Brunori, Paolo; Ferreira, Francisco H. G.; Neidhöfer, Guido

Working Paper

Inequality of Opportunity and Intergenerational Persistence in Latin America

IZA Discussion Papers, No. 17202

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Brunori, Paolo; Ferreira, Francisco H. G.; Neidhöfer, Guido (2024) : Inequality of Opportunity and Intergenerational Persistence in Latin America, IZA Discussion Papers, No. 17202, Institute of Labor Economics (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/305644

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

DISCUSSION PAPER SERIES

IZA DP No. 17203

Inequality of Opportunity and Intergenerational Persistence in Latin America

Paolo Brunori
Francisco Ferreira
Guido Neidhöfer

AUGUST 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17203

Inequality of Opportunity and Intergenerational Persistence in Latin America

Paolo Brunori
University of Florence and London School of Economics

Francisco Ferreira
London School of Economics and IZA

Guido Neidhöfer
Turkish-German University and ZEW

AUGUST 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Inequality of Opportunity and Intergenerational Persistence in Latin America*

How strong is the transmission of socio-economic status across generations in Latin America? To answer this question, we first review the empirical literature on intergenerational mobility and inequality of opportunity for the region, summarizing results for both income and educational outcomes. We find that, whereas the income mobility literature is hampered by a paucity of representative datasets containing linked information on parents and children, the inequality of opportunity approach – which relies on other inherited and pre-determined circumstance variables – has suffered from arbitrariness in model selection. Two new data-driven approaches – one aligned with the ex-ante and the other with the ex-post conception of inequality of opportunity – are introduced to address this shortcoming. They yield a set of new inequality of opportunity estimates for twenty-seven surveys covering nine Latin American countries over various years between 2000 and 2015. In most cases, more than half of the current generation's inequality is inherited from the past – with a range between 44% and 63%. We argue that on balance, given the parsimony of the population partitions, these are still likely to be underestimates.

JEL Classification: D31, I39, J62, O15

Keywords: inequality of opportunity, intergenerational mobility, 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 Nancy Daza Báez, Matias Ciaschi, and Luiz Eduardo Barajas Prieto for excellent research assistance. We are also very grateful to Leonardo Gasparini for sharing the data for this project with us; and to François Bourguignon and Pedro Salas-Rojo for invaluable advice. Finally, we thank Sergio Firpo, Forhad Shilpi, Florencia Torche, two anonymous referees, and participants at the UNU-WIDER and LACEA Conferences in 2022, and at the LACIR-Cartagena workshop in 2023, for comments on earlier versions. An earlier version of this paper was circulated as a UNU-WIDER working paper (Brunori et al. 2023), and the present version supersedes it. All remaining errors are ours alone.

1. Introduction

The *nature* of inequality matters as much as, if not more than, its *amount*. If, as Friedman and Friedman (1962) hypothesized, high cross-sectional income inequality at a point in time was accompanied by considerable mobility – over time or across generations – perhaps it would not be of great concern. If, on the other hand, economic advantage is persistent across generations, so that the same people or lineages always enjoy wealth and privilege, while others are systematically excluded from them, then we may be considerably more inequality averse. Others have argued that when inequality reflects differences in personal effort and responsibility, it is less objectionable than inequality due to inherited circumstances that people cannot control, such as race, sex, or family background (e.g., Roemer, 1998). According to this view, income inequality is more of a problem in a society with greater inequality of opportunity, driven by pre-determined circumstances, than in one where people face a level playing field and outcome differences reflect only differences in effort.[2]

Empirically, it turns out that these hypothetical examples of "unproblematic" inequality of outcomes seem to be very rare, at best. Countries with greater income inequality also tend to display less intergenerational income mobility (Corak, 2013) and more inequality of opportunity (Brunori et al., 2013). These positive correlations between income inequality on the one hand, and intergenerational persistence (the opposite of mobility) or inequality of opportunity on the other, are fairly robust findings (DiPrete, 2020; Durlauf et al., 2022). But they are certainly not deterministic: there is variation around the regression lines and, furthermore, these indices have not been computed over long-enough periods for a sufficient number of countries for us to know how stable the associations are. This is particularly true for developing countries, where the data constellations are more challenging.

In this paper we investigate the extent and nature of inequality of opportunity and intergenerational persistence in Latin America, one of the world's most unequal regions in income terms. The next section reviews the empirical literature for Latin America and is organized into two subsections: (i) intergenerational mobility / persistence of income and education; and (ii) inequality of opportunity (IOp), also for income and education.

Although we have learned a fair amount about *educational* mobility across generations in Latin America, studies of intergenerational *income* mobility in the region have been hampered by severe data shortcomings, primarily due to the absence of data on parental incomes that can be linked to the incomes of their adult children in an unbiased way. In that context, using alternative family background variables that are more widely available such as parental education and occupation, as the IOp studies do, can be a valuable addition. However, these latter studies have also suffered from their own shortcomings, including the use of *ad hoc* selections of circumstance variables and categories with which to partition the

[2] There are at least two different justifications for this view. The first is ethical: one may believe that individuals are responsible for the effort they exert and therefore deserve to keep the return to their effort. The second is agnostic about what people deserve but acknowledges that rewarding effort may allow societies to generate more output, making it easier to achieve any desired welfare allocation.

population into "types".[3] IOp measures are sensitive to the type partition and the choice of that partition trades off two opposing biases: a downward omitted variable bias and an upward overfitting bias (Brunori et al., 2019).

Section 3 therefore adopts a new approach to the measurement of inequality of opportunity, in both its ex-ante and ex-post varieties. The key characteristic of this approach is that it lets the data determine the optimal partition of the sample, in a well-defined statistical sense. Both the partitioning algorithm and the computation of the summary IOp index differ between the ex-ante and the ex-post cases: the ex-ante indices are computed using conditional inference trees or forests, which rely exclusively on information about subgroup (or "type") means. This is in keeping with the ex-ante approach of measuring inequality of opportunity as inequality between the expected values of the opportunity sets of different types. The ex-post indices are computed using transformation trees, which use information on the entire quantile function of each type, in keeping with the ex-post view of inequality of opportunity as an aggregation of inequality across conditional quantiles. See below for details.

Section 4 describes the data used for the estimation, which comes from 27 household surveys covering nine Latin American countries. Section 5 presents results for the ex-ante measures, including the summary indices, the tree structure, and a Shapley decomposition of the relative importance of individual circumstance variables. Section 6 does the same for the ex-post measures. In both cases, we use the recursive partitioning of the sample (the 'trees') not only as a means to obtain the optimal partition of the population into final nodes – the types – and the summary measure of inequality among them, but also as informative of the structure of opportunity in these societies. Section 7 briefly compares the ex-ante and ex-post results to one another, but also to previous mobility estimates from the literature reviewed in Section 2. Section 8 concludes.

2. A review of the literature

Before reviewing the literature on intergenerational mobility (or persistence) and on inequality of opportunity in Latin America, it is useful to briefly reflect on the relationship between the two concepts. Mobility and equality of opportunity are closely related, both theoretically and empirically. Although mobility can mean different things and be measured in different ways, the kind of mobility we associate with "origin independence" (Fields, 2000) is typically measured by (the complement of) some indicator of the association between outcomes – e.g., incomes or education levels – across generations. One of the simplest such indicators is the Pearson correlation coefficient between, say, the income of a parent and the income of their child.

Inequality of opportunity can also be defined and measured in different ways (see Ferreira and Peragine, 2016, for a survey) but one common approach is to measure it as the share of inequality in an outcome variable that can be accounted for by all pre-determined circumstances over which individuals have no

[3] A type is a subgroup of the population that is homogeneous in terms of all circumstance variables used in the partition (Roemer, 1998).

control. One simple measure might be, say, the R-squared of a regression of the (adult) child's income on those circumstances. Of course, if the only circumstance were parental income, then that measure would be a monotonic transformation (the square) of the Pearson correlation coefficient, our earlier measure of mobility.

Indeed many, if not most, commonly used measures of intergenerational mobility and of inequality of opportunity share a similar structure: they are ratios (or functions of ratios) of inequality in predicted incomes to inequality in observed incomes – $I(\hat{y})/I(y)$ – where "predicted" means predicted by inherited circumstances.[4] This common structure relies on estimates of how well inherited variables – parental income, education, occupation, and so on – *predict* current incomes and makes the two concepts isomorphic. See Brunori, Ferreira, and Salas-Rojo (2023) for a discussion.

That said, intergenerational mobility and inequality of opportunity are not *precisely* the same thing. Unless parental income is a sufficient statistic for all pre-determined circumstances, they will differ if we consider other circumstances. And different concepts of mobility – particularly absolute concepts, such as the proportion of people doing better than their parents – are much less closely related to inequality of opportunity. But relative measures of intergenerational persistence and inequality of opportunity *are* closely related conceptually and turn out to be strongly correlated in practice (Brunori et al, 2013).

Nonetheless, in Latin America as elsewhere, most studies of intergenerational persistence have focused on either one concept or the other. This section is therefore organized in two parts: (i) a review of the literature on intergenerational mobility in the region, both for education and income; and (ii) a review of the literature on inequality of opportunity for the same two variables.

2.1 The literature on intergenerational mobility in Latin America

Some excellent reviews of this topic are already available. Torche (2014), for example, provides a comprehensive review of the early literature on intergenerational mobility in Latin America and mainly subdivides it into a first generation of social mobility research in the 1960s and 1970s, and a second generation starting from the 1990s. While the first generation of mobility studies was heavily dominated by sociological research and focused on occupational mobility, economists started to study the subject more extensively in the second generation. Another difference between the two periods is that, while in the first generation the topic was mostly studied with ad-hoc surveys, in specific (urban) areas, or with rather limited samples, the use of representative household surveys became much more common in the second generation. To minimize duplication with Torche (2014), we focus on contributions estimating intergenerational mobility of education and income and belonging to the second generation, as well as

[4] So, for example, the Pearson correlation coefficient between parental and child income can be written as $\hat{\rho} = \sqrt{\frac{I(\hat{y}_M)}{I(y)}}$, where $I(x) = Var \log x$ and $\hat{y}_M = e^{\hat{\alpha}+\hat{\beta}\log y_p+\sigma^2/2}$, whereas an ex-ante parametric measure of relative inequality of opportunity is given by $IOR_{EA} = \frac{I(\hat{y}_{EA})}{I(y)}$, where $\hat{y}_{EA} = e^{\hat{\alpha}+C\hat{\gamma}}$, and $I(x)$ can be any meaningful inequality measure.

on more recent contributions using more extensive samples, new data sources and different dimensions of mobility, which might indeed define a third generation of intergenerational mobility research in Latin America.[5]

Owing to the nature of the household survey data available at the time, early second-generation studies were usually cross-sectional and followed different methodological approaches. Behrman, Gaviria and Székely (2001) study intergenerational mobility of schooling and occupational status for Brazil, Colombia, Mexico and Peru using regression analysis on household survey samples that contained retrospective questions on parents' education and occupation. Behrman, Birdsall and Székely (1999) instead analyse intergenerational mobility in 16 Latin American countries on a sample of children co-residing with their parents. Their measure of mobility is the degree of association between family background and the schooling gap, i.e. the number of years of schooling that an individual would have if he or she entered school at age six and advanced one grade every year, minus the number of years of school that he or she actually has. Dahan and Gaviria (2001) work with a similar sample of co-residents but propose a different methodology based on sibling correlations. For 16 Latin American countries, they first compute an indicator of socio-economic failure for children, which is similar to the schooling gap and defined with respect to the median years of schooling of the cohort. Then, the sibling correlation is based on the proportion of the variance in that indicator that can be explained by differences between families as opposed to differences within families.

There were also a number of studies for single countries, addressing intergenerational mobility either directly or in a broader sense, namely as the association between parental socio-economic status and their children's education or labour market outcomes. Examples include Behrman and Wolfe (1987) for Nicaragua; Binder and Woodruff (2002) for Mexico; Lam and Schoeni (1993) for Brazil; and Heckman and Hotz (1986) for Panama. The general conclusion of this early literature, which focused primarily on educational outcomes, is that in Latin America family background was a strong predictor of individual educational success, and intergenerational mobility was low when compared, for instance, to the US. For instance, Behrman, Gaviria and Székely (2001) obtain intergenerational regression coefficients for years of schooling of around 0.7 for Brazil and Colombia, 0.5 for Mexico and Peru, and 0.35 for the United States. These are measures of persistence – the opposite of mobility – so that higher numbers characterize less mobile societies.

More recent contributions spanning multiple countries—and also mostly using nationally representative household surveys that include retrospective questions on parental education to avoid co-residency bias—highlight that the degree of intergenerational mobility differs significantly across countries (Daude and Robano, 2015; Ferreira et al., 2013; Neidhöfer, Serrano and Gasparini, 2018). For the 1964-1967 cohort, for example, Neidhöfer, Serrano and Gasparini (2018) obtain an average regression coefficient of

[5] Space constraints preclude us from also reviewing the literature on occupational mobility. For a comprehensive review of second-generation studies measuring occupational mobility in Latin America, see Torche (2014). Updated estimates of occupational mobility in Brazil and Mexico, along with a comparison to the US, can be found in Torche (2021).

parents' schooling for 18 Latin American countries around 0.5, ranging from 0.34 in Venezuela and 0.37 in Costa Rica, to estimates around 0.6 or even higher in Bolivia, Brazil, El Salvador and Guatemala.[6] Additional findings include that intergenerational mobility in Latin America is negatively associated with income inequality and economic crises, and positively associated 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; Torche, 2010).

Comparative studies of intergenerational educational mobility worldwide mainly confirm the patterns highlighted by these contributions: They classify Latin America as one of the regions with the lowest average levels of intergenerational mobility (Ahsan et al., 2023; Hertz et al., 2008; Narayan et al., 2018; Van der Weide et al., 2024). However, intergenerational mobility *trends* – as opposed to levels – paint a somewhat more encouraging picture. Neidhöfer et al. (2018) show that the advantage (for children's education) associated with one additional year of parental education shrank from 0.6 years for people born in the 1940s to 0.4 for the 1980s cohorts on average for the region. Also, the likelihood of individuals with low-educated parents completing secondary education was more than twice as high for people born in the 1980s than for those born in the 1940s, reaching levels of more than 50% in many countries. Hence, while the educational mobility of older cohorts is indeed rather low, the mobility of younger cohorts is more similar to that of developed countries. The regression coefficients between 0.33-0.35 estimated for the 1980s cohorts in Argentina, Brazil, Costa Rica and Venezuela are comparable to those obtained for Italy (0.33), Spain (0.31) and the US (0.33) (Narayan et al., 2018).[7]

On the other hand, not all countries in the region show the same pattern. In some countries, such as Guatemala, Honduras and Nicaragua, educational upward mobility remains at very low levels and virtually unchanged over time. In those countries, even in the 1980s cohort, only around one out of ten children with low-educated parents completes secondary education (Neidhöfer et al., 2018). Furthermore, persistence at the top of the educational distribution—measured as the likelihood of individuals whose parents completed secondary education to complete secondary education themselves—is remarkably strong in most countries, between 70% and 80%, and stable over time.

Most research on intergenerational mobility in Latin America focuses on education as proxy for the socio-economic status of parents and children. This is not only meaningful in itself—since education is a very important dimension of current and future well-being, and arguably less correlated with preferences than income or occupation—but also has some practical advantages. Education is less volatile over the life cycle than income or earnings and is completed by individuals relatively early in life (usually between the ages of 18 and 30). Hence, it provides a stable and consistent indicator for the socio-economic status of individuals that can be easily measured in most datasets for parents and children. However, focusing on education alone may also provide only a partial and imperfect picture of economic mobility. As highlighted by Torche (2021), among others, increases in absolute educational mobility may not necessarily lead to a

[6] The website https://mobilitylatam.website/ includes data visualization tools with maps and trends for several measures of educational mobility for 18 Latin American countries.

[7] But see below on why regression coefficients must be interpreted with care in this context.

substantial improvement in equality of opportunity, particularly in a context of broad educational expansions such as those experienced in most Latin American countries over the past decades. This is true even of a comparison of the Galtonian regression coefficient, β, with the correlation coefficient, ρ. Using σ_p to denote the standard deviation of years of schooling in the parents' generation, and σ_c the standard deviation in the children's generation, it is well known that $\beta = \rho \frac{\sigma_c}{\sigma_p}$. Given ceiling effects, educational expansions tend to reduce dispersion in the distribution of years of schooling, i.e. to lower σ_c relative to σ_p. One cannot therefore infer that a lower β necessarily implies a reduction in the margin-independent, pure measure of association, ρ.

Indeed, the rising trend in upward absolute educational mobility in Latin America, which was more pronounced than in most other regions of the world during the latter part of the 20th Century, was not accompanied by an increase in relative mobility. As shown by Neidhöfer et al. (2018), while in Latin America the likelihood of children from low-educated families to complete secondary schooling improved steadily, relative mobility in education, for instance measured by the rank correlation, remained largely stable over the same period. In addition, it is not clear whether the improvement in educational opportunities experienced by the region translated into equality of opportunity in the labour market or for income generation.

Turning to incomes, the study of *income* mobility in Latin America is particularly challenging. Ideally, valid measures of income mobility require longitudinal data with several income spells to avoid bias (see Jäntti and Jenkins, 2015). While some studies dedicated to *intra*generational income mobility in Latin America provide consistent estimates based on one generation—e.g., Fields et al. (2007); Cuesta, Ñopo, and Pizzolitto (2011) using synthetic panels; and more recently Beccaria et al. (2022)—the additional hurdle to access several income spells for both parents and children makes the study of income mobility in several countries as yet almost impossible. Estimates based on directly observed links between parents' and children's lifetime incomes are available for relatively few countries (e.g., Canada, France, Norway, Sweden, the United Kingdom, the US). They are based either on administrative data or on panel data that includes multiple income observations for parents and children, which are usually unavailable in Latin American countries. Researchers have therefore often tried to assess intergenerational mobility of income and earnings in Latin America with the two-sample-two-stage least squares (TSTSLS) method, following Björklund and Jäntti (1997).[8] Some examples are: Jimenez (2016) for Argentina; Ferreira and Veloso (2006) and Dunn (2007) for Brazil; Nunez and Miranda (2010) for Chile; Grawe (2004) for Peru and Ecuador; Doruk et al. (2022) for Brazil and Panama; and Daza Báez (2021) for Mexico.

Important recent exceptions are the studies by Leites et al. (2022) for Uruguay and Britto et al. (2022) for Brazil, which provide, for the first time, intergenerational income mobility estimates for developing countries based on administrative data (tax and social security records). These studies highlight a very

[8] In this approach, one first identifies a set of parental characteristics which are observed in the main survey (for the children's generation). Then an earnings (or income) regression on those characteristics is run in an earlier sample, selected so that it is representative of the parents' generation during its prime earning years. The coefficients from this "first-stage" regression in the earlier sample are then used to predict parental income in the main survey.

important aspect, which arguably is negligible for the study of income mobility in developed countries but of high importance in developing countries, namely that a large part of earnings and household income may derive from the informal sector for a considerable number of individuals. Hence, in administrative records, several income spells for individuals with less attachment to the formal labour market might be missing. Leites et al. (2022) implement a set of strategies to mitigate the bias resulting from this situation. Their results suggest that the degree of intergenerational persistence is significantly higher when considering families with less attachment to the formal labour market. Britto et al. (2022) account for informal income by imputing it based on survey data and come to the same conclusion.

One implication of these more recent studies is that the new frontier of intergenerational mobility research in Latin America (as well as in developing countries more generally) should probably involve a combination of administrative records, other novel data sources and well-established nationally representative surveys, in particular those including retrospective questions on parents' socio-economic status. Recent examples include Muñoz (2021), who uses census data to estimate educational mobility for various Latin American countries at a very granular geographical level; Neidhöfer et al. (2023), who compute intergenerational mobility trends for subnational regions and estimate the relationship between social mobility and future economic development; Ciaschi, Marchionni and Neidhöfer (2023), who estimate the association between parental social status and their children's education and income rank adopting the Lubotski-Wittenberg method (Lubotsky and Wittenberg, 2006); Ahsan et al. (2023), who provide estimates for sibling correlations in schooling for a large number of developing countries, including several Latin American countries, using DHS surveys; Neidhöfer, Ciaschi and Gasparini (2022), who estimate intergenerational mobility of economic well-being with Latinobarometro data by exploiting information about homeownership, goods that the household owns, and other measures for socio-economic status; and Gabrielli (2022), who measures intergenerational mobility of self-perceived socio-economic status of respondents and their parents.

Finally, the literature in Latin America, as elsewhere, has moved towards estimating associations over three generations (i.e., from grandparents to grandchildren) rather than just over two generations (i.e. from parents to children). The main aim of this branch of the literature is to estimate long-run patterns of intergenerational mobility and to test the hypothesis that the intergenerational transmission of advantage follows an AR(1) process. That would imply that children's outcomes depend directly only on the outcomes of their parents, and not of earlier generations (for a review of the literature, see Anderson et al., 2018). Contributions that estimate educational mobility over three generations for Latin American countries include Celhay and Gallegos (2015) for Chile, and Celhay and Gallegos (2023) for six Latin American countries. They find, first, that educational mobility over three generations is lower than the AR(1) model would predict, with a much larger difference for Latin America than for developed countries, and, second, that compulsory schooling laws contribute to explaining long-run mobility patterns.[9]

[9] However, Moreno (2021) finds that, using Mexican data, grandparental education has no effect once parental education is considered. This finding is in line with one part of the international literature on the topic, which argues that a significant coefficient for grandparental outcomes could be a statistical artifact caused by omitted variable bias (e.g. Solon, 2014).

2.2 The literature on inequality of opportunity in Latin America

An alternative approach for assessing the extent to which inherited factors determine children's outcomes is to use pre-determined variables other than parental income (generally referred to as "circumstances"), such as parental education and occupation; place of birth; race or ethnicity; and biological sex at birth. This is what the literature on inequality of opportunity does. Bourguignon, Ferreira, and Menendez (2007) were the first to offer empirical estimates of inequality of opportunity for Latin America, by analysing the role of circumstances in accounting for income inequality in Brazil. They found that the share explained by observed circumstances amounts to about 25% of total inequality. Subsequent analyses by Ferreira and Gignoux (2011) for seven Latin American countries, and by Núñez and Tartakowsky (2011) for Chile show that inequality of opportunity for income in other Latin American countries was broadly similar or even higher. Ferreira and Gignoux's (2011) estimates for the share of total income inequality accounted for by inequality of opportunity ranges from 23% (in Colombia) to 36% (in Guatemala). The shares were higher for consumption inequality, from 24% to 53%, again in Colombia and Guatemala respectively. Although interpreted as lower-bound estimates, these shares are relatively high compared to the estimates obtained for developed countries, as shown, for instance, by the comparative multi-country study by Brunori, Ferreira and Peragine (2013). Interestingly, parental education typically ranks as the single circumstance with the strongest influence.

In the same spirit, researchers have also estimated inequality of opportunity in educational achievements in the region. Andersen (2003) estimates the importance of family background in determining the schooling gap for children in 18 Latin American countries. Her results rank Guatemala and Brazil as the countries with the highest levels of inequality of opportunity, and Chile, Argentina, Uruguay and Peru as those with the lowest levels. Gamboa and Waltenberg (2012) use data from the 2006 and 2009 PISA surveys to estimate inequality of educational opportunities in the six Latin American countries included in the survey. Pooling all Latin American pupils, and adding the pupil's country as further circumstance, their results confirm a degree of inequality of opportunity of 21-27%. However, their findings also highlight substantial heterogeneity across countries, years, and specification of circumstances. Brazil stands out as the country with the highest inequality of opportunity. Parental education, again, has the strongest influence. Furthermore, school type (public or private) shows up as a circumstance significantly influencing individual opportunities for educational success.

Using data from the same PISA surveys, Ferreira and Gignoux (2014) analyse inequality of educational opportunities for a larger set of countries worldwide. They find that the six Latin American countries in the PISA dataset are among those with the highest levels of inequality of opportunity. Paes de Barros et al. (2009) develop the Human Opportunity Index for children, which includes access to education as one important dimension. Other dimensions are access to basic services, such as water and electricity. Their results mainly confirm the ranking of countries found by other studies on inequality of opportunity in income and education in Latin America.

3. A new approach

The previous section highlighted some of the difficulties faced by researchers trying to estimate intergenerational income mobility in Latin America: chiefly the absence of datasets that allow for a direct link between the reliably recorded incomes of parents and their (adult) children for a representative sample (e.g., one free of co-residency bias). Alternative approaches, such as TSTSLS estimation can help, but they face their own shortcomings (Emran and Shilpi, 2019; Olivetti and Paserman, 2015; Santavirta and Stuhler, 2020; Bloise et al., 2021). Recent studies using administrative datasets are promising, but (i) they are too few in number to allow for regional coverage, and (ii) they still struggle with the absence of informal workers from the data. As a result, much of what we know about intergenerational persistence in this region still comes from the analysis of educational transmission.

While the education work is highly valuable in itself, it clearly does not answer all the questions one might have about the intergenerational reproduction of inequality. It is possible, for example, that there is movement in the education distribution but that this is transmitted only slowly, or partly, or not at all, to the distribution of incomes. We know that there are other mechanisms for income persistence, such as the intergenerational transmission of employers (Corak and Piraino, 2011), of social networks, or of socio-emotional skills – all of which might weaken the connection between changes in educational persistence and income persistence.

As noted above, an alternative approach that has had some success in examining the persistence of income inequality is the inequality of opportunity approach, where a number of non-income variables replace parental income on the right-hand side of the estimating equation, so to speak. The objective in this literature is to quantify the amount of (present day) inequality that is due exclusively to pre-determined circumstances – variables over which people have no control or for which they cannot be held responsible. Empirically, this can be done in a number of different ways. Conceptually, though, there are two main approaches: ex-ante IOp and ex-post IOp[10]

The ex-ante approach seeks to measure the inequality between types, that is: between population groups that share the same circumstance variables. It requires that the researcher choose a way to measure the value of the opportunity set corresponding to each type, and then compute the (population-weighted) inequality in the distribution of those values (see, e.g., Ferreira and Gignoux, 2011). The ex-post approach sees inequality of opportunity as all inequality between people who exert the same degree of effort. If one is willing to assume that all inequality within types is due to effort and that outcome is monotonically increasing with effort, then the quantiles of the type-specific income distribution would be indicators of the relative degree of effort expended by individuals in those positions. Inequality across types for each quantile, subsequently aggregated across quantiles, would be the right measure of inequality of opportunity (see, e.g., Checchi and Peragine, 2010).

[10] See Fleurbaey and Peragine (2013) for definitions and an analysis of the theoretical distinctions between the two approaches.

Slightly more formally, consider a population $\mathcal{P}$: $\{i, i = 1 \dots, N\}$ of *N* individuals indexed by i. Let each individual *i* be fully characterized by a scalar measure of advantage y_i, such as income, and by a vector of pre-determined circumstances, C_i.[11] The vector of circumstances C_i – which takes a form such as (male; ethnicity: Aymara; born in El Alto; mother's education: primary; father's education: secondary,…) – defines the type to which individual *i* belongs. A type C is a set of individuals who share identical circumstances.

Denote the set of all possible types by $\mathbb{C}$. In any given population there is a finite set of types $\Gamma, \Gamma \in \mathbb{C}$, which is, by definition, a partition of the population: The intersection of any two types $C \in \Gamma$ is empty, and the union of all $C \in \Gamma$ is $\mathcal{P}$. For each $C \in \Gamma$, there is a type-specific income distribution, $F(y_c|C = c)$, with mean μ_c and quantile function $y_c = F^{-1}(q|C = c)$.

If we are prepared to use the expected value of a type's income distribution as a measure of the value of the opportunity set of type *c* (van de Gaer, 1993), then one class of ex-ante measures of inequality of opportunity is given simply by $IO_a = I(w_c\mu_c)$, where w_c denotes the population share of type C, and *I* is a suitable inequality measure, such as the Gini coefficient or the mean logarithmic deviation, defined over the vector of population-weighted type means, $w_c\mu_c$, the dimension of which is the number of types in the partition (the cardinality of Γ).

Alternatively, if we are prepared to assume that all inequality within types is due to effort, then one class of ex-post measures of inequality of opportunity is given by $IO_p = \int_{q=0}^{1} I_q \left(\frac{\mu}{\mu_q} y_{qc}\right)$, where μ_q denotes the mean income (across types) at quantile q. Here, one is computing inequality across types at each individual quantile of the type-specific quantile functions, and then aggregating those inequality estimates across quantiles. Relative measures of IOp are simply either of the above expressions divided by total inequality in the population, $I(y_i)$.

It is important to note that both the ex-ante and ex-post approaches share the same first step: to select a partition $\Gamma(\in \mathbb{C})$ of the sample into subgroups that share identical circumstances (i.e., types). The choice of partition is not unique and inevitably involves decisions by the researcher. Consider the example of our Bolivian sample, which includes as potential circumstance variables the sex of the respondent (two categories); ethnicity (seven categories); occupation of the father or mother, whichever is more highly ranked (eleven categories); and education of the father and mother (four categories each). So, if one used the finest possible partition, there would be 2 x 7 x 11 x 4 x 4 = 2,464 potential types. Once the sample restrictions which are discussed in the next section are applied, the sample contains 5,265 individuals, just over twice the number of potential types. 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. (See Brunori, Peragine and Serlenga, 2018).

[11] So, each individual is fully characterized by $\{y_i, C_i\}$.

This pitfall was recognized by the early papers in this literature, so that much more parsimonious partitions were typically used. Ferreira and Gignoux (2011) did not have data for Bolivia but, using similar kinds of data on circumstances for other Latin American countries, they restricted their partitions to 54 or 108 types for each country, by arbitrarily grouping subcategories into coarser groupings. They recognized that this would lead to an omitted variable bias, arising both from the absence of other, truly unobserved circumstances (such as parental income), and from the loss of variation from those circumstance variables or categories that were observed, but not used in the partition.

The choice of partition for IOp estimation, given the available variables and categories in any given data set, therefore inevitably involves a trade-off between reducing the downward omitted circumstance variable bias (by enlarging the number of types) and reducing the upward "overfitting" bias (by reducing the number of types). This has been the main recent challenge in this literature: to find an optimal, non-arbitrary way of splitting the sample or population in the first step of the estimation – be it for an ex-ante or ex-post IOp analysis (see Brunori, Peragine, and Serlenga, 2018, and Brunori, Ferreira, and Salas-Rojo, 2023)

There are a number of possible ways to try to address this challenge. Here we follow the proposals by Brunori, Hufe, and Mahler (2023) for the ex-ante case, and by Brunori, Ferreira, and Salas-Rojo (2023) for the ex-post case. These two studies rely on different (but related) machine-learning algorithms to obtain the most relevant partition given the data under consideration, consistent with a preselected level of statistical significance.

3.1 Estimating Ex-Ante IOp using Conditional Inference Trees

The ex-ante approach of Brunori, Hufe, and Mahler (2021) employs the conditional inference trees and random forests developed by Hothorn, Hornik and Zeileis (2006). A conditional inference tree consists of a set of terminal nodes (leaves) obtained by recursive binary splitting, as follows:

1. Choose a critical significance level α for hypothesis testing.

2. Given a set of circumstance variables and categories, compute the correlation coefficient between the outcome variable and each circumstance. If the Bonferroni-adjusted p-value of all correlation tests are higher than the chosen critical value α, one exits the algorithm.

3. If the null hypothesis is rejected, the variable whose correlation with the outcome has the smallest Bonferroni-adjusted p-value 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 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. [C]* is chosen as $[C]^* = \{[C]: argmin\ p^{[C]}\}$ That is to say: when there are n > 2 categories for a particular circumstance variable, the categories are divided into the two groups that are least likely to have the same mean.

6. Repeat steps 2 – 5 for each node (sub-sample), until one has exited everywhere. [12]

When one has exited everywhere, the output consists of a partition of the sample or population. We treat each terminal node of the tree as a type, for which we compute the population weight $\hat{w}_c$ and the mean $\hat{\mu}_c$. The ex-ante estimate of inequality of opportunity is then $\widehat{IO}_a = I(\hat{w}_c\hat{\mu}_c)$. In addition to the estimate itself, this approach has the considerable added benefit that the partitioning process, as embodied in the tree itself, contains interesting information on the structure of inequality of opportunity in the particular society. The conditional inference tree for Bolivia in 2008 is shown as an example in Section 5 below, where we will return to this interpretation.

Among machine learning algorithms, regression trees are known to be characterized by low bias but to suffer from high variance and conditional inference regression trees are no exception. This means that the opportunity tree which is first estimated on its own is rather sensitive to the particular sample observed and that an equally representative but slightly different sample might lead to a different partition. As is standard in the machine learning literature when dealing with high variance learners, one can alleviate this kind of problem by bagging: constructing subsamples of the original data and computing trees for each one. Under the appropriate aggregation procedures, this process generates what is known as *a random forest*. Following Hothorn, Hornik and Zeileis (2006) we obtain our conditional inference random forest by using fivefold cross validation to tune the two main parameters, namely the significance level α and the number of circumstances permuted at each split.[13]

Because a conditional inference tree chooses partitions on the basis of differences in a single statistic of interest in each group or type, it is particularly well-suited to the ex-ante approach, so long as that statistic of interest is a suitable measure of the value of the opportunity set for each type, e.g., the mean of its conditional income distribution.

3.2 Estimating Ex-Post IOp using Transformation Trees

Brunori, Ferreira and Salas Rojo (2023) suggest that an alternative (but related) algorithm is better suited to an ex-post IOp interpretation. That algorithm is based on the transformation trees first proposed by

[12] We set $\alpha = 0.01$ and impose the additional requirement that each terminal node must have a country-specific minimum of N_j^{min} observations. This is chosen for each country *j so that* $N_j^{min}/S_j = \min_c N_{cJ}^*/S_J$, where N_{cJ}^* denotes the number of observations of type c in country J when the partitioning algorithm is run with no min-bucket restriction (in addition to setting $\alpha = 0.01$), and country *j=J* is the country with the lowest sample size, S_J. All other parameters are the default parameters in the "*ctree*" R function.

[13] The number of circumstances permuted at each split is the integer nearest to the square root of the number of available circumstances. All other tuning parameters are set to the default values in the "*cforest*" R function.

Hothorn and Zeileis (2021). In essence, a transformation tree algorithm is analogous to a conditional inference tree except that, instead of comparing a single statistic (e.g., the mean) across all possible partitions in $\mathbb{C}$ to choose the split with the lowest p-value for the null hypothesis that the statistics are identical on both sides, the algorithm estimates full distribution functions for each possible partition. It then chooses to split the sample (in a binary fashion at each step) between the two groups whose distribution functions are least likely to be identical.

Just as type means are the key ingredients for an ex-ante estimation of IOp, type-specific distribution functions are the key ingredients for an ex-post estimation. The key assumption underlying the transformation tree algorithm is that the true functions $F(y_c|C = c)$ can be sufficiently well-represented by parametric approximations $F\left(y_{qc}, \theta(c)\right)$. $\theta(c)$, known as the conditional parameter function, maps from the set of all possible type partitions, $\mathbb{C}$, on to the set of possible distributional parameters, Θ. Under this assumption, the problem of estimating the conditional distributions for all types in the optimal partition reduces to the problem of selecting the optimal parameter estimates, $\hat{\theta}$, given the data *{y, C}*.

Hothorn and Zeileis (2021) propose an adaptive local likelihood maximization approach for that purpose. Specifically, they select $\hat{\theta}$ as:

$$\hat{\theta}^N(c) = \arg\max_{\theta \epsilon \Theta} \sum_{i=1}^{N} w_i(c, \theta)\ell_i(\theta) \tag{2}$$

where $i \in \{1, \dots, N\}$ denotes each observation in the data set and $\ell_i(\theta)$ denotes the log-likelihood contribution of *i*, when the parameters are given by θ. The recursive binary splitting process that creates a transformation tree is implemented by choosing weights:

$$w_i(c) = \sum_{b=1}^{B} I(c \in \mathcal{B}_b \wedge c_i \in \mathcal{B}_b) \tag{3}$$

The indicator function in (3) – and therefore the weight it defines – take the value 1 when observation c_i is "sufficiently close" to *c*, and zero otherwise. In other words, the optimal weights define the cells, or nodes, of the (optimal) partition. At the terminal nodes, $\mathcal{B}_b$ corresponds to a type so the maximization process given by eqs. (2) and (3) allocates each observation to a type and sums the local likelihood functions across types. The type partition and the parameter vector θ are chosen so as to maximize that sum of likelihoods. That is, given the available data *{y, C}* and the recursive splitting approach to weights, the likeliest set of types and income distributions conditional on type is the one given by $F\left(y_{qc}, \hat{\theta}^N(c)\right)$.

In practice, $\hat{\theta}$ are chosen from the class of Bernstein polynomials, using the "*trefotree*" R function developed by Hothorn and Zeileis (2021). We set the critical significance level $\alpha = 0.01$ and the minimum number of observations at each terminal node as before. Unlike with conditional inference trees,

transformation trees require the econometrician to choose the order of the Bernstein polynomial used to approximate the type-specific conditional distribution functions. We choose that order by setting a minimum improvement in the aggregate out-of-sample log-likelihood of 0.1% to justify a higher order.

The output of the estimation consists once again of a partition, but now including a parametric estimate of each type's cumulative income distribution function, based on the polynomials estimated as just described. These parametric conditional distributions can then be inverted to yield the predicted type quantile functions $\hat{y}_{qc} = F^{-1}\left(q,\ \hat{\theta}(c)\right)$, from which a measure of ex-post inequality of opportunity can be computed as $\widehat{IOp} = \int_{q=0}^{1} I_q\left(\frac{\hat{\mu}}{\hat{\mu}_q}\hat{y}_{qc}\right)$. Just as in the ex-ante case, the transformation tree itself is of additional intrinsic interest, beyond being a means to the end of generating the ex-post IOp estimate. The "family" of type-specific parametric cumulative distribution functions (CDFs) can be displayed directly, as in Figure 3 for Bolivia in Section 5.

Just as conditional inference random forests seek to add robustness to the estimation of conditional inference trees, Hothorn and Zeileis (2021) propose an algorithm to estimate transformation forests. That algorithm obtains individual predictions from a forest of transformation trees without assigning individuals to a particular terminal node (or type), which is problematic in the context of IOp measurement because the counterfactual distribution used to assess unequal opportunities is based on the type-specific expected CDFs (ECDFs).

Yet, like CI trees, transformation trees are also high variance estimators, and researchers might therefore be tempted to implement aggregation methods to enhance the robustness of ex-post IOp estimates. Unfortunately, approaches like bagging or others aimed at reducing algorithm variance end up introducing a significant downward bias in the case of ex-post IOp. In each iteration, the researcher would, in fact, obtain a different partition among types (similarly to what occurs with random forests in ex-ante IOp). Furthermore, individuals would be observed at different quantiles within their type-specific distributions. Estimating individual advantages across iterations becomes exceedingly noisy and aggregating these noisy measures results in a reduction of explained variability.

In the last step of our analysis, we address the question of the relative importance of the different observed circumstances in contributing to inequality of opportunity as measured. Just as 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, or any decomposition thereof. Nonetheless, the various circumstances contribute differently to the overall IOp estimate and quantifying those differences is of descriptive interest.

Since there is no guarantee (or likelihood) that the contributions of all circumstance variables are additively separable, the correct approach to identifying individual contributions is through a Shapley decomposition (see Shapley, 1953, and Shorrocks, 2013). Intuitively, a Shapley decomposition calculates the overall contribution of a variable *x* to some outcome function *y* as the average decline in *y* across all

possible combinations of ways in which y can be generated without x.[14] More precisely, we follow Brunori, Ferreira and Salas-Rojo (2023) and obtain Shapley value decompositions as follows:

1. Draw a sub-sample of the full sample.[15]
2. Estimate IOp in the sub-sample, in either the ex-ante or ex-post fashion, as above.
3. Re-estimate IOp in the sub-sample for all possible elimination sequences for each circumstance. (Each elimination consists of replacing the relevant circumstance with a constant vector **1**.)
4. After each elimination sequence, the tree and the resulting IOp measure are estimated and the IOp values after elimination are stored.
5. Average IOp across all elimination sequences for circumstance c. The difference between the overall IOp in the subsample and this average is the contribution of c.
6. Repeat steps 1-5 one hundred times.
7. The final estimate of the contribution of c to IOp is the average contribution across the 100 iterations.[16]

We should emphasize that, although it is based on the aggregation of many overfitted trees, we do not expect our evaluation of the relative importance of each circumstance to suffer from any bias. The focus is not on the absolute level of estimated IOp, but rather on the relative contribution of each circumstance. Consequently, the fact that each tree is overfitted, obtained from a subsample of the original data and then aggregated entails the typical advantage of bagging weak learners without affecting the robustness of the relative importance estimates of each circumstance. This is true both in the ex-ante and ex-post approaches.

The next section describes the data sets to which we apply these two data-driven approaches to the estimation of ex-ante and ex-post inequality of opportunity.

4. Data

Our basic data requirements consist of datasets containing information on $\{y_i, C_i\}, i = 1, \dots, N$ for a sample that is representative of a well-defined population: either nationally or, say, for all urban areas. Our unit of analysis is the individual and the income concept attached to each individual is, in all cases,

[14] It is important to remember that this value represents the *average* contribution of a circumstance to the total predictive power. It should not be confused with the *marginal* effect of a specific category. For example, in a society where 99% of individuals identify as white and 1% as black, the Shapley value for the circumstance "race" will be low, regardless of the level of discrimination against black individuals. Although the marginal effect of being black may be large and negative, the ability to predict income based on race will be low for the majority of respondents, who are white. This is because the average income of white individuals is very close to the overall population average.

[15] The default in the "*cforest*" R algorithm (Hothorn et al., 2006) is for a subsample share of 0.632. When the overall sample is less than 7,000 observations, we replace this with 0.9, so as to preserve sufficient sample size at each iteration to allow different circumstances to play a role in determining the structure of the tree.

[16] The contributions of each circumstance are reported in relative terms to adjust for the fact that sample sizes are smaller in the 100 replications than in the original sample.

age-adjusted equivalized household income, using the square-root equivalence scale. The age adjustment is intended to account for income variations driven by lifecycle factors, as an alternative to considering the date of birth as an additional circumstance variable. It is conducted by regressing each person's equivalized household income *x* on her age and age squared and using the regression constant plus residual as our outcome variable.[17]

The candidate vector of circumstances varies slightly across countries but always consists of at least four of the following six individual circumstance variables: sex; race or ethnicity; place of birth; father's *and* mother's education; and father's *or* mother's occupation (whichever ranks highest).[18] The specific categories within each of the last five (all but sex) vary from survey to survey. We use twenty-seven household surveys for nine Latin American countries, fielded between 2000 and 2015, that satisfy these requirements. These surveys are listed first in Table 1, which reports only the country, survey name and the corresponding acronym.

Table 1: Household surveys used in our analysis

Country	Survey Name	Acronym
Argentina	Encuesta Nacional sobre la Estructura Social	ENES
Bolivia	Encuesta de Hogares	EH
Brazil	Pesquisa Nacional por Amostra de Domicílios	PNAD
Chile	Encuesta de Caracterización Socioeconómica Nacional	CASEN
Colombia	Encuesta Nacional de Condiciones de Vida	ECV
Ecuador	Encuesta de Condiciones de Vida	ECV
Guatemala	Encuesta Nacional sobre Condiciones de Vida	ENCOVI
Panama	Encuesta de Niveles de Vida	ENV
Peru	Encuesta Nacional de Hogares	ENAHO

Survey waves available for different years for the same country are always waves of the survey named above. Table 2 therefore identifies each survey only by the country name and survey wave. All datasets were obtained from the SEDLAC harmonized database maintained by CEDLAS at the University of La Plata in Argentina. The final samples used for our analysis differ from the full samples in SEDLAC in three ways. First, we only include surveys that include retrospective questions on parental education and occupation.

[17] Specifically, we regress $\log x = \beta_0 + \beta_1 age + \beta_2 age^2 + \varepsilon$. All of the analysis described in Section 3 is then carried out using $y = \exp(\beta_0 + \varepsilon)$.

[18] We combine father's and mother's occupation into a single "highest parental occupation" variable, in order to reduce the number of observations for which these variables are missing. The ranking of occupations is based on ISCO codes where those are available. For the four countries where ISCO-coded occupations are not available (Chile, Ecuador, Guatemala and Panama), we rank them by employment category as follows: employer, employee, self-employed, laborer, domestic service, and other.

Second, from each household we include only household heads and spouses (if any).[19] The youngest 1% and the oldest 1% of these individuals is then removed, as are those living in households reporting negative or zero incomes. Third, observations with missing values for income or *any* circumstance variable are also excluded. For each survey, Table 2 reports the list of available circumstance variables; the age range in the final sample; as well as the final sample size, both in absolute numbers and as a share of the original sample size.

Table 2: Basic description of the household survey data

Country	Survey Year	Circumstances	Age Range	Final Sample Size	Relative sample size
Argentina	2014	Sex, race or ethnicity, place of birth, father's and mother's education, father's occupation	21 – 85	6,532	50.4%
Bolivia	2008	Sex, race or ethnicity, father's and mother's education, father's and mother's occupation	19 – 65	5,265	93.7%
Brazil	2014	Sex, race or ethnicity, place of birth, father's and mother's education, father's and mother's occupation	20 – 86	22,707	49.2%
Chile	2006	Sex, race or ethnicity, place of birth, father's and mother's education, father's and mother's occupation (only 2009)	23 – 85	82,555	68.6%
	2009		23 – 86	64,613	56.5%
	2011		23 – 86	55,398	59.6%
	2013		23 – 87	58,713	56.4%
	2015		23 – 87	75,789	59.1%
Colombia	2010	Sex, race or ethnicity, place of birth, father's and mother's education	20 – 84	16,946	74.2%
Ecuador	2006	Sex, race or ethnicity, place of birth, father's and mother's education, father's and mother's occupation	20 – 83	18,971	84.7%
	2014		20 – 85	39,229	83.0%
Guatemala	2000	Sex, race or ethnicity, place of birth, father's and mother's education, father's and mother's occupation (only 2000)	19 – 79	11,617	93.5%
	2006		19 – 81	20,234	87.3%
	2011		19 – 83	20,058	88.0%
Panama	2003	Sex, race or ethnicity (except 2008), place of birth, father's and mother's education, father's	21 – 84	8,789	86.2%
	2008		21 – 85	8,627	77.2%

[19] Children are omitted from the sample since the equivalized income in their households is closer to a circumstance than to an outcome for them. Other adults are excluded since they may be temporary residents or non-family members, with a more tenuous relationship to the household's situation.

		and mother's occupation (except 2003)			
Peru	2001	Sex, race or ethnicity, place of birth, father's and mother's education	21 – 83	23,852	87.3%
	2006		22 – 86	14,641	72.8%
	2007		22 – 86	16,516	76.0%
	2008		22 – 86	15,616	74.3%
	2009		23 – 86	15,836	74.6%
	2010		23 – 87	15,568	74.0%
	2011		24 – 87	17,699	73.1%
	2012		23 – 87	18,134	73.8%
	2013		24 – 87	21,382	71.8%
	2014		23 – 88	21,580	71.3%
	2015		23 – 87	22,716	72.2%

5. Ex-Ante Inequality of Opportunity

As described in Section 3, the main outputs from our estimation of ex-ante IOp for each country/year are: (i) a conditional inference (C.I.) tree; (ii) a partition of the population (consisting of the terminal nodes of that tree), with population share and mean income[20] for each type; (iii) estimates of IOp from both the tree and the associated random forest. As an illustration, Figure 1 below depicts the C.I. tree for Bolivia, 2008, with the type partition at the bottom. Population shares are expressed in percent and type means are expressed as multiples of the overall mean (of US$ 636.96 per month at 2011 PPP exchange rates.) Trees for the other eight countries are presented in Appendix 1, for the most recent available survey for each country.

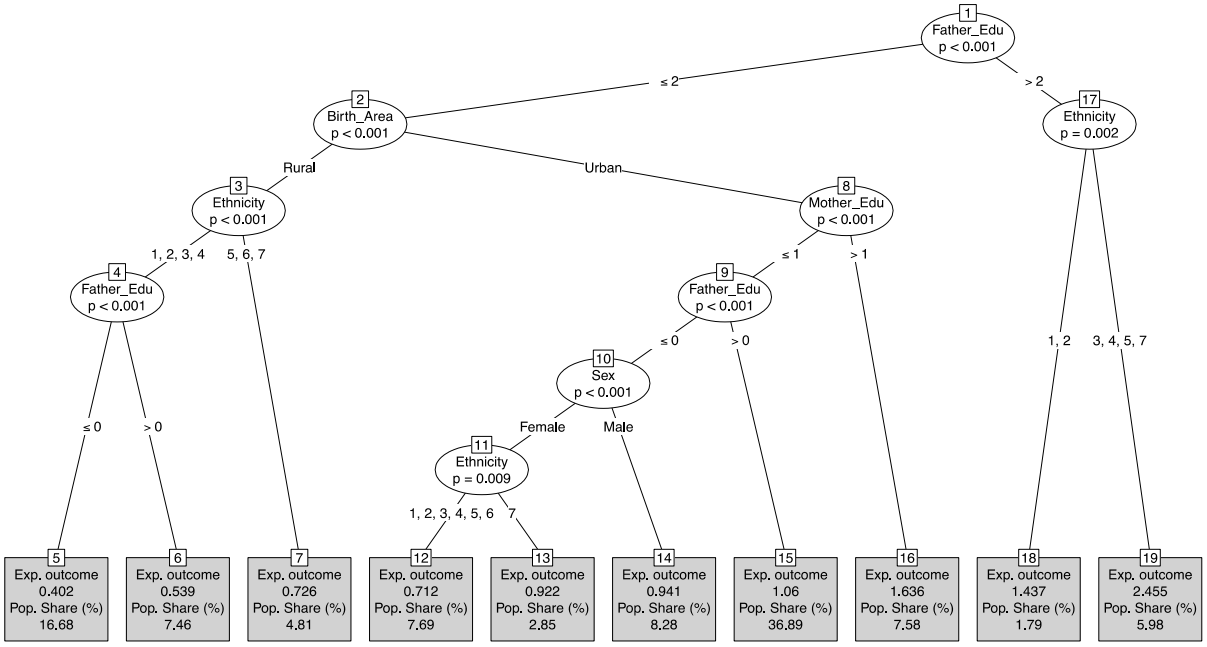
Starting from a sample of 5,265 individuals, with information on income, sex, ethnicity, father's and mother's education and highest parental occupation, the algorithm yields a final partition of the population into ten types. As noted earlier, this compares with a maximum of 2,464 potential types, arising from the combination of two categories for sex of the respondent, seven categories for ethnicity, eleven categories for the occupation of the father *or* mother (whichever is more highly ranked), and four categories each for the education of the father and mother.[21]

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

[21] The seven ethnicity categories are: 1=Quechua, 2=Aymara, 3=Guarani, 4=Chiquitano, 5=Mojeño, 6=other indigenous, and 7=not indigenous. The occupational categories are armed forces; managers; professionals; technicians and associate professionals; clerks; service and sales workers; agricultural, forestry and fishery workers; craft and trade workers; plant and machine operators; elementary occupations; and unemployed. The four parental education categories are: 1=no education or incomplete primary; 2=primary complete; 3=secondary education; and 4=tertiary education.

The ten types range in size from less than 2% to almost 37% of the population, and in income from 40% to 246% of mean income. The measures of inequality of opportunity arising from this partition are a Gini coefficient of 0.218 and a mean logarithmic deviation (MLD) of 0.091. Given the overall inequality levels in the same sample (a Gini of 0.490 and an MLD of 0.477), these results imply that inequality of opportunity in Bolivia accounts for 45% of overall inequality when measured by the Gini coefficient, or 19% when measured by the mean log deviation. The analogous figures from the random forest are a Gini coefficient of 0.243 (50% of the overall Gini) and an MLD of 0.102 (21% of the overall MLD).

Figure 1: Conditional Inference Tree for Bolivia, 2008



Source: Authors' elaboration using data from Bolivia's Encuesta de Hogares 2008.

These results call for two remarks. First, although a single tree may suffer from high variance, and one should therefore never place too much emphasis on its exact structure, the full tree structure is nonetheless informative. The most salient cleavage it identifies in the Bolivian society – in terms of the statistical significance of the difference in equivalized incomes between any two groups – is between those whose fathers went to university, and those whose fathers did not. The first group is only subdivided once more, by ethnicity, yielding two of the country's three richest types, with average incomes 1.4 and

2.5 times the national average, respectively. Together, these two types account for 7.8% of the population.

Those whose fathers did not attend university – the remaining 92% of Bolivians – are then split by rural and urban areas of birth. The rural types are next split into the main indigenous groups on one side (Quechua, Aymara, Guarany and Chiquitano), and the non-indigenous (and two very small groups, the Mojeños and "others") in another. This is basically the "whites" group. Down the main indigenous branch, father's education appears again as splitting circumstance. For the urban types, mother's and father's education, sex and ethnicity all appear as splitting circumstances. The very poorest type, with incomes 40% of the national average, are indigenous people born in rural areas to fathers with no formal education. Comprehensive tables presenting the poorest and richest types in all nine countries, utilizing both the ex-ante and ex-post approaches, can be found in Appendix 2.

Second, the relative IOp measures (those expressed as shares of total inequality, $I(y_i)$) obtained from both the tree and the random forest are much higher for the Gini coefficient than for the mean log deviation. This is not specific to Bolivia, nor indeed to using a machine-learning approach for selecting the partition (see Brunori, Palmisano and Peragine, 2019, and Brunori, Ferreira, and Salas-Rojo, 2023). Instead, this fact reflects the different sensitivities of the two measures: Whereas the mean log deviation is particularly sensitive to the tails of the distribution, the Gini coefficient is more sensitive to gaps closer to the mean of the distribution. Type means are the result of averaging within sizable groups and are, therefore, clustered closer to the overall mean by the law of large numbers. The share of overall inequality accounted for by these differences is therefore greater, for the same distribution, when measured by an index that is center-sensitive than by one which is tail-sensitive.

It should also be noted that the fact that, unlike MLD, the Gini coefficient is not exactly decomposable by population subgroups is not material for our analysis. Although the early empirical literature on inequality of opportunity placed great importance in selecting fully decomposable inequality indices (typically member of the Generalized Entropy Class and, in particular, the MLD), an exact decomposition is not essential for our purposes here. We do not interpret the difference between the overall and the IOp Gini coefficients as an aggregated within-type Gini. Indeed, the fact that the residual of the Gini decomposition is always positive means that the true between-type Gini is no lower, and possibly higher, than the IOp measure we adopt (see Ferreira and Peragine, 2016). In what follows, we therefore present our main results using the Gini coefficient. All corresponding estimates using the MLD are presented in Appendix 3.

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-seven surveys covering Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guatemala, Panama, and Peru. Table 3 below presents the main results for the Gini coefficient, from both the conditional inference trees and the associated random forest. Column 1 reports the number of types in each tree partition and Column 2 lists the overall Gini coefficient for each country/year. Columns 3 and 4 give the inequality of opportunity estimates from the tree in absolute and relative terms respectively,

whereas columns 5 and 6 report the absolute and relative IOp estimates from the random forest. The number of types ranges from a low of ten (in Bolivia, 2008) to a high of 32 (in Chile, 2013). The overall Gini coefficient for age-adjusted equivalized incomes ranges from 0.39 (in Argentina, 2014) to 0.56 in Colombia (2010).[22]

Table 3: Conditional inference tree results for 27 surveys

Country/Year	# Types	Total Gini	IOp Gini (Trees)	Relative IOp Gini (Trees)	IOp Gini (Forest)	Relative IOp Gini (Forest)
Argentina 2014	14	0.3918	0.1715	0.4377	0.1731	0.4418
Bolivia 2008	10	0.4901	0.2181	0.4450	0.2433	0.4965
Brazil 2014	25	0.5157	0.3037	0.5889	0.3039	0.5893
Chile 2006	21	0.5347	0.2806	0.5248	0.2844	0.5319
Chile 2009	28	0.5524	0.3034	0.5492	0.2527	0.4574
Chile 2011	27	0.5285	0.2966	0.5613	0.2794	0.5287
Chile 2013	32	0.5262	0.2767	0.5259	0.2594	0.4930
Chile 2015	29	0.5003	0.2537	0.5071	0.2614	0.5225
Colombia 2010	12	0.5588	0.2460	0.4402	0.2640	0.4724
Ecuador 2006	18	0.5295	0.2883	0.5445	0.2850	0.5383
Ecuador 2014	18	0.4643	0.2053	0.4422	0.2103	0.4530
Guatemala 2000	11	0.5454	0.2957	0.5421	0.2933	0.5377
Guatemala 2006	16	0.5329	0.3296	0.6185	0.3189	0.5984
Guatemala 2011	11	0.5311	0.2711	0.5104	0.2479	0.4667
Panama 2003	14	0.5430	0.2998	0.5521	0.2748	0.5061
Panama 2008	13	0.5122	0.2630	0.5135	0.2717	0.5305
Peru 2001	17	0.5087	0.2790	0.5485	0.2778	0.5461
Peru 2006	19	0.4962	0.2996	0.6038	0.2812	0.5667
Peru 2007	18	0.4933	0.2827	0.5731	0.2736	0.5547
Peru 2008	20	0.4673	0.2594	0.5551	0.2620	0.5607
Peru 2009	21	0.4635	0.2474	0.5337	0.2463	0.5314
Peru 2010	17	0.4495	0.2265	0.5039	0.2357	0.5244
Peru 2011	17	0.4501	0.2281	0.5068	0.2218	0.4928
Peru 2012	17	0.4432	0.2188	0.4937	0.2217	0.5003
Peru 2013	23	0.4416	0.2217	0.5021	0.2271	0.5143
Peru 2014	21	0.4255	0.2182	0.5128	0.2268	0.5330
Peru 2015	23	0.4293	0.2199	0.5122	0.2298	0.5353

[22] The lowest overall mean log deviation is also found in Argentina, 2014 (0.28), but the highest is 0.65 for Panama, 2003. See Appendix Table A1.

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 just over 0.30 in Brazil (2014), Chile (2009) and Guatemala (2006). Random forest estimates are remarkably similar, also ranging from 0.17 in Argentina (2014) to just over 0.30 in Brazil and Guatemala (2006), though they are somewhat lower for Chile, 2009.[23] For comparison, the Gini coefficient for the entire population of the Slovak Republic (in 2019) is 0.23.[24] In fact, besides Slovakia, the opportunity Gini coefficient for Brazil (2014), which reflects income differences between just 25 subgroups of the country's population, is higher than the overall population Gini coefficients of Belgium, Croatia, the Czech Republic, Denmark, Finland, Iceland, the Netherlands, Norway, Slovenia, and the United Arab Emirates[25] – counting only countries for which the World Bank's World Development Indicators report inequality for income, rather than consumption distributions.

As a share of total inequality, inequality of opportunity as measured by the Gini coefficient accounts for between 44% (in Argentina, 2014) and 62% (in Guatemala, 2006) when estimated by the conditional inference tree, and between 44% (in Argentina, 2014) and 60% (in Guatemala, 2006) when estimated by the ex-ante random forest. The correlation between these two series (relative Ginis from trees and forests) is 0.78. These are very large estimates of inequality of opportunity: More often than not, the shares are greater than 50%, including for Brazil, Chile, Ecuador, Guatemala, Panama and Peru, in various years. We are not aware of any previous estimate of inequality of opportunity for income that is greater than half of total national inequality.[26]

Descriptively, how important is each of the circumstance variables in accounting for these inequality of opportunity estimates? Table 4 presents the results of the Shapley value decomposition of the Opportunity Gini coefficients for the latest available survey year for each of our nine countries, as well as a simple average across countries. Note that parental occupation is missing (for both parents) in Chile, Colombia, Guatemala and Peru; mother's occupation is not used to construct parental occupation in Argentina; and ethnicity is missing in Panama – all of which makes cross-country comparisons perilous. That said, the circumstances that account for the largest share of inequality of opportunity as measured by the Gini coefficient are mother's and father's education which, together represent almost 60% of the total for the simple LAC average. Parental education is followed by birthplace (20%); and parental

[23] 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, Hufe, Mahler (2023), 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.

[24] For household per capita income. Source World Development Indicators online (21 August 2022).

[25] The UAE estimate is based on grouped, rather than unit-record, data.

[26] As noted earlier, inequality shares are much lower for the opportunity mean log-deviation, ranging from 17% (Argentina, 2014) to 32% (Peru, 2006), but the ranking is remarkably consistent with that of the relative Gini (correlation = 0.97). Tree and forest estimates of the mean log deviation are also quite similar, with a correlation coefficient of 0.82. See Appendix Table A1.

occupation (19%). Race or ethnicity account for 9.5% of overall inequality of opportunity and sex accounts for 2.6%.[27]

A number of country-specific results are worthy of mention, although some comparisons are perilous, as noted. The clearest example of the latter are comparisons of the parental education share between countries with and countries without information on parental occupation. These shares are much higher in Chile and Colombia, for example, than in Argentina or Bolivia, but this is likely to reflect, at least in part, the fact that they are capturing part of the effect of the omitted parental occupation variable in the former.[28] On the other hand, the comparison between Panama and Bolivia, which does not suffer from this problem, is informative, with parental education representing over 62% of IOp in Panama, but 42% in Bolivia. There is a great deal of variation in the importance of race and ethnicity as well, which ranges from 0.5% in Argentina to 18% in Bolivia – a country with a large ethnically indigenous population. It is almost as high in Peru and Guatemala, which are similar in that regards, and 12% in Brazil, where more than half the population identifies as black or mixed-race. In Peru, a country for which we have access to eleven survey waves, the contribution of ethnicity has increased steadily since 2001.

Table 4: Ex-ante Shapley value decompositions

Variable	Argentina 2014	Bolivia 2008	Brazil 2014	Chile 2015	Colombia 2010	Ecuador 2014	Guatemala 2011	Panama 2008	Peru 2015	AVG
Sex	2.01	3.04	1.61	6.78	2.71	2.46	2.70	0.78	1.35	2.61
Birthplace	35.22	18.61	14.15	13.83	25.27	2.38	28.70	17.26	24.07	19.94
Ethnicity	0.54	17.56	11.85	2.73	2.55	9.10	15.60		15.63	9.45
Father's Education	23.29	21.78	25.67	38.13	29.99	32.32	24.43	28.18	28.49	28.03
Mother's Education	22.36	20.46	25.33	38.54	39.49	32.78	28.56	34.23	30.45	30.24
Parental Occupation	16.57	18.55	21.39			20.95		19.54		19.4

[27] 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 have the same equivalized household income, so intra-household inequality is ignored. 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 caveat 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% for both ex-ante and ex-post IOp, over 90% of respondents do not report belonging to any ethnic minority. But, in addition, when interpreting Shapley values, it is important to understand that the structure of opportunities we observe today—reflected in the joint predictive power of the observed circumstances—is the result of historical evolution. It is quite possible that the distribution of parental education and occupation reflects the importance of ethnicity in earlier periods. Various ascriptive characteristics have historically influenced the lack of opportunities. Identifying a causal link or even the historical mechanisms of evolution that have shaped different countries in Latin America and the Caribbean since colonization is beyond the scope of this analysis.
[28] 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.

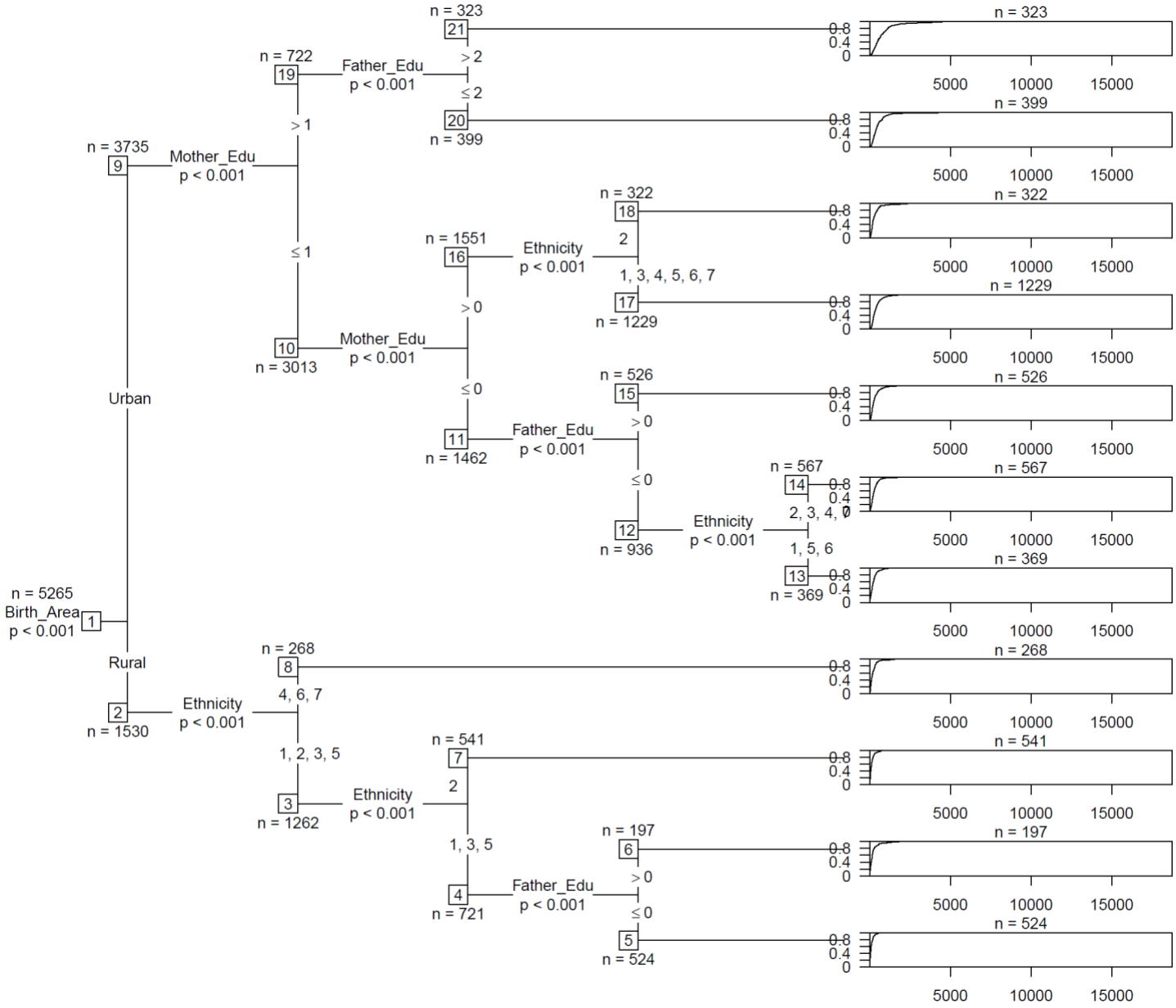
6. Ex-post Inequality of Opportunity

We now turn to the ex-post IOp estimates, computed as described in Section 3.2. In this case, besides (i) the transformation tree, (ii) the population partition obtained from the tree, (iii) summary IOp estimates obtained from the partition, and (iv) the Shapley value decomposition by individual circumstances, there is one additional output that conveys information about the conditional distribution within types, namely estimates of the expected cumulative distribution function (*ECDF*) for each type. As in Section 5, we present the transformation tree for Bolivia (2008) as an illustration, as well as the corresponding type-specific *ECDFs*, before reporting the comparative results for all countries. Transformation trees for the other eight countries are presented in Appendix 4 for the most recent available survey in each country.

Whereas the (ex-ante) conditional inference tree for Bolivia yielded a partition into ten types, the (ex-post) transformation tree in Figure 2 below partitions the sample into eleven types. The key difference between the two algorithms is, as noted earlier, that the ex-post approach reported in this section does not look for the most statistically significant difference between means to define sample splits; it looks for the most statistically significant difference between the full expected *CDF's*. It is therefore sensitive to differences across types in higher moments, and features such as within-type inequality, skewness, kurtosis, etc.

Therefore, if the differences in the 'shape' of the distributions across types are substantial, there is no presumption that the two trees should yield identical, or even very similar, results. In the particular case of Bolivia, the biggest difference is the demotion of father's education from first splitting variable to third. There are some marked similarities as well: the urban/rural dichotomy, which was a second splitting variable (for the bulk of the population) in the ex-ante case, is the first splitting variable in the ex-post case. The ECDFs for the four types that are exclusively rural (numbered 5, 6, 7 and 8) can be seen clearly to the left side of Figure 3. Although there isn't always first-order stochastic dominance, they are systematically poorer than the urban types. Beyond birthplace, ethnicity remains the next fundamental circumstance in rural areas, whereas mother's education becomes the most important circumstance in urban areas. In urban areas, father's education and ethnicity appear at the third and fourth levels and re-appear further below. At the extremes, the poorest type – type 5, consisting of Quechua, Guarany, or Mojeño individuals born in rural areas to parents with no formal education – is first-order dominated by all other types. Similarly, the richest type – type 21, consisting of urban-born individuals with fathers with secondary education or higher and mothers with primary education of higher – first-order stochastically dominate all other types.

Figure 2: Transformation tree for Bolivia, 2008.

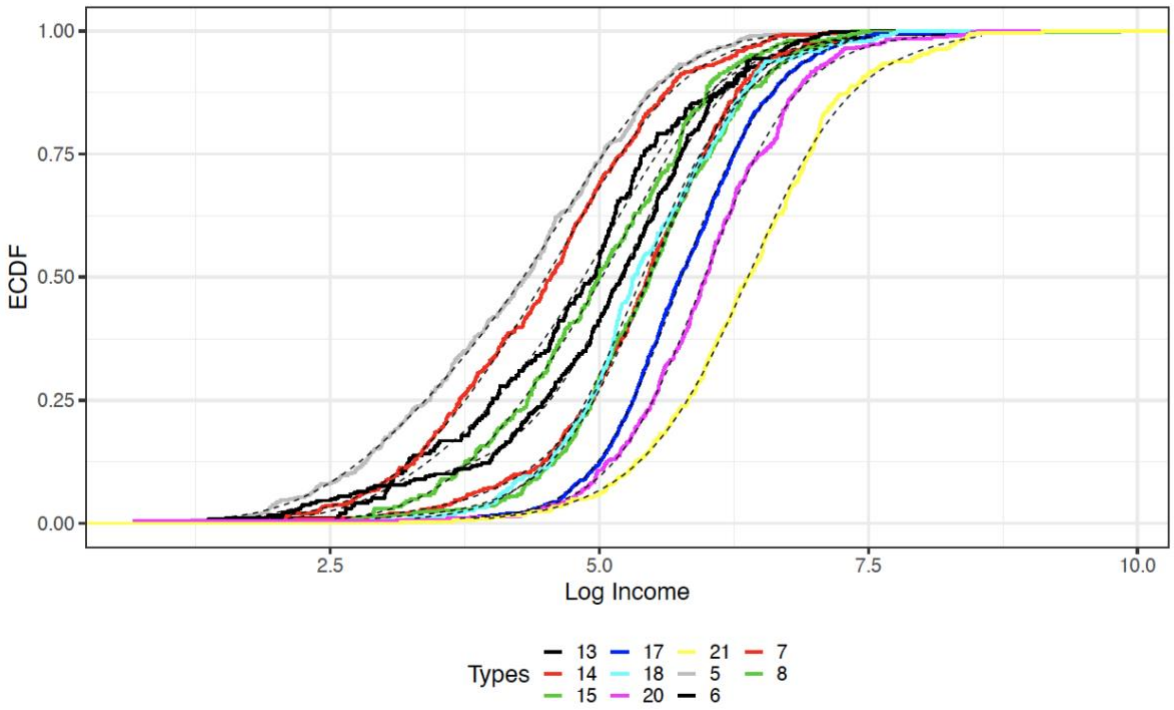


Source: Authors' elaboration using data from Bolivia's Encuesta de Hogares 2008.

The opportunity Gini coefficient arising from this eleven-type partition is 0.286, or 58% of the overall national income Gini of 0.49.[29] A Gini coefficient of 0.286 – obtained here by eliminating all inequality within these eleven population subgroups and considering only the inequality (across quantiles) between them – is roughly equal to that of the entire population of Norway, or of the Netherlands. It is higher than Belgium's (0.27).

[29] The opportunity mean log deviation is 0.153, or 32% of the overall national MLD of 0.477. See Appendix 3.

Figure 3: Type-specific expected cumulative distribution functions for Bolivia, 2008



Source: Authors' elaboration using data from Bolivia's Encuesta de Hogares 2008.

Turning now to the comparative results for ex-post inequality of opportunity for our 27 surveys from nine countries, Table 5 presents the main results. Analogously to Table 3, it reports the number of types in each transformation tree partition (column 1), overall national Gini coefficients in each sample (column 2), tree-based IOp Gini indices in absolute levels and as shares of overall inequality (columns 3 and 4). Although the exact same samples are used for the ex-ante and ex-post estimation, the latter tends to yield slightly finer partitions (i.e., with a greater number of types) than the former. The average number of types in Table 5 is 20.9, as compared to 18.9 in Table 3, and the increases are sometimes substantial, as in the case of Guatemala, 2011, where the ex-ante partition consists of 11 types, and the ex-post consists of 23. But there are also exceptions, such as Argentina, 2014 or Chile, 2009.

Table 5: Transformation tree results for 27 surveys

Country/Year	# Types	Overall national Gini	IOp Gini	Relative IOp Gini
Argentina 2014	12	0.3918	0.1735	0.4428
Bolivia 2008	11	0.4901	0.2858	0.5840
Brazil 2014	25	0.5157	0.2980	0.5779
Chile 2006	21	0.5347	0.2639	0.4936
Chile 2009	25	0.5524	0.2557	0.4629
Chile 2011	32	0.5285	0.2608	0.4935

Chile 2013	26	0.5262	0.2485	0.4723
Chile 2015	31	0.5003	0.2612	0.5221
Colombia 2010	10	0.5588	0.2668	0.4774
Ecuador 2006	22	0.5295	0.3053	0.5766
Ecuador 2014	18	0.4643	0.2197	0.4732
Guatemala 2000	14	0.5454	0.2936	0.5383
Guatemala 2006	23	0.5329	0.3352	0.6290
Guatemala 2011	23	0.5311	0.2493	0.4694
Panama 2003	13	0.5430	0.3096	0.5701
Panama 2008	11	0.5122	0.2930	0.5721
Peru 2001	27	0.5087	0.2741	0.5389
Peru 2006	25	0.4962	0.2967	0.5980
Peru 2007	24	0.4933	0.2801	0.5678
Peru 2008	21	0.4673	0.2679	0.5733
Peru 2009	19	0.4635	0.2564	0.5532
Peru 2010	19	0.4495	0.2382	0.5299
Peru 2011	22	0.4501	0.2324	0.5163
Peru 2012	18	0.4432	0.2362	0.5330
Peru 2013	21	0.4416	0.2383	0.5397
Peru 2014	22	0.4255	0.2347	0.5515
Peru 2015	29	0.4293	0.2481	0.5779

Despite the conceptual differences between the two approaches, the IOp estimates from ex-ante and ex-post trees are actually quite similar: the average ex-ante IOp Gini coefficient is 0.259 (Table 3), whereas its ex-post counterpart is 0.264 (Table 5). The correlation coefficient between the two tree-based (absolute) IOp Gini series, ex-ante and ex-post, is 0.812.[30] Although the two approaches solve different algorithms that search for distinct differences among the type distributions and typically yield different tree structures, the summary measures of inequality of opportunity are clearly similar. In both the ex-ante and ex-post analyses, the two countries with the lowest levels of inequality of opportunity (measured by the tree-based IOp Gini) were Argentina and Ecuador, both in 2014. At the upper end, Brazil and Guatemala both figure among the highest tree-based IOp countries in both the ex-ante and ex-post cases, although Ecuador (2006) and Panama (2003) are higher than Brazil (2014) in the ex-post case. Relative ex-post tree-based IOp Gini shares range from 44% in Argentina to 63% in Guatemala (2006) – a range identical to the ex-ante case.

The final step in the ex-post analysis, as in the ex-ante, is an effort to gauge the relative importance of individual circumstance variables by means of a Shapley value decomposition. Table 6 below reports the results, analogously to Table 4 in Section 5. Once again, parental occupation is missing in Chile, Colombia,

[30] The analogous correlation for the tree-based ex-ante and ex-post MLD estimates in Appendix 3 is 0.82.

Guatemala and Peru; mothers' occupation is not used for Argentina and ethnicity is missing in Panama. The birthplace variable for Bolivia and Colombia is an urban/rural dummy only. As before, the average contribution for parental occupation is computed excluding countries where it is missing (which is why the average column does not add up to 100%).

Table 6: Ex-post Shapley value decompositions

Variable	Argentina 2014	Bolivia 2008	Brazil 2014	Chile 2015	Colombia 2010	Ecuador 2014	Guatemala 2011	Panama 2008	Peru 2015	AVG
Sex	2.70	2.57	3.05	8.60	3.21	2.74	4.54	0.74	1.92	3.34
Birth Place	36.98	11.00	15.12	19.08	31.18	2.13	29.34	14.84	23.70	20.38
Ethnicity	0.00	16.03	12.20	2.79	3.21	8.22	10.80		13.30	8.32
Father's Education	20.64	25.15	24.07	35.50	24.34	32.23	26.16	28.62	28.49	27.25
Mother's Education	20.62	21.67	23.14	34.03	38.05	32.01	29.15	34.24	32.60	29.50
Parents Occupation	19.10	23.58	22.42			22.67		21.55		21.86

Results are very similar to those from the ex-ante case: Mother's and father's education make the greatest contributions to the ex-post inequality of opportunity Gini estimates, followed by birthplace and parental occupation. Ethnicity/race and gender are less important on average, thought ethnicity is quite important in Bolivia, Peru and Brazil. The consistency with the ex-ante results is reassuring.

7. How persistent is inequality in Latin America: a comparison across approaches

Having reviewed the earlier literature on intergenerational mobility and inequality of opportunity in Latin America, and then presented some new results based on two novel data-driven approaches that address the critical model or partition selection issue, we now briefly compare results and discuss implications. First of all, it is useful to illustrate the difference between the partitioning process followed by the ex-ante approach (using conditional inference trees) and the ex-post (using transformation trees) by means of an example. As noted earlier, the essential difference is that the C.I. splits the sample at each node so as to maximize the statistical significance of differences in means, whereas the transformation tree takes the shape of the entire distribution function into account.

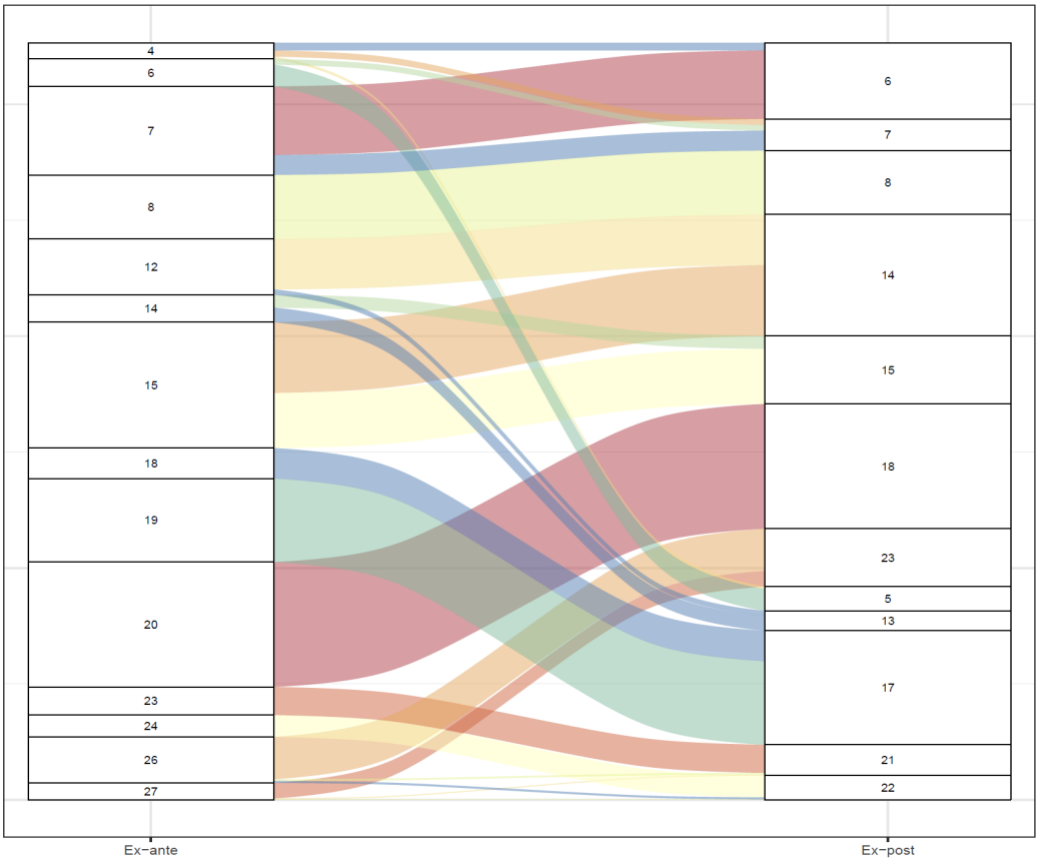
Figure 4 below illustrates the difference by means of an alluvial diagram (Panel A) that maps individuals according to the types to which they belong in the ex-ante partition (on the left) and in the ex-post (on the right), for Argentina (2014). There are fourteen ex-ante types and twelve ex-post and, in most cases, there is a clear correspondence across partitions. For example, the composition of Type 8 is essentially identical across the two. Ex-ante type 20 clearly corresponds to ex-post type 18. And so on. But there are also differences. Consider, for instance, ex-ante Type 7 (People born in provinces 1, Gran Buenos Aires, and 5, Patagonia, to fathers with education between categories 3 and 9, with occupations ranked 4 or higher). Individuals belonging to this type are divided into two types in the ex-post partition: Types 6 (born

in Gran Buenos Aires with a father in medium-low occupations) and 7 (born in Patagonia), incorporating a further split by area of birth. These two types are shown in red and green respectively, in the collection of ECDFs for Argentina in Panel B of Figure 4. The two types have relatively similar means (and cross near the median), but clearly different inequality levels, with the ECDF of type 7 being considerably flatter than that of Type 6. This likely reflects the fact that type 6 contains only individuals whose father had a medium-low level of education and a medium-low occupation while Type 7 contains individuals whose father had a medium-low level of education, irrespective of their occupation level.

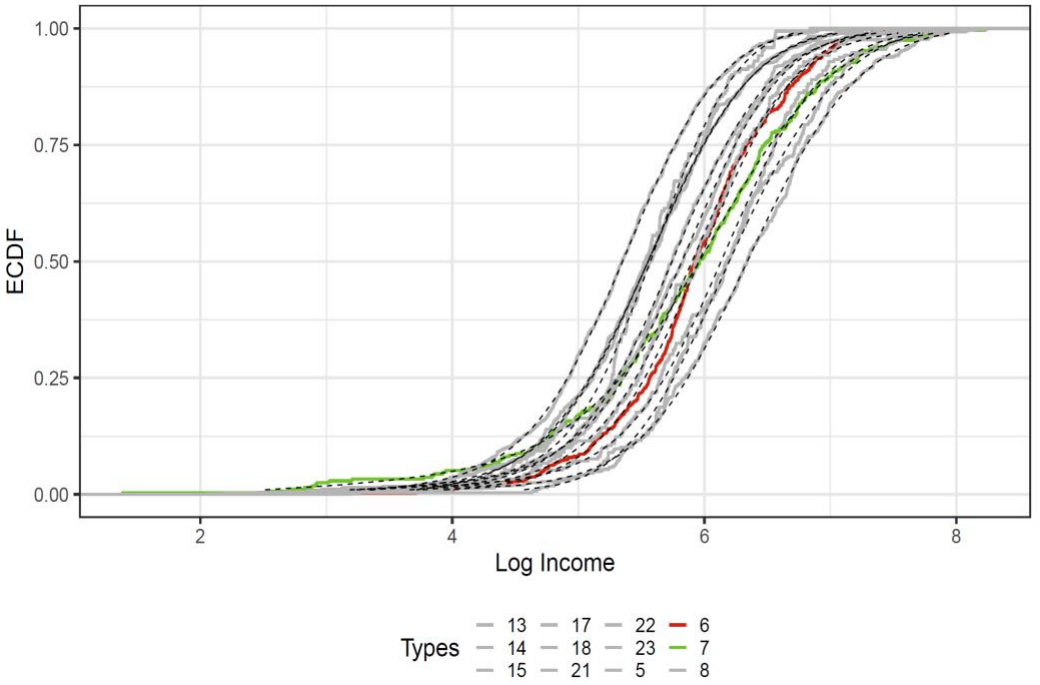
Despite these differences, we have seen in Sections 5 and 6 that IOp Gini coefficients across ex-ante trees and forests and ex-post trees are closely correlated. They measure the shares of the inequality observed in one generation that can be explained by factors inherited from the previous generation: a good measure of how strong the transmission of socioeconomic status is across generations in Latin America. As noted in Section 2, they are conceptually equivalent to the squares of intergenerational correlation coefficients for income or education. The main difference is that the latter estimates, commonly associated with intergenerational mobility studies, use a single variable to proxy for all factors inherited by the previous generation.

Figure 4: Type transitions between ex-ante and ex-post trees: the case of Argentina

Panel A



Panel B



Source: Authors' elaboration using data from Argentina's ENES 2014.

In Table 7 below we reproduce the ex-ante and ex-post tree-based Gini shares from Tables 3 and 5, in columns 1 and 2. Column 3 reports the squares of the correlation coefficients between the years of schooling of fathers and sons, computed from data in the same surveys. In column 4 we report approximations to correlation coefficients between parental and child income, which would be implied by the regression coefficients from some of the income mobility studies surveyed in Section 2. These studies typically do not report correlation coefficient estimates, so we approximate them by assuming that the variance of log incomes is the same in the parents' and children's generations. If that were the case, that slope coefficient would be equal to the correlation coefficient, which we then square and report in column 4 below.[31] The estimates are for specific years listed in footnote 31, and not for the survey years listed in the table. They are entered in the first row corresponding to the country they are from. Grawe's (2004) estimate for Ecuador in 1994 is greater than 1, which is clearly suspect. When reporting averages for the last column, we therefore also include (in brackets) an estimate excluding that observation.

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

Country/year	Relative Gini: ex-ante tree	Relative Gini: ex-post tree	Education ρ^2	Approx Income ρ^2
Argentina_2014	0.4377	0.4428	0.239	0.591
Bolivia_2008	0.4450	0.5840		
Brazil_2014	0.5889	0.5779	0.253	0.302
Chile_2006	0.5248	0.4936	0.248	0.325
Chile_2009	0.5492	0.4629	0.283	
Chile_2011	0.5613	0.4935	0.321	
Chile_2015	0.5259	0.4723	0.312	
Colombia_2010	0.5071	0.5221	0.200	
Ecuador_2006	0.4402	0.4774	0.370	1.277
Ecuador_2014	0.5445	0.5766	0.314	
Guatemala_2000	0.4422	0.4732	0.354	
Guatemala_2006	0.5421	0.5383	0.354	
Guatemala_2011	0.6185	0.6290	0.286	
Panama_2003	0.5104	0.4694	0.331	
Panama_2008	0.5521	0.5701	0.337	0.078

[31] The studies from which we draw estimates of the Galtonian income regression coefficient are: Jiménez (2016), who estimated the intergenerational elasticity (IGE) for Argentina in 2010 to be 0.769; Doruk et al. (2022) who estimated the IGE for Panama in 2010 to be 0.28; Núñez and Miranda (2010) who estimated the IGE for Chile in 2006 to be 0.57. Grawe (2004) estimated the IGE for Ecuador in 1994 to be 1.13 and for Peru in 1985 to be 0.67. For Brazil, Britto et al. (2022) estimated income-rank correlations using tax data and imputing informal income, finding a rank-rank correlation estimate of 0.55 for the 1988-90 cohort.

Peru_2001	0.5135	0.5721	0.311	0.449
Peru_2005	0.5485	0.5389	0.304	
Peru_2006	0.6038	0.5980	0.298	
Peru_2007	0.5731	0.5678	0.285	
Peru_2008	0.5551	0.5733	0.294	
Peru_2009	0.5337	0.5532	0.284	
Peru_2010	0.5039	0.5299	0.292	
Peru_2011	0.5068	0.5163	0.271	
Peru_2012	0.4937	0.5330	0.265	
Peru_2013	0.5021	0.5397	0.283	
Peru_2014	0.5128	0.5515	0.285	
Peru_2015	0.5122	0.5779	0.273	
Average	**0.524**	**0.535**	**0.294**	**0.504 (0.345)**

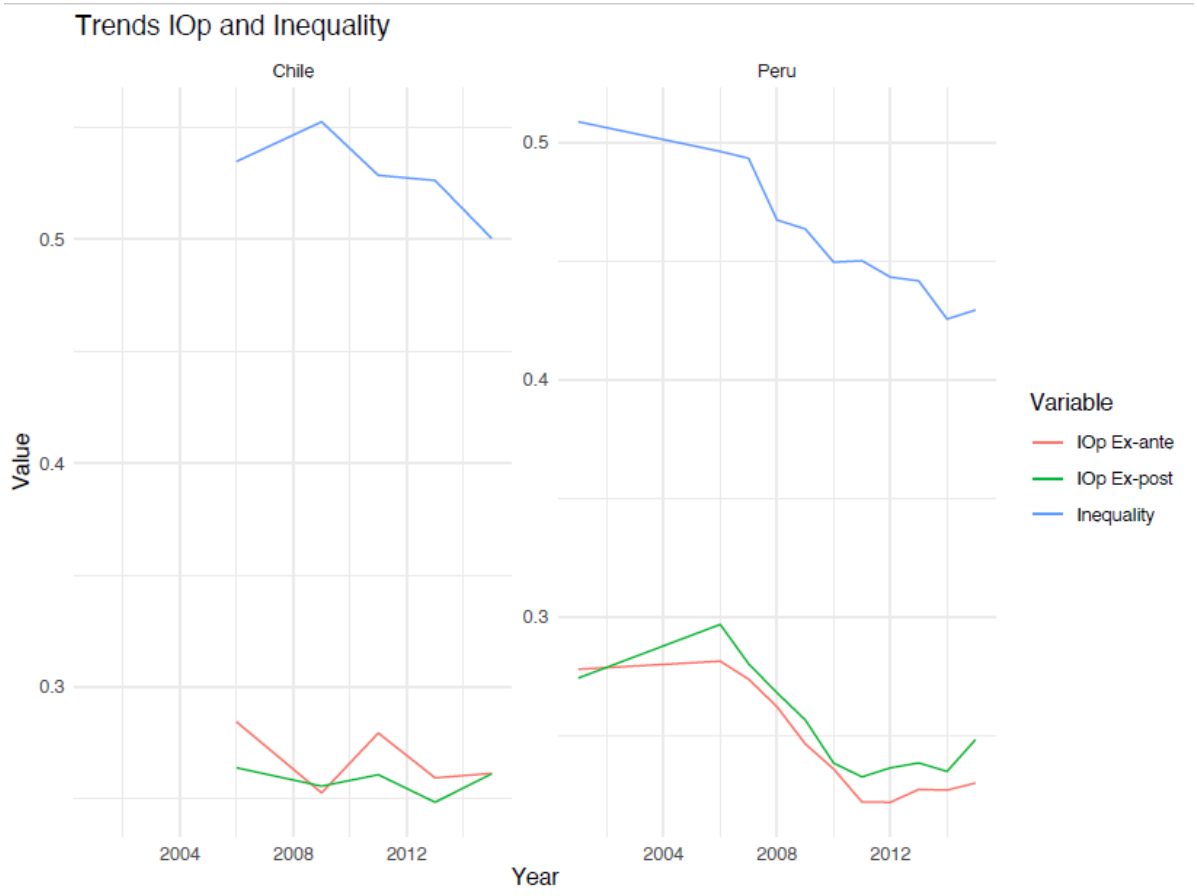
Three conclusions arise from the results summarized in Table 7. First, the transmission of socioeconomic status across generations in Latin America is remarkably strong. This also implies that inequality is highly persistent across generations. Columns 1 and 2 indicate that, on average, over half of all income inequality observed across our twenty-seven surveys can be accounted for by variation associated with inherited, pre-determined circumstances such as sex, race, birthplace, and family background. This result holds whether one takes an ex-ante or an ex-post view of inequality of opportunity; that is, whether partitions of the population are chosen so as maximize differences between type means or full type-specific quantile functions.

Second, these two approaches do not yield exactly the same results – because differences in higher moments of the type distributions matter in the ex-post case – but nor are they at complete loggerheads. As noted earlier, the Pearson correlation coefficient between the ex-ante and ex-post tree-based absolute IOp Gini estimates is 0.81. For the relative series reported in columns 1 and 2 of Table 7, it is 0.56. Brazil and Guatemala – subject to some temporal variation – are high IOp countries by both criteria, while Argentina is consistently at the bottom of the table in both cases. Some countries, like Bolivia, do perform quite differently across the two approaches. But overall, a consistent and striking picture emerges: inequality of opportunity – that is, inherited inequality – defined on the basis of income differences between as few as ten and no more than 32 population subgroups, accounts for no less than 43% - and as much as 63% - of all inequality measured in Latin America.

For those countries with enough surveys available over time, such as Chile and Peru, there is consistency in terms of trends, as well as levels. Figure 5 below plots three Gini measures of inequality over time for those two countries: the blue line shows overall income inequality, while the orange and green lines show our absolute ex-ante and ex-post IOp estimates, respectively. The two IOp estimates move in tandem in both cases. In Peru, overall inequality declines throughout the period, and quite markedly between 2007 and 2010. All of that decline is accounted for by falling inequality of opportunity. In Chile, by contrast, the

decline between 2009 and 2015 is not mirrored by a matching reduction in inequality of opportunity and must thus have been driven by income differences within types.

Figure 5 Inequality and IOp dynamics over time in Chile and Peru

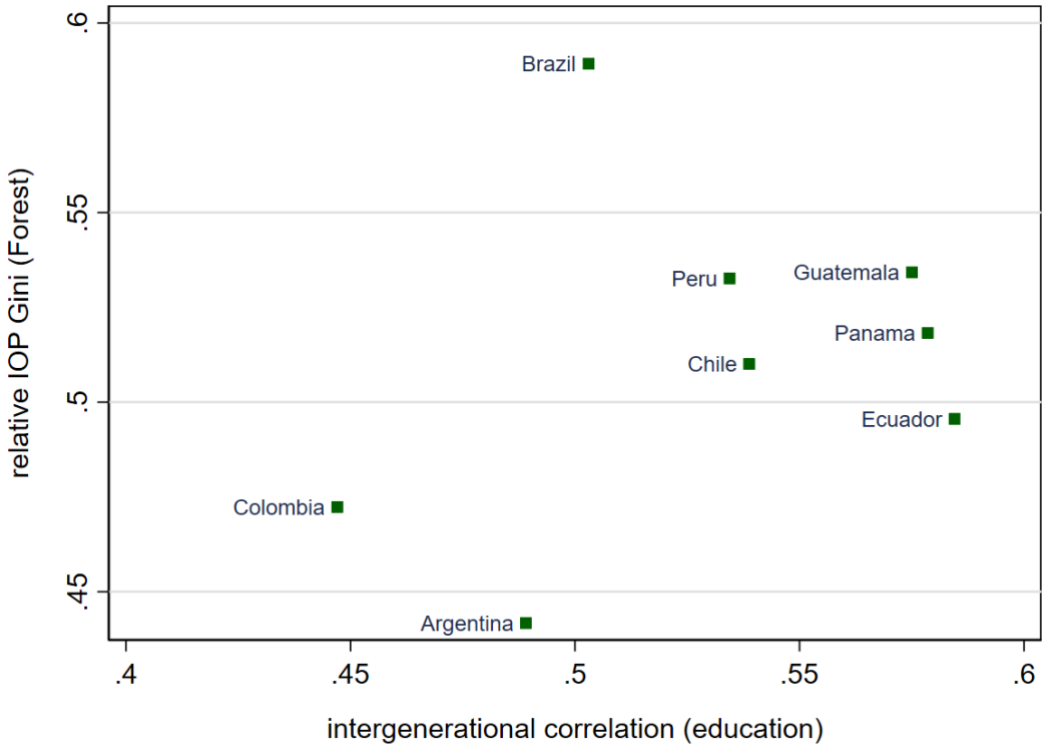


Source: Authors' elaboration using data from Chile's CASEN and Peru's ENAHO surveys, various years.

Third, these results for age-adjusted equivalized income, using a set of family background and other inherited circumstances, are considerably higher than those implied by the measures of educational persistence reported in column 3 for the same surveys. These squared correlation coefficients are measures of the share of the variance in years of schooling of the children's generation accounted for by years of schooling in the parents' generation, and thus conceptually comparable to the shares reported in columns 1 and 2. These numbers average to 29%, considerably lower than the IOp measures. This may reflect three different factors: (i) education is not income, and persistence in the two measures is driven by different mechanisms and can vary substantially; (ii) the variance and the Gini coefficient are not the same measure of dispersion, and this too can make a difference; and (iii) the inclusion of multiple circumstances in the IOp calculations captures more of the sources of socio-economic persistence than parental education alone.

Nevertheless, the two measures are also positively correlated. Figure 6 plots the relative ex-ante IOp Gini coefficient (averaged across surveys over time) for our nine countries, against their education ρ^2s. Once again, the rough taxonomy of countries is consistent with what we have seen earlier: Brazil and Guatemala are high-persistence countries, as are Ecuador and Panama (particularly if more weight is placed on the educational results). Below that 'outer envelope', Chile and Peru appear as perhaps slightly less inequitable, while Argentina and Colombia provide a 'lower envelope' in terms of intergenerational persistence for our sample.

Figure 6: Relative ex-ante IOp from Random Forests and education ρ



Source: The coordinates are the education ρ estimates used to produce the squared correlation coefficients included in Table 7, and averages across all available years for IOp.

The squared income persistence measures shown in column 4 of Table 7 are even less comparable to the IOp estimates in columns 1 and 2. They are not derived from our surveys, years, or samples. They are drawn from different studies, for different years and using different methods. They are all based on estimates of intergenerational regression coefficients, and thus very roughly approximate correlation coefficients under the very strong (and indeed, almost certainly false) assumption that inequality in the marginal distributions was constant. They can therefore be taken as no more than very roughly indicative of orders of magnitude. Yet these too are lower than our IOp estimates; quite markedly so if the outlying estimate for Ecuador (1994) is excluded.

8. Conclusions

Building both on a review of the literature on intergenerational mobility and inequality of opportunity, as well as on new analysis of twenty-seven representative household surveys covering nine Latin American countries over the 2000-2015 period, we find that the intergenerational transmission of socioeconomic status in the region is extremely strong.

We use two data-driven approaches to obtain optimal partitions of the population into types, in well-defined statistical senses. The conditional inference tree and forest approach closely corresponds to the ex-ante definition of inequality of opportunity, while the transformation tree approach corresponds to the ex-post definition. Although the conceptual differences do yield different trees, type partitions and insights, they largely agree on the overall share of current inequality that is accounted for by inherited factors: 52-54% on average, ranging from 44% in Argentina to around 62-63% in Guatemala.

Descriptively, family background variables such as parental education and occupation account for the lion's share of the process of inequality reproduction. The geography and ethnicity of one's birth also matter considerably, particularly among families from lower socioeconomic backgrounds. Ethnicity is markedly more important in countries with a historical legacy of conquest of large indigenous populations (such as Bolivia and Guatemala) or of slavery, such as Brazil. Using household equivalized incomes as our welfare concept - and therefore ignoring all intrahousehold inequality - biological sex is relatively unimportant everywhere.

Overall, our study found that individuals with "better" family backgrounds have significantly better outcomes than those from low socioeconomic status families, with ethnicity and birth area becoming more significant factors among the latter group. The rural-urban divide is also important, but the greatest variation is found within urban areas. Ethnicity matters too, and there is some evidence of variation over time. For instance, the Shapley decompositions suggest that in Chile and Peru, the two countries with the largest time horizon available in our analysis, the role of ethnicity in shaping the income distribution has increased over time. As inspection of the trees reveals, inequalities between ethnic groups tend to appear at the bottom of the distribution, among the types with the lowest levels of income, more often than at the top – although there are exceptions.

Methodologically, our findings suggest that, at least in the absence of detailed data linking objectively measured incomes across generations, and for more than one period in each, it is worth exploring other circumstance variables that are more widely available – and perhaps more accurately measured – in the kinds of data frequently available to analysts. When doing so, one should avoid ad-hoc and arbitrary partitions of the population, which are always susceptible to different combinations of (downward) omitted variable and (upward) overfitting biases. Data-driven approaches such as those used here can be argued to strike the right statistical balance, while remaining true to the theoretical concepts of inequality of opportunity that one is trying to estimate.

References

Ahsan, Md. Nazmul; Emran, M. Shahe; Jiang, Hanchen; Han, Qingyang; & Shilpi, Forhad J. (2023). 'Growing Up Together: Sibling Correlation, Parental Influence, and Intergenerational Educational Mobility in Developing Countries'. World Bank Policy Research Working Paper No. WPS 10285.

Andersen, Lykke E. 2003. 'Social Mobility in Latin America: Links with Adolescent Schooling'. in *Critical Decisions at a Critical Age: Adolescents and Young Adults in Latin America.*, edited by S. Duryea, A. Cox Edwards, and M. Ureta. Inter-American Development Bank.

Anderson, Lewis, Paula Sheppard, and Christiaan Monden. 2018. 'Grandparent Effects on Educational Outcomes: A Systematic Review'. *Sociological Science* 5:114–42. doi: 10.15195/v5.a6.

Beccaria, Luis, Roxana Maurizio, Martin Trombetta, and Gustavo Vázquez. 2022. 'Short-Term Income Mobility in Latin America in the 2000s: Intensity and Characteristics'. *Socio-Economic Review* 20(3):1039–67. doi: 10.1093/ser/mwaa043.

Behrman, Jere, Nancy Birdsall, and Miguel Székely. 1999. 'Intergenerational Mobility in Latin America: Deeper Markets and Better Schools Make a Difference'. *Carnegie Endowment for International Peace, Global Policy Program* Vol. 3.

Behrman, Jere R., Alejandro Gaviria, Miguel Székely, Nancy Birdsall, and Sebastián Galiani. 2001. 'Intergenerational Mobility in Latin America [with Comments]'. *Economía* 2(1):1–44.

Behrman, Jere R., and Barbara L. Wolfe. 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):395–419. doi: 10.1016/0304-3878(87)90024-1.

Binder, Melissa, and Christopher Woodruff. 2002. 'Inequality and Intergenerational Mobility in Schooling: The Case of Mexico'. *Economic Development and Cultural Change* 50(2):249–67. doi: 10.1086/322882.

Birdsall, Nancy, Jere R. Behrman, and Miguel Székely. 1998. 'Intergenerational Schooling Mobility and Macro Conditions and Schooling Policies in Latin America'. *SSRN Scholarly Paper*. doi: 10.2139/ssrn.1817183.

Bloise, Francesco, Paolo Brunori, and Patrizio Piraino. 2021. 'Estimating Intergenerational Income Mobility on Sub-Optimal Data: A Machine Learning Approach'. *The Journal of Economic Inequality* 19(4):643–65. doi: 10.1007/s10888-021-09495-6.

Bourguignon, François, Francisco H. G. Ferreira, and Marta Menéndez. 2007. 'Inequality of Opportunity in Brazil'. *Review of Income and Wealth* 53(4):585–618. doi: 10.1111/j.1475-4991.2007.00247.x.

Britto, Diogo G. C., Alexandre de Andrade Fonseca, Paolo Pinotti, Breno Sampaio, and Lucas Warwar. 2022. 'Intergenerational Mobility in the Land of Inequality'. *SSRN Scholarly Paper*. doi: 10.2139/ssrn.4237631.

Brunori, Paolo, Francisco H. G. Ferreira, and Guido Neidhöfer. 2023. 'Inequality of Opportunity and Intergenerational Persistence in Latin America', UNU-WIDER Working Paper 2023/39.

Brunori, Paolo, Francisco H. G. Ferreira, and Vito Peragine. 2013. 'Inequality of Opportunity, Income Inequality, and Economic Mobility: Some International Comparisons'. in *Getting Development Right: Structural Transformation, Inclusion, and Sustainability in the Post-Crisis Era*, edited by E. Paus. New York: Palgrave Macmillan US.

Brunori, Paolo, Francisco H. G. Ferreira, and Pedro Salas-Rojo. 2023. 'Inherited Inequality: A general framework and an application to South Africa'. *LSE III Working Paper 107.*

Brunori, Paolo, Paul Hufe, and Daniel Gerszon Mahler. 2022. 'The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees'. *The Scandinavian Journal of Economics*, First published: 20 February 2023 doi: 10.1111/sjoe.12530

Brunori, Paolo, Flaviana Palmisano, and Vitorocco Peragine. 2019. 'Inequality of Opportunity in Sub-Saharan Africa'. *Applied Economics* 51(60):6428–58. doi: 10.1080/00036846.2019.1619018.

Brunori, Paolo, Vito Peragine, and Laura Serlenga. 2019. 'Upward and Downward Bias When Measuring Inequality of Opportunity'. *Social Choice and Welfare* 52(4):635–61. doi: 10.1007/s00355-018-1165-x.

Celhay, Pablo and Sebastian Gallegos. 2023. 'Educational Mobility Across Three Generations in Latin American Countries'. *CAF Working Papers*.

Celhay, Pablo, and Sebastián Gallegos. 2015. 'Persistence in the Transmission of Education: Evidence across Three Generations for Chile'. *Journal of Human Development and Capabilities* 16(3):420–51. doi: 10.1080/19452829.2015.1048789.

Checchi, Daniele, and Vito Peragine. 2010. 'Inequality of Opportunity in Italy'. *The Journal of Economic Inequality* 8(4):429–50. doi: 10.1007/s10888-009-9118-3.

Ciaschi, Matías, Mariana Marchionni, and Guido Neidhöfer. 2023. 'Intergenerational Mobility in Latin America: The Multiple Facets of Social Status and the Role of Mothers'. *CEDLAS Working Paper*.

Corak, Miles. 2013. 'Income Inequality, Equality of Opportunity, and Intergenerational Mobility'. *Journal of Economic Perspectives* 27(3):79–102. doi: 10.1257/jep.27.3.79.

Corak, Miles, and Patrizio Piraino. 2011. 'The Intergenerational Transmission of Employers'. *Journal of Labor Economics* 29(1):37–68. doi: 10.1086/656371.

Cuesta, Jose, Hugo Ñopo, and Georgina Pizzolitto. 2011. 'Using Pseudo-Panels to Measure Income Mobility in Latin America'. *Review of Income and Wealth* 57(2):224–46. doi: 10.1111/j.1475-4991.2011.00444.x.

Dahan, Momi, and Alejandro Gaviria. 2001. 'Sibling Correlations and Intergenerational Mobility in Latin America'. *Economic Development and Cultural Change* 49(3):537–54. doi: 10.1086/452514.

Daude, Christian, and Virginia Robano. 2015. 'On Intergenerational (Im)Mobility in Latin America'. *Latin American Economic Review* 24(1):9. doi: 10.1007/s40503-015-0030-x.

Daza Báez, Nancy. 2021. 'Intergenerational Earnings Mobility in Mexico'. *UCL DoQSS Working Paper*.

DiPrete, Thomas A. 2020. 'The Impact of Inequality on Intergenerational Mobility'. *Annual Review of Sociology* 46(1):379–98. doi: 10.1146/annurev-soc-121919-054814.

Doruk, Ömer Tuğsal, Francesco Pastore, and Hasan Bilgehan Yavuz. 2022. 'Intergenerational Mobility: An Assessment for Latin American Countries'. *Structural Change and Economic Dynamics* 60:141–57. doi: 10.1016/j.strueco.2021.11.005.

Dunn, Christopher E. 2007. 'The Intergenerational Transmission of Lifetime Earnings: Evidence from Brazil'. *The B.E. Journal of Economic Analysis & Policy* 7(2). doi: 10.2202/1935-1682.1782.

Durlauf, Steven N., Andros Kourtellos, and Chih Ming Tan. 2022. 'The Great Gatsby Curve'. *Annual Review of Economics* 14(1):571–605. doi: 10.1146/annurev-economics-082321-122703.

Emran, M. Shahe, and Forhad Jahan Shilpi. 2019. *Economic Approach to Intergenerational Mobility: Measures, Methods, and Challenges in Developing Countries*. *Working Paper*. 2019/98. WIDER Working Paper. doi: 10.35188/UNU-WIDER/2019/734-7.

Ferreira, Francisco H. G., and Jérémie Gignoux. 2011. 'The Measurement of Inequality of Opportunity: Theory and an Application to Latin America'. *Review of Income and Wealth* 57(4):622–57. doi: 10.1111/j.1475-4991.2011.00467.x.

Ferreira, Francisco H. G., and Jérémie Gignoux. 2014. 'The Measurement of Educational Inequality: Achievement and Opportunity1'. *The World Bank Economic Review* 28(2):210–46. doi: 10.1093/wber/lht004.

Ferreira, Francisco H. G., Julian Messina, Jamele Rigolini, Luis-Felipe López-Calva, Maria Ana Lugo, and Renos Vakis. 2013. *Economic Mobility and the Rise of the Latin American Middle Class*. Washington, DC: World Bank.

Ferreira, Francisco H. G., and Vito Peragine. 2016. 'Individual Responsibility and Equality of Opportunity'. P. 0 in *The Oxford Handbook of Well-Being and Public Policy*, edited by M. D. Adler and M. Fleurbaey. Oxford University Press.

Ferreira, Sergio Guimarães, and Fernando A. Veloso. 2006. 'Intergenerational Mobility of Wages in Brazil'. *Brazilian Review of Econometrics* 26(2):181–211. doi: 10.12660/bre.v26n22006.1576.

Fields, Gary S. 2000. 'Income Mobility: Concepts and Measures'. in *New Markets, New Opportunities? Economic and Social Mobility in a Changing World*, edited by N. Birdsall and C. L. Graham. Brookings Institution Press.

Fields, Gary S., Robert Duval Hernández, Samuel Freije, María Laura Sánchez Puerta, Omar Arias, and Juliano Assunção. 2007. 'Intragenerational Income Mobility in Latin America [with Comments]'. *Economía* 7(2):101–54.

Fleurbaey, Marc, and Vito Peragine. 2013. 'Ex Ante Versus Ex Post Equality of Opportunity'. *Economica* 80(317):118–30. doi: 10.1111/j.1468-0335.2012.00941.x.

Friedman, Milton, and Rose D. Friedman. 1962. *Capitalism and Freedom*. University of Chicago Press.

Gabrielli, M. V. 2022. 'Perceived Intergenerational Mobility: New Evidence from Latin America'. Paris School of Economics.

Gamboa, Luis Fernando, and Fábio D. Waltenberg. 2012. 'Inequality of Opportunity for Educational Achievement in Latin America: Evidence from PISA 2006–2009'. *Economics of Education Review* 31(5):694–708. doi: 10.1016/j.econedurev.2012.05.002.

Grawe, Nathan D. 2004. 'Intergenerational Mobility for Whom? The Experience of High- and Low-Earning Sons in International Perspective'. Pp. 58–89 in *Generational Income Mobility in North America and Europe*, edited by M. Corak. Cambridge: Cambridge University Press.

Heckman, James J., and V. Joseph Hotz. 1986. 'An Investigation of the Labor Market Earnings of Panamanian Males Evaluating the Sources of Inequality'. *The Journal of Human Resources* 21(4):507–42. doi: 10.2307/145765.

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina. 2008. 'The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends'. *The B.E. Journal of Economic Analysis & Policy* 7(2). doi: 10.2202/1935-1682.1775.

Hothorn, Torsten, Kurt Hornik, and Achim Zeileis. 2006. 'Unbiased Recursive Partitioning: A Conditional Inference Framework'. *Journal of Computational and Graphical Statistics* 15(3):651–74. doi: 10.1198/106186006X133933.

Hothorn, Torsten, and Achim Zeileis. 2021. 'Predictive Distribution Modeling Using Transformation Forests'. *Journal of Computational and Graphical Statistics* 30(4):1181–96. doi: 10.1080/10618600.2021.1872581.

Jäntti, Markus, and Stephen P. Jenkins. 2015. 'Chapter 10 - Income Mobility'. in *Handbook of Income Distribution*. Vol. 2, *Handbook of Income Distribution*, edited by A. B. Atkinson and F. Bourguignon. Elsevier.

Jiménez, Maribel. 2016. 'Movilidad intergeneracional del ingreso en Argentina. Un análisis de sus cambios temporales desde el enfoque de igualdad de oportunidades'. *CEDLAS Working paper*.

Lam, David, and Robert F. Schoeni. 1993. 'Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil'. *Journal of Political Economy* 101(4):710–40.

Leites, Martín, Xavier Ramos, Cecilia Rodríguez, and Joan Vilá. 2022. 'Intergenerational Mobility along the Income Distribution: Estimates Using Administrative Data for a Developing Country'. *Instituto de Economía, Facultad de Ciencias Económicas y Administración, Universidad de La República, Uruguay. Working Paper*.

Lubotsky, Darren, and Martin Wittenberg. 2006. 'Interpretation of Regressions with Multiple Proxies'. *The Review of Economics and Statistics* 88(3):549–62.

Marteleto, Letícia, Denisse Gelber, Celia Hubert, and Viviana Salinas. 2012. 'Educational Inequalities among Latin American Adolescents: Continuities and Changes over the 1980s, 1990s and 2000s'. *Research in Social Stratification and Mobility* 30(3):352–75. doi: 10.1016/j.rssm.2011.12.003.

Moreno, Hector. 2021. 'The Influence of Parental and Grandparental Education in the Transmission of Human Capital'. *ECINEQ Working Paper 2021/588*.

Munoz, Ercio. 2021. 'The Geography of Intergenerational Mobility in Latin America and the Caribbean'. *SSRN Scholarly Paper*. doi: 10.2139/ssrn.3807350.

Narayan, Ambar, Roy Van der Weide, Alexandru Cojocaru, Christoph Lakner, Silvia Redaelli, Daniel Gerszon Mahler, Rakesh Gupta N. Ramasubbaiah, and Stefan Thewissen. 2018. *Fair Progress?: Economic Mobility Across Generations Around the World*. Washington, DC: World Bank.

Neidhöfer, Guido. 2019. 'Intergenerational Mobility and the Rise and Fall of Inequality: Lessons from Latin America'. *The Journal of Economic Inequality* 17(4):499–520. doi: 10.1007/s10888-019-09415-9.

Neidhöfer, Guido, Matías Ciaschi, and Leonardo Gasparini. 2022. 'Intergenerational Mobility of Economic Well-Being in Latin America'. *CEDLAS Working Paper*.

Neidhöfer, Guido, Matías Ciaschi, Leonardo Gasparini, and Joaquín Serrano. 2023. 'Social Mobility and Economic Development'. *Journal of Economic Growth, 1-33*. doi: 10.1007/s10887-023-09234-8

Neidhöfer, Guido, Joaquín Serrano, and Leonardo Gasparini. 2018. 'Educational Inequality and Intergenerational Mobility in Latin America: A New Database'. *Journal of Development Economics* 134:329–49. doi: 10.1016/j.jdeveco.2018.05.016.

Nunez, Javier I., and Leslie Miranda. 2010. 'Intergenerational Income Mobility in a Less-Developed, High-Inequality Context: The Case of Chile'. *The B.E. Journal of Economic Analysis & Policy* 10(1). doi: 10.2202/1935-1682.2339.

Núñez, Javier, and Andrea Tartakowsky. 2011. 'The Relationship between Income Inequality and Inequality of Opportunities in a High-Inequality Country: The Case of Chile'. *Applied Economics Letters* 18(4):359–69. doi: 10.1080/13504851003636172.

Olivetti, Claudia, and M. Daniele Paserman. 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. doi: 10.1257/aer.20130821.

Paes de Barros, Ricardo, Francisco H. G. Ferreira, Jose R. Molinas Vega, and Jaime Saavedra Chanduvi. 2009. *Measuring Inequality of Opportunities in Latin America and the Caribbean*. Washington, DC: World Bank.

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

Santavirta, Torsten, and Jan Stuhler. 2020. 'Name-Based Estimators of Intergenerational Mobility: Evidence from Finnish Veterans'. *Stockholm University, Mimeo*.

Shapley, Lloyd S. 1953. '17. A Value for N-Person Games'. Pp. 307–18 in *Contributions to the Theory of Games (AM-28)*. Vol. II, edited by H. W. Kuhn and A. W. Tucker. Princeton: Princeton University Press.

Shorrocks, Anthony F. 2013. 'Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value'. *The Journal of Economic Inequality* 11(1):99–126. doi: 10.1007/s10888-011-9214-z.

Solon, Gary. 2014. 'Theoretical Models of Inequality Transmission across Multiple Generations'. *Research in Social Stratification and Mobility* 35:13–18. doi: 10.1016/j.rssm.2013.09.005.

Torche, Florencia. 2010. 'Economic Crisis and Inequality of Educational Opportunity in Latin America'. *Sociology of Education* 83(2):85–110. doi: 10.1177/0038040710367935.

Torche, Florencia. 2014. 'Intergenerational Mobility and Inequality: The Latin American Case'. *Annual Review of Sociology* 40(1):619–42. doi: 10.1146/annurev-soc-071811-145521.

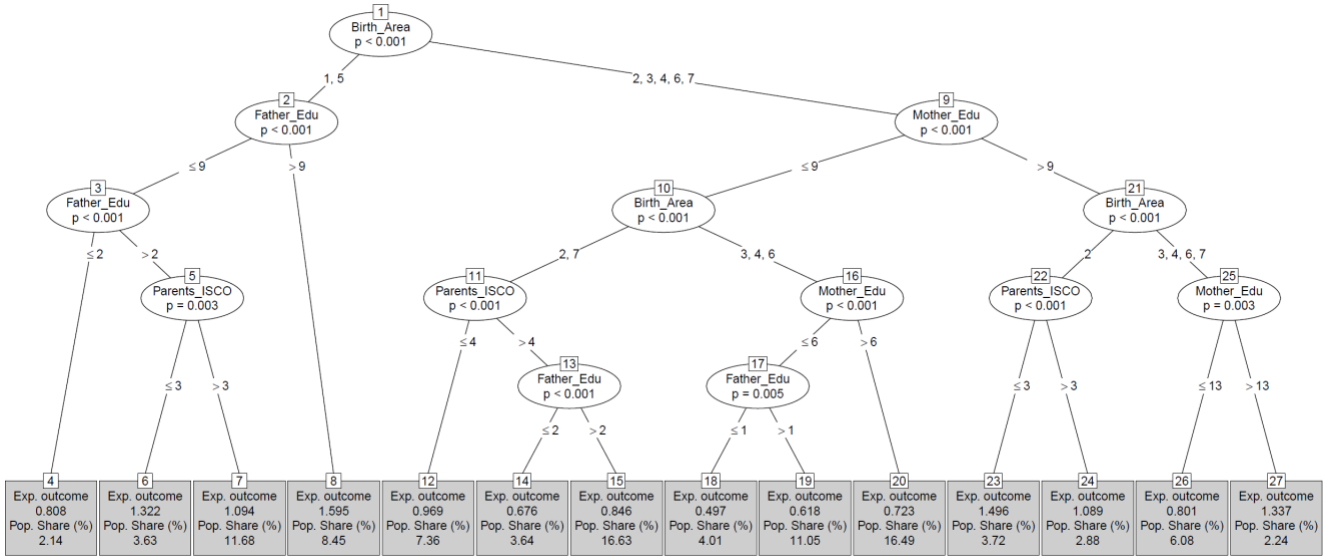
Torche, Florencia. 2021. 'Intergenerational Mobility in Latin America in Comparative Perspective'. *UNDP LAC Working Paper No. 02.*

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

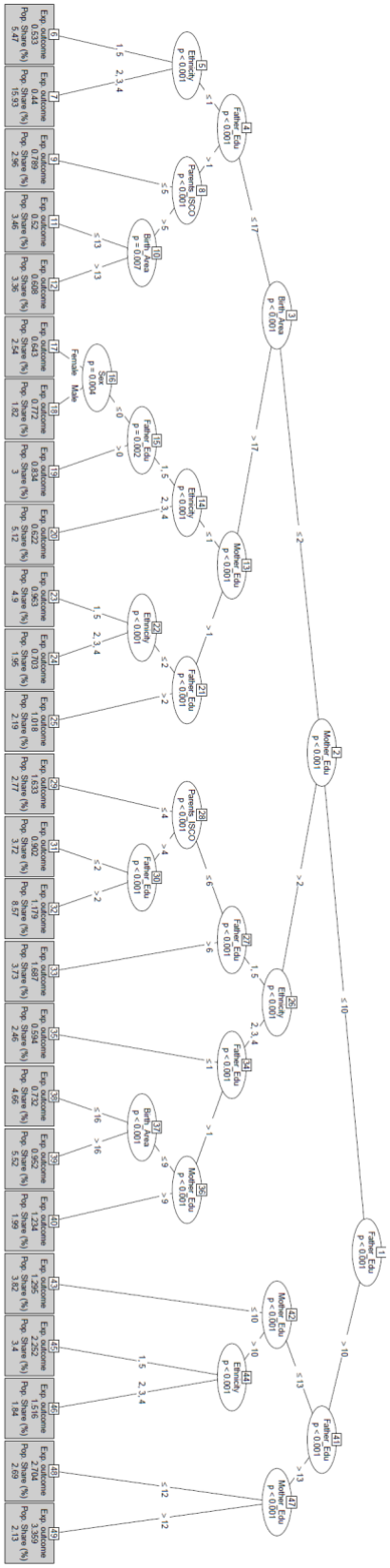
Van der Weide, Roy, Christoph Lakner, Daniel Gerszon Mahler, Ambar Narayan, and Rakesh Ramasubbaiah. 2024. 'Intergenerational Mobility Around the World'. *Journal of Development Economics*.

Appendix 1: Conditional inference (ex-ante) trees for the most recent surveys in eight Latin American countries.

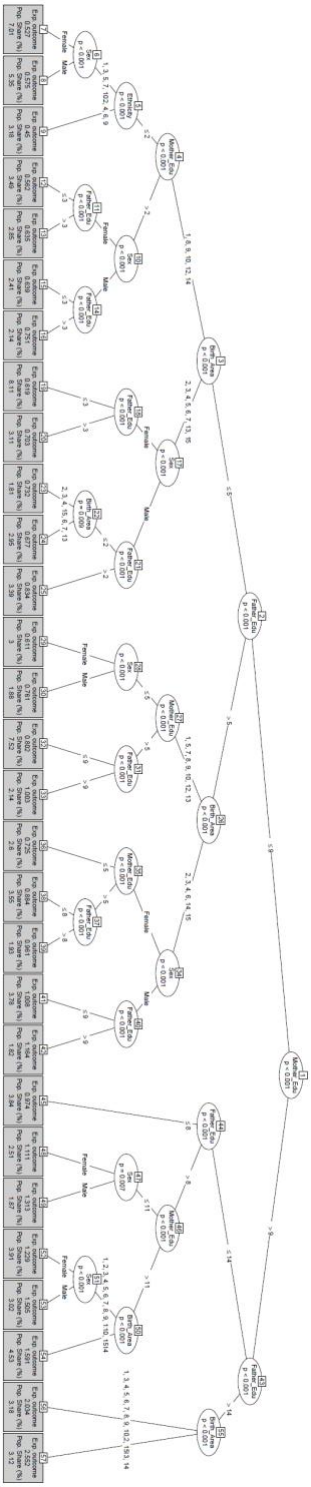
Argentina, 2014



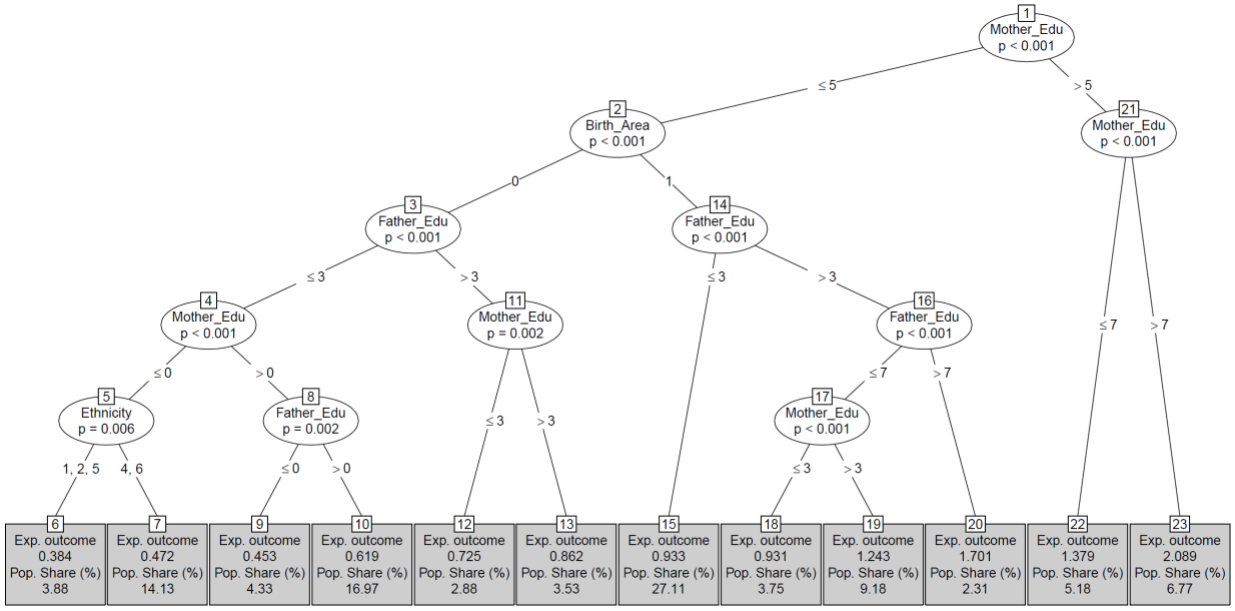
Brazil 2014



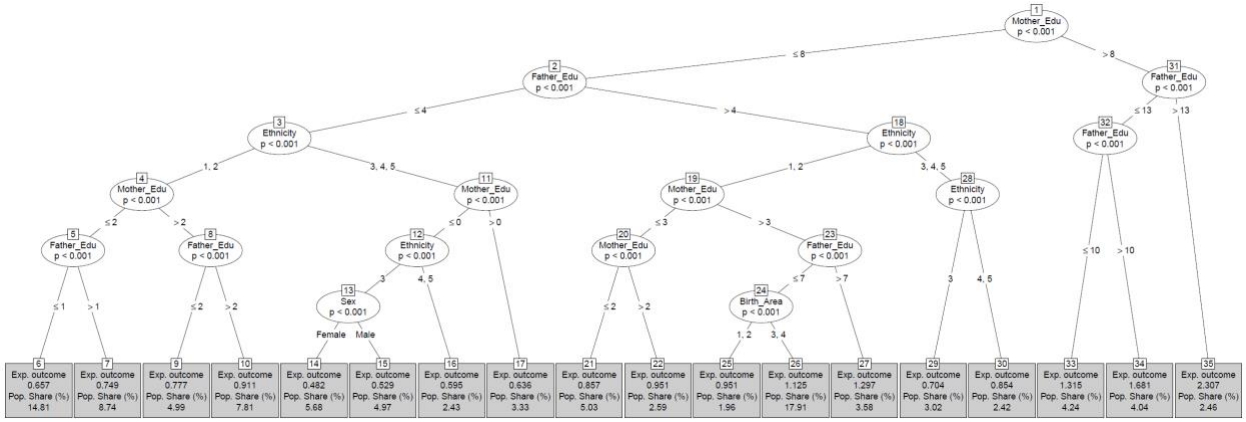
Chile 2015



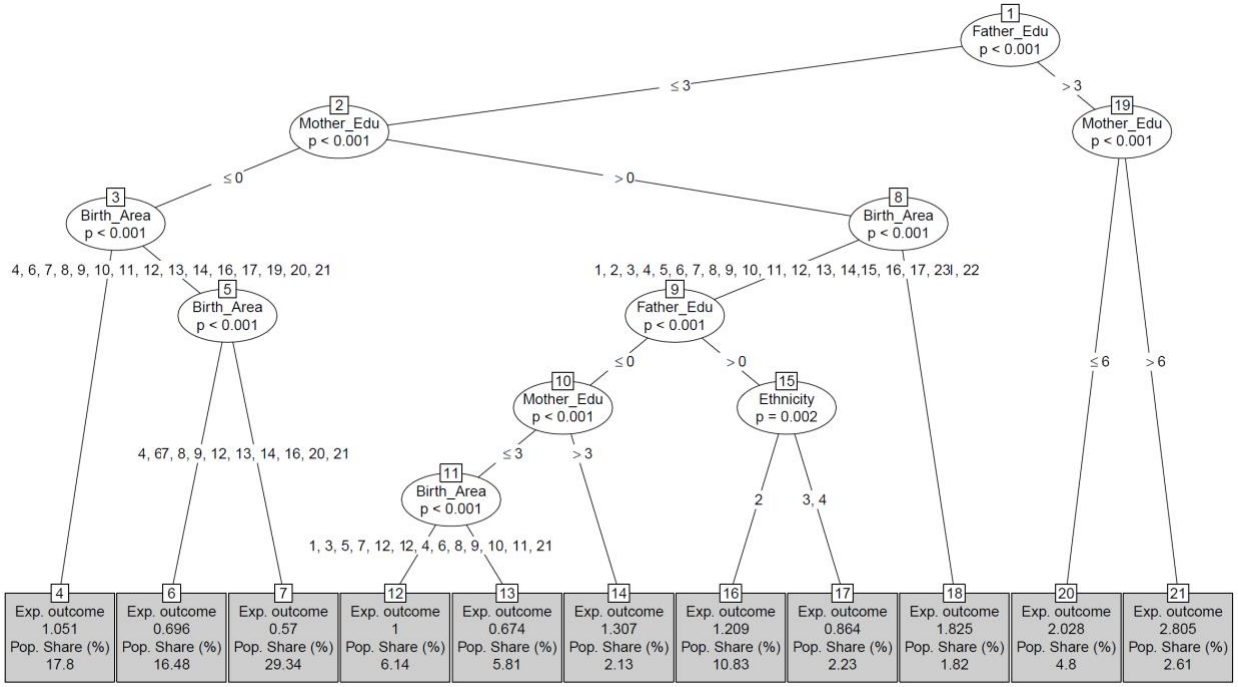
Colombia 2010



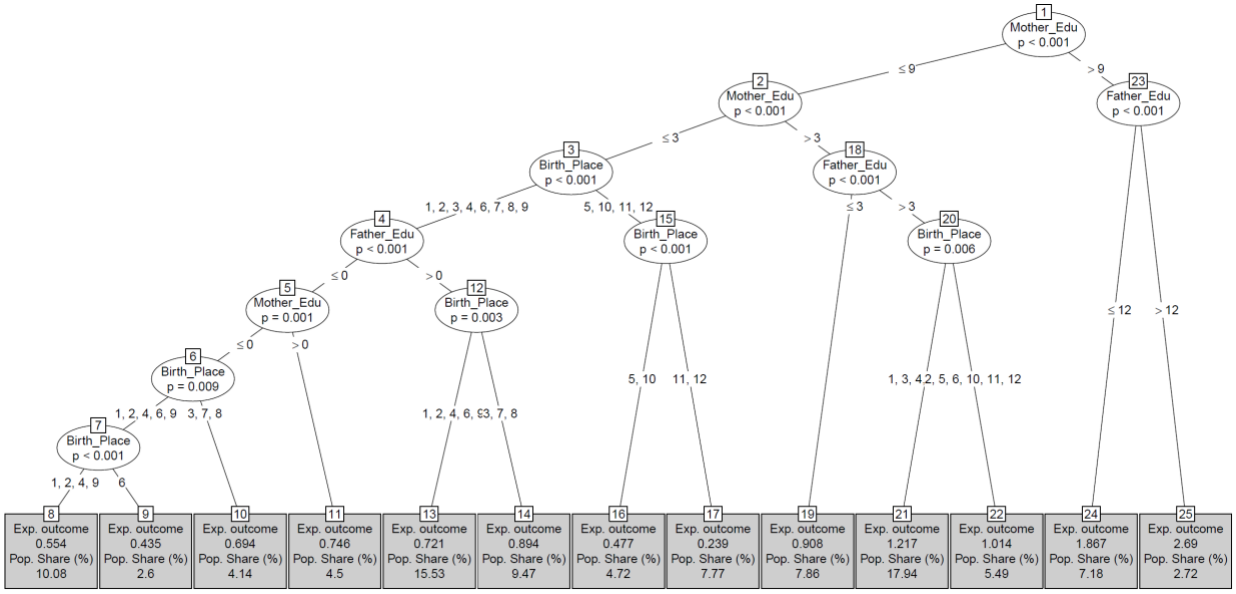
Ecuador 2014



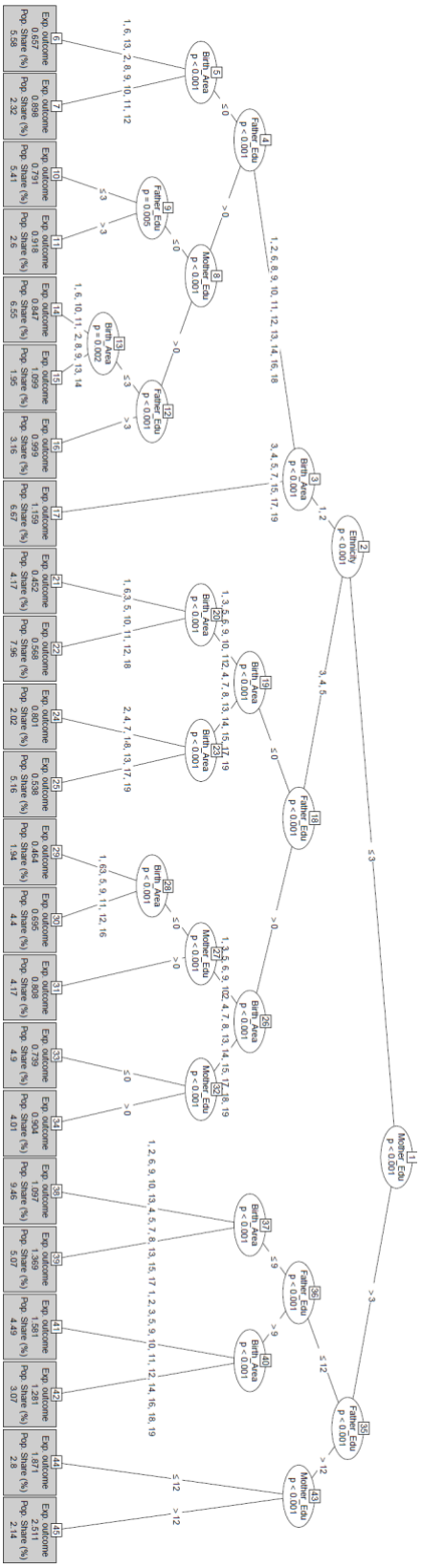
Guatemala 2011



Panama 2008



Peru 2015



Appendix 3: Ex-ante estimates using the mean logarithmic deviation

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

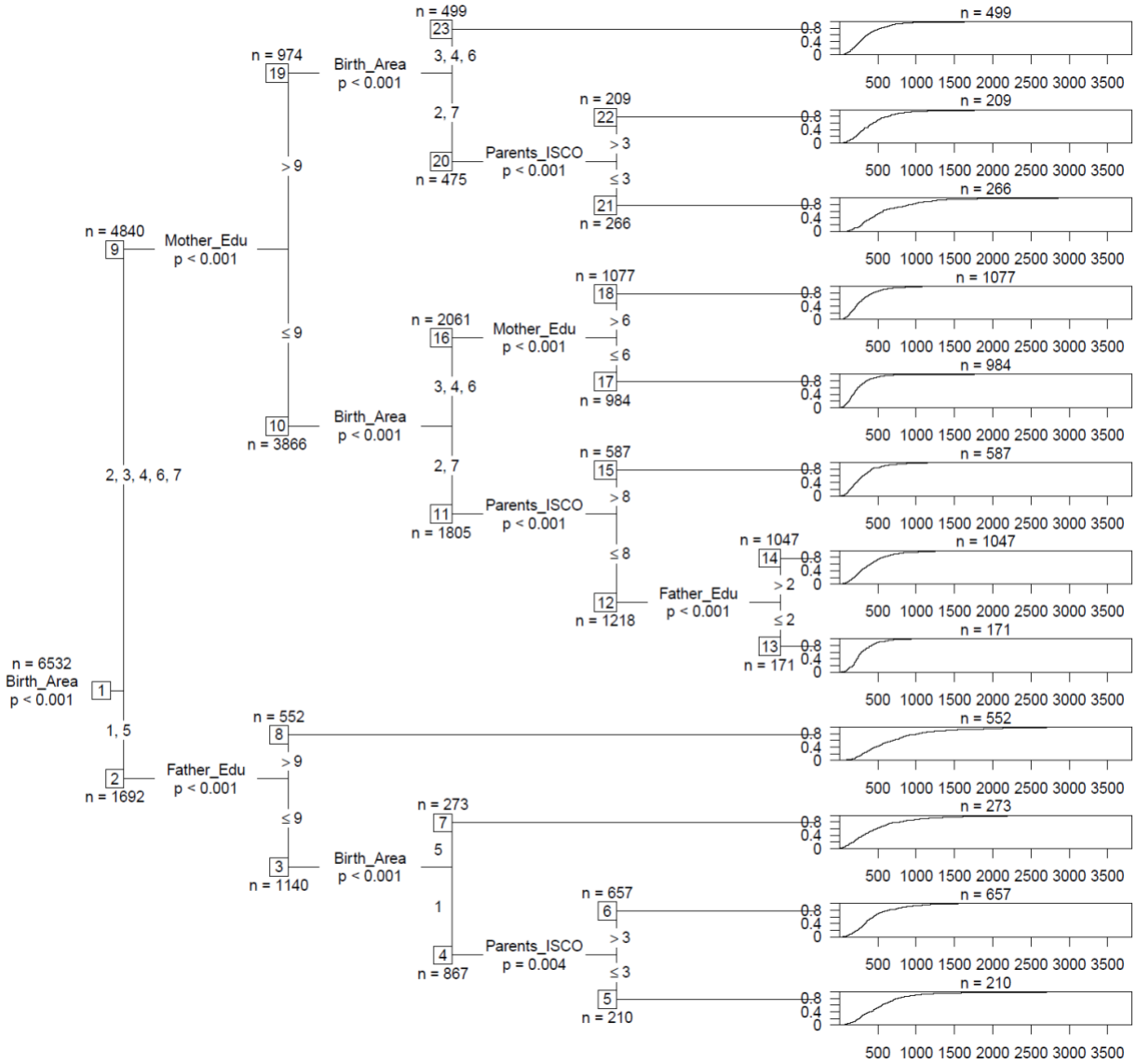
Country/Year	# Types	Total MLD	IOp MLD (Trees)	Relative IOp MLD (Trees)	IOp MLD (Forest)	Relative IOp MLD (Forest)
Argentina 2014	14	0.2812	0.0464	0.1650	0.0466	0.1657
Bolivia 2008	10	0.4766	0.0912	0.1913	0.1020	0.2140
Brazil 2014	25	0.4760	0.1452	0.3051	0.1444	0.3034
Chile 2006	21	0.5054	0.1251	0.2475	0.1271	0.2515
Chile 2009	28	0.5453	0.1448	0.2655	0.1010	0.1852
Chile 2011	27	0.4933	0.1386	0.2810	0.1219	0.2471
Chile 2013	32	0.4896	0.1192	0.2435	0.1043	0.2130
Chile 2015	29	0.4760	0.1001	0.2103	0.1058	0.2222
Colombia 2010	12	0.5974	0.1010	0.1691	0.1120	0.1875
Ecuador 2006	18	0.5600	0.1381	0.2466	0.1309	0.2337
Ecuador 2014	18	0.3895	0.0672	0.1725	0.0694	0.1782
Guatemala 2000	11	0.5641	0.1441	0.2555	0.1383	0.2452
Guatemala 2006	16	0.5412	0.1743	0.3221	0.1623	0.2999
Guatemala 2011	11	0.5252	0.1183	0.2253	0.0976	0.1858
Panama 2003	14	0.6540	0.1471	0.2249	0.1211	0.1852
Panama 2008	13	0.5373	0.1212	0.2256	0.1212	0.2256
Peru 2001	17	0.4864	0.1255	0.2580	0.1215	0.2498
Peru 2006	19	0.4528	0.1462	0.3228	0.1242	0.2743
Peru 2007	18	0.4548	0.1263	0.2777	0.1184	0.2604
Peru 2008	20	0.4064	0.1081	0.2660	0.1082	0.2662
Peru 2009	21	0.3984	0.0969	0.2432	0.0955	0.2397
Peru 2010	17	0.3700	0.0803	0.2170	0.0863	0.2332
Peru 2011	17	0.3724	0.0809	0.2172	0.0765	0.2054
Peru 2012	17	0.3628	0.0749	0.2064	0.0765	0.2108
Peru 2013	23	0.3559	0.0775	0.2178	0.0803	0.2256
Peru 2014	21	0.3296	0.0756	0.2294	0.0798	0.2421
Peru 2015	23	0.3338	0.0769	0.2303	0.0821	0.2459

Table A2: MLD measures of inequality and ex-post IOp

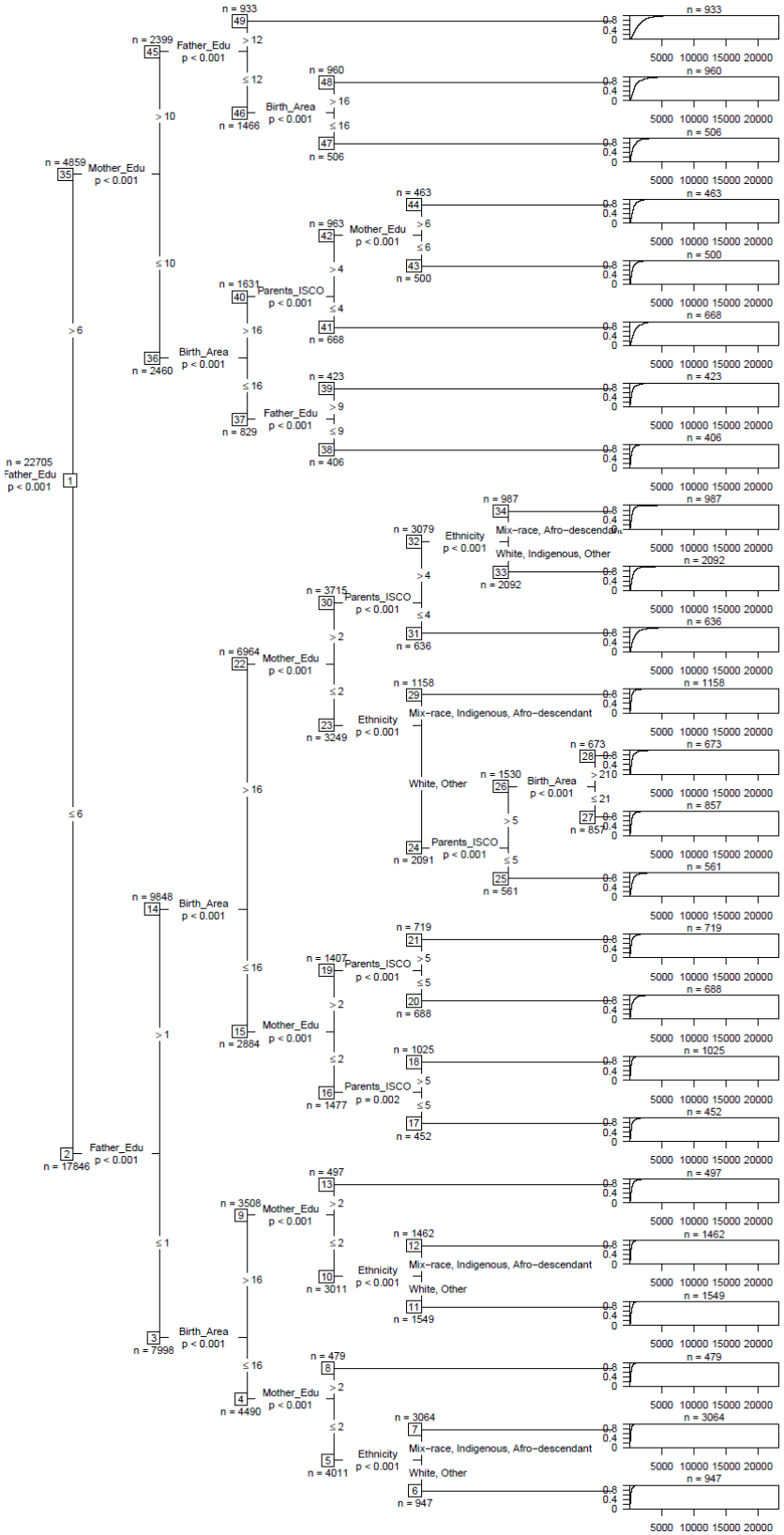
Country/Year	# Types	Total MLD	IOp MLD (Trees)	Relative IOp MLD (Trees)	IOp MLD (Forest)	Relative IOp MLD (Forest)
Argentina 2014	12	0.2812	0.0468	0.1664	0.0215	0.0765
Bolivia 2008	11	0.4766	0.1528	0.3221	0.0853	0.1790
Brazil 2014	25	0.4760	0.1426	0.2996	0.0796	0.1672
Chile 2006	21	0.5054	0.1110	0.2196	0.0614	0.1215
Chile 2009	25	0.5453	0.1121	0.2056	0.0450	0.0825
Chile 2011	32	0.4933	0.1129	0.2289	0.0431	0.0874
Chile 2013	26	0.4896	0.1011	0.2065	0.0465	0.0950
Chile 2015	31	0.4760	0.1118	0.2349	0.0422	0.0886
Colombia 2010	10	0.5974	0.1121	0.1876	0.0694	0.1161
Ecuador 2006	22	0.5600	0.1530	0.2732	0.0884	0.1578
Ecuador 2014	18	0.3895	0.0758	0.1946	0.0377	0.0968
Guatemala 2000	14	0.5641	0.1400	0.2482	0.0627	0.1112
Guatemala 2006	23	0.5412	0.1832	0.3385	0.0808	0.1493
Guatemala 2011	23	0.5252	0.0994	0.1893	0.0399	0.0760
Panama 2003	13	0.6540	0.1663	0.2543	0.1009	0.1543
Panama 2008	11	0.5373	0.1471	0.2738	0.0841	0.1565
Peru 2001	27	0.4864	0.1211	0.2490	0.0592	0.1217
Peru 2006	25	0.4528	0.1415	0.3125	0.0837	0.1848
Peru 2007	24	0.4548	0.1290	0.2837	0.0608	0.1337
Peru 2008	21	0.4064	0.1167	0.2871	0.0683	0.1680
Peru 2009	19	0.3984	0.1074	0.2696	0.0548	0.1376
Peru 2010	19	0.3700	0.0918	0.2481	0.0484	0.1308
Peru 2011	22	0.3724	0.0893	0.2398	0.0473	0.1270
Peru 2012	18	0.3628	0.0895	0.2467	0.0509	0.1403
Peru 2013	21	0.3559	0.0891	0.2504	0.0527	0.1481
Peru 2014	22	0.3296	0.0861	0.2612	0.0466	0.1414
Peru 2015	29	0.3338	0.0976	0.2924	0.0502	0.1504

Appendix 4: Transformation (ex-post) trees for the most recent surveys in eight Latin American countries.

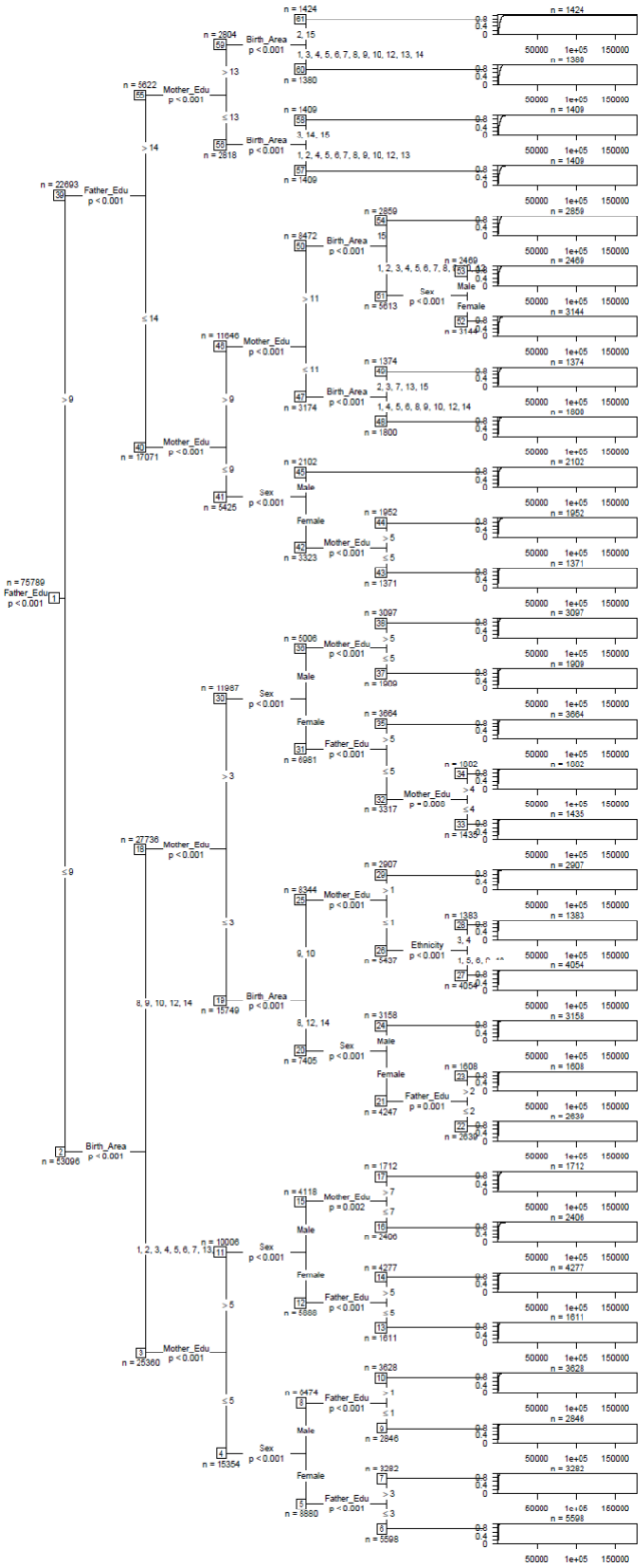
Argentina 2014



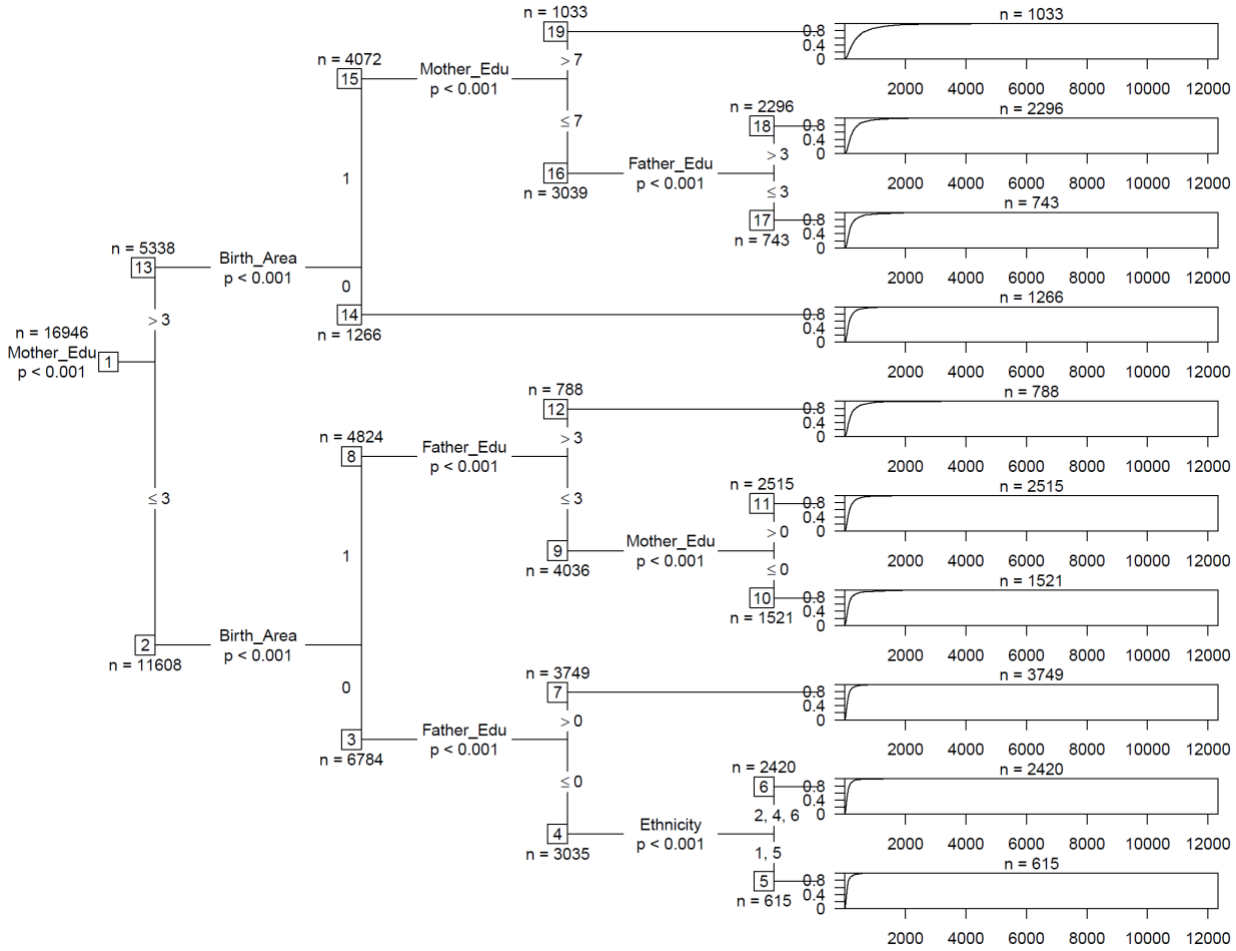
Brazil 2014


n = 22705
Father_Edu
p < 0.001
1
≤ 6
n = 17846
2
Father_Edu
p < 0.001
≤ 1
n = 7998
3
Birth_Area
p < 0.001
≤ 16
n = 4490
4
Mother_Edu
p < 0.001
≤ 2
n = 4011
5
Ethnicity
p < 0.001
White, Other
n = 947
6
Mix−race, Indigenous, Afro−descendant
n = 3064
7
n = 479
8
> 2
n = 3011
10
Mother_Edu
p < 0.001
> 16
n = 3508
9
Mix−race, Indigenous, Afro−descendant
White, Other
n = 1549
11
n = 1462
12
n = 497
13
> 1
n = 9848
14
Birth_Area
p < 0.001
≤ 16
n = 1477
16
Parents_ISCO
p = 0.002
> 5
≤ 5
n = 452
17
n = 1025
18
> 2
n = 2884
15
Mother_Edu
p < 0.001
≤ 2
n = 688
20
Parents_ISCO
p < 0.001
> 5
≤ 5
n = 1407
19
n = 719
21
> 16
n = 6964
22
Mother_Edu
p < 0.001
≤ 2
n = 3249
23
Ethnicity
p < 0.001
White, Other
n = 2091
24
Parents_ISCO
p < 0.001
> 5
≤ 5
n = 561
25
n = 1530
26
Birth_Area
p < 0.001
> 210
≤ 21
n = 857
27
n = 673
28
Mix−race, Indigenous, Afro−descendant
n = 1158
29
> 2
n = 3715
30
Parents_ISCO
p < 0.001
> 4
≤ 4
n = 636
31
n = 3079
32
Ethnicity
p < 0.001
White, Indigenous, Other
n = 2092
33
Mix−race, Afro−descendant
n = 987
34
n = 4859
35
Mother_Edu
p < 0.001
> 10
n = 2460
36
Birth_Area
p < 0.001
> 16
≤ 16
n = 829
37
Father_Edu
p < 0.001
> 9
≤ 9
n = 406
38
n = 423
39
≤ 10
n = 1631
40
Parents_ISCO
p < 0.001
> 4
≤ 4
n = 668
41
n = 963
42
Mother_Edu
p < 0.001
> 6
≤ 6
n = 500
43
n = 463
44
n = 2399
45
Father_Edu
p < 0.001
> 12
≤ 12
n = 1466
46
Birth_Area
p < 0.001
> 16
≤ 16
n = 506
47
n = 960
48
n = 933
49
0.8
0.4
0
5000 10000 15000 20000

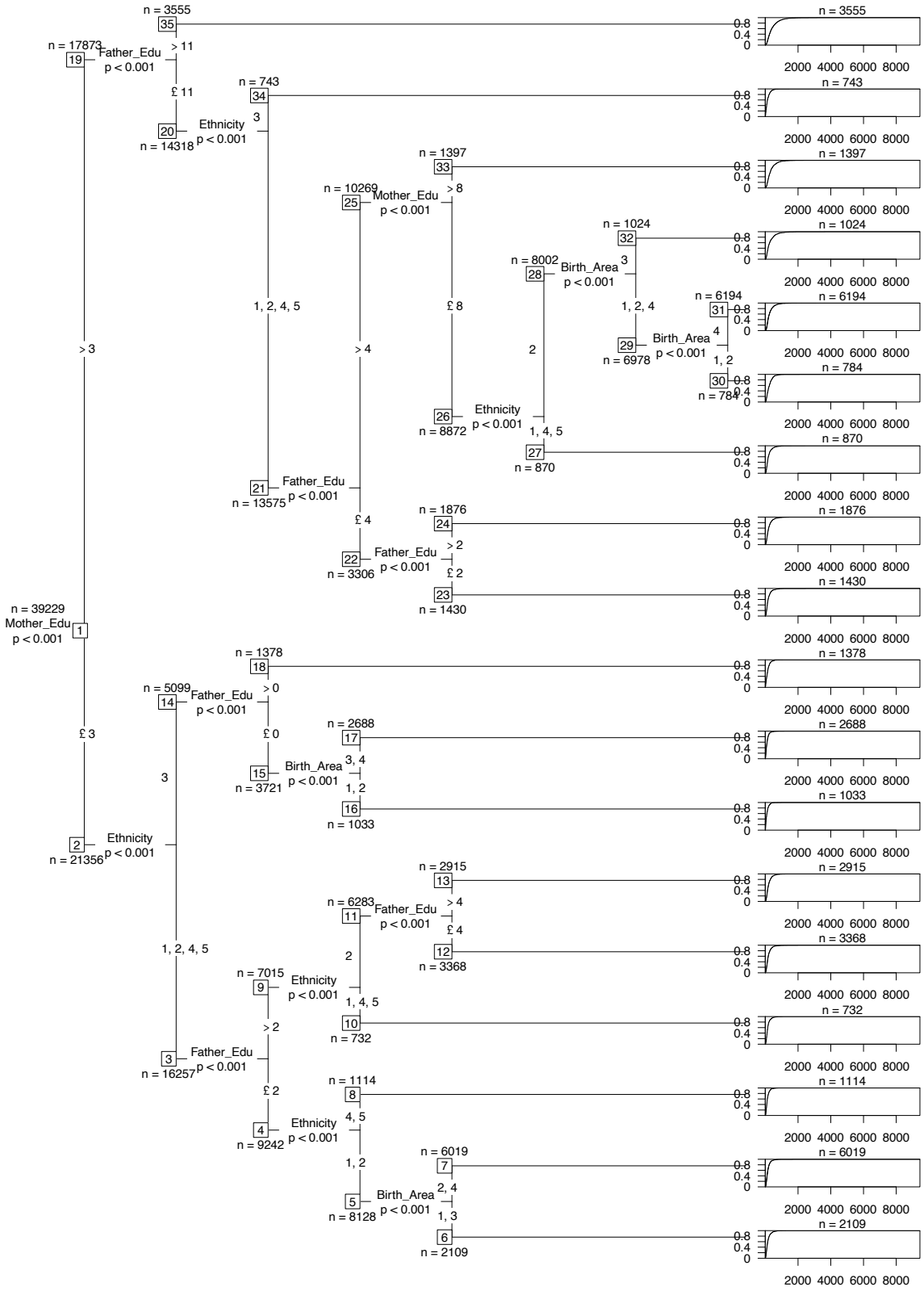
Chile 2015


n = 75789
Father_Edu
p < 0.001
1
≤ 9
> 9
2
n = 53096
Birth_Area
p < 0.001
3
n = 25360
Mother_Edu
p < 0.001
4
n = 15354
Sex
p < 0.001
5
n = 8880
Father_Edu
p < 0.001
6
n = 5598
7
n = 3282
8
n = 6474
Father_Edu
p < 0.001
9
n = 2946
10
n = 3628
11
n = 10006
Sex
p < 0.001
12
n = 5888
Father_Edu
p < 0.001
13
n = 1611
14
n = 4277
15
n = 4118
Mother_Edu
p = 0.002
16
n = 2406
17
n = 1712
18
n = 27736
Mother_Edu
p < 0.001
19
n = 15749
Birth_Area
p < 0.001
20
n = 7405
Sex
p < 0.001
21
n = 4247
Father_Edu
p = 0.001
22
n = 2639
23
n = 1608
24
n = 3158
25
n = 8344
Mother_Edu
p < 0.001
26
n = 5437
Ethnicity
p < 0.001
27
n = 4054
28
n = 1383
29
n = 2907
30
n = 11987
Sex
p < 0.001
31
n = 6981
Father_Edu
p < 0.001
32
n = 3317
Mother_Edu
p = 0.008
33
n = 1435
34
n = 1882
35
n = 3664
36
n = 5006
Mother_Edu
p < 0.001
37
n = 1909
38
n = 3097
39
n = 22693
Father_Edu
p < 0.001
40
n = 17071
Mother_Edu
p < 0.001
41
n = 5425
Sex
p < 0.001
42
n = 3323
Mother_Edu
p < 0.001
43
n = 1371
44
n = 1952
45
n = 2102
46
n = 11646
Mother_Edu
p < 0.001
47
n = 3174
Birth_Area
p < 0.001
48
n = 1800
49
n = 1374
50
n = 8472
Birth_Area
p < 0.001
51
n = 5613
Sex
p < 0.001
52
n = 3144
53
n = 2469
54
n = 2859
55
n = 2818
Birth_Area
p < 0.001
56
n = 1409
57
n = 1409
58
n = 1409
59
n = 2804
Birth_Area
p < 0.001
60
n = 1380
61
n = 1424
Male
Female
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14
2, 15
3, 14, 15
1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13
1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14
1, 2, 3, 4, 5, 6, 7, 8
15
2, 3, 7, 13, 15
1, 4, 5, 6, 8, 9, 10, 12, 14
1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13
8, 9, 10, 12, 14
9, 10
3, 4
1, 5, 6, 7, 8, 9, 10
1, 2, 3, 4, 5, 6, 7, 13
50000
1e+05
150000

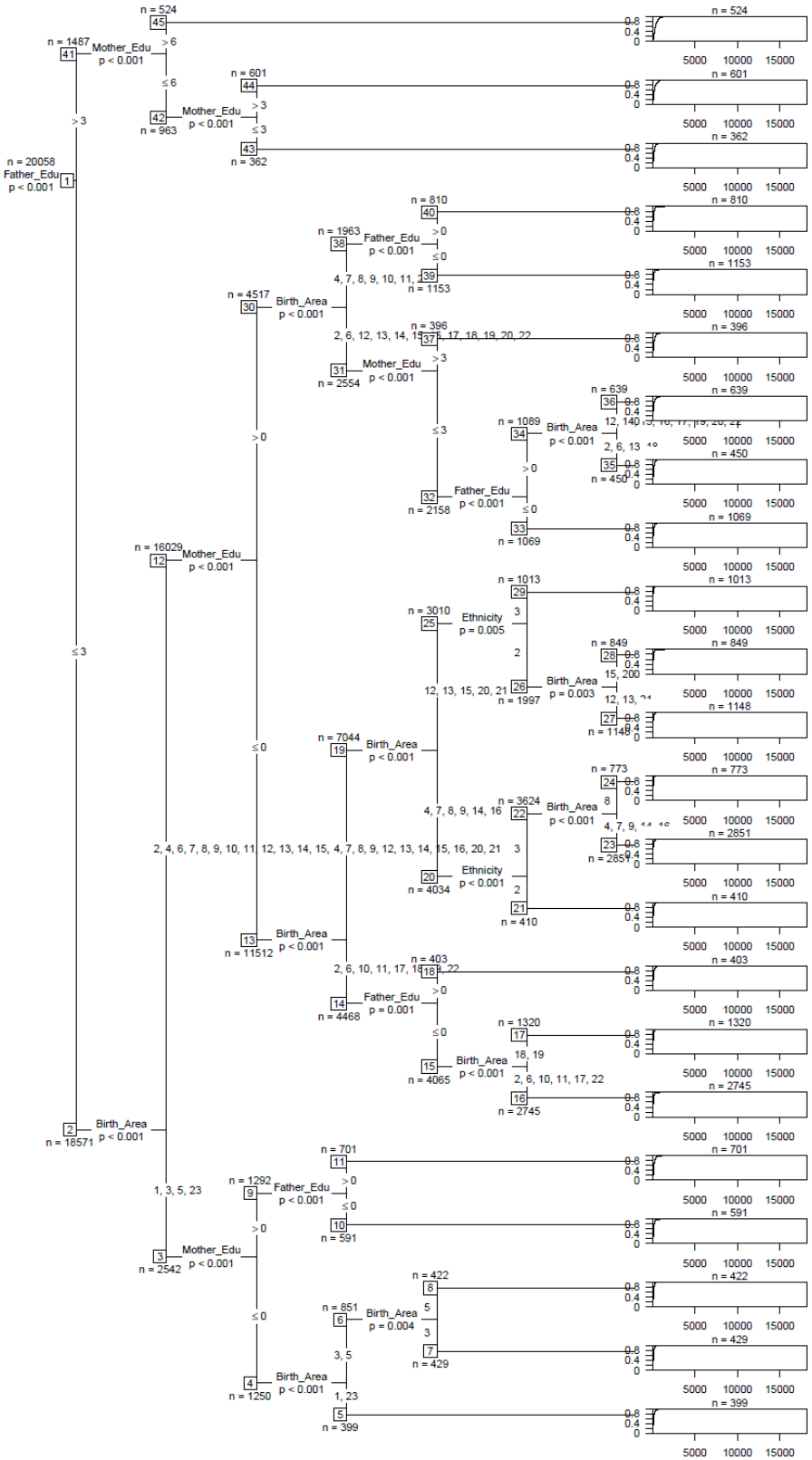
Colombia 2010



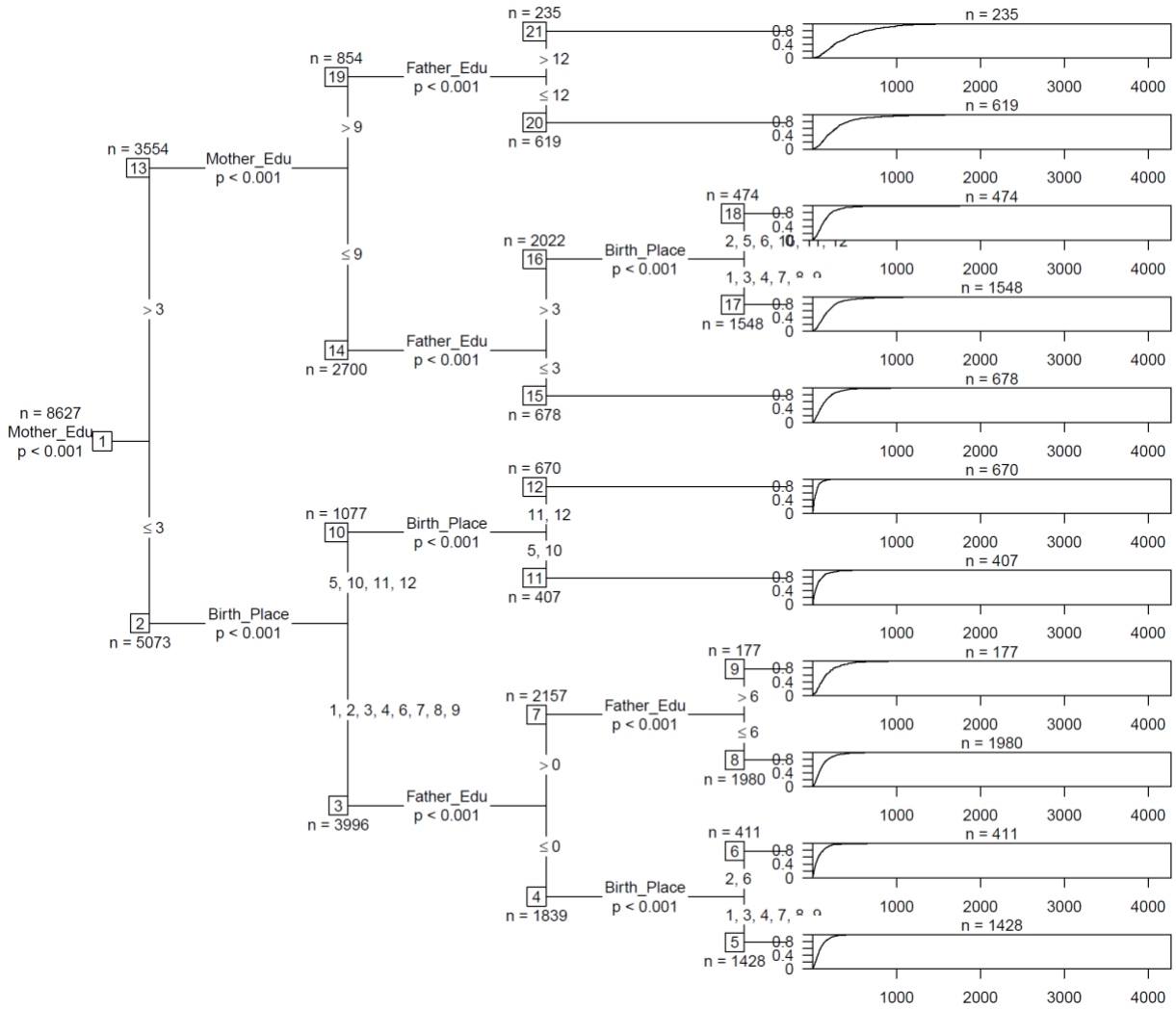
Ecuador 2014



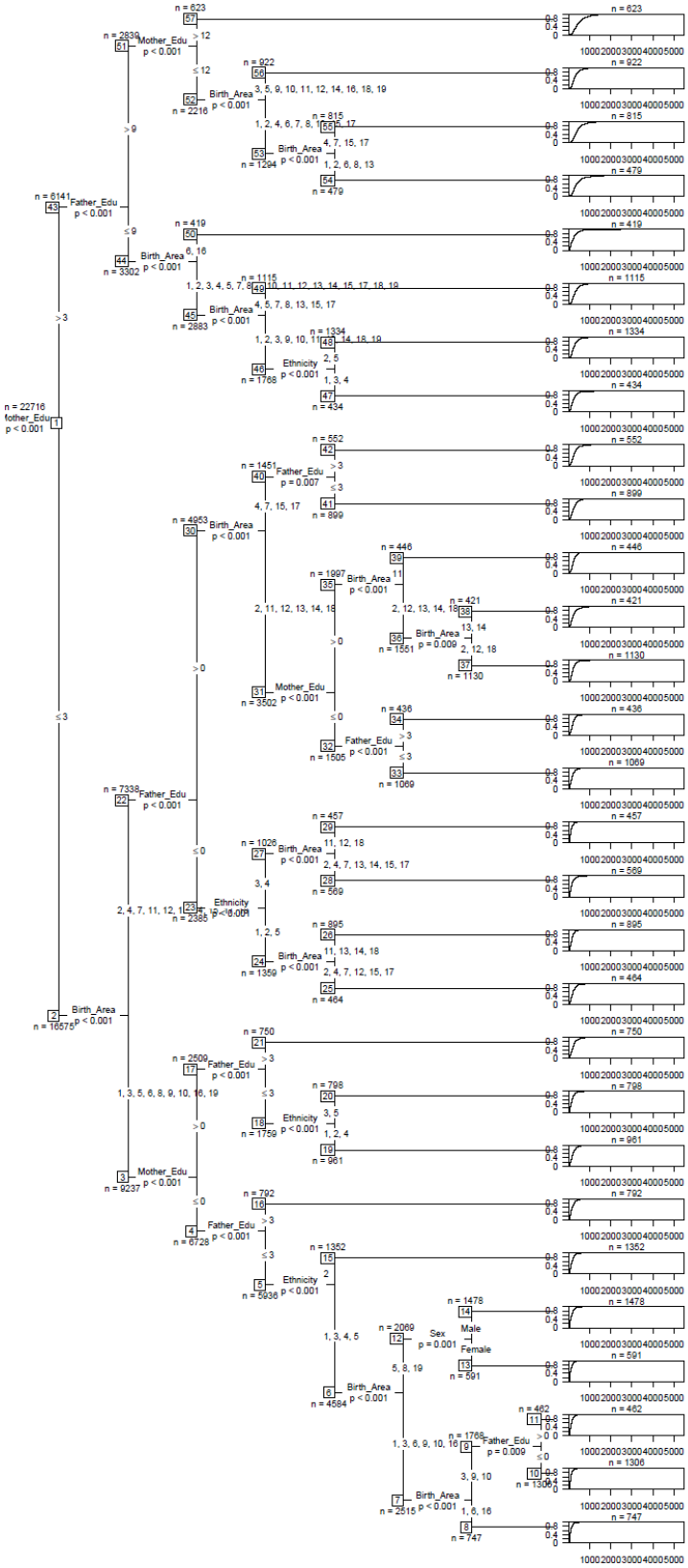
Guatemala 2011



Panama 2008


n = 8627
Mother_Edu
p < 0.001
1
≤ 3
> 3
n = 5073
2
Birth_Place
p < 0.001
n = 3554
13
Mother_Edu
p < 0.001
n = 1077
10
Birth_Place
p < 0.001
5, 10, 11, 12
1, 2, 3, 4, 6, 7, 8, 9
n = 3996
3
Father_Edu
p < 0.001
n = 854
19
Father_Edu
p < 0.001
> 9
≤ 9
n = 2700
14
Father_Edu
p < 0.001
n = 235
21
> 12
≤ 12
n = 619
20
n = 2022
16
Birth_Place
p < 0.001
> 3
≤ 3
n = 678
15
n = 670
12
11, 12
5, 10
n = 407
11
n = 2157
7
Father_Edu
p < 0.001
> 0
≤ 0
n = 1839
4
Birth_Place
p < 0.001
n = 474
18
2, 5, 6, 10, 11, 12
1, 3, 4, 7, 8, 9
n = 1548
17
n = 177
9
> 6
≤ 6
n = 1980
8
n = 411
6
2, 6
1, 3, 4, 7, 8, 9
n = 1428
5
0.8
0.4
0
1000
2000
3000
4000

Peru 2015



Appendix 2: Richest and poorest types across countries and time

Ex-ante

Country	Year	Richest type	Income level	Sample	Pop. Share	Poorest type	Income level	Sample	Pop. Share
Argentina	2014	Individuals born in Gran Buenos Aires or Patagonia with fathers with incomplete secondary education or more	719.45	552	8.45%	Individuals born in Cuyo, NOA and NEA regions with fathers with no formal education and mothers with complete primary education or less	221.29	262	4.01%
Bolivia	2008	Individuals with fathers with secondary education or more and from Guarany, Chiquitano, Mojeño or non-indigenous ethnicity	899.21	315	5.98%	Individuals born in a rural area with fathers with no formal education and from Quechua, Aymara, Guarany or Chiquitano ethnicity	129.86	878	16.68%
Brazil	2014	Individuals with fathers and mothers with complete secondary education or more	1617.12	483	2.13%	Individuals born in Rondônia, Acre, Amazonas, Roraima, Pará, Amapá, Tocantins, Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia or Minas Gerais; with fathers and mothers with incomplete primary education or less, and mix-race, Indigenous or Afro-descendant ethnicity	221.78	3617	15.93%
Chile	2015	Individuals born in Antofagasta, Arica or Perinacota with fathers with complete tertiary education, and mothers with incomplete secondary education or more	2801.29	2368	3.12%	Individuals born in Tarapcá, Biobío, Araucanía, Los Lagos, Magallanes or Los Ríos with father with complete primary education or less, mothers with incomplete primary education or less and from Rapa Nui, Mapuche, Coya or Diaguita ethnicity	448.71	2411	3.18%

Colombia	2010	Individuals with mothers with incomplete secondary education or more	484.11	1148	6.77%	Individuals born in a rural area with fathers with incomplete primary education or less, mothers with no formal education and from indigenous, Gypsi (Rom) or Afro-descendant ethnicity	95.74	658	3.88%
Ecuador	2014	Individuals with fathers with complete secondary education or more and mothers with incomplete secondary education or more	574.98	964	2.46%	Females with fathers with incomplete primary education or less, mothers with no formal education and from Indigenous ethnicity	112.34	2227	5.68%
Guatemala	2011	Individuals with fathers with incomplete primary education or more and mothers with incomplete secondary education or more	385.19	524	2.61%	Individuals born in Solola, Totonicapan, Quetzaltenango, San Marcos, Huehuetenango, Quiche, Alta Verapaz, Chiquimula or Jalapa with fathers with incomplete primary education or less and mothers with no formal education	90.14	5886	29.34%
Panama	2008	Individuals with fathers with complete secondary education or more, and mothers with incomplete secondary education or more	433.98	235	2.72%	Individuals born in Comarca Embera-Wounaan or Comarca Ngobe Bugle with mothers with incomplete primary education or less	45.58	670	7.77%
Peru	2015	Individuals with fathers and mothers with complete secondary education or more	564.32	487	2.14%	Individuals born in Amazonas, San Martin, Ucayali, Cajamarca, Huánuco or Loreto with fathers with no formal education, mothers with incomplete primary education or less and from Indigenous, Afro-descendant or Other ethnicity	102.73	948	4.17%

Ex-post

Country	Year	Richest type	Income level	Sample	Pop. Share	Poorest type	Income level	Sample	Pop. Share
Argentina	2014	Individuals born in Gran Buenos Aires or Patagonia with fathers with incomplete secondary education or more	719.45	552	8.45%	Individuals born in Cuyo, NOA or NEA regions with mothers with complete primary education or less	250.85	984	15.06%
Bolivia	2008	Individuals born in an urban area with fathers with secondary education or more and mothers with primary education or more	871.76	323	6.13%	Individuals born in a rural area with father with no formal education and from Quechua, Guarany or Mojeño ethnicity	125.15	524	9.95%
Brazil	2014	Individuals with fathers with complete secondary education or more and mothers with incomplete secondary education or more	1545.19	933	4.11%	Individuals born in Rondônia, Acre, Amazonas, Roraima, Pará, Amapá, Tocantins, Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe or Bahia with fathers with no formal education, mothers with incomplete primary education or less and from Mix-race, Indigenous or Afro-descendant ethnicity	214.70	3064	13.49%
Chile	2015	Individuals from Antofagasta, Arica or Perinacota with fathers with complete tertiary education and mothers with incomplete tertiary education or more	2813.45	1424	1.88%	Individuals born in Araucanía or Los Lagos with fathers with incomplete secondary education or less and mothers with incomplete primary education or less and from Quechua or Mapuche ethnicity	413.20	1383	1.82%
Colombia	2010	Individuals born in an urban area with mothers with incomplete secondary education or more	510.17	1033	6.10%	Individuals born in a rural area with fathers with no formal education, mothers with incomplete priaary education or less and from Indigenous or Afro-descendant ethnicity	94.55	615	3.63%
Ecuador	2014	Individuals with fathers with complete secondary education or more and mothers with incomplete primary education or more	444.69	3555	9.06%	Individuals born in North or Center region with fathers with no formal education, mothers with incomplete primary education or less and from Indigenous ethnicity	109.59	1033	2.63%

Guatemala	2011	Individuals with fathers with incomplete primary education or more and mothers with incomplete secondary education or more	385.19	524	2.61%	Individuals born in San Marcos, Huehuetenango or Jalapa with fathers with incomplete primary education or less, mothers with no formal education and from Mix-race ethnicity	85.74	1148	5.72%
Panama	2008	Individuals with fathers with incomplete tertiary education or more and mothers with incomplete secondary education or more	433.98	235	2.72%	Individuals born in Comarca Embera-Wounaan or Comarca Ngobe Bugle with mothers with incomplete primary education or less	45.58	670	7.77%
Peru	2015	Individuals with fathers with incomplete secondary education or more and mothers with incomplete tertiary education or more	537.05	623	2.74%	Individuals born in Amazonas, San Martin, Ucayali, Cajamarca or Loreto with fathers with incomplete primary or less, mothers with no formal education, and from White, Indigenous, Afro-descendant or Other ethnicity	111.81	747	3.29%