specifically, we favour age windows with the highest lower bound, and, in case of ties, those with the highest median age. To break the final ties, we favour estimates that use the most common parents' income definition in this pool of studies, which is father's earnings.

Table 2 – Income mobility in Latin America: Selected intergenerational elasticity and rank-rank correlation estimates.

Country	*Intergenerational income elasticity*	*Rank-rank correlation in income*
ECU	0.23 (1992)	0.272 (1992)
CHL	0.263 (1988)	0.235 (1988)
PAN	0.336 (1980)	
BRA	0.479 (1966)	0.546 (1994)
BOL	0.479* (1979)	
ARG	0.493* (1975)	

URY	0.552* (1982)	0.276 (1978)
PER	0.699* (1972)	
COL	0.738* (1971)	
MEX	0.768* (1968)	0.315* (1974)
GTM	0.934* (1972)	

*Notes: Countries are sorted by ascending order of intergenerational income elasticity. The * symbol indicates estimates obtained using the TSTSLS approach. The median birth cohort of the study sample is indicated in parentheses. Intergenerational income elasticities are from Araya (2019) [URY], Cortes Orihuela et al. (2024) [CHL], Del Pozo and Moreno (2024) [ECU], Doruk et al. (2022) [BRA, PAN], Jiménez (2016) [ARG], Munoz and van der Weide (2025) [BOL, COL, GTM, MEX, PER]. Rank-rank correlations in income are from Britto et al. (2022) [BRA], Cortes Orihuela et al. (2024) [CHL], Daza Báez (2025) [MEX], Del Pozo and Moreno (2024) [ECU], Leites et al. (2022) [URY].*

The selected set of estimates presented in Table 2 still shows very large cross-country variations in the intergenerational elasticity, from 0.23 in Ecuador to 0.934 in Guatemala. While the latter is particularly high by international standards, the former is particularly low. Intergenerational elasticities in northern European countries and Australia are typically close to 0.2 (Bratberg et al., 2017, Deutscher and Mazumder, 2021, Hjorth-Trolle and Landersø, 2025), and amount to 0.31 in Canada (Connolly et al., 2019). In the United States, where intergenerational persistence is considered high, estimates range from 0.344 (Chetty et al., 2014) up to 0.7 (Mitnik, 2020). Similarly to what is observed within countries in Figure 4, estimates based on the TSTSLS approach are systematically higher than those based on linked administrative data.

Rank-rank correlations in income vary much less, from 0.272 in Ecuador to 0.546 in Brazil. This pattern is not specific to Latin America. For comparison, the rank-rank slope in the United States is estimated at 0.341 (Chetty et al., 2014), while in Australia, Canada, France, Italy, Spain, and the Scandinavian countries it ranges between 0.19 and 0.30 (Acciari et al., 2022; Bratberg et al., 2017; Connolly et al., 2019; Deutscher and Mazumder, 2020; Heidrich, 2017; Helsø, 2021; Kenedi and Sirugue, 2023; Soria-Espin and Medina, 2025).

Variations in rank-rank correlations cannot be directly compared to variations in the elasticities, since rank-based estimates are not affected by changes in cross-sectional inequality across generations. Also, despite our effort to select the most comparable estimates across countries, differences in sample coverage, income definitions, and treatment of informality imply that these variations must be interpreted with considerable circumspection.

Multigenerational Mobility

Another recent development in the literature in Latin America, as in other regions, has been to investigate associations across three generations (i.e., from grandparents to grandchildren) rather than just two generations (i.e., from parents to children). The primary aim of this research is to estimate long-term patterns of intergenerational mobility and to test the hypothesis that the transmission of advantage follows an AR(1) process. According to this hypothesis, children's outcomes depend directly only on the outcomes of their parents, not on those of earlier generations (for a review, see Anderson et al., 2018).

Key contributions to this area include Celhay and Gallegos (2015), who examined educational mobility over three generations in Chile, Celhay and Gallegos (2025), who extended the analysis to overall six Latin American countries, and Moreno (2021) for Mexico. Their findings reveal two important points. First, they find that educational mobility over three generations is lower than the AR(1) model predicts, with a much larger deviation in Latin America compared to developed countries. Second, compulsory schooling laws play a significant role in explaining long-term mobility patterns.

On the other hand, Moreno (2021), found that grandparental education has no effect on grandchildren's outcomes once parental education is accounted for. This finding aligns with a segment of the international literature which argues that significant coefficients for grandparental outcomes might result from omitted variable bias, i.e., that grandparental education does not have an influence on individuals' education per se but rather captures that of other unobserved factors (Solon, 2014).

New Directions in Intergenerational Mobility Research in Latin America

One suggestive implication of this review of the literature, and particularly of the more recent studies, is that the next frontier of intergenerational mobility research in Latin America—and in developing countries more broadly—is likely to benefit from exploring a variety of data sources, including administrative records, opinion surveys, and well-established nationally representative surveys, particularly those that include retrospective questions on parents' socio-economic status. Recent studies are also searching for greater spatial granularity and increasingly exploring heterogeneity, be it along the income distribution, across genders and racial groups, or across occupations.

For instance, Neidhöfer et al. (2024) use harmonised, nationally representative household surveys from ten Latin American countries containing retrospective information on parental education to map subnational patterns of intergenerational mobility and examine their relationship with subsequent economic development. In addition to documenting considerable heterogeneity in both the level and evolution of mobility across regions, the study finds that higher intergenerational mobility has a positive long-term effect on economic growth and development, as reflected in indicators such as income per capita, night-time luminosity, poverty reduction, and other measures of well-being.

Muñoz (2024) employs harmonised census data to estimate educational mobility across several Latin American countries at a highly granular geographical level, thereby extending the analysis to a broader spatial scale. The study confirms pronounced heterogeneity both across and within countries and further reveals a narrowing of the mobility gap between urban and rural populations, minor gender differences, and a general increase in upward mobility over time.

Ciaschi, Marchionni, and Neidhöfer (2025) apply the Lubotsky–Wittenberg method (Lubotsky and Wittenberg, 2006), which integrates multiple proxy measures of a latent variable—in this case, parental background—into a unified framework. Using harmonized household survey data for five Latin American countries, they estimate the degree of association between parents' social status and their children's education and income rank. The results show that incorporating parents' occupation as an additional proxy for family background – which moves in the direction of inequality of opportunity estimation, by enlarging the set H of inherited characteristics – increases intergenerational persistence estimates substantially compared with using only parental education.

The finding of a relatively stable trend in economic mobility (as in Figure 2, panel B above) is further supported by Neidhöfer, Ciaschi, and Gasparini (2022), and extended beyond the domain of education alone. Using Latinobarómetro data with retrospective questions, they estimate intergenerational mobility in economic well-being based on indicators such as homeownership, job stability, and access to household goods, which are combined into a composite index. Their results indicate that, despite rising absolute upward mobility in education, the association between parents' educational position and their children's relative well-being did not weaken but rather strengthened over time, possibly reflecting decreasing returns to education and increasing returns to parental background.[12]

In addition, several recent studies have estimated global or regional trends in alternative measures of mobility that include Latin American countries. Ahsan et al. (2023) analyse Demographic and Health Survey (DHS) data on co-residing siblings for 53 developing countries, including several in Latin America, to estimate sibling correlations in schooling. Their results indicate that intergenerational educational mobility is lowest in Latin America and the Caribbean, where the average sibling correlation reaches 0.65, compared with an average of 0.41 in developed countries.

Finally, Genicot et al. (2024) compute the upward mobility measure developed by Genicot and Ray (2023) for 122 countries using aggregate income data by quantiles from the World Inequality Database. For the eight included Latin American countries, the study finds that overall, over the period 1990-2018, both upward and relative mobility was somewhat higher than the average. This pattern contrasts with regions such as Europe, Oceania, and North America, where stagnant or declining relative mobility is observed over the same period.

[12] Gabrielli (2025) studies intergenerational mobility using the same Latinobarómetro data but a different indicator of economic wellbeing, namely self-reported economic status, from 2000 to 2020.

In interpreting this comparative result, which is strongly at odds with the rest of the literature, it is important to note that, while the measure by Genicot and Ray (2023) is presented as an indicator of upward mobility, arguably it does not capture "mobility" in the conventional sense. In particular, its reliance on instantaneous growth as the central element of the index effectively reintroduces anonymity, making it more akin to a measure of change in social welfare than of mobility per se. Because it is computed from an anonymous growth incidence curve rather than a non-anonymous one, it does not track the movement of the same units over time and does not establish a link between origins and destinations at the individual or household level, a defining feature of most mobility measures.

4. Inequality of Opportunity in Latin America: a review

An alternative way to study the inheritance of inequality is the inequality of opportunity approach. As discussed in the Introduction to this chapter, the literature on inequality of opportunity originated to contribute to a normative field, namely the assessment of unfair inequalities. It relies on two broad normative principles – the idea that inequalities due to factors outside people's control, or beyond their sphere of responsibility, are unfair and should be compensated; and the idea that inequalities that arise as a result of individual effort or responsibility are, at least to some degree, fair. These are the principles of compensation and reward, of which there are many versions.

But if one restricts the set of circumstances – variables outside the realm of personal responsibility – to characteristics that are inherited at birth, then one is looking at a somewhat more restrictive notion than IOp – which Ferreira and Brunori (2024) call "inherited inequality" – that becomes closely analogous to that class of mobility measures which is "origin-independent" (Fields, 2000). Indeed, and as illustrated in Section 2, there is a mathematical isomorphism between these IGM and relative IOp measures: they can be written as ratios (or functions of ratios) of inequality in predicted incomes to inequality in observed incomes, where "predicted" refers to incomes predicted by inherited circumstances.

Of course, intergenerational mobility and inequality of opportunity are not identical. If, as seems likely, parental income is not a sufficient statistic for all predetermined circumstances, these measures will differ when other circumstances are included. Different concepts of mobility—especially absolute concepts, like the proportion of individuals surpassing their parents' status—are much less aligned with inequality of opportunity. Nonetheless, relative measures of intergenerational persistence and inequality of opportunity are conceptually aligned and strongly correlated in practice (Brunori et al., 2013).

In Latin America, given the paucity of data sources containing suitable estimates of both parents' and children's income, measures of inequality of opportunity have been estimated using inherited characteristics other than parental income — typically including parental education and occupation, place of birth, race or ethnicity, and biological sex at birth. Because all these variables are pre-determined and inherited at birth, IOp estimates in this region fall under the category of inherited inequality.

The fact that all these circumstance variables are categorical has implications for the choice of prediction function $\hat{y} = \hat{f}(H)$ used to compute the absolute and relative measures of inherited inequality or IOp: $I(\hat{y})$ and $I(\hat{y})/I(y)$. Most of the early studies described in this section used either one of two functional forms. First was a fully non-parametric version:

$$\hat{f}_N(H) = E(y|H = h) \tag{3}$$

where the conditional expectation is taken for each possible vector h of the observed circumstances. This corresponds to partitioning the population into all possible cells containing only individuals with identical circumstances and taking averages within those cells. Alternatively, a simple parametric version was also used:

$$\hat{f}_P(H) = \hat{\alpha} + H\hat{\beta} \tag{4}$$

where the estimated parameters are obtained from an OLS regression of y on H, typically without any interactions. Naturally, if a full set of interactions were added, (4) would converge towards (3).

Bourguignon, Ferreira, and Menendez (2007) provided the first empirical estimates of inequality of opportunity in Latin America by examining the predictive power of these inherited characteristics in Brazil. They found that observable circumstances accounted for about 25% of total inequality, when measured by the mean logarithmic deviation. These authors had not yet fully embraced the purely predictive nature of the exercise. They attempted to use bounding methods to estimate a 'structural' model where incomes depended on circumstances and observed efforts, and efforts in turn depended on circumstances – although they also presented the reduced form equations which later became standard in the literature.

Subsequent studies by Ferreira and Gignoux (2011) for seven Latin American countries and by Núñez and Tartakowsky (2011) for Chile also found high levels of inequality of opportunity for income. Ferreira and Gignoux (2011) show that the share of total income inequality attributable to inequality of opportunity in Latin America ranges from 23% in Colombia to 36% in Guatemala. These shares were even higher for consumption inequality, ranging from 24% to 53% in Colombia and Guatemala, respectively. Despite being interpreted as lower-bound estimates – because not all inherited circumstances are observed – these figures are relatively high compared to those obtained for developed countries, as demonstrated by the comparative multi-country study by Brunori, Ferreira, and Peragine (2013). Notably, parental education often emerges as the most influential single circumstance.

In a similar vein, researchers have sought to measure inequality of opportunity for education in the region. Andersen (2003) evaluated the impact of family background on the schooling gap for children in 18 Latin American countries. Her findings indicate that—in line with the general patterns highlighted by the literature on intergenerational mobility summarized in Table 1—Guatemala and Brazil have the highest levels of inequality of opportunity, while Chile, Argentina, Uruguay, and Peru have the lowest levels. It is important to note, however that, because it uses the schooling gap measure – defined as in Behrman, Birdsall and Székely (1999), above – as the dependent variable, rather than the child's years of schooling, this study is not strictly comparable to the those discussed below.

Turning from attainment to achievements, Gamboa and Waltenberg (2012) used data from the 2006 and 2009 PISA surveys to estimate inequality of educational opportunities in six Latin American countries. However, their findings also highlighted significant variation across countries, years, and circumstances. Again, Brazil stood out with the highest inequality of opportunity. Once again, parental education had the strongest influence, and school type (public or private) significantly affected individual opportunities for educational success. Using data from the same PISA surveys, Ferreira and Gignoux (2014) analysed inequality of educational opportunities for a larger set of countries globally, finding that the six Latin American countries in the PISA dataset had relatively high levels of inequality of opportunity, particularly when correcting for sample selection into test takers.[13]

5. Inequality of Opportunity in Latin America: the state of the art

As noted in the Introduction, since inequality of opportunity is estimated as $I(\hat{y})$ or $I(\hat{y})/I(y)$, where $\hat{y} = \hat{f}(C)$, three methodological decisions are key: choosing the set of predictors, C; choosing the specific prediction function $\hat{f}(\quad)$; and the specific inequality index $I(\quad)$.[14] In what follows, we will treat the set of circumstance variables (or predictors) C to be determined by data availability. In the Latin American context, where relatively few circumstance variables are observed in most household surveys, yielding a fairly narrow set that is present across countries,

---

[13] Relatedly, Paes de Barros et al. (2009) developed the Human Opportunity Index for children, which includes access to education as a key dimension, along with access to basic services such as water and electricity.

[14] The same is true, of course, of intergenerational mobility. In that literature, $H = \{y_p\}$, typically. But $\hat{f}(\quad)$ can and does vary widely, as various functional forms can be used to estimate a Galtonian regression, or indeed measures can be computed from a transition matrix or copula in other ways. Similarly, the information from these models can be summarized using different scalar indices, such as regression coefficients, Pearson correlation coefficients, rank-rank correlation coefficients, etc.

this is not a bad assumption. In fact, the literature has largely coincided in setting $C = H = \{$father's and mother's educational attainment (in years of schooling or grades completed); father's and mother's occupation; place of birth; race or ethnicity; and biological sex at birth.$\}$ Again, because all elements of the circumstance vector are inherited or determined at birth, we write $C = H$ and treat IOp estimates as inherited inequality.

The next step is the choice of prediction function, $\hat{f}(\quad)$, which in fact consists of two related aspects. The first is how exactly to partition the population on the basis of the different categories of *H.* Consider the example of the Brazilian sample, to which we will turn below. It includes as potential circumstance variables the sex of the respondent (two categories); ethnicity (five categories); occupation of the father and mother (eleven categories each); and education of the father and mother (fourteen categories each), and place of birth (28 categories). So, if one used the finest possible partition, there would be 2 x 5 x 11 x 11 x 14 x 14 x 28= 6,640,480 potential types. Once the sample restrictions which are discussed in the next section are applied, the sample contains 41,031 individuals. Any estimate of IOp based on this "fine partition" would obviously be plagued by an upward "overfitting" bias that arises when there are "too many" types, so that sampling error becomes too large within each type and sampling noise is erroneously treated as IOp (Brunori, Peragine and Serlenga, 2019).

This issue was recognized in the early studies within this literature, prompting the adoption of more parsimonious partitions. For instance, Ferreira and Gignoux (2011) used comparable data on circumstances from six Latin American countries and restricted their partitions to 54 or 108 types per country, by arbitrarily merging subcategories into broader groups. Table 3 below – which reproduces their Table 3 – shows the categories into which each variable was divided to reach those numbers. Notice, for example, that father's occupation was coded into only two categories, whereas many more would have been possible. Similarly, places of birth were grouped into three categories, where once again a much finer partition would have been possible. These decisions, made largely on an ad-hoc basis, implied not one but two levels of arbitrariness. With fathers' occupation, for instance, it is not only arbitrary to choose two rather than, say, three categories. It is also arbitrary how to split the ten possible classifications into the two groups. Ferreira and Gignoux tended to choose 'agricultural workers' versus all other occupations, based on observed frequency and a 'sense' of social structures in these countries. But this is a combination of 10 occupations into groups of 2, so another 44 possibilities would have been possible, and their choice was ultimately arbitrary.

Ferreira and Gignoux (2011) also acknowledged that their parsimonious approach to the number of cells in the partition could introduce omitted variable bias—both from the exclusion of genuinely unobserved circumstances (such as parental income) and to the loss of variation due to the aggregation or omission of observed circumstance variables. In fact, we now understand clearly that the selection of a partition for the measurement of IOp based on the available variables and categories in a dataset inevitably involves a trade-off. On the one hand, increasing the number of types reduces downward omitted variable bias by incorporating more circumstances. On the other hand, reducing the number of types mitigates upward "overfitting" bias.

Table 3: Choice of categories for individual circumstances leading to the type partition in Ferreira and Gignoux (2011)

		BRAZIL	COLOMBIA	ECUADOR	GUATEMALA	PANAMA	PERU
Ethnicity							
	category 1	self reported white ethnicity	Other	self-reported ethnicity: white, mixed blood ("mestizo") or other	European maternal language	Other	European maternal language
	category 2	self reported black ("negro") and mixed blood ("pardo") ethnicity	self-reported minority ethnicity: "indígena, gitano, archipiélago o palenquero"	self-reported ethnicity: indigenous, black ("negro" or "mulato").	indigenous maternal language	speaks indigenous language	indigenous maternal language
Father's occupation							
	category 1	agricultural worker	Missing	agricultural worker or domestic worker	agricultural worker	agricultural worker	missing
	category 2	Other		Other	Other	Other	
Mother's and father's education							
	category 1	none or unknown	none or unknown	none or unknown	none or unknown	none or unknown	none or unknown
	category 2	completed grade 1 to 4	primary incomplete	Primary	primary incomplete	Primary	primary incomplete
	category 3	completed grade 5 or more	primary complete or more	secondary or more	primary complete or more	secondary or more	primary complete or more
Birth region							
	category 1	Sao Paulo & Federal district	departments at the periphery	Sierra & Amazonia provinces	Guatemala city, North-East departments and El Petén	cities and intermediate urban centers	inland non-southern departments
	category 2	South East, Center-West & South	central departments(a)	Costa & Insular provinces	North & North-West departments	other urban centers	Southern and other costal departments
	category 3	North-East, North or missing	Bogota, San Andres and Providencia islands and foreign country	Pichincha province (with Quito) & Azuay province	South-East, South-West & Center departments	rural areas	Arequipa, Callao & Lima

(a) Central departments are Boyaca, Caldas, Caqueta, Cundinamarca, Huila, Meta, Norte de Santander, Quindio, Risaralda, Santander, Tolima, and Valle del Cauca.

Source: The table reproduces Table 3 in Ferreira and Gignoux (2011).

The key challenge in recent research in this area has been to develop a justifiable, non-arbitrary method for partitioning the sample or population in the initial estimation step. Once that is done, the second aspect of the choice of prediction function refers to the functional form itself. Should the categories be entered in a linear regression? In a saturated model, or in a more parsimonious specification? What about non-linearities?

Some recent papers have attempted to address these two aspects of the choice of prediction function in an integrated manner, by adopting data-driven methods that select the empirical partition according to algorithms designed to maximize the predictive power of observed variables in a data set, given its size and characteristics. These are typically supervised machine learning algorithms with inbuilt cross-validation steps that aim to maximize predictive power out of sample. One such approach, which uses conditional inference trees and random forests, and then effectively applies Eq. (3) to the tree nodes, was proposed by Brunori, Hufe and Mahler (2023). An alternative, which uses transformation trees instead of CITs, was suggested by Brunori, Ferreira and Salas-Rojo (2023). Both approaches have been applied to a set of Latin American countries by Brunori, Ferreira, and Neidhöfer (2025).

Finally, once $f(H)$ has been specified and estimated, and a counterfactual distribution or vector $\widehat{\boldsymbol{y}}$ has been computed, what inequality index should be used to compute $I(\hat{y})$? Here too, a trade-off turns out to exist. The early literature – the Latin American portion of which was summarized in Section 3 above – chose to use decomposable inequality indices, perhaps on the basis that these would permit a clear separation between inequality of opportunity $I(\hat{y})$, and "fair inequality", $I(y) - I(\hat{y})$. Except that, even then, authors were aware that this latter component could not be properly interpreted as fair inequality, because of the omitted circumstance variable problem. These unobserved circumstances might contaminate that second component, which should then be more properly regarded as a 'residual component', while $I(\hat{y})$ should be seen as a lower-bound estimate, at least if overfitting is successfully avoided.

Be that as it may, the early literature nonetheless preferred to use decomposable inequality measures, such as the two Theil indices, both of which are members of the Generalized Entropy class, *E(0)* and *E(1).* The first of these has particularly appealing properties in that it is the only inequality measure anchored on mean incomes that is path-independently decomposable (Foster and Shneyerov, 2000), and was used, for example, by Checchi and Peragine (2010) and Ferreira and Gignoux (2011). However, as is well-known, inequality indices differ in their sensitivity to different parts of the distribution (Atkinson, 1970), and the mean log deviation happens to be particularly sensitive to the bottom tail. While that property makes it an appealing measure to those who are averse to inequality arising from particularly low incomes, it also makes it less suitable for assessing between-group inequality which, being inequality among averages (see Eq. 3), necessarily tend towards the middle of the distribution, by the central limit theorem.

Brunori, Palmisano and Peragine (2019) therefore proposed using the Gini coefficient as the index of choice, since it is more sensitive to income gaps towards the middle of the distribution and hence has greater discerning power over the range of interest for IOp measurement. The Gini cannot be exactly decomposable into a between- and a within-group term, as Generalized Entropy measures can. Many alternative Gini decompositions have been proposed, all of which tend to have a third term in addition to the standard within and between terms. Bhattacharya and Mahalanobis (1967) and Lambert and Aronson (1993) show that this third term, which is driven by the degree of overlaps between groups and hence include elements of both within-ness and between-ness, is always positive. It follows that, when the Gini coefficient is used as $I(\quad)$, $I(y) - I(\hat{y})$ should be interpreted as a residual term that contains all within-group inequality, but also some elements of between-group inequality. This was already the appropriate interpretation even with the MLD, because of the possibility of omitted circumstances, but the argument is now reinforced, in the case of the Gini, whenever the supports of the groups overlap. While we follow

Brunori, Palmisano and Peragine (2019) in using the Gini coefficient for our headline estimates below, we also present results using the mean logarithmic deviation in Appendix Table A1.

Data

In the remainder of the chapter, we revisit, update and extent the work of Brunori, Ferreira and Neidhöfer (2025). In doing so, we produce new estimates of inherited inequality in Latin America, which are in line with the latest Global Estimates of Opportunity and Mobility (GEOM) database, as well as presenting some new analysis.[15] GEOM, and therefore the analysis that follows, is based solely on nationally representative datasets. The unit of analysis is the individual, with income measured as age-adjusted equivalized household income, calculated using the square-root equivalence scale. All incomes are standardized to 2017 USD and adjusted for both CPI and PPP. To account for lifecycle-related income variations, we apply an age adjustment by regressing each individual's income on their age and age squared. The adjusted income measure used as our outcome variable is derived from the regression constant and residual.

The vector of circumstances varies slightly across countries but always consists of at least five of the following seven circumstance variables, as discussed above: sex; race or ethnicity; place of birth; father's and mother's education; father's and mother's occupation. The specific categories within each of the last six (all but sex, coded in two categories in all surveys) vary from country to country.

We use twenty-six household surveys for ten Latin American countries, fielded between 2000 and 2017, that satisfy these variable inclusion requirements. For Argentina, data come from the Encuesta Nacional sobre la Estructura Social (ENES), while for Bolivia the source is the Encuesta de Hogares (EH). In Brazil, information is drawn from the Pesquisa Nacional por Amostra de Domicílios (PNAD), and in Chile from the Encuesta de Caracterización Socioeconómica Nacional (CASEN). For Colombia and Ecuador, the data originate from their respective Encuesta Nacional de Condiciones de Vida and Encuesta de Condiciones de Vida (both abbreviated ECV). Guatemala contributes data from the Encuesta Nacional sobre Condiciones de Vida (ENCOVI), while Mexico uses the Encuesta ESRU de Movilidad Social en México (EMOVI). Finally, data for Panama are obtained from the Encuesta de Niveles de Vida (ENV), and for Peru from the Encuesta Nacional de Hogares (ENAHO).

Table 4 provides, for each country, the survey waves we use, the circumstances available in them, and the sample size used in the analysis after dropping observations with missing information on incomes, household size or any circumstance variable. All datasets were sourced from the Socio-Economic Database for Latin America and the Caribbean (SEDLAC) harmonized database, maintained by Center for Distributive, Labor and Social Studies (CEDLAS) at the University of

[15] The GEOM database can be accessed through this link: https://geom.ecineq.org/.

La Plata in Argentina. The final samples used in our analysis differ from the full SEDLAC samples as follows.

Table 4: Basic description of the household survey data

Country	Year	Circumstances	Final Sample Size
Argentina	2014	Sex, race or ethnicity, place of birth, father’s and mother’s education, father’s occupation	6,632
Bolivia	2008	Sex, race or ethnicity, place of birth, father’s and mother’s education, father’s and mother’s occupation	4,197
Brazil	2014	Sex, race or ethnicity, place of birth, father’s and mother’s education, father’s and mother’s occupation	41,031
Chile	2006	Sex, race or ethnicity, place of birth, father’s and mother’s education	97,558
	2011		64,756
	2013		68,473
	2015		88,309
Colombia	2010	Sex, race or ethnicity, place of birth, father’s and mother’s education	25,924
Ecuador	2006	Sex, race or ethnicity, place of birth, father’s and mother’s education, father’s and mother’s occupation	25,465
	2014		51,007
Guatemala	2000	Sex, race or ethnicity, place of birth, father’s and mother’s education, father’s and mother’s occupation (only 2000)	14,690
	2006		28,640
	2011		29,008
Mexico	2017	Sex, ethnicity, place of birth, father’s and mother’s education, father’s and mother’s occupation.	13,475
Panama	2003	Sex, race or ethnicity, place of birth, father's and mother's education	13,557
Peru	2001	Sex, race or ethnicity, place of birth, father's and mother's education	31,648
	2006		23,722
	2007		26,074
	2008		25,287
	2009		26,016
	2010		25,270
	2011		28,610
	2012		29,545
	2013		34,717

2014	34,532
2015	34,607

Source: Own elaboration.

First, we included only surveys that contained self-reported sex and retrospective questions on parental education, parental occupation, and a country-specific variable describing "origin" (such as place of birth, race and/or ethnicity). This resulted in the exclusion of some survey waves that lacked this information. Second, we restricted the sample to adult individuals (aged 18 or older) living in households with non-negative incomes. Third, observations with missing data for income or any of the circumstance variables were excluded.

Finally, for the Mexican data, the cleaning process included an imputation stage designed to address the lack of comprehensive income information in the ESRU-EMOVI survey. Income was imputed from the ENIGH survey into the ESRU-EMOVI data following a cross-survey imputation method that ensures consistency in rank order and improves the reliability of estimates of inequality of opportunity. Details and validation tests can be found in Torres-Lopez et al. (2025).

New estimates of inequality of opportunity for income

Following Brunori, Ferreira and Neidhöfer (2025) and Brunori, Hufe and Mahler (2023), we employ the conditional inference trees (CIT) and random forests developed by Hothorn et al. (2006) to obtain the most relevant partition of the sample given the data under consideration, consistent with a preselected level of statistical significance. A conditional inference tree consists of a set of terminal nodes (leaves) obtained by recursive binary splitting, according to the following algorithm:

1. Choose a critical significance level $(1 - \alpha)$ for hypothesis testing, e.g. 0.99.
2. Compute the correlation coefficient between the outcome variable and each and all observed circumstances. If the Bonferroni-adjusted *p-value* of all correlation tests are higher than the chosen α, exit the algorithm.
3. If the null hypothesis is rejected for at least one circumstance, the variable producing the most significant correlation is selected as the first splitting variable $[c]$.
4. The algorithm then considers how circumstance $[c]$ can be used to partition the sample into two subsamples [C]. For all possible resulting binary partitions, it computes the p-value for the null hypothesis that the statistic of interest (e.g., the mean) in the two sub-samples is identical.
5. The binary split that produces the smallest p-value in the test, [C]*, is chosen.
6. Repeat steps 2 – 5 for each node (sub-sample), until stopping at step 2. [16]

[16] We set $\alpha = 0.01$ and impose one additional requirement: each terminal node must have a minimum of 1% of the observations in the sample (or 50 if the sample size is smaller than 5,000). All other parameters are the default

Once the algorithm has exited everywhere, the output consists of a partition of the sample or population. Denote each terminal node of the tree by h and, as in equation (3), let predicted incomes be given by the mean income at each node: $\hat{y} = E(y|H = h)$. Consider the full counterfactual distribution $\hat{\mathbf{y}}$, a vector with the same dimension of the original vector $\mathbf{y}$, but where each individual income has been replaced by its prediction $\hat{y}$.[17] The absolute estimate of inequality of opportunity is then $\widehat{IO}_a = I(\hat{\mathbf{y}})$.

Regression trees are known among machine learning algorithms for their low bias but high variance, and conditional inference regression trees are no exception. This means that the opportunity tree initially estimated is sensitive to the specific sample used, such that a slightly different but equally representative sample could produce a different partition. As a result, researchers are generally cautious not to overinterpret the exact structure of the trees. Moreover, to address the high variance typically associated with this algorithm, a standard approach in supervised machine learning is to aggregate multiple trees into random forests. In essence, this involves creating multiple subsamples of the original data without replacement and estimating trees for each subsample using a subsample of the controls at each split. Following Hothorn et al. (2006), we construct our conditional inference random forest by estimating 200 Ctrees and using default values for most of the tuning parameters.[18] See also Brunori, Ferreira and Neidhöfer (2025).

Although our preferred estimate of IOp is the estimate based on random forests, as an illustration, Figure 5 below depicts a "root" tree for Brazil (2014), with the type partition at the bottom. This example is illustrative both because Ctrees can be used themselves to measure ex-ante IOp and because random forests are essentially an aggregation of trees, so observing the structure of a single tree allows us to better understand the estimation procedure. Population shares are expressed as percentage of the total population, and type means are expressed as multiples of the overall mean income (US$ 12,882).[19]

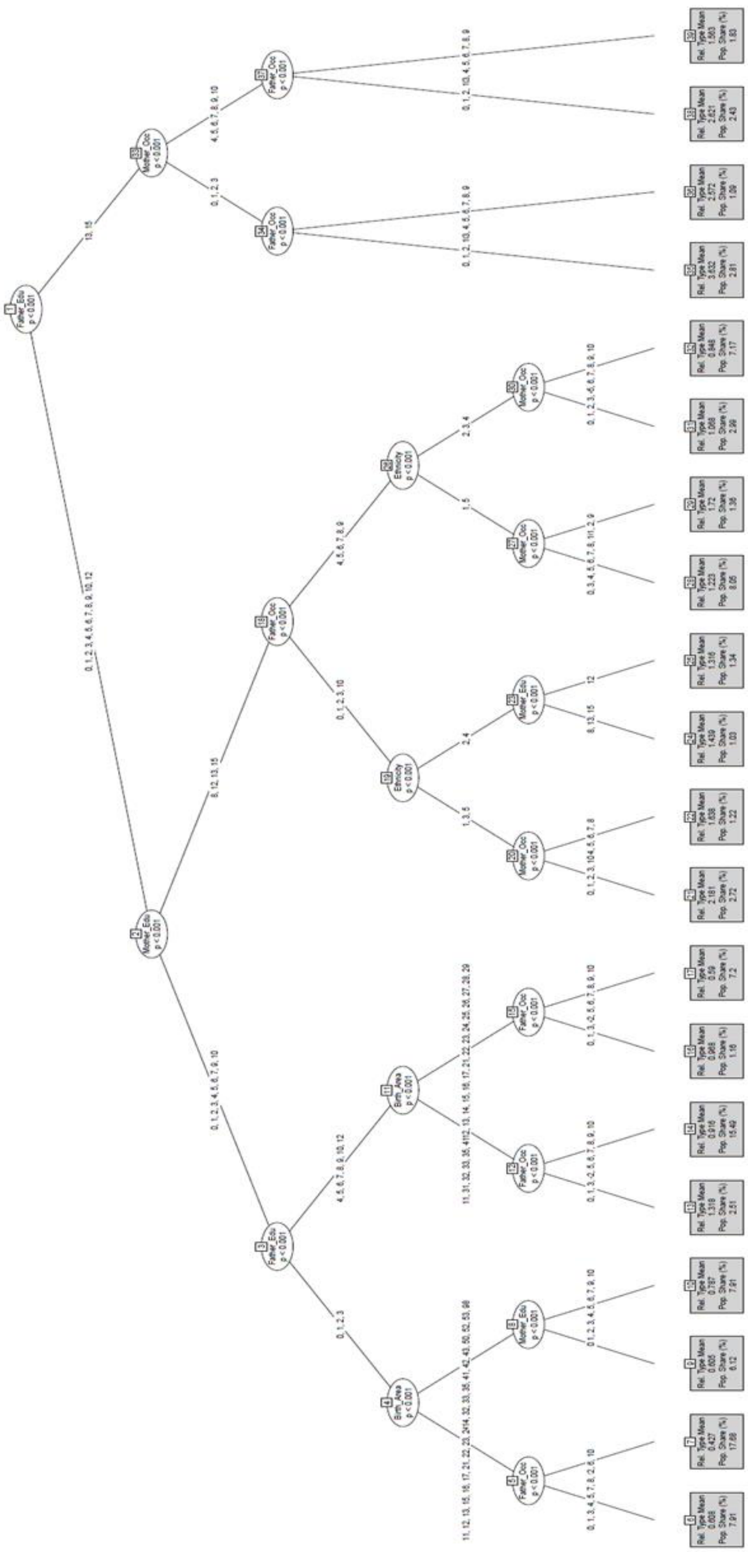
parameters in the "*ctree*" R function. Note also that we do not use weights to determine splits. Instead, weights are used to calculate the values of the counterfactual distribution and to estimate IOp. The R functions used to produce these results are publicly available in https://github.com/pedrosalasrojo/GEOM.

[17] This is what Foster and Shneyerov (2000) call a *smoothed distribution*.

[18] In the random forest algorithm, the minimum number of observations that we allow in each terminal node is 0.1% of the sample size, with the aim of maximizing comparability across surveys with different sample size (or 10, if the sample size was smaller than 1,000). All other tuning parameters are set to the default values in the "*cforest*" R function in the package "*partykit*".

[19] For brevity, we henceforth write "income" to mean age-adjusted equivalized household income per individual, as defined earlier.

Figure 5: Conditional inference "root" tree for Brazil (2014)



Source: PNAD (2014). Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$*).*

In our main analysis, the results of which are presented in Table 5, we set $\alpha = 0.01$, requiring 99% confidence that cell means are different in order to conduct a split. That is on the demanding end of conventional test size criteria, but it nonetheless leads to a tree with 52 nodes, which hampers visualization. For presentational purposes, therefore, Figure 5 sets an even higher confidence level at 99.99%, thereby 'pruning' the tree, and generating what we call a "root tree". With this confidence parameter and starting from a sample of 41,031 individuals, with information on income, sex, ethnicity, region of birth, father's and mother's education and father's and mother's occupation, the algorithm yields a final partition of the population into twenty types. The types range in size from 1.0% to 17.7% of the population, and in income from 43% to 363% of mean income.

Although single trees are high-variance estimators, the structure of the tree is nonetheless informative. The first split – in some sense the most salient cleavage it identifies in Brazilian society – is between people whose fathers had at least some university education, with 13-15 years of schooling, and those whose fathers had less than that. In fact, the importance of parental education and occupation in this Brazilian tree is remarkable. No other variable is used to split the sample in the first three levels of the tree, and these four variables account for fifteen of the nineteen splits.[20] The only other variables used for a split are ethnicity and place of birth, twice each.

The richest type in this partition – Type 35 – consists of people with university-educated fathers who worked as managers, professionals or army officers, and whose mothers had similar professions (but also including technicians and associate professionals). This group accounts for just 2.8% of the population, and its average income is 3.6 times the national average. At the other extreme, the poorest type in this tree is Type 7, with average incomes just 43% of the national average. They consist of people born in the North and Northeast of the country to fathers with 0-3 years of schooling and working in agricultural, forestry and fisheries or unemployed, and whose mothers have less than completed secondary school (0-10 years of schooling).[21] This poorest type is also the largest in the partition, accounting for almost 18% of the population.

The full conditional inference tree for Brazil, with 52 nodes (of which the tree in Figure 5 is a pruned, "root" version) yields an absolute IOp Gini coefficient $\widehat{IO}_a = I(\hat{\boldsymbol{y}})$ of 0.32. Given the overall inequality level in the same sample (0.49 Gini points), this result implies that relative inequality of opportunity in Brazil is 66% of overall inequality. The preferred estimates from the random forest are almost identical in this case: an absolute IOp Gini of 0.32, or 66% of overall

[20] In Chile, mother's or father's education – but not occupation - have a similar importance, accounting for five of the six splits in levels 1-3 of the tree, and appearing eight times altogether. Sex appears five times, and birth area,

[21] The set of father's occupations of people in this poorest type also include 'professionals'. CITs do occasionally generate counterintuitive groupings like this, generally for variables with multiple categories, where the 'surprising' category is particularly rare conditional on previous splits. In this case, these would be professional fathers who had 0-3 years of schooling.

inequality (see Table 5). "Root" trees for the latest available survey wave in each of the other nine countries are presented in the Appendix Figures FA1 to FA9.[22]

Having examined an example of conditional inference tree and the basic nature of the results that are obtained from it and from the associated random forest, we now turn to the comparative results for the full set of twenty-six surveys covering Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guatemala, Mexico, Panama, and Peru. Table 5 below presents the main results for the Gini coefficient, from both the conditional inference trees and the associated random forest. Column 2 reports the number of types in each tree partition, and Column 3 lists the overall Gini coefficient for each country/year. Columns 4 and 5 give the inequality of opportunity estimates from the tree in absolute and relative terms respectively, whereas columns 6 and 7 report the absolute and relative IOp estimates from the random forest.[23] The number of types of ranges from a low of ten (in Bolivia, 2008) to a high of 52 (in Brazil, 2014). The overall Gini coefficient for incomes range from 0.39 (in Argentina, 2014) to 0.55 in Guatemala (2006).[24]

Table 5: Inequality of Opportunity in Latin America

Wave	Number of Types	Total Gini	Absolute IOp Gini (Tree)	Relative IOp Gini (Tree)	Absolute IOp Gini (Forest)	Relative IOp Gini (Forest)
Argentina (2014)	11	0.388	0.167	43.0	0.179	46.0
Bolivia (2008)	10	0.500	0.278	55.6	0.294	58.8
Brazil (2014)	52	0.488	0.322	66.0	0.320	65.7
Chile (2006)	29	0.518	0.269	51.8	0.267	52.1
Chile (2011)	32	0.514	0.284	55.3	0.271	52.8
Chile (2013)	31	0.507	0.265	52.3	0.249	49.0
Chile (2015)	40	0.492	0.239	48.5	0.248	50.3
Colombia (2010)	18	0.535	0.245	45.8	0.257	48.0
Ecuador (2006)	30	0.520	0.295	56.7	0.292	56.1
Ecuador (2014)	34	0.455	0.227	49.9	0.229	50.3
Guatemala (2000)	21	0.544	0.318	58.4	0.310	56.9
Guatemala (2006)	23	0.546	0.361	66.1	0.340	62.2
Guatemala (2011)	13	0.526	0.298	56.6	0.291	55.4
Mexico (2017)	20	0.533	0.303	56.8	0.298	55.9
Panama (2003)	17	0.527	0.306	58.0	0.286	54.3
Peru (2001)	28	0.497	0.297	59.8	0.302	60.8

[22] As in the Brazilian case, "root trees" are subtrees containing only splits significant at 99.99% confidence level, for the last wave of each country. The full trees can be found in the "Country Profile" Section in the Global Estimates of Opportunity and Mobility (https://geom.ecineq.org/).
[23] Analogous estimates using the mean logarithmic deviation (MLD) are reported in Table A1 of the Appendix.
[24] The lowest overall mean log deviation is also found in Argentina, 2014 (0.28), but the highest is 0.64 for Panama, 2003. See Appendix Table A1.

Peru (2006)	38	0.477	0.316	66.2	0.306	64.9
Peru (2007)	35	0.476	0.314	66.1	0.313	65.9
Peru (2008)	32	0.447	0.282	63.0	0.286	64.1
Peru (2009)	32	0.447	0.278	62.3	0.280	62.5
Peru (2010)	33	0.441	0.265	60.0	0.268	60.7
Peru (2011)	33	0.437	0.258	59.0	0.255	58.3
Peru (2012)	30	0.428	0.252	58.8	0.253	59.2
Peru (2013)	36	0.426	0.247	58.1	0.256	60.2
Peru (2014)	37	0.416	0.252	60.7	0.260	62.4
Peru (2015)	36	0.423	0.260	61.6	0.267	63.3

Source: Data from ENES, EH, PNAD, CASEN, ECV, ENCOVI, EMOVI, ENV, ENAHO. See more details in Table 1 and Table 2

Of greatest interest to us, of course, are the summary measures of inequality of opportunity. The opportunity Gini coefficient from the trees ranges from 0.17 in Argentina (2014) to 0.36 Guatemala (2006). Random forest estimates are similar, also ranging from 0.18 in Argentina (2014) to just over 0.34 in Guatemala (2006).[25] For context, the Gini coefficient – for incomes, not opportunities – for the entire population of Germany (in 2016) is 0.31.[26]

As a share of total inequality, inequality of opportunity as measured by the Gini coefficient accounts for between 43% (in Argentina, 2014) and 66% (in Brazil, 2014; Guatemala, 2006; and Peru, 2006 and 2007) when estimated by the conditional inference tree, and between 46% (in Argentina, 2014) and 66% (in Brazil, 2014 and Peru, 2007) by the random forests. The correlation between these two series (relative Ginis from trees and forests) is 0.95. These values are very large: only three forest-based relative IOp Ginis are below 50%: Argentina, Chile 2013, Colombia.[27] For 11 of the 26 country-year combinations, they are 60% or greater. Because they arise from algorithms designed to avoid overfitting, working with a relatively limited set of circumstances, and operating with high confidence levels in survey-sized samples, it is very likely that these numbers are underestimates – suggesting that at least half, and in some cases two-thirds of all income inequality we observe in these countries is inherited at birth.

[25] Note that all estimates reported in this chapter, regardless of the approach followed, the algorithm used, or the inequality measure chosen, may partly depend on sample size. A larger sample size implies higher power in the test performed to split the sample. Therefore, ceteris paribus, a deeper tree with a higher expected level of inequality of opportunity. However, as shown in Brunori et al. (2023a), the sensitivity of conditional inference trees and random forests to sample size is no greater than the sensitivity of more standard regression-based econometric approaches.

[26] For household per capita income. Source World Bank: https://databank.worldbank.org/GINI-index-by-country/id/c7a387ee .

[27] Citing correspondence from the Colombian Statistical Agency (DANE), Shaikh and Gomez Tamayo (2025) claim that the 2010 wave of Colombia's ECV is "subject to incorrect classification of parental education categories resulting from a change in response categories", leading to a significant underestimate of IOp. They report higher estimates for other years in the series. Although their paper is as yet unpublished, we have no reason to suspect that their information and analysis are incorrect, so we urge readers to treat all Colombian IOp estimates in this Section as potential underestimates.

The relative contribution of individual circumstances

Next, we turn to assessing the relative importance of the different circumstances in contributing to inequality of opportunity. Just as standard measures of intergenerational mobility cannot be interpreted causally – since all variables (other than parental education or income) that contribute to determining the child's outcome are omitted – neither can IOp measures, nor any decomposition thereof. Nonetheless, the various circumstances contribute differently to the overall IOp estimate and quantifying those differences may be of descriptive interest. We aim to assess both the average and the marginal contribution of each variable, as defined below.

Regarding the average contribution, since there is neither a guarantee nor an expectation that the contributions of all circumstance variables are additively separable, the appropriate method for identifying individual contributions is a Shapley-Shorrocks decomposition (see Shapley, 1953, and Shorrocks, 2013). Intuitively, a Shapley decomposition estimates the overall contribution of a variable x to an outcome function y by calculating the average reduction in y across all possible combinations of ways in which y can be generated without x. In this case the target variable is IOp, the variability in income that can be predicted by circumstances. Following Brunori, Ferreira and Salas-Rojo (2023a), we compute Shapley value decompositions as follows:

1. Draw a sub-sample from the observed sample.[28]
2. Estimate IOp in the sub-sample using a Ctree.
3. Re-estimate IOp in the sub-sample for all possible elimination sequences for each circumstance.
4. After each elimination sequence, the tree and the resulting IOp measure are estimated and the IOp values are stored.
5. Average IOp across all elimination sequences for each circumstance *c*. The difference between the overall IOp in the subsample and this average is the contribution of *c*.
6. Repeat steps 1-5 100 times.
7. Calculate the average contribution to IOp across the 100 iterations.

It is important to note that, although this part of the analysis is based on the aggregation of numerous overfitted trees, we do not expect any bias in the evaluation of the relative importance of each circumstance. Our focus is not on the absolute level of estimated IOp but on the relative contribution of each circumstance. As such, the overfitting of individual trees, which are constructed from subsamples of the original data and then aggregated, provides the usual advantage of bagging weak learners without compromising the robustness of the relative importance estimates for each circumstance.

[28] The sub-sample consists of 5,000 observations or 90% of the original sample size, whichever is greater.

These Shapley values represent the average contribution of a given circumstance to the overall predictive power. They should not be interpreted as the marginal effect of specific circumstance categories. For instance, in a society where 99% of individuals identify as white and only 1% as indigenous, the Shapley value for the circumstance "race/ethnicity" will be low, even in the presence of significant discrimination against indigenous individuals. While the marginal effect of being indigenous may be large and negative, the predictive power of race will remain low for most respondents, who are non-indigenous. This occurs because the proportion of individuals for whom the circumstance is a strong predictor is minimal, while for the majority, the circumstance provides little information—their conditional income remains close to the sample average.

For this reason, we complement Shapley values with Partial Dependence Plots (PDPs). PDPs, originally introduced by Friedman (2001), are visual tools designed to help interpret machine learning outputs. They show how changes in a specific predictor variable affect the predicted outcome while holding the role of other variables constant. The values plotted in a PDP represent the partial dependence function at a particular feature value, which is the average prediction if we were to force all data points to assume that feature value. PDPs provide an interesting complement to measures of average importance because their focus is on the marginal contribution of each characteristic, independent of its marginal distribution. Still, they should be interpreted with care, especially when highly correlated predictors are present or when categories/values are observed in very few cases.

Descriptively, how important is each of the circumstance variables in accounting for the inequality of opportunity estimates reported above? Table 6 presents the results of the Shapley decomposition of the Opportunity Gini coefficients for the latest available survey year for each of our ten countries, as well as a simple unweighted average across countries. Note that parental occupation is missing (for both parents) in Chile, Colombia, Guatemala, Panama and Peru; and mother's occupation is not used in Argentina. This makes cross-country comparisons perilous, particularly between these six countries and the other four. Comparisons within subgroups of countries with the same sets of inherited characteristics should be valid.

Across the sample, the largest contributions to inequality of opportunity as measured by the Gini coefficient come from the mother's and father's education which, together, represent between 32% and 55% of the total for countries in which parental occupation is observed and between 56% and 75% for countries in which parental occupation is not reported. Parental occupation, again taking mothers and fathers together, is the second most important predictor, accounting for an average of 32% of the total in the countries where it is observed. Place of birth is also important, with a 19% average contribution. Race or ethnicity account for 7% on average of overall inequality of opportunity and sex accounts for 4% on average, with Mexico a clear outlier.

Table 6: Relative contributions of individual circumstances: Shapley value decompositions

Circumstances	ARG 2014	BOL 2008	BRA 2014	CHL 2015	COL 2010	ECU 2014	GTM 2011	MEX 2017	PAN 2003	PER 2015	Ave-rage
Birth Area	33.74	15.02	12.81	14.57	26.56	1.88	27.91	15.04	21.97	23.10	19.26
Ethnicity	0.10	15.81	9.48	2.43	3.38	7.16	12.40	4.38	2.76	11.68	6.96
Father Education	23.07	15.99	22.09	37.86	29.44	27.41	28.82	20.92	36.23	31.57	27.34
Father Occupation	18.82	17.08	19.23	.	.	19.84	.	18.22	.	.	18.64
Mother Education	21.72	16.65	21.14	37.13	37.28	27.41	27.8	18.02	37.51	31.69	27.64
Mother Occupation	.	17.52	14.06	.	.	13.81	.	8.63	.	.	13.51
Sex	2.53	1.89	1.18	7.97	3.32	2.34	3.01	14.75	1.51	1.94	4.04

Source: Data from ENES, EH, PNAD, CASEN, ECV, ENCOVI, EMOVI, ENV, ENAHO. See Table 4 for more details on the datasets. All values are relative (%) contributions to random forest IOp estimates in Table 5.

It is important to recall that sex is a variable at the individual level and that the income concept is age-adjusted equivalized household income, not individual income or earnings. All individuals in a given household are allocated the same equivalized household income, so intra-household inequality is ignored entirely. As measured here, the contribution of sex to inequality of opportunity therefore reflects only differences in household composition, including the number and incomes of single-sex household.

A second remark concerns the (perhaps surprisingly) small impact of ethnicity. In some cases, this can be explained by the relatively homogeneous populations living in some countries today. For example, in Argentina and Chile, where the Shapley value for ethnicity is below 3%, over 90% of respondents do not report belonging to any ethnic minority. A more informative measure of the importance of ‘rare’ characteristics for those who happen to have them is provided by the Partial Dependency Plots presented below. But, in addition, when interpreting Shapley values, one should bear in mind that the structure of opportunities we observe today - reflected in the joint predictive power of the observed circumstances - is the result of historical evolution, and it is probable that the distributions of parental education and occupation themselves reflect the importance of ethnicity in earlier periods. Various ascriptive characteristics have historically influenced the lack of opportunities. Identifying a causal link or the historical mechanisms of evolution that shaped different countries in Latin America and the Caribbean since colonization is beyond the scope of this descriptive analysis.

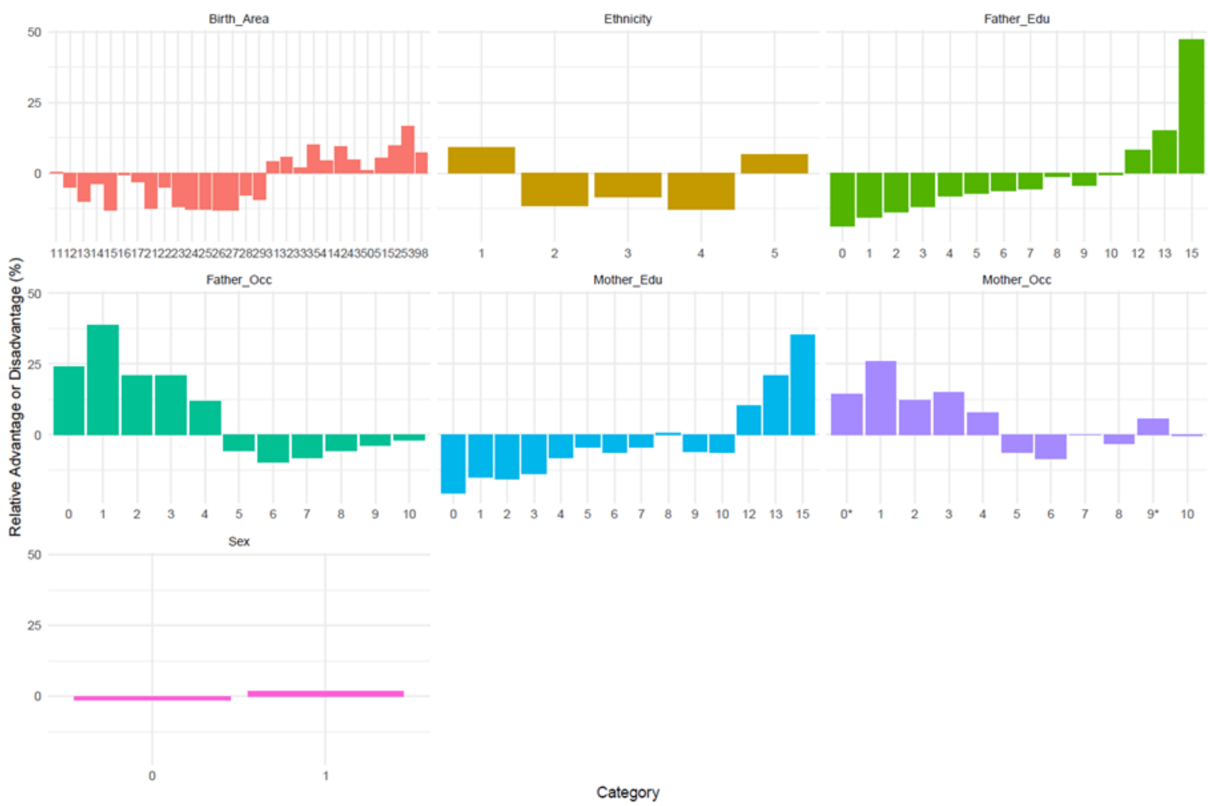
There are several country-specific results worth highlighting, although some comparisons should be made with caution, as noted earlier. A clear example of this is the comparison of the parental education share between countries with and without data on parental occupation. For instance, these shares are notably higher in Chile and Panama compared to Argentina or Bolivia. This discrepancy is very likely due, at least in part, to the fact that the parental education variable in Chile and Panama may be capturing some of the effect of the omitted parental occupation variable.[29] There is also considerable variation in the role of race and ethnicity, ranging from 0.1% in Argentina to 15.8% in Bolivia, a country with a significant indigenous population. At about 12%, the share of race and ethnicity is also high in Peru and Guatemala, which share similar demographic characteristics, and stands at 9.5% in Brazil, where more than half the population identifies as black or mixed-race. In Peru, where we have data from eleven survey waves, the contribution of ethnicity has gradually but consistently increased from 8.9% in 2001 to 11.7% in 2015.

As discussed, the Shapley value decomposition helps us disentangle the relative contribution of circumstances to IOp but says nothing about their marginal effects. To this end, the Partial Dependency Plots show the relative marginal advantage or disadvantage associated with each circumstance category. Figure 6 shows the PDP values for Brazil in 2014. To avoid over-interpreting the partial effect of a trait observed in very few respondents, a star above a category indicates that such a characteristic is observed in less than 50 respondents; in Brazil this only occurs for mothers whose occupations were in the Armed Forces.[30]

The picture arising from these partial dependence plots complements the analysis of average contributions from the Shapley decompositions. In most cases, it confirms the main messages from the earlier analysis: Biological sex is not an important predictor of household equivalized income (which, as noted earlier, ignores intra-household inequality by construction). Father's and mother's education are powerfully predictive, and the PDPs show the heterogeneous impacts by year of schooling more granularly: for both parents, impacts are only positive for completed secondary and some tertiary education, with particularly large effects for completing a university education. The predictive power of place of birth confirms longstanding impressions of the regional distribution of opportunities in Brazil, with being born in the North and Northeast having negative marginal effects and being born elsewhere having positive effects. The Centre-West belongs firmly to the latter category, with effects similar to or greater than those of the South and Southeast. The largest positive effect accrues to those born in the capital city of Brasília and the surrounding Federal District. Having parents who worked as managers, professionals, or technicians predicts higher incomes, and more so for fathers than for mothers. Being indigenous, black or mixed-race have negative partial effects and these are largest for Afro-Brazilians, followed closely by people of mixed race.

[29] Birthplace comparisons should also be informed by the fact that in Bolivia and Colombia this variable is a simple dummy for rural or urban birth, whereas in other countries it refers to a regional partition.
[30] Figures analogous to Figure 6 for the other nine countries are shown in Appendix Figures A10-18.

Figure 6: Partial Dependence Plots for Brazil (2014)



Source: PNAD, 2014. Notes: Categories are coded as follows. Sex: 0=Female; 1=Male. Father's and mother's education: Number of years of schooling, with 14 and above grouped together as 15. Father's and mother's education: 0=Armed Forces; 1=Managers; 2=Professionals; 3=Technicians and associate professionals; 4=Clerks; 5=Services and sales workers; 6=Agriculture, forestry and fishery workers; 7=Craft and trade workers; 8=Plant and machinery operators and assemblers; 9=Elementary occupations; 10=Unemployed. Ethnicity: 1=White; 2=Mixed race; 3=Indigenous; 4=Afro-Brazilian; 5=Other. Place of birth: 11-29=States in the North and Northeast regions; 31-52=States in the South, Southeast and Center West; 53=Federal District; 98=Foreign-born.

New estimates of inequality of opportunity for education

As we saw in Section 3, the earlier literature on inequality of opportunity in Latin America included studies where education was the outcome variable of interest. Some, like Andersen (2003) focused on measures of educational attainment, such as the schooling gap, whereas others, like Ferreira and Gignoux (2014) considered measures of educational achievement, such as standardized PISA test scores for 15-year-olds. Much as their counterparts looking at IOp for incomes, these pioneering studies adopted rather ad-hoc approaches to the choice of prediction function $\hat{f}(\ )$. In this subsection, we report some new results on IOp for education, both for attainment (now measured directly as years of schooling) and achievement, once again measured by PISA test scores.

Table 7 reports relative inequality of opportunity measures, $I(\hat{y})/I(y)$ for three different outcome variables. Column 1 simply reproduces the last column of Table 5, reporting relative IOp Gini coefficients for income estimated from Random Forests. Column 2 reports original estimates of relative IOp measures for years of schooling, calculated by applying the Random Forest method described above to reported years of schooling in the surveys listed in Table 4. The method used is identical to that described for the income results in Column 1, with two exceptions: the outcome variable y is educational *attainment*, measured by years of schooling, and the inequality index $I(\quad)$ is the Variance, a translation-invariant inequality measure. Column 3 reports relative IOp for educational *achievement*, where the outcome of interest y denotes standardized Mathematics test scores from the PISA dataset. These estimates, which are only available for a subset of the countries and years in Table 8, are reproduced from Brunori, Gil-Hernandez and Triventi (2025). They are also estimated by conditional inference random forests based on a set of circumstances comparable to the one used to produce the other two measures. This column also uses the Variance as the inequality indicator of choice, for the reasons discussed by Ferreira and Gignoux (2014).[31]

Table 7: The inherited share of inequality: a comparison of estimates

Wave	Relative Income IOp (Forest)	Relative Education IOp (Forest)	Relative Education "IOp" (PISA)
Argentina (2014)	46.0	28.9	23.8
Bolivia (2008)	58.8	51.6	
Brazil (2014)	65.7	40.6	21.1
Chile (2006)	52.1	36.0	30.8
Chile (2011)	52.8	42.6	34.6
Chile (2013)	49.0	41.5	
Chile (2015)	50.3	40.9	28.6
Colombia (2010)	48.0	34.4	30.9
Ecuador (2006)	56.1	47.4	
Ecuador (2014)	50.3	40.6	
Guatemala (2000)	56.9	47.3	
Guatemala (2006)	62.2	44.1	
Guatemala (2011)	55.4	39.5	
Mexico (2017)	55.9		20.0

[31] While the estimates in Column 3 can be interpreted as measures of *inherited inequality* exactly as the earlier income estimates, readers taking the more traditional view of *inequality of opportunity* as a part of inequality that lies beyond people's responsibilities may be unwilling to read much into these numbers. If one thinks of the "age of responsibility" - an age threshold below which we do not regard people as fully responsible for their actions – as being higher than 15, one might consider all inequality in test scores at 15 as being inequality of opportunity.

Panama (2003)	54.3	42.7	
Peru (2001)	60.8	42.6	
Peru (2006)	64.9	43.7	
Peru (2007)	65.9	42.7	
Peru (2008)	64.1	44.6	
Peru (2009)	62.5	44.4	42.2
Peru (2010)	60.7	44.3	
Peru (2011)	58.3	41.8	
Peru (2012)	59.2	41.1	32.8
Peru (2013)	60.2	41.6	
Peru (2014)	62.4	42.5	
Peru (2015)	63.3	41.2	29.9
Average	57.3	41.8	29.5

Source: Data for column 2 and 3 comes from surveys listed in Table 4. Data for column 3 comes from Brunori et al. (2025) and uses the following PISA waves: for Argentina, 2012; for Brazil, 2015; for Chile, 2006, 2012 and 2015; for Colombia, 2009; for Mexico, 2018; and for Peru, 2009, 2012, and 2015.

Table 7 suggests a couple of insights. First, while it is difficult to compare Gini ratios (in column 1) with the Variance ratios in columns 2 and 3, the latter two columns can be properly compared. Such a comparison indicates that the share of variance in achievement (test scores) that can be predicted by inherited characteristics is systematically lower than the share of attainment (years of schooling). This suggests that family background, race and sex are more predictive of the timing when a person leaves school or college than of how well they do academically, conditional on being at school at age 15. Could it be that the latter depends to a greater extent on innate ability, which may be less intergenerationally persistent than other factors? It is impossible to tell from these data, of course, but the pattern is intriguing.

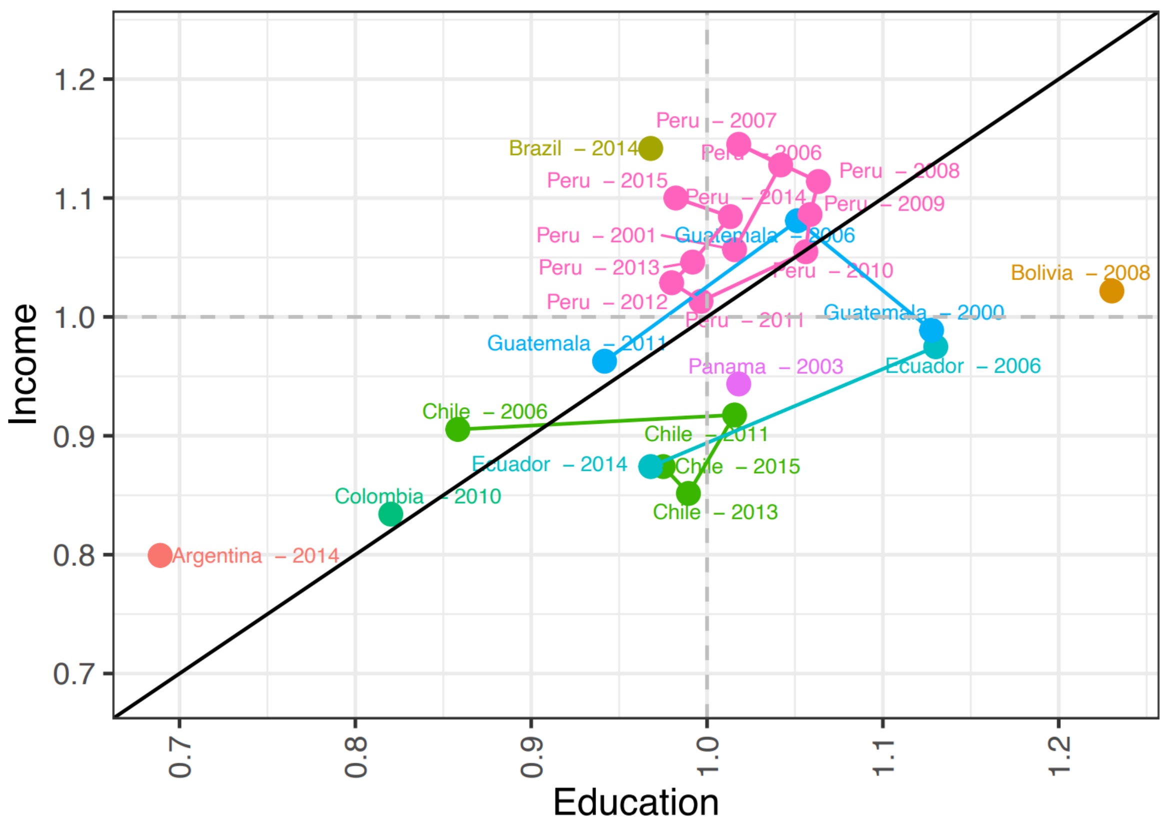
Second, the cross-survey Pearson correlation coefficient between the first two columns is positive, but not particularly strong, at 0.53.[32] Since these series come from the same data sets and use the same circumstance variables, this suggests that the processes determining the inheritability of educational attainment and incomes, while related, are far from being the same. It is likely that, while early childhood upbringing and schooling are part of both processes, labour market matching and remuneration outcomes can mediate how circumstances predict incomes, but not years of schooling. Similarly for assortative matching processes in family formation.

Figure 7 presents a scatterplot depicting the relationship between these two variables in columns 1 and 2 of Table 7. To alleviate the comparability problem between different inequality measures, the IOp measures are expressed relative to the mean. The positive but imperfect correlation is

[32] Excluding Mexico, where "own education" is a variable used to impute income, creating a mechanically strong correlation between the two variables.

visually clear. Enhancing the scatter plot, observations for the same country are linked chronologically. While IOp for educational attainment fell in Guatemala between 2000 and 2011, the changes in IOp for income were mixed, and ultimately small. Chile seems to have seen a small decline in IOp for income between 2006 and 2013, but an increase in IOp for education. Peru's trajectory looks like a random walk. Finally, the vertical distance from the 45-degree line indicates the degree to which income IOp exceeds education IOp, relative to the mean in the series. Brazil and Argentina stand out above the line, while Bolivia and Ecuador are some distance below it. This suggests that Brazil and Argentina exhibit a higher level of inequality of opportunity in income than would be anticipated given their levels of education IOp, while the reverse is true for Bolivia and Ecuador.

Figure 7: Income and education opportunity

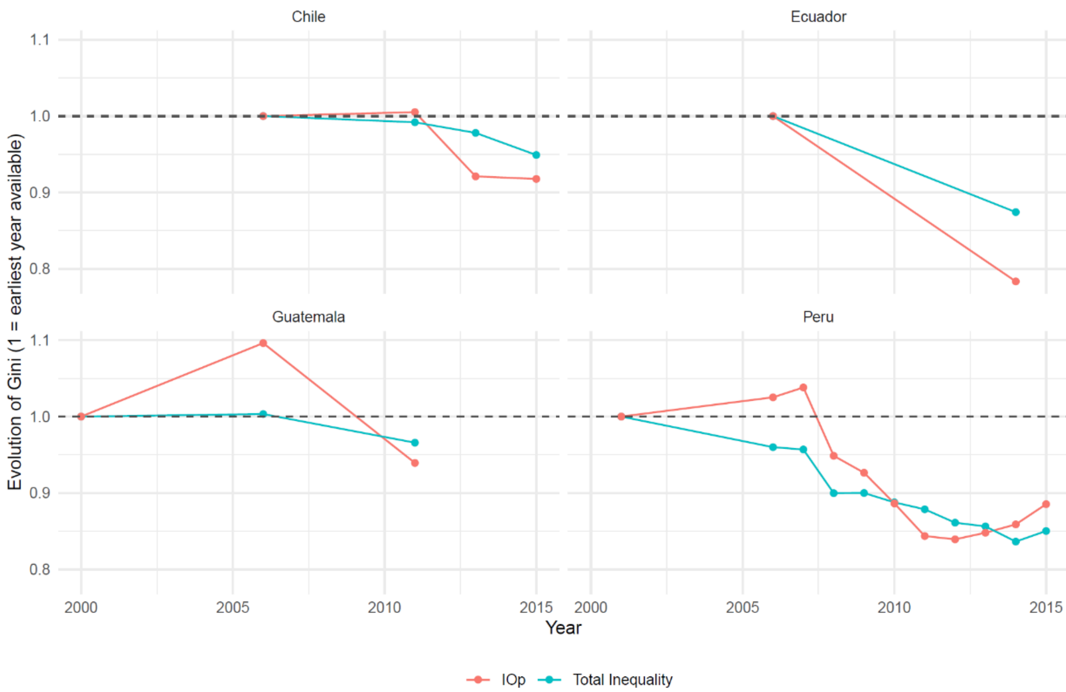


Source: Data comes from surveys listed in Table 4.

Inequality of Opportunity Dynamics

For the four countries that we observe at multiple points in time, namely Chile, Guatemala, Ecuador, and Peru, Figure 8 summarizes the dynamics of income inequality and inequality of opportunity. The income Gini series is shown in blue, while the opportunity Gini series is shown in red. In all series, the first available value is normalized to 1.

Figure 8: IOp dynamics in Chile, Ecuador, Guatemala and Peru



Source: Data comes from Table 7, with the initial values in all series normalized to one.

In Chile, both overall inequality and IOp remain relatively stable between 2006 and 2011 but decline thereafter, with IOp declining more markedly in proportional terms. Similarly, in both Guatemala and Peru, overall inequality and IOp estimates tend to move in tandem. In Peru, overall inequality decreases throughout the period, with a particularly sharp decline between 2007 and 2010, followed by a reversal of the trend only between 2014 and 2015. Notably, this entire decline is driven by falling inequality of opportunity. In Ecuador there is a rather sharp decline (of more than 20%) in inequality of opportunity, which seems to have contributed to a (proportionately) smaller decline in overall income inequality.

6. Conclusions

This chapter has reviewed estimates of inherited inequality in income and education in Latin America, drawing on the intergenerational mobility and the inequality of opportunity literatures. Though the two approaches have different origins, both relative IOp and relative IGM measures can be interpreted as capturing the share of contemporaneous inequality that can be predicted by inherited factors: parental income (or education) in the mobility literature, and a vector of inherited circumstances in the opportunity literature. Although, the circumstance vector can in principle include parental income – and indeed has done in studies in other regions[33] – in Latin America,

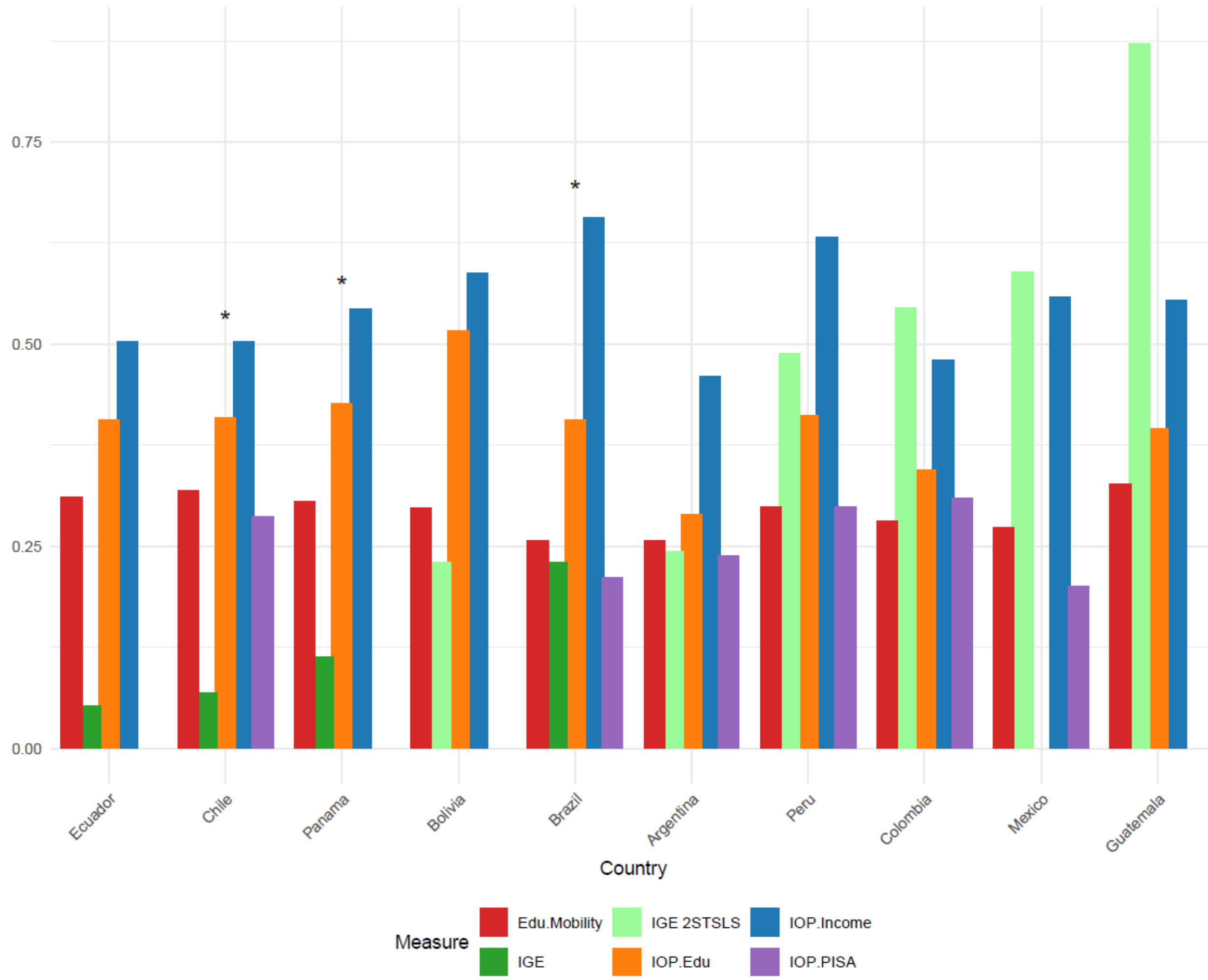
[33] See, e.g., Björklund, Jäntti and Roemer (2012).

the circumstances have typically been categorical variables such as sex, race, place of birth, parental education and parental occupation.

Figure 9 provides a visual comparison of five different kinds of estimates discussed in the chapter, for ten countries, which are arranged in increasing order of the income IGE estimate.[34] Starting with education, three regularities appear to hold. First, squared correlation coefficients of years of schooling are sizable, and relatively similar across countries, suggesting that 26% to 33% of the variance in children's years of schooling can be predicted by the schooling attainment of their parents. These correspond to correlation coefficients in the 0.51 – 0.57 range, very high by international standards. Second, and nonetheless, an even larger share of the variance in years of schooling can be predicted by inherited characteristics when other circumstances are added to parental education as predictors. The IOp in Education estimates in the orange bars range from 29% (in Argentina) to 52% (in Bolivia). This suggests that interpreting the correlation of years of schooling across generations as a full measure of how much variation in the schooling of today's generation is inherited would lead to considerable underestimation. Third, the share in the variance of PISA test scores – a measure of actual learning, or educational achievement – that can be predicted by family background and inherited characteristics is lower than that for years of schooling. Represented in purple in the Figure, this estimate is only available for six of the ten countries, and it ranges from 20% to 30%. The background characteristics we observe seem to be less predictive of overall learning variation at age 15 than of the variation in the quantity of schooling attained.

[34] The data underlying this Figure is in Appendix Table A2.

Figure 9: A comparison of inherited inequality estimates for ten Latin American countries

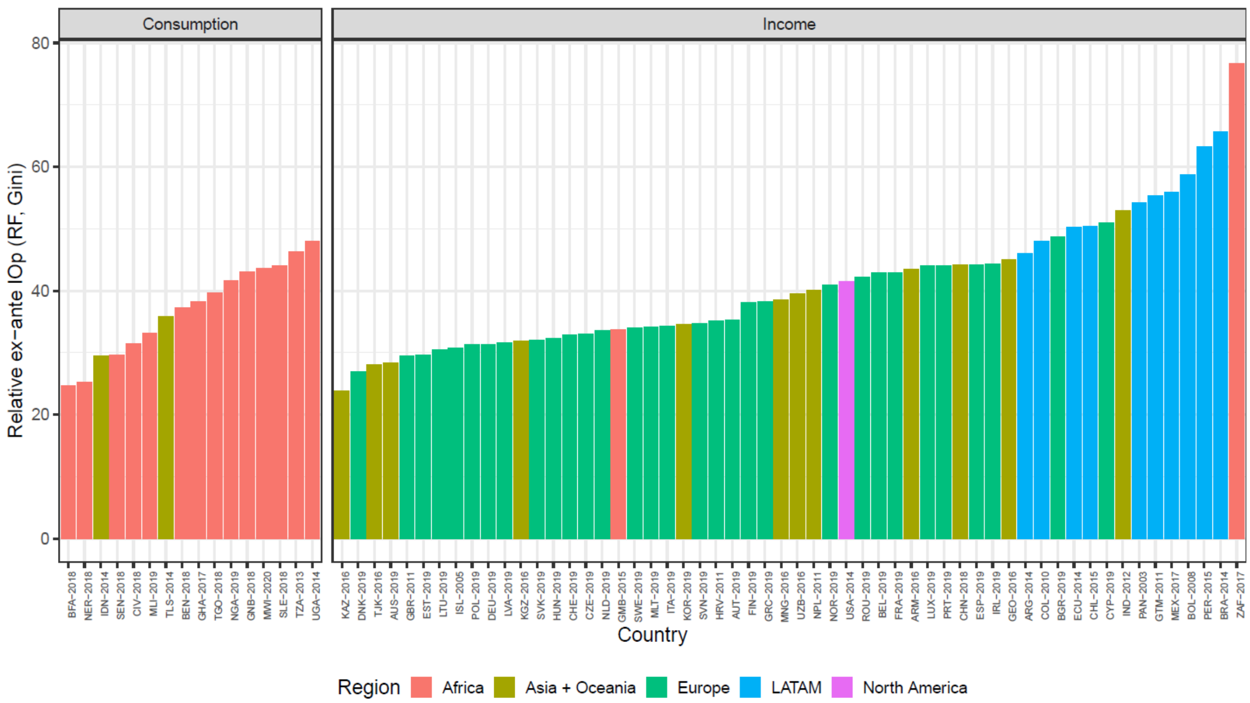


*Notes: All education mobility estimates are squared correlation coefficients of years of schooling produced by using the data from Neidhöfer et al. (2018). The corresponding regression coefficient estimates using the same samples are in Column 2 in Table 1. IGE estimates are the squares of our preferred estimates of income IGM, from column 1 in Table 2. * denotes IGE estimates reported in a published peer-reviewed journal article; other IGE estimates come from working papers. Dark green denotes OLS estimates of IGE, whereas light green denotes TSTSLS estimates of IGE. IOp.Education are Random Forest estimates from column 2 in Table 7. IOp.Income are Random Forest estimates from Column 1 in Table 7. IOp.PISA are estimates from Column 3 in Table 7 and were originally produced by Brunori et al. (2025). The data for this figure is available in table format in Appendix Table A2.*

For income persistence, on the other hand, the picture is less clear. One regularity evident in the figure is that the share of predicted income variation (measured by the Gini coefficient) is even higher than for years of schooling. The comparison is far from perfect, since the variance and the Gini coefficient are different measures of dispersion. Nonetheless, it is instructive that the inherited characteristics or circumstances we observe can predict between 46% (in Argentina) and 66% (in Brazil) of the variation in current incomes. Figure 10 plots these estimates alongside analogous figures for another 62 countries, calculated using identical empirical protocols for the Global

Estimates of Opportunity and Mobility database. The sixteen bars in the left panel refer to consumption expenditures, rather than income, and are thus not comparable with the fifty-six country estimates on the right. Amongst these, Latin America stands out as a region with particularly high levels of inequality of opportunity. Only South Africa exceeds the top five Latin American countries, and only Bulgaria, Cyprus and India fall within the range of Latin American estimates.

Figure 10: Random Forest estimates of Inequality of Opportunity for Income Acquisition across 72 countries



Source: Global Estimates of Opportunity and Mobility database (https://geom.ecineq.org/). Notes: These seventy-two estimates are obtained from random forest estimates exactly analogous to those presented in this chapter, which are included in the figure as the blue bars. They are for the latest years available in the database. The estimates on the left panel measure IOp for consumption expenditures, rather than income, and are thus not comparable to those on the right panel, which use the definition of income described in Section 4 above.

But the other series of income estimates plotted in Figure 9, namely the squared IGE estimates shown in green, tell a somewhat different story. First, note that the IGE estimates are squared so that they can be compared to the ρ^2 and relative IOp indices in the Figure and interpreted as shares of variation predicted by parental income. Ideally, one would have used squared correlation coefficients, which would then be equal to the R^2 of the Galtonian income regressions used to estimate the IGE coefficient. These R^2 are not available for all relevant studies, so we use squared

βs instead. This corresponds to assuming that the standard deviation of incomes is the same for parents and children – clearly not a plausible assumption, but nonetheless probably the best way to convert IGEs to the metric of explained shares, so they can be compared to the other inherited inequality estimates in Figure 9.

Unlike the other estimates in Figure 9, these IGE squares vary widely across countries, from 0.05 in Ecuador, to 0.87 in Guatemala – spanning almost the full (0, 1) theoretical range for the estimator. If taken at face value, these numbers would suggest that parental income can only predict 5% of the variance in the logarithms of children's income in Ecuador, but almost 90% in Guatemala. The corresponding income IOp estimates are 50% for Ecuador and 55% for Guatemala. Behind this extreme cross-country variation is the wide variation in methods and data sources used in the region and shown in Figure 4. The IGE estimates for Ecuador in Figure 4, admittedly an extreme case, range from 0.23 to Grawe's (2004) estimate of approximately 1.10. Estimates for Mexico also range from about 0.25 to almost 0.80.

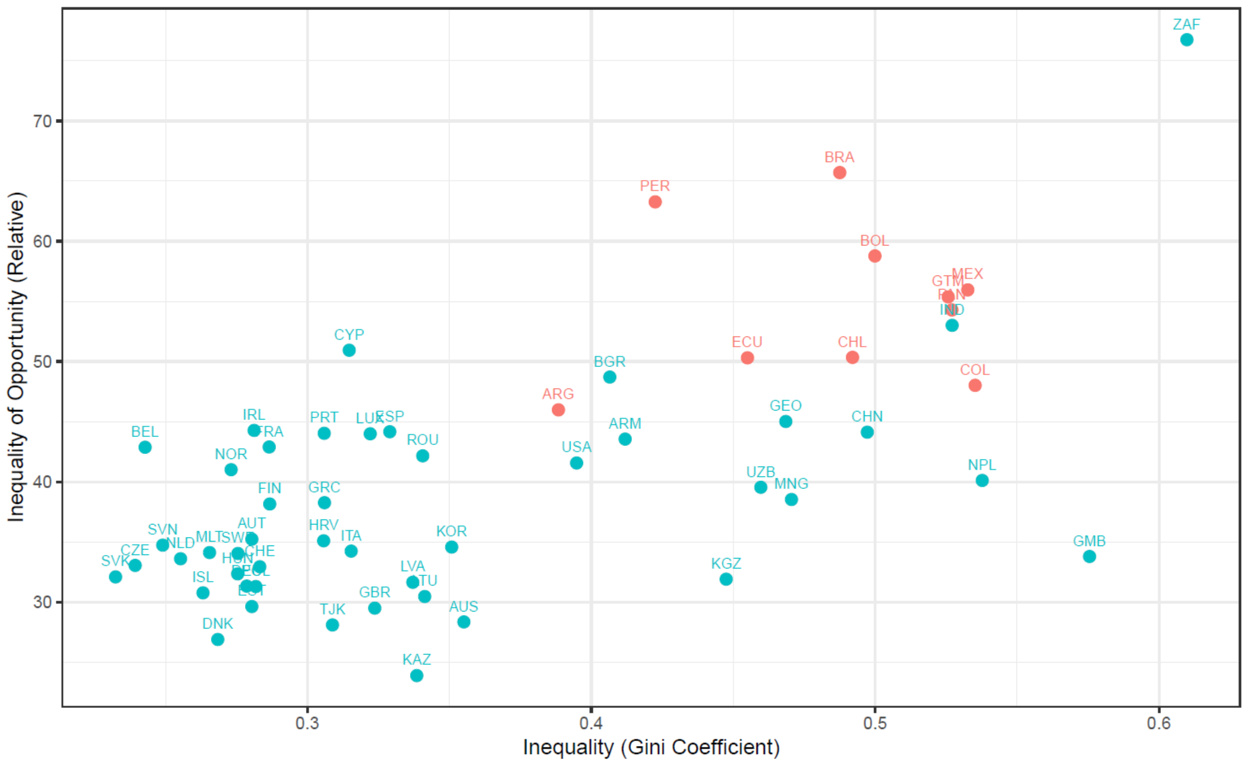
The uncertainty revealed by these disparities ultimately reflects the data limitations discussed in Section 3. On the one hand, most studies for Latin America use data that do not contain observed incomes for the two generations, and must therefore rely on TSTSLS estimators, which are known to be potentially biased, and likely upwardly biased. This concern is exacerbated when the set of parental characteristics common to both surveys on which the estimator relies is very limited, as in the estimate for Guatemala from Muñoz and van der Weide (2025).

On the other hand, there are now a few studies that have attempted to match observed incomes from parents and children available in administrative data. Because OLS estimates on observed data are generally preferred to TSTSLS estimates, we built that preference into the selection criteria leading from the range of coefficients in Figure 4 to our preferred IGE estimates in Table 2. The table therefore contains four OLS estimates: for Brazil, Chile, Ecuador and Panama. As shown in Figure 4, these four estimates are all lower than any TSTSLS for the same countries, sometimes markedly so. Unfortunately, however, these studies face a (different) data limitation of their own: incomes from parents and children in the informal economy – which sometimes accounts for more than half of the labour force – are not observed. They must be imputed, and the imputation methods vary from study to study. In the case of Chile, for example, no imputation is attempted, and lineages where either the parent or the child worked in the informal sector are therefore simply missing from the analysis.

It is therefore not clear, at least at this point, that each and every one of these new studies based on administrative data succeeds in capturing the 'true' IGE. The estimates for Ecuador and Chile, which imply that parental income can only predict 5% and 7% (respectively) of the variation in children's income, do not appear credible, particularly given the IOP and education numbers in Figure 4. They suggest that these methods are not yet fully mature, and that works remains to be done. Given the uncertainty arising from this combination of data and methodological challenges in the estimation of income IGEs, we regard the IOp measures reported in Section 5 and shown in Figures 9 and 10 above as likely to provide more reliable estimates of inherited income inequality in the region.

Using these income IOp estimates, Figure 11 plots our preferred measure of relative inequality of opportunity against total income inequality, measured by the Gini coefficient, for the latest available years in the GEOM database. Given the close relationship between relative mobility and inequality of opportunity which we have discussed, this graph is closely analogous to the well-known 'Great Gatsby curve', in which Corak (2013) plotted intergenerational elasticities against cross-section inequality.

Figure 11: The Opportunity Great Gatsby Curve



Source: Global Estimates of Opportunity and Mobility database (https://geom.ecineq.org/). Notes: These fifty-six estimates were obtained from random forest estimates exactly analogous to those presented in this chapter. They are for the latest years available in the database. Countries where consumption expenditures were used to measure IOp are omitted. The horizontal axis represents overall inequality, measured by the Gini Coefficient, and the vertical axis represents the relative measure of Inequality of Opportunity. The correlation coefficient between both measures is 0.64.

As in the original Gatsby curve, Figure 11 shows a clear positive association between cross-sectional inequality and the degree of intergenerational persistence, or the inheritability of inequality. This association – which has been described as "one of the most visible stylized facts in contemporary inequality research" (Durlauf et al. 2022, p. 572) – reflects the fact that unequal outcomes today shape unequal opportunities for the next generation and, in turn, unequal opportunities beget unequal outcomes. The ten Latin American countries in the Figure lie in its Northeast quadrant, the area of high and highly inheritable inequality, where income differences among families are not only large, but persistent across generations.

References

Acciari, P., Polo, A. and Violante, G.L., 2022. And yet it moves: Intergenerational mobility in Italy. *American Economic Journal: Applied Economics*, 14(3), 118–163.

Ahsan, M.N., Emran, M.S., Jiang, H., Han, Q. and Shilpi, F., 2023. Growing up together: Sibling correlation, parental influence, and intergenerational educational mobility in developing countries. GLO Discussion Paper No. 1123. Global Labor Organization (GLO).

Alvaredo, F., Bourguignon, F., Ferreira, F.H.G. and Lustig, N., 2025. Inequality bands: seventy-five years of measuring income inequality in Latin America. *Oxford Open Economics*, 4(S1), pp.i9–i35.

Andersen, L.E., 2003. Social mobility in Latin America: links with adolescent schooling, in Duryea, S., Cox Edwards, A. and Ureta, M. (eds), *Critical Decisions at a Critical Age: Adolescents and Young Adults in Latin America*, Inter-American Development Bank.

Anderson, L.R., Sheppard, P., and Monden, C.W.S., 2018. Grandparent effects on educational outcomes: A systematic review. Sociological Science, 5(6), 114–142.

Araya, F., 2019. Evidence about intergenerational earnings mobility for a developing country: The Uruguayan case. *El Trimestre Económico*, 86(342), 265–305.

Attanasio, O., de La O, A.L., Ferreira, F.H.G., Ibáñez, A.M. and Messina, J., 2025. Inequality in Latin America and the Caribbean: a wide-ranging review, *Oxford Open Economics*, 4(S1), pp. i1–i8.

Atkinson, A.B., 1970. On the measurement of inequality. *Journal of Economic Theory*, 2(3), 244–263.

Bhattacharya, N. and Mahalanobis, B., 1967. Regional disparities in household consumption in India, *Journal of the American Statistical Association*, 62(317), pp. 143–161.

Beccaria, L., Maurizio, R., Trombetta, M. and Vázquez, G., 2022. Short-term income mobility in Latin America in the 2000s: Intensity and characteristics. *Socio-Economic Review*, 20(3), 1039–1067.

Behrman, J.R., Birdsall, N. and Székely, M., 1999. *Intergenerational mobility in Latin America: Deeper markets and better schools make a difference* (Vol. 3). Carnegie Endowment for International Peace, Global Policy Program.

Behrman, J.R., Gaviria, A., Székely, M., Birdsall, N. and Galiani, S., 2001. Intergenerational mobility in Latin America [with comments]. *Economía*, 2(1), 1–44.

Behrman, J.R. and Wolfe, B.L., 1987. Investments in schooling in two generations in pre-revolutionary Nicaragua: The roles of family background and school supply. *Journal of Development Economics*, 27(1–2), 395–419.

Binder, M. and Woodruff, C., 2002. Inequality and intergenerational mobility in schooling: The case of Mexico. *Economic Development and Cultural Change*, 50(2), 249–267.

Björklund, A. and Jäntti, M., 1997. Intergenerational income mobility in Sweden compared to the United States. *The American Economic Review*, 87(5), 1009–1018.

Björklund, A., Jäntti, M. and Roemer, J., 2012. Equality of opportunity and the distribution of long-run income in Sweden. *Social Choice and Welfare*, 39(2–3), 675–696.

Bloise, F., Brunori, P. and Piraino, P., 2021. Estimating intergenerational income mobility on sub-optimal data: A machine learning approach. *The Journal of Economic Inequality*, 19(4), 643–665.

Bourguignon, F., Ferreira, F.H.G. and Menéndez, M., 2007. Inequality of opportunity in Brazil, *Review of Income and Wealth*, 53, pp. 585–618.

Bratberg, E., Davis, J., Mazumder, B., Nybom, M., Schnitzlein, D.D. and Vaage, K., 2017. A comparison of intergenerational mobility curves in Germany, Norway, Sweden, and the US, *Journal of Economic Inequality*, 15(1), 72-101.

Britto, D.G.C., Fonseca, A., Pinotti, P., Sampaio, B. and Warwar, L., 2022. Intergenerational mobility in the land of inequality, *CESifo Working Paper Series* 10004, CESifo.

Brunori, P., Ferreira, F.H.G. and Peragine, V., 2013. Inequality of opportunity, income inequality, and economic mobility: Some international comparisons. In: E. Paus, ed. *Getting Development Right: Structural Transformation, Inclusion, and Sustainability in the Post-Crisis Era*. New York: Palgrave Macmillan US.

Brunori, P., Ferreira, F.H.G. and Neidhöfer, G., 2025. Inequality of opportunity and intergenerational persistence in Latin America, *Oxford Open Economics*, 4(S1), pp. i167–i199.

Brunori, P., Ferreira, F. and Salas-Rojo, P., 2023a. Inherited inequality: A general framework and an application to South Africa. LSE III Working Paper No. 107.

Brunori, P., Gil-Hernández, C. and Triventi, M., 2025. *Unfair educational inequality worldwide: A theoretically driven machine learning approach.* International Inequalities Institute, London School of Economics and Political Science, *mimeo*.

Brunori, P., Hufe, P. and Mahler, D.G., 2023b. The roots of inequality: Estimating inequality of opportunity from regression trees. *The Scandinavian Journal of Economics, 125(3), 900-932.*

Brunori, P., Palmisano, F. and Peragine, V., 2019. Inequality of opportunity in Sub-Saharan Africa, *Applied Economics*, 51(60), pp. 6428–6458.

Brunori, P., Peragine, V. and Serlenga, L., 2019. Upward and downward bias when measuring inequality of opportunity, *Social Choice and Welfare*, 52(4), pp. 635–661.

Campos-Vázquez, R.M., Delgado Barrera, V.H. and Vélez-Grajales, R., 2021. Intergenerational economic mobility in Mexico. *Estudios Económicos (México, DF)*, 36(1), 151–176.

Carranza, R., De Rosa, M. and Flores, I., 2025. Wealth inequality in Latin America (2000–2020): data, facts and conjectures, *Oxford Open Economics*, 4(Supplement_1), pp. i36–i54.

Celhay, P. and Gallegos, S., 2015. Persistence in the transmission of education: Evidence across three generations for Chile. *Journal of Human Development and Capabilities*, 16(3), 420–451.

Celhay, P. and Gallegos, S., 2025. Schooling mobility across three generations in six Latin American countries. *Journal of Population Economics*, 38(1), 23–57.

Checchi, D. and Peragine, V., 2010. Inequality of opportunity in Italy, *The Journal of Economic Inequality*, 8(4), pp. 429–450.

Chetty, R., Hendren, N., Kline, P., Saez, E. and Turner, N., 2014. Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *American Economic Review*, 104(5), 141–147.

Ciaschi, M., Marchionni, M. and Neidhöfer, G., 2025. Intergenerational mobility in Latin America: The multiple facets of social status and the role of mothers. *Journal of Population Economics*.

Connolly, M., Corak, M. and Haeck, C., 2019. Intergenerational mobility between and within Canada and the United States. *Journal of Labor Economics*, 37(S2), S595–S641.

Corak, M., 2013. Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3), pp.79–102.

Corak, M., 2020. The Canadian geography of intergenerational income mobility. *The Economic Journal*, 130(631), 2134–2174.

Cortés Orihuela, J., Díaz, J.D., Gutiérrez Cubillos, P. and Troncoso, P.A., 2025. The two-sample two-stage least squares method to estimate the intergenerational earnings elasticity. *Journal of Economic Inequality*, 23(2), 549–570.

Cortés Orihuela, J., Díaz, J.D., Gutiérrez Cubillos, P., Troncoso, P.A. and Villarroel, G.I., 2024. Intergenerational earnings mobility in Chile: The tale of the upper tail. *Empirical Economics*, 67(5), 2411–2447.

Cuesta, J., Ñopo, H. and Pizzolitto, G., 2011. Using pseudo-panels to measure income mobility in Latin America. *Review of Income and Wealth*, 57(2), 224–246.

Dahan, M. and Gaviria, A., 2001. Sibling correlations and intergenerational mobility in Latin America. *Economic Development and Cultural Change*, 49(3), 537–554.

Daude, C. and Robano, V., 2015. On intergenerational (im)mobility in Latin America. *Latin American Economic Review*, 24(1), 9–34.

Daza Báez, N., 2025. Intergenerational earnings mobility in Mexico. *Journal of Income Distribution*.

Del Pozo, D. and Moreno, L., 2024. *Intergenerational mobility in Ecuador: A first approach.* National Institute of Statistics and Censuses Working paper No. 19 – December 2024

Deutscher, N. and Mazumder, B., 2020. Intergenerational mobility across Australia and the stability of regional estimates. *Labour Economics*, 66, 101861.

Deutscher, N. and Mazumder, B., 2023. Measuring intergenerational income mobility: A synthesis of approaches. *Journal of Economic Literature*, 61(3), 988–1036.

Doruk, Ö.T., Pastore, F. and Yavuz, H.B., 2022. Intergenerational mobility: An assessment for Latin American countries. *Structural Change and Economic Dynamics*, 60, 141–157.

Dunn, C.E., 2007. The intergenerational transmission of lifetime earnings: Evidence from Brazil. *The BE Journal of Economic Analysis and Policy*, 7(2).

Durlauf, S., Kourtellos, A. and Tan, C.M., 2022. The Great Gatsby curve. *Annual Review of Economics*, 14, 571–605.

Emran, M.S., Greene, W. and Shilpi, F., 2018. When measure matters: Coresidency, truncation bias, and intergenerational mobility in developing countries. *Journal of Human Resources*, 53(3), 589–607.

Eslava, F. and Valencia Caicedo, F., 2025. Origins of Latin American inequality, *Oxford Open Economics*, 4(S1), pp. i595–i614.

Fernández, R., Pagés, C., Székely, M. and Acevedo, I., 2025. Education inequalities in Latin America and the Caribbean, *Oxford Open Economics*, 4(S1), pp. i55–i76.

Ferreira, F.H.G. and Brunori, P., 2024. Inherited inequality, meritocracy, and the purpose of economic growth, III Working Paper (147), International Inequalities Institute, LSE.

Ferreira, F.H.G. and Gignoux, J., 2011. The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of Income and Wealth*, 57(4), 622–657

Ferreira, F.H.G. and Gignoux, J., 2014. The measurement of educational inequality: Achievement and opportunity. *The World Bank Economic Review*, 28(2), 210–246.

Ferreira, F.H.G., Messina, J., Rigolini, J., López-Calva, L.-F., Lugo, M.A. and Vakis, R., 2013. *Economic mobility and the rise of the Latin American middle class.* Washington, DC: World Bank.

Ferreira, F.H.G. and Peragine, V., 2016. Individual responsibility and equality of opportunity, in Adler, M.D. and Fleurbaey, M. (eds), *The Oxford Handbook of Well-Being and Public Policy*, Oxford University Press, Oxford.

Ferreira, S.G. and Veloso, F.A., 2006. Intergenerational mobility of wages in Brazil. *Brazilian Review of Econometrics*, 26(2), 181–211.

Fields, G.S., 2000. Income mobility: concepts and measures, in Birdsall, N. and Graham, C.L. (eds), *New Markets, New Opportunities? Economic and Social Mobility in a Changing World*, Brookings Institution Press.

Fields, G.S., Hernández, R.D., Freije, S., Puerta, M.L.S., Arias, O. and Assunção, J., 2007. Intragenerational income mobility in Latin America [with comments]. *Economía*, 7(2), 101–154.

Fleurbaey, M., 1994. On fair compensation, *Theoretical Decisions*, 36, pp. 277–307.

Foster, J.E. and Shneyerov, A.A., 2000. A general class of additively decomposable inequality measures. *Journal of Economic Theory*, 91(2), 199–222.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.

Gabrielli, M. V. 2025. Climbing the Ladder? How Latin Americans see their economic mobility. Unpublished manuscript. Paris School of Economics.

Gáfaro, M., Ibáñez, A.M., Sánchez-Ordóñez, D. and Ortiz, M.C., 2025. Farm size and income distribution of Latin American agriculture: new perspectives on an old issue, *Oxford Open Economics*, 4(S1), pp. i148–i166.

Gaentzsch, A. and Zapata Román, G., 2018. More educated, less mobile? Diverging trends in income and educational mobility in Chile and Peru. *SSRN Working Paper.*

Gamboa, L.F. and Waltenberg, F.D., 2012. Inequality of opportunity for educational achievement in Latin America: evidence from PISA 2006–2009, *Economics of Education Review*, 31(5), pp. 694–708.

Genicot, G and Ray, D., 2023. Measuring upward mobility. *American Economic Review*, 113(11), 3044–3089.

Genicot, G. ⓡ, Ray, D. ⓡ and Concha-Arriagada, C., 2024. Upward mobility in developing countries. *NBER Working Paper* No. 32391, Grawe, N.D., 2004. Intergenerational mobility for whom? The experience of high- and low-earning sons in international perspective. In: *Generational income mobility in North America and Europe*, pp. 58–89.

Heckman, J.J. and Hotz, V.J., 1986. An investigation of the labor market earnings of Panamanian males: evaluating the sources of inequality. *Journal of Human Resources*, pp.507–542.

Heidrich, S., 2017. Intergenerational mobility in Sweden: a regional perspective. *Journal of Population Economics*, 30(4), 1241–1280.

Helsø, A.L., 2021. Intergenerational income mobility in Denmark and the United States. *The Scandinavian Journal of Economics*, 123(2), 508–531.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N. and Verashchagina, A., 2008. The inheritance of educational inequality: international comparisons and fifty-year trends. *The BE Journal of Economic Analysis and Policy*, 7(2).

Hjorth-Trolle, A. and Landersø, R., 2025. The different sources of intergenerational income mobility in high- and low-income families. *Journal of Applied Econometrics. Online 21 August 2025.*

Hothorn, T., Hornik, K. and Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674.

Jäntti, M. and Jenkins, S.P., 2015. Income mobility. In: *Handbook of Income Distribution*, Vol. 2, pp.807–935. Elsevier.

Jiménez, M., 2016. Movilidad intergeneracional del ingreso en Argentina: un análisis de sus cambios temporales desde el enfoque de igualdad de oportunidades. *Documentos de Trabajo del CEDLAS.*

Kenedi, G. and Sirugue, L., 2023. Intergenerational income mobility in France: a comparative and geographic analysis. *Journal of Public Economics*, 226, 104974.

Lam, D. and Schoeni, R.F., 1993. Effects of family background on earnings and returns to schooling: evidence from Brazil. *Journal of Political Economy*, 101(4), 710–740.

Lambert, P.J. and Aronson, J.R., 1993. Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal*, 103(420), 1221–1227.

Leites, M., Ramos, X., Rodríguez, C. and Vilá, J., 2022. Intergenerational mobility along the income distribution: estimates using administrative data for a developing country. *Serie Documentos de Trabajo*, 05/22.

Lubotsky, D. and Wittenberg, M., 2006. Interpretation of regressions with multiple proxies. *The Review of Economics and Statistics*, 88(3), 549–562.

Marteleto, L., Gelber, D., Hubert, C. and Salinas, V., 2012. Educational inequalities among Latin American adolescents: continuities and changes over the 1980s, 1990s and 2000s. *Research in Social Stratification and Mobility*, 30(3), 352–375.

Meneses, F. and Asahi, K., 2025. Intergenerational mobility in Chile: a year-to-year analysis of a national cohort of students. In: *Opportunity, Mobility and Inequality*, Vol. 31, pp.5–40. Emerald Publishing Limited.

Mitnik, P.A., 2020. Intergenerational income elasticities, instrumental variable estimation, and bracketing strategies. *Sociological Methodology*, 50, 1–46.

Moreno, H., 2021. The influence of parental and grandparental education in the transmission of human capital. *ECINEQ Working Paper*, 588.

Muñoz, E., 2024. The geography of intergenerational mobility in Latin America and the Caribbean. *Economía*, 23(1), 333–354.

Muñoz, E. and van der Weide, R., 2025. Intergenerational income mobility around the world. *Policy Research Working Paper*. World Bank Group.

Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Mahler, D.G., Ramasubbaiah, R.G.N. and Thewissen, S., 2018. *Fair progress? Economic mobility across generations around the world.* World Bank Publications.

Neidhöfer, G., 2019. Intergenerational mobility and the rise and fall of inequality: lessons from Latin America. *The Journal of Economic Inequality*, 17(4), 499–520.

Neidhöfer, G., Ciaschi, M. and Gasparini, L., 2022. Intergenerational mobility of economic well-being in Latin America. *CEDLAS Documento de Trabajo*, No. 303.

Neidhöfer, G., Ciaschi, M., Gasparini, L. and Serrano, J., 2024. Social mobility and economic development. *Journal of Economic Growth*, 29(2), 327–359.

Neidhöfer, G., Serrano, J. and Gasparini, L., 2018. Educational inequality and intergenerational mobility in Latin America: a new database. *Journal of Development Economics*, 134, 329–349.

Núñez, J. and Miranda, L., 2010. Intergenerational income mobility in a less-developed, high-inequality context: the case of Chile, *The B.E. Journal of Economic Analysis and Policy*, 10(1), pp. 1–17.?

Núñez, J. and Tartakowsky, A., 2011. The relationship between income inequality and inequality of opportunities in a high-inequality country: the case of Chile, *Applied Economics Letters*, 18(4), pp. 359–369.

Olivetti, C. and Paserman, M.D., 2015. In the name of the son (and the daughter): intergenerational mobility in the United States, 1850–1940. *American Economic Review*, 105(8), 2695–2724.

Paes de Barros, R., Ferreira, F.H.G., Molinas Vega, J.R. and Saavedra Chanduvi, J., 2009. Measuring inequality of opportunities in Latin America and the Caribbean, Washington, DC: World Bank.

Peragine, V., Palmisano, F. and Brunori, P., 2014. Economic growth and equality of opportunity. *The World Bank Economic Review*, 28(2), 247–281.

Ramírez Zuluaga, J.S., 2016. *Movilidad social intergeneracional por ingresos en Colombia.* Doctoral dissertation.

Ramos, X. and van de Gaer, D., 2016. Approaches to inequality of opportunity: principles, measures and evidence. *Journal of Economic Surveys*, 30(5), 855–883.

Ravallion, M. and Chen, S., 2003. Measuring pro-poor growth. *Economics Letters*, 78(1), 93–99.

Roemer, J.E., 1993. A pragmatic theory of responsibility for the egalitarian planner, *Philosophy and Public Affairs*, 22(2), pp. 146–166.

Roemer, J.E., 1998. *Equality of Opportunity.* Cambridge, MA: Harvard University Press.

Roemer, J.E. and Trannoy, A., 2016. Equality of opportunity: theory and measurement, *Journal of Economic Literature*, 54(4), pp. 1288–1332.

Shaikh, M. and Gomez Tamayo, S. 2025. “Inequality of Opportunity in Colombia: Assessing the impact of family background on adult incomes, 1997-2023”. Unpublished manuscript, Vienna University of Economics and Business.

Shapley, L.S., 1953. A value for n-person games. In: Kuhn, H.W. and Tucker, A.W. (Eds.), *Contributions to the Theory of Games (AM-28), Vol. II*, pp.307–318. Princeton: Princeton University Press.

Shorrocks, A.F., 2013. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *The Journal of Economic Inequality*, 11(1), 99–126.

Sokoloff, K.L. and Engerman, S.L., 2000. Institutions, factor endowments, and paths of development in the New World*, Journal of Economic Perspectives*, 14(3), pp. 217–232.

Solon, G., 2014. Theoretical models of inequality transmission across multiple generations. *Research in Social Stratification and Mobility*, 35, 13–18.

Soria-Espin, J. and Medina, O. (2025). Intergenerational mobility in Spain: geographic analysis and causal neighbourhood effects. World Inequality Lab, working paper nº2025/12.

Torche, F., 2014. Intergenerational mobility and inequality: the Latin American case. *Annual Review of Sociology*, 40(1), 619–642.

Torche, F. and Ribeiro, C.C., 2010. Pathways of change in social mobility: industrialization, education and growing fluidity in Brazil. *Research in Social Stratification and Mobility*, 28(3), 291–307.

Torres-López, P.J., Monroy-Gómez-Franco, L.A. and Vélez-Grajales, R., 2025. Survey-to-survey imputation when external covariates matter: estimating inequality of opportunity in Mexico. *Documento de Trabajo CEEY*, No. 01/2025. Centro de Estudios Espinosa Yglesias (CEEY).

van de Gaer, D., 1993. *Equality of Opportunity and Investment in Human Capital.* Ph.D. Thesis, Katholieke Universiteit Leuven.

Van der Weide, R., Lakner, C., Mahler, D.G., Narayan, A. and Gupta, R., 2024. Intergenerational mobility around the world: a new database. *Journal of Development Economics*, 166, 103167.

Table Appendix

Table A1: MLD and Variance of logs measures of inequality and ex-ante IOp

Wave	Total MLD	Absolute IOp MLD (Forest)	Relative IOp MLD (Forest)
Argentina (2014)	0.2759	0.0499	18.09
Bolivia (2008)	0.5048	0.1442	28.57
Brazil (2014)	0.4266	0.1626	38.12
Chile (2006)	0.4725	0.1150	24.34
Chile (2011)	0.4641	0.1155	24.89
Chile (2013)	0.4520	0.0963	21.31
Chile (2015)	0.4577	0.0954	20.84
Colombia (2010)	0.5471	0.1063	19.43
Ecuador (2006)	0.5413	0.1375	25.40
Ecuador (2014)	0.3768	0.0822	21.82
Guatemala (2000)	0.5651	0.1521	26.92
Guatemala (2006)	0.5768	0.1829	31.71
Guatemala (2011)	0.5187	0.1362	26.26
Mexico (2017)	0.4954	0.1419	28.64
Panama (2003)	0.6359	0.1317	20.71
Peru (2001)	0.4724	0.1441	30.50
Peru (2006)	0.4229	0.1533	36.25
Peru (2007)	0.4301	0.1582	36.78
Peru (2008)	0.3777	0.1320	34.95
Peru (2009)	0.3739	0.1254	33.54
Peru (2010)	0.3587	0.1132	31.56
Peru (2011)	0.3528	0.1025	29.05
Peru (2012)	0.3417	0.1021	29.88
Peru (2013)	0.3351	0.1042	31.10
Peru (2014)	0.3182	0.1065	33.47
Peru (2015)	0.3270	0.1130	34.56

Source: Data from ENES, EH, PNAD, CASEN, ECV, ENCOVI, EMOVI, ENV, ENAHO. See more details in Table 4.

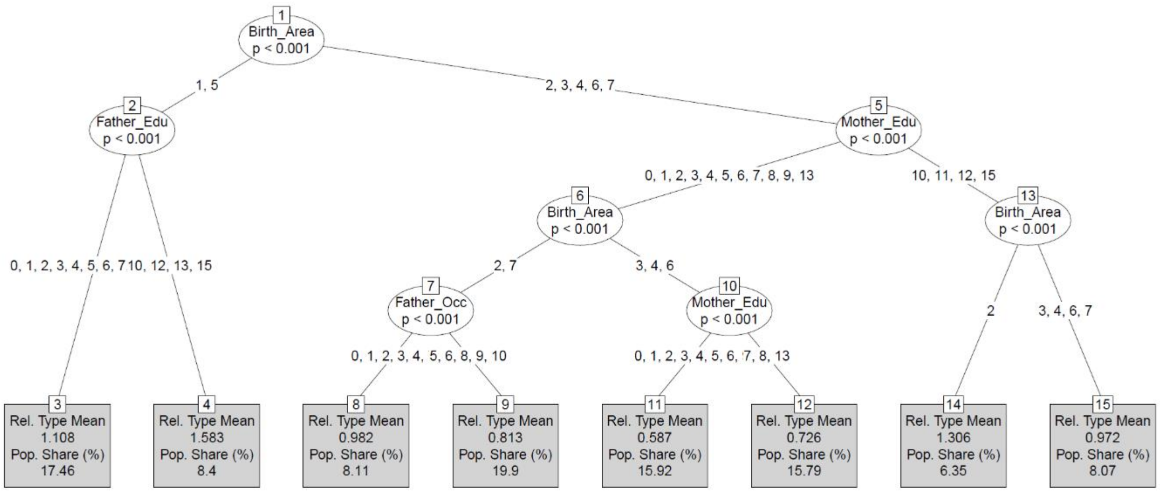
Table A2: A comparison of inherited inequality estimates for ten Latin American countries

Country	Year (IOp)	IOp (income)	IOp (edu)	IOp (edu) PISA	IGE^2	IGE (TSTSLS)	IGE Published article	Edu mobility (ρ^2)
Argentina	2014	0.460	0.289	0.238	0.243	YES	NO	0.257
Bolivia	2008	0.588	0.516	.	0.229	YES	NO	0.298
Brazil	2014	0.657	0.406	0.211	0.229	NO	YES	0.257
Chile	2015	0.503	0.409	0.286	0.069	NO	YES	0.319
Colombia	2010	0.480	0.344	0.309	0.545	YES	NO	0.281
Ecuador	2014	0.503	0.406	.	0.053	NO	NO	0.311
Guatemala	2011	0.554	0.395	.	0.872	YES	NO	0.327
Mexico	2017	0.559	.	0.200	0.590	YES	NO	0.273
Panama	2003	0.543	0.427	.	0.113	NO	YES	0.305
Peru	2015	0.633	0.412	0.299	0.489	YES	NO	0.299

*Notes: All education mobility estimates are squared correlation coefficients of years of schooling produced by using the data from Neidhöfer et al. (2018). The corresponding regression coefficient estimates using the same samples are in Column 2 in Table 1. IGE estimates are the squares of our preferred estimates of income IGM, from column 1 in Table 2. * denotes IGE estimates reported in a published peer-reviewed journal article; other IGE estimates come from working papers and ongoing research. Dark green denotes OLS estimates of IGE, whereas light green denotes TSTSLS estimates of IGE. IOp.Education are Random Forest estimates from column 2 in Table 8. IOp.Income are Random Forest estimates from Column 1 in Table 8. IOp.PISA are estimates from Column 3 in Table 8 and were originally produced by Brunori et al. (2025).*

Figure Appendix

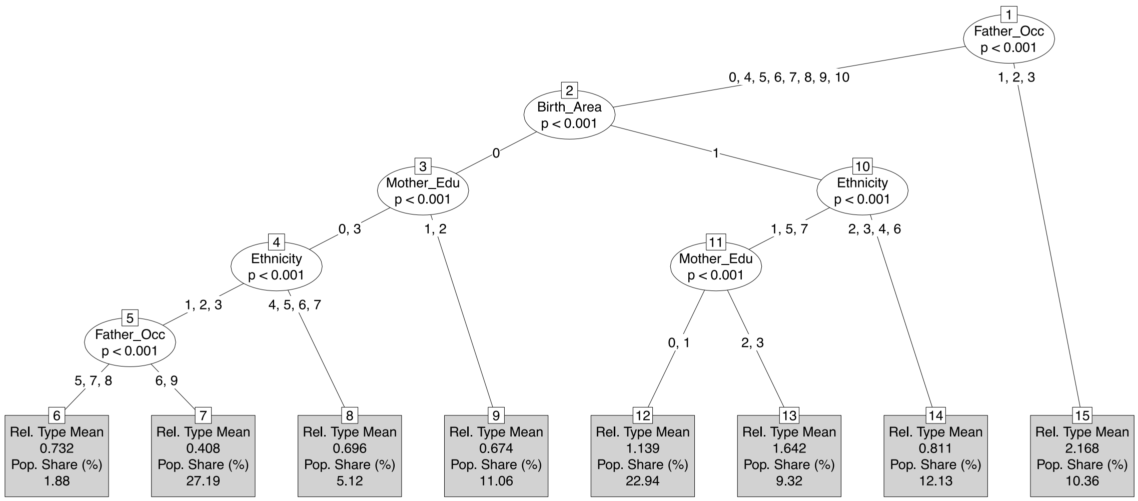
FA1: Ex-ante Root Tree in Argentina (2014)



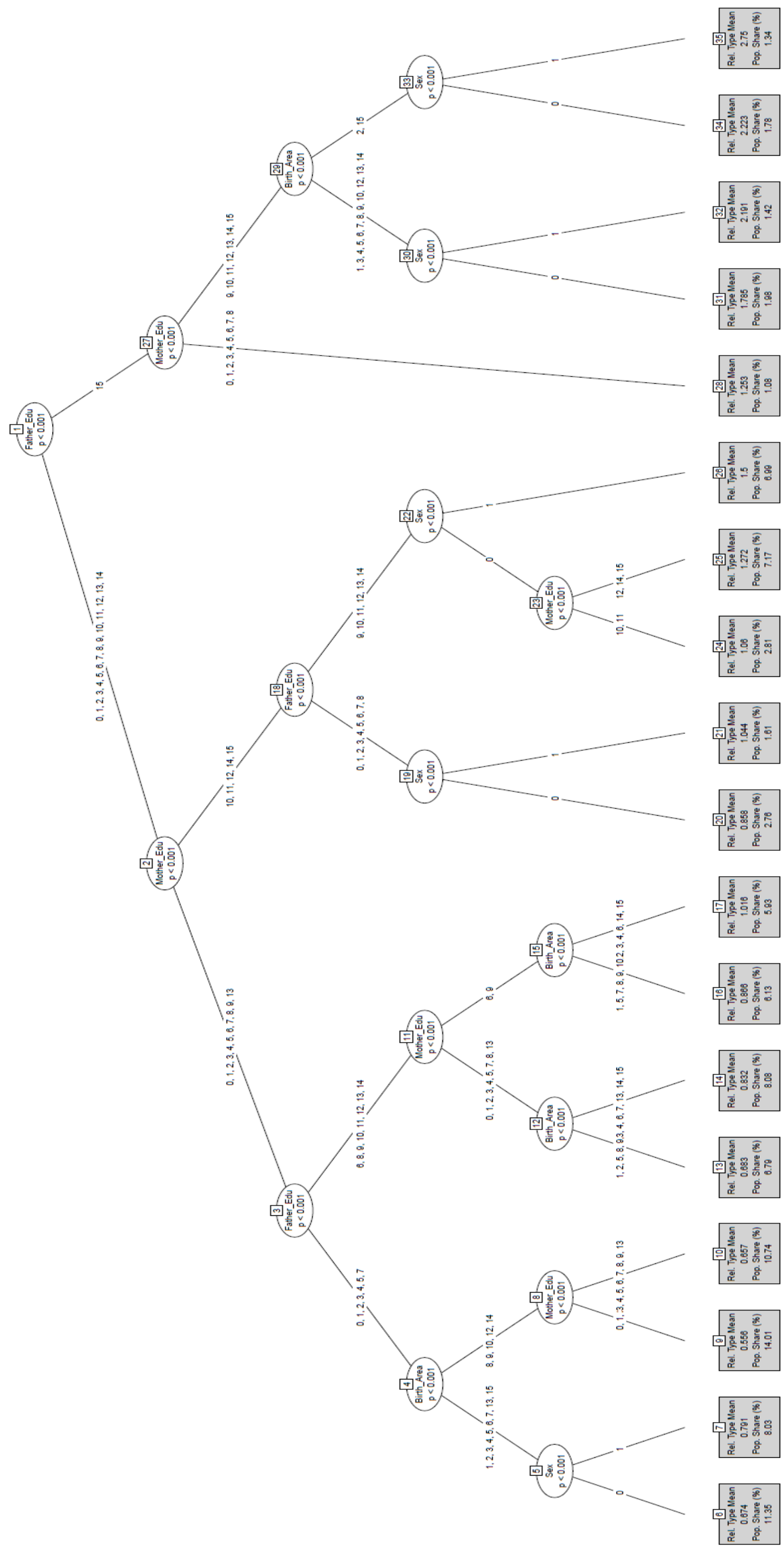
Source: ENES (2014). Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$).

FA2: Ex-ante Root Tree in Bolivia (2008)

Source: EH, 2008. Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$).

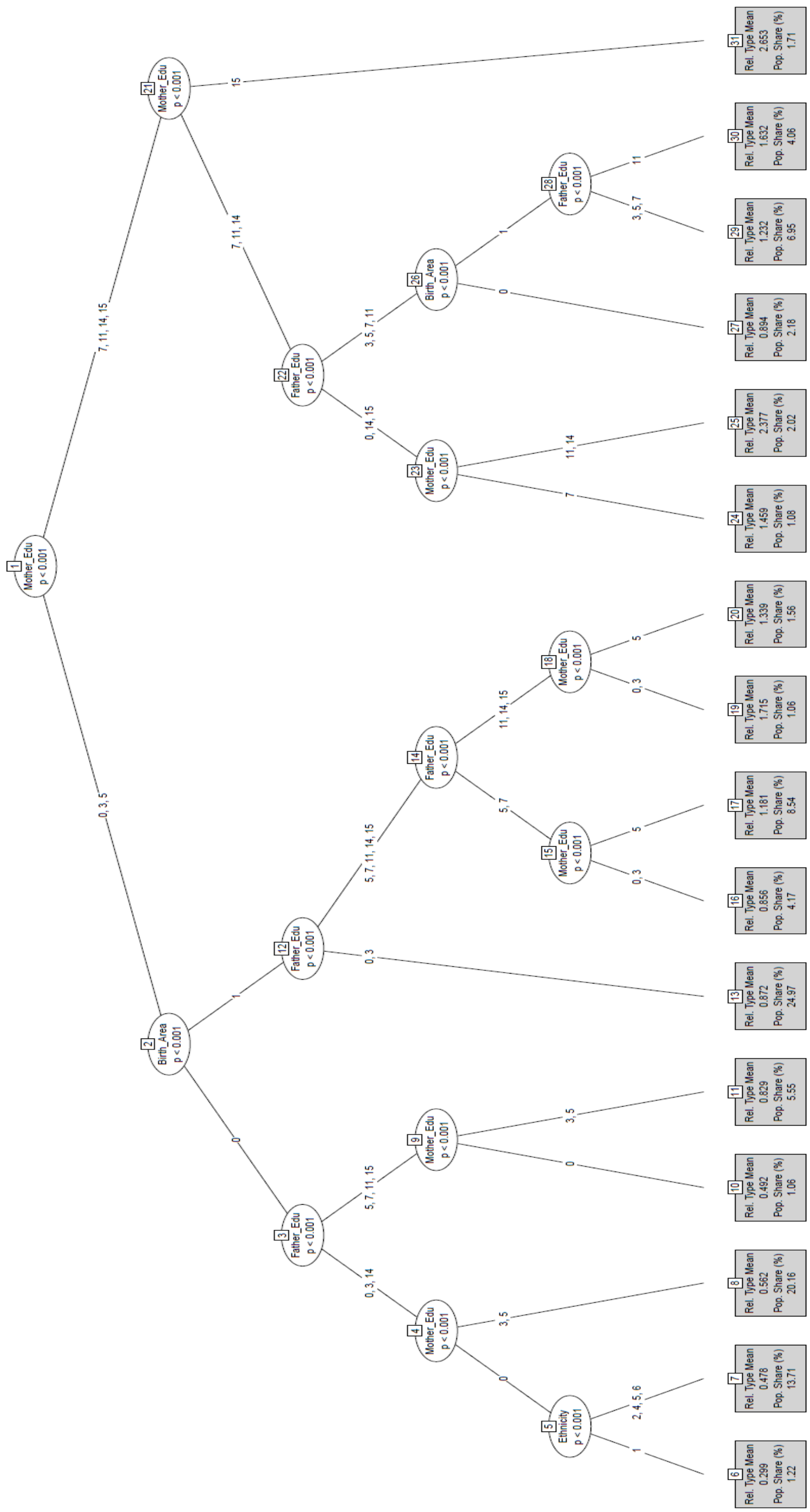


FA3: Ex-ante Root Tree in Chile (2015)



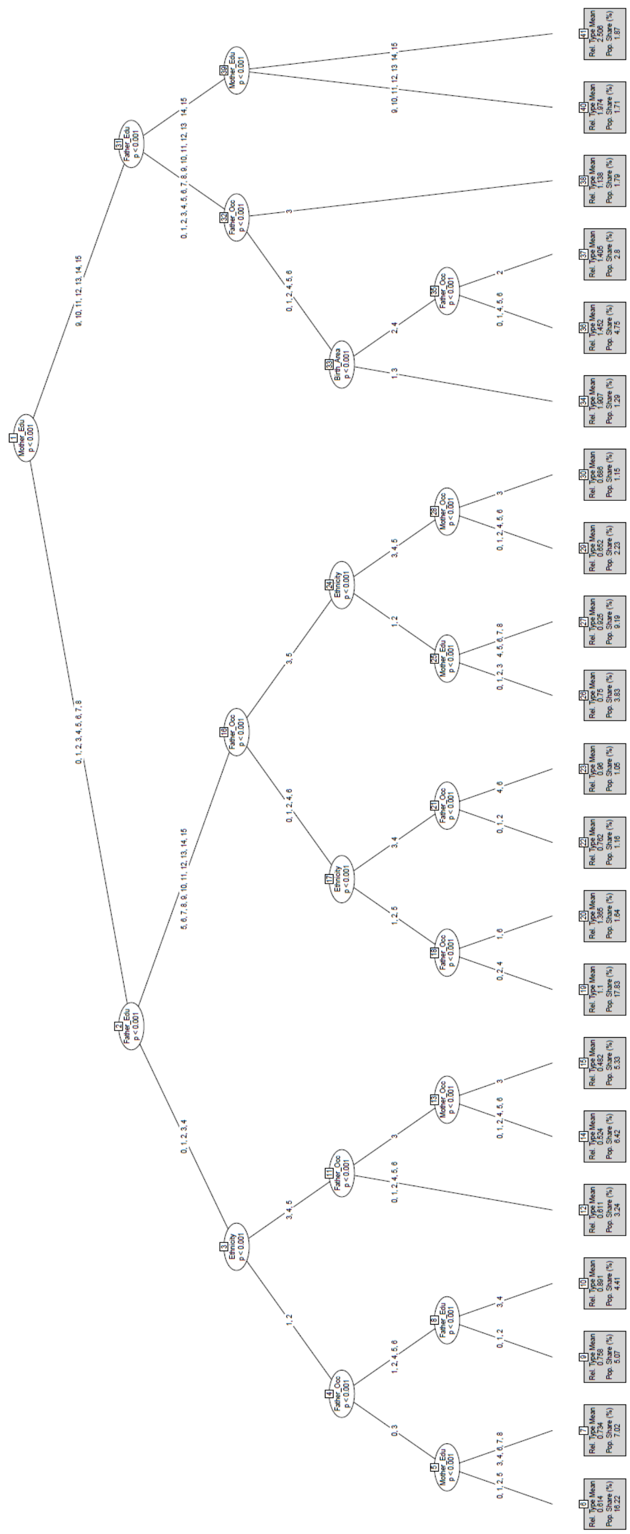
Source: CASEN (2015). Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$).

FA4: Ex-ante Root Tree in Colombia (2010)



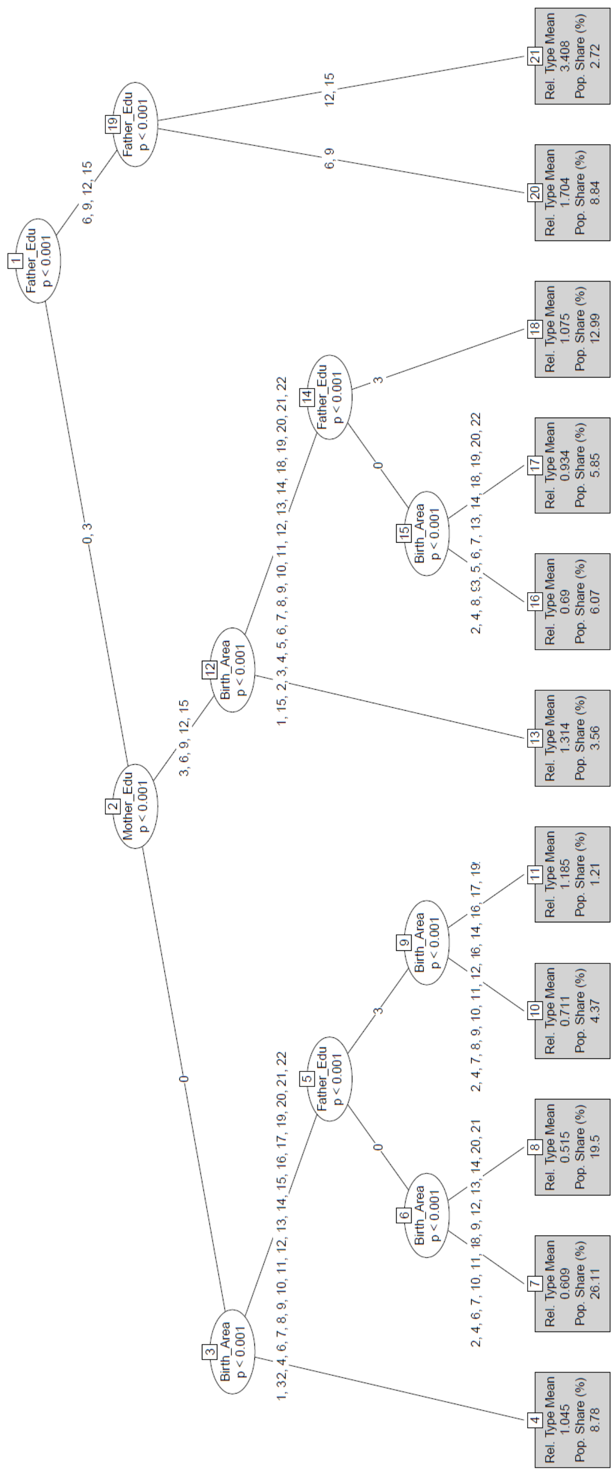
Source: ECV (2010). Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$).

FA5: Ex-ante Root Tree in Ecuador (2014)



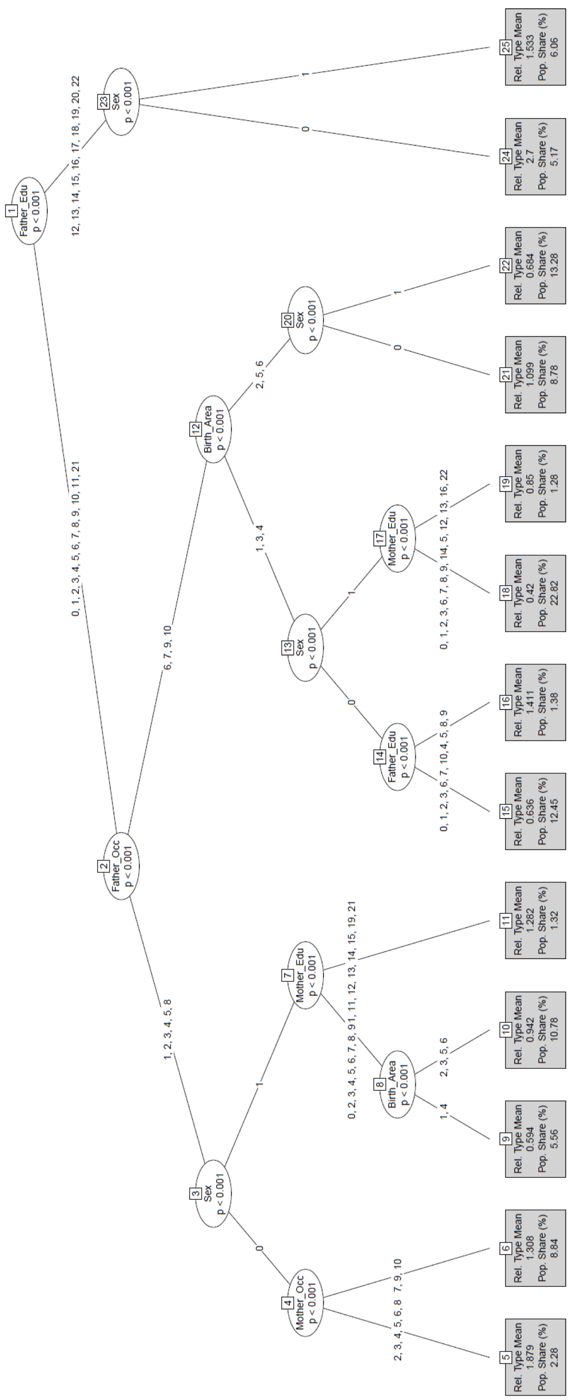
Source: ECV (2014). Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$).

FA6: Ex-ante Root Tree in Guatemala (2011)



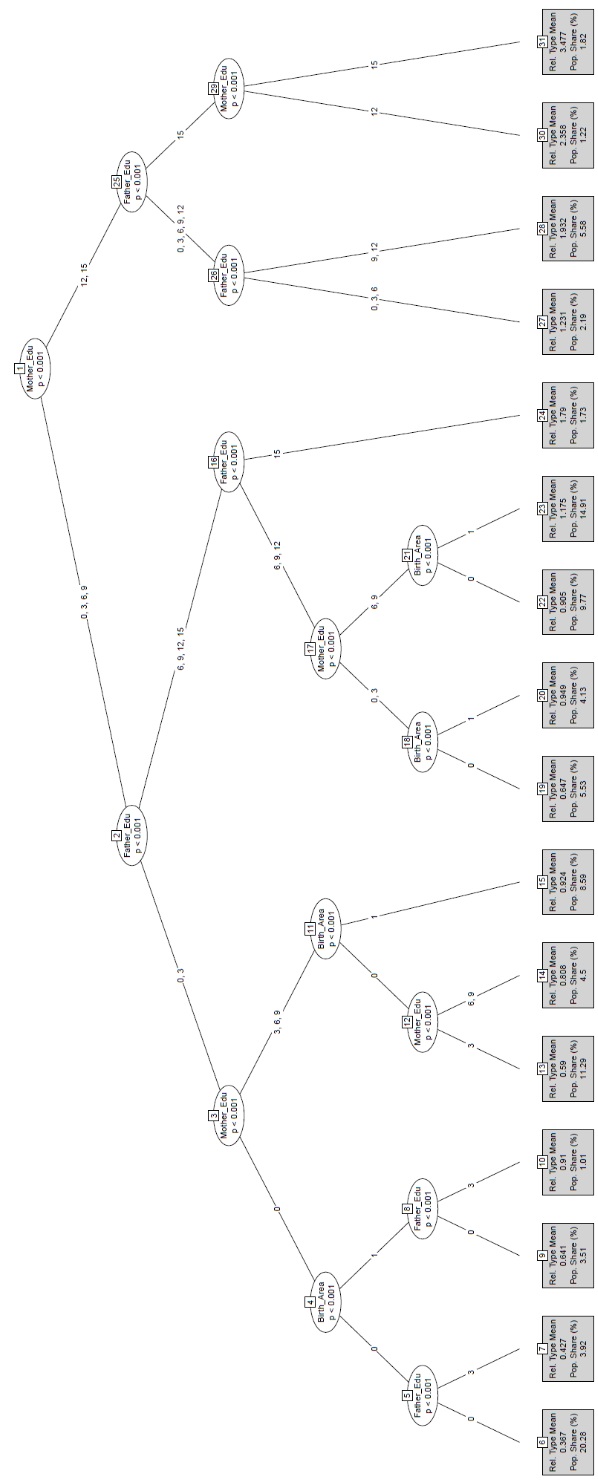
Source: ENCOVI (2011). Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$*).*

FA7: Ex-ante Root Tree in Mexico (2017)



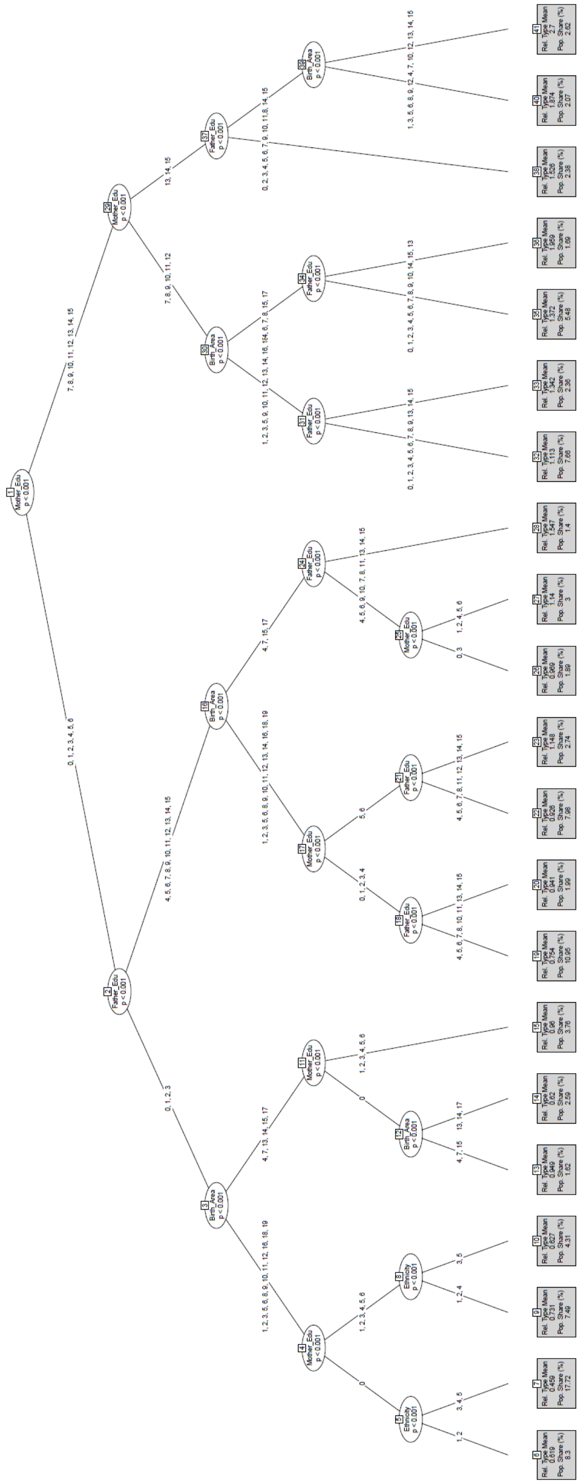
Source: EMOVI (2017). Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$*).*

FA8: Ex-ante Root Tree in Panama (2003)



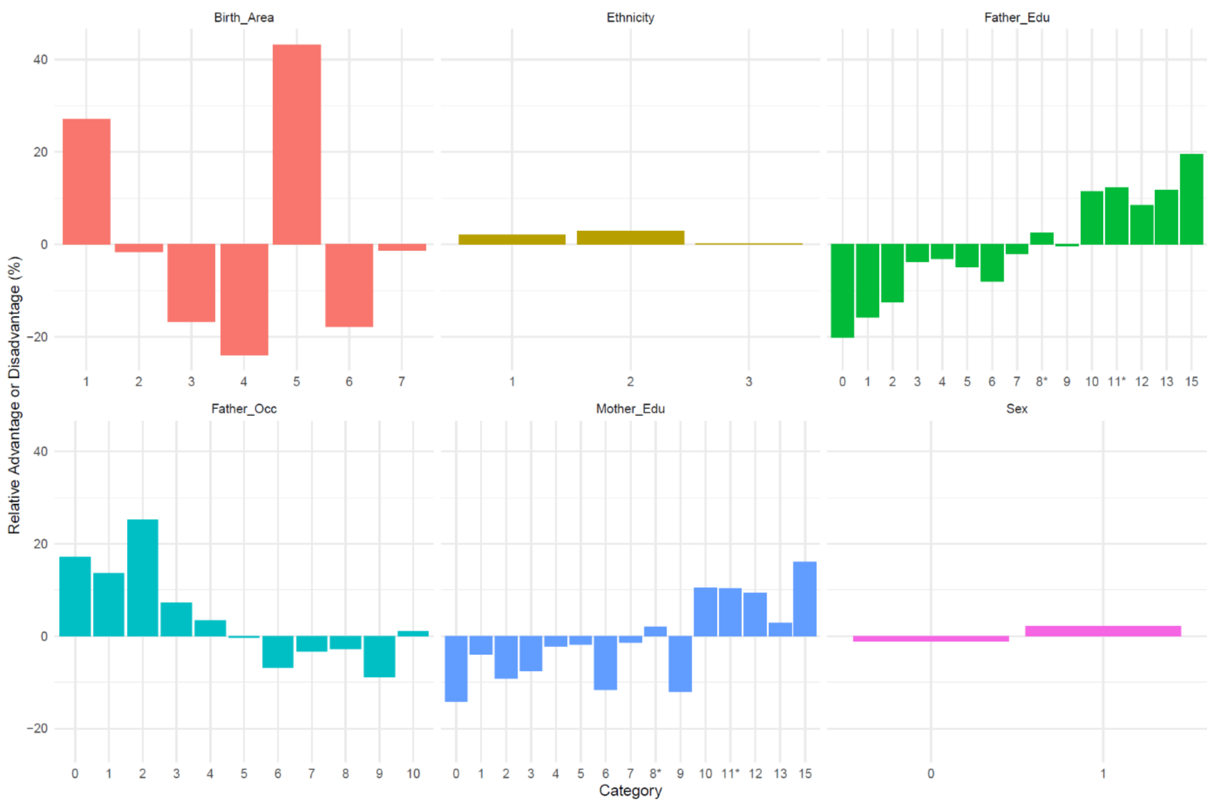
Source: ENV (2003). Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$*).*

FA9: Ex-ante Root Tree Peru (2015)



Source: ENAHO (2015). Tree obtained setting statistical significance at 99.99% ($\alpha = 0.0001$*).*

FA10: Partial Dependence Plot in Argentina (2014)

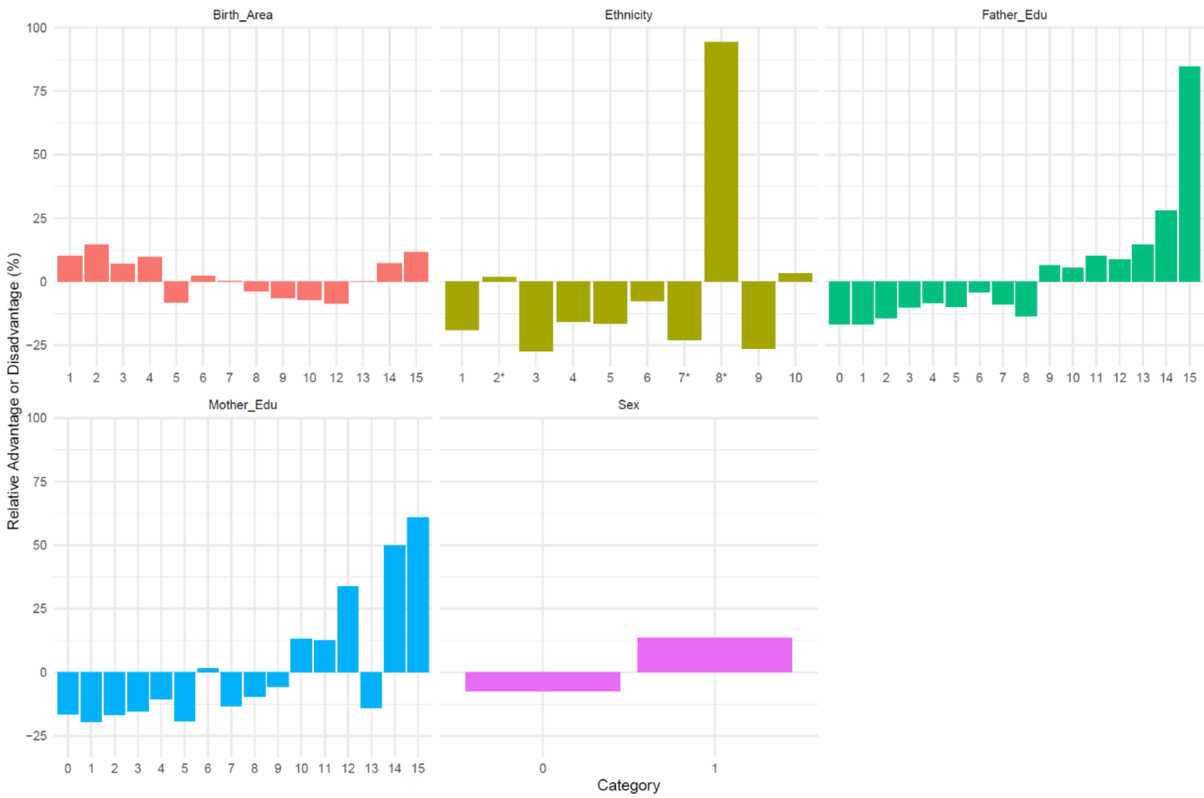


Source: ENES (2014).

FA11: Partial Dependence Plot in Bolivia (2008)

Source: EH (2008).

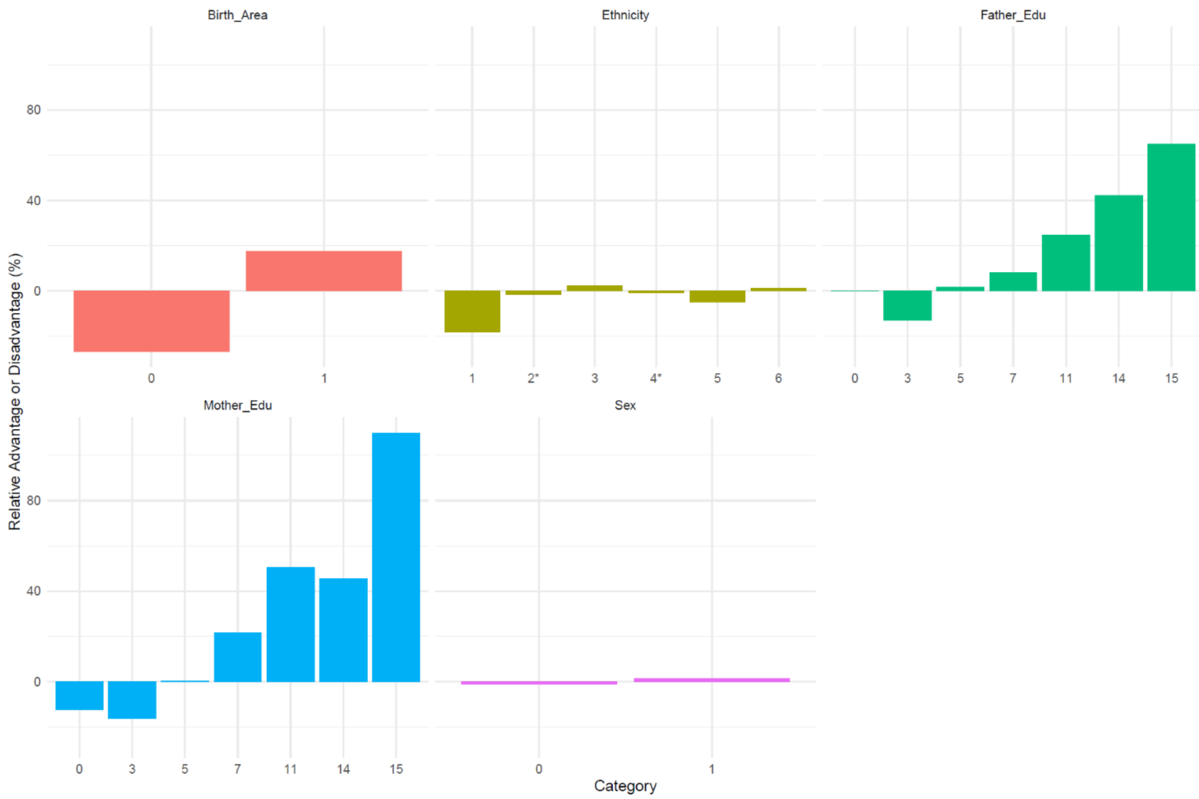
FA12: Partial Dependence Plot in Chile (2015)



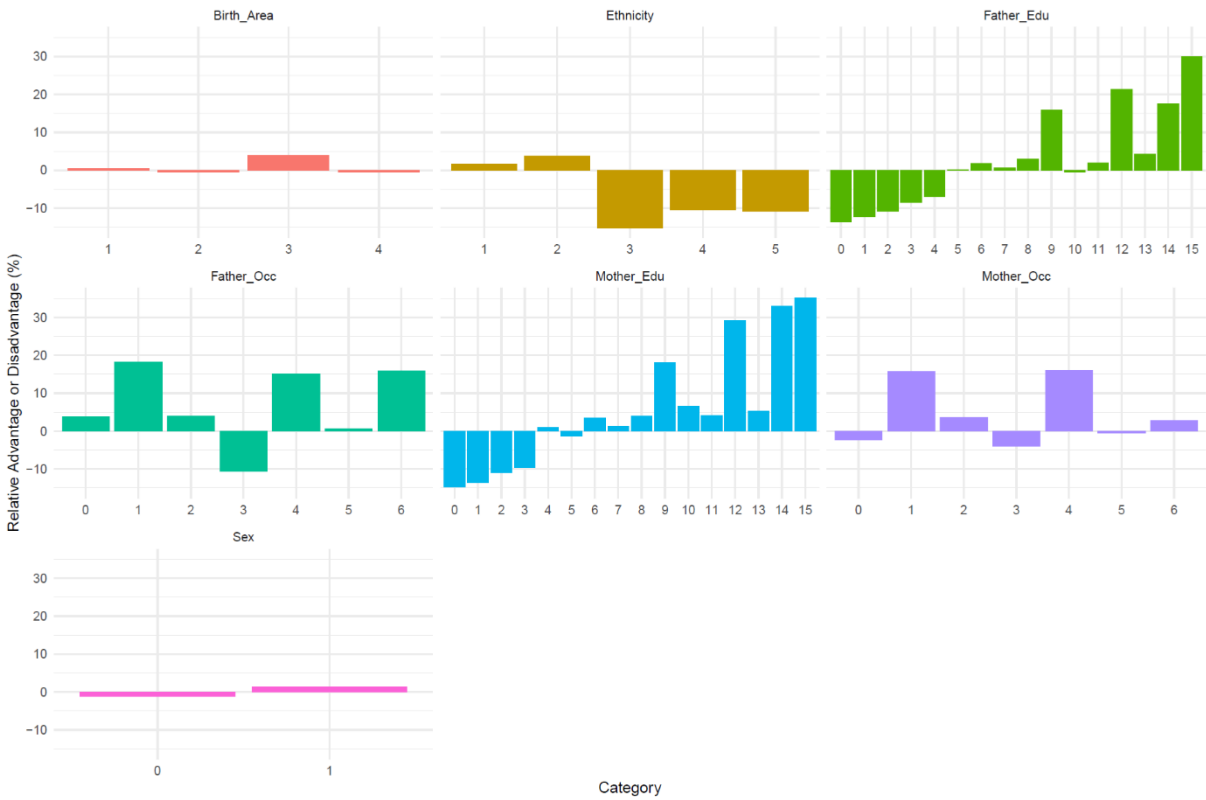
Source: CASEN (2015).

FA13: Partial Dependence Plot in Colombia (2010)

Source: ECV (2010).



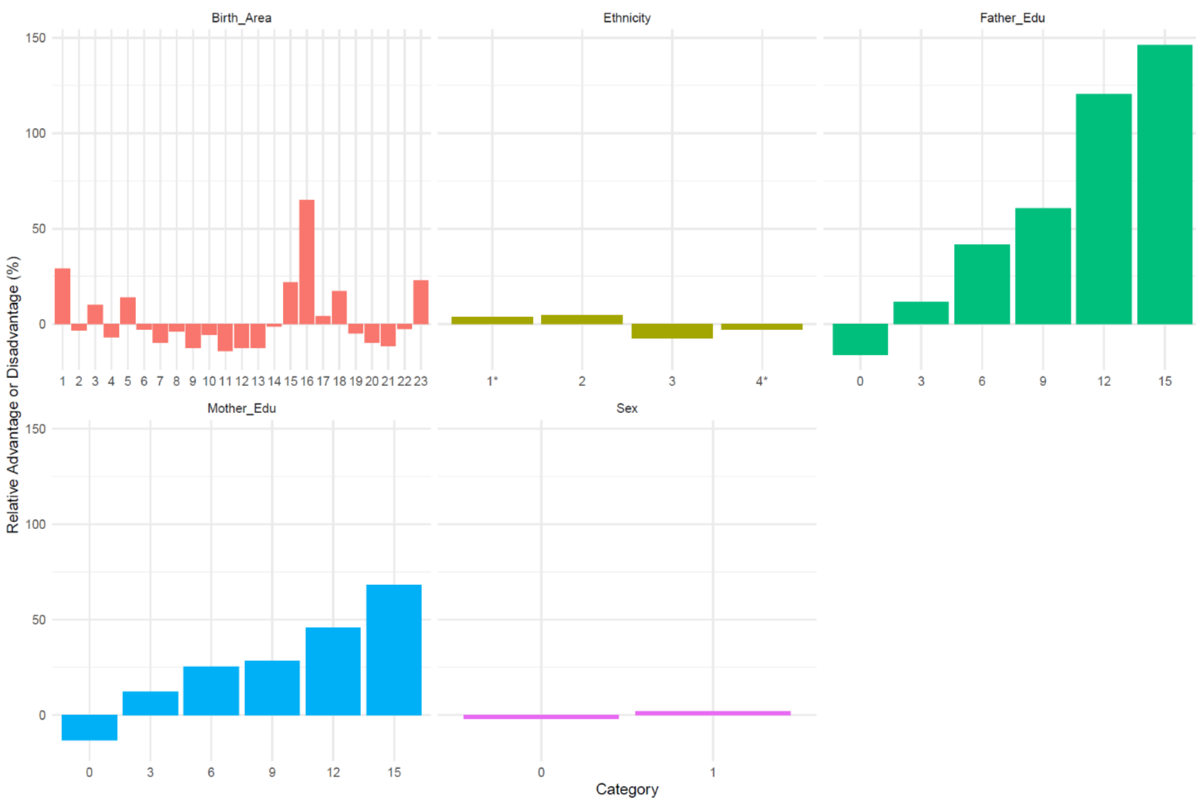
FA14: Partial Dependence Plot in Ecuador (2014)



Source: ECV (2014).

FA15: Partial Dependence Plot in Guatemala (2011)

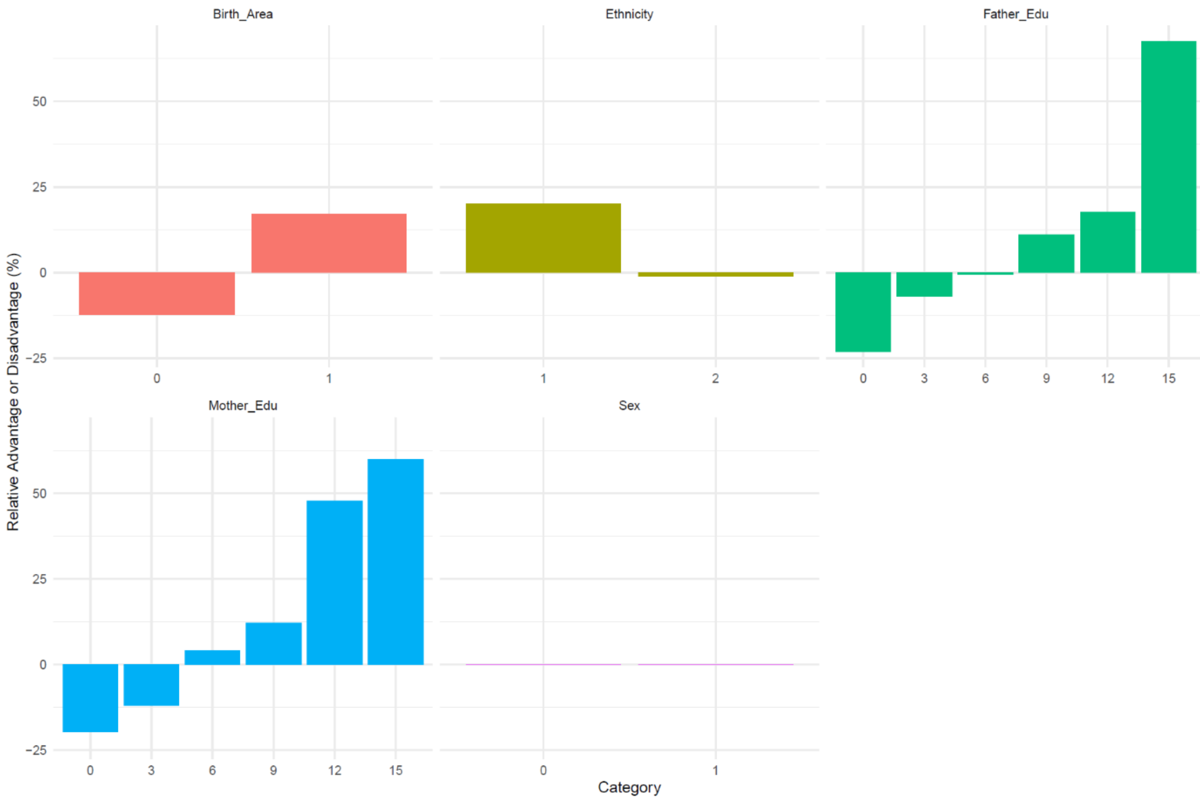
Source: ENCOVI (2011).



FA16: Partial Dependence Plot in Mexico (2017)

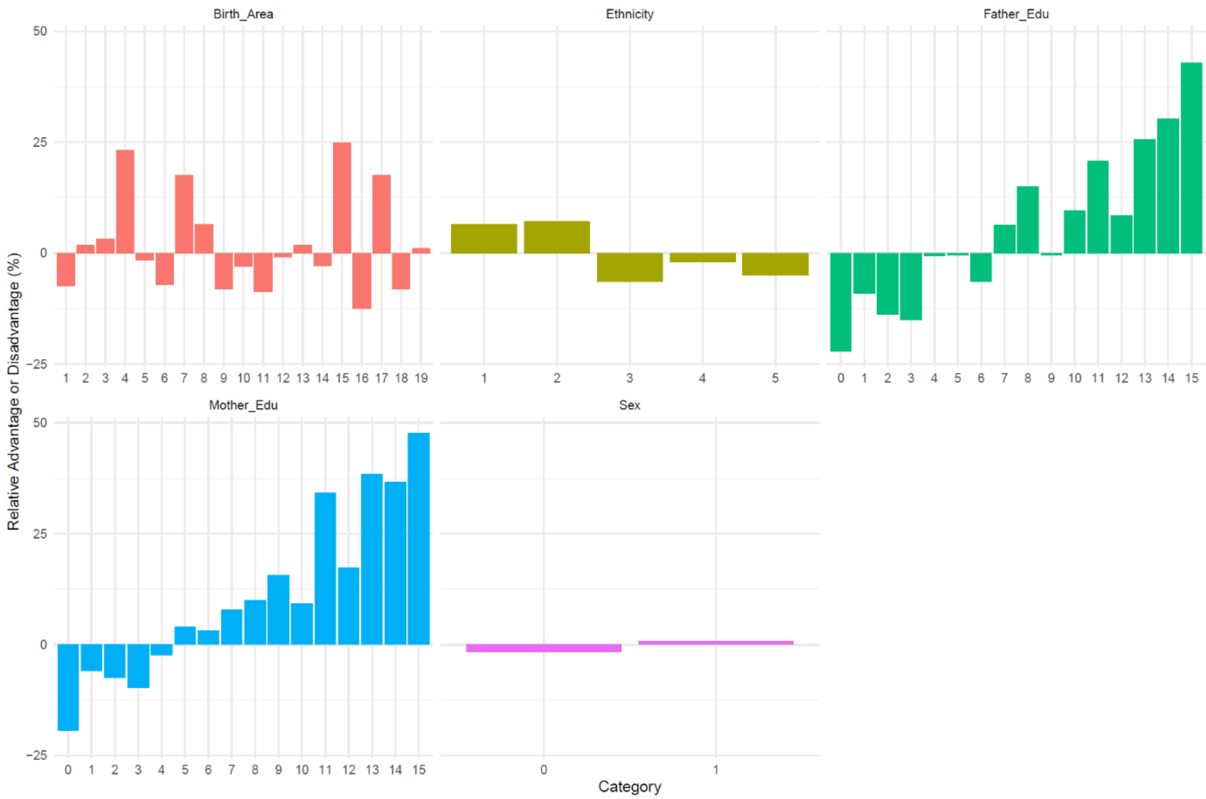
Source: EMOVI (2017).

FA17: Partial Dependence Plot in Panama (2003)



Source: ENV (2003).

FA18: Partial Dependence Plot in Peru (2015)



Source: ENAHO (2015).