ECINEQ

Society for the Study of Economic Inequality

Working Paper Series

A Random Forest approach of the Evolution of Inequality of Opportunity in Mexico

Thibaut Plassot

Isidro Soloaga

Pedro J. Torres

ECINEQ 2022 614

ECINEQ 2022 614
2022 June
www.ecineq.org

A Random Forest approach of the Evolution of Inequality of Opportunity in Mexico

Thibaut Plassot
Universidad Iberoamericana, Mexico City: Department of Economics
Isidro Soloaga
Universidad Iberoamericana, Mexico City: Department of Economics
Pedro J. Torres
Universidad Iberoamericana, Mexico City: Department of Economics

Abstract

This work presents the trend of Inequality of Opportunity (IOp) and total inequality in wealth in Mexico for the years 2006, 2011 and 2017, and provides estimations using both an ex-ante and ex-post compensation criterion. We resort on a data-driven approach using supervised machine learning models to run regression trees and random forests that consider individuals' circumstances and effort. We find an intensification of both total inequality and IOp between 2006 and 2011, as well as a reduction of these between 2011 and 2017, being absolute IOp slightly higher in 2017 than in 2006. From an ex-ante perspective, the share of IOp within total inequality slightly decreased although using an ex-post perspective the share remains stable across time. The most important variable in determining IOp is household´s wealth at age 14, followed by both, father´s and mother´s education. Other variables such as the ability of the parents to speak an indigenous language proved to have had a lower impact over time.

Keyword: Inequality Of Opportunity, Mexico, Shapley Decomposition, Random Forests

JEL Cassification: C14, C81, D31, D63

A Random Forest approach of the Evolution of Inequality of Opportunity in Mexico

Thibaut Plassot♦ Isidro Soloaga♠ Pedro J. Torres L.♣

Universidad Iberoamericana, Mexico City: Department of Economics

June 2022

Abstract

This work presents the trend of Inequality of Opportunity (IOp) and total inequality in wealth in Mexico for the years 2006, 2011 and 2017, and provides estimations using both an ex-ante and ex-post compensation criterion. We resort on a data-driven approach using supervised machine learning models to run regression trees and random forests that consider individuals' circumstances and effort. We find an intensification of both total inequality and IOp between 2006 and 2011, as well as a reduction of these between 2011 and 2017, being absolute IOp slightly higher in 2017 than in 2006. From an ex-ante perspective, the share of IOp within total inequality slightly decreased although using an ex-post perspective the share remains stable across time. The most important variable in determining IOp is household´s wealth at age 14, followed by both, father´s and mother´s education. Other variables such as the ability of the parents to speak an indigenous language proved to have had a lower impact over time.

Keywords: Inequality Of Opportunity, Mexico, Shapley Decomposition, Random Forests

JEL Classification: C14, C81, D31, D63

♦ Email: thiplass@gmail.com; ORCID: https://orcid.org/0000-0002-5652-3279

♠ Email: isidro.soloaga@gmail.com

♣ Email: pedroj.tol@gmail.com; ORCID: https://orcid.org/0000-0003-2762-2355

1. Introduction

Academics such as Arneson (1989), Cohen (2015), Fleurbaey (1995) and Roemer (1998) focused on developing a theory of Equality of Opportunity based on the principles of Rawls's (1971) theory where total inequality can be decomposed into inequality derived from circumstances over which individuals have no control, named Inequality of Opportunity (IOp), and inequality derived from the effort that individuals exhort.

One of the major interests for researchers and policymakers in monitoring IOp levels is to evaluate the scope of the social determinisms within total inequality, to identify bottlenecks and evaluate the impact of public policies, economic conjuncture, and the role of institutions on social stratification across regions and over time. Obtaining country-level or subnational measures afford to evaluate how macroeconomic factors and policies in a specific territory or period of time can lead to reduce inequality. There is also a need to identify the circumstances that contribute more to IOp, and how this contribution changes along time, with the aim of adjusting the actions on the principal barriers to equality of opportunity. In this sense, this article provides comparable measures of IOp across time, and calculates the weight of different circumstances into IOp.

The concept of IOp combines the ethical principles of compensation and reward. The compensation principle advocates for transfers to correct inequalities derived from circumstances and for which the society does not hold the individual responsible, whereas the reward principle sustains to maintain the inequality that arises from differences in personal responsibility or effort. Roemer proposed to classify individuals into types according to their circumstances, and into tranches according to the effort they exert. Equality of opportunity is reached when "all those who expend the same degree of effort, regardless of their type, have the same chances of achieving the objective" (Roemer, 2004, p.49). To identify degrees or tranches of effort, it is required to previously identify types. Indeed, Roemer underlines that the effort level is influenced by the circumstances and recommends identifying degrees of effort as a personal responsibility factor free from the effect of circumstances. To do this, researchers generate for each type-specific distribution quantiles as a measurement for the degrees of effort. In the IOp literature two approaches for the compensation principle have been identified. The ex-ante compensation, which is applied without considering effort and focuses on inequality between-types, and the ex-post compensation, which is applied after observing effort and focuses on inequality within-tranches.

There has been an increasing literature on this topic in Mexico where researchers present measures of ex-ante IOp for 2006 (Wendelspiess Chávez Juárez, 2015), for 2011 (Vélez-Grajales et al., 2019), for 2016 (Monroy-Gómez-Franco et al., 2021) and for 2017 (Monroy-Gómez-Franco & Corak, 2019; Plassot et al., 2019), and ex-post IOp for 2017 (Plassot et al., 2022) [1]. These studies obtained measures for separate years only, with different sets of circumstances and outcome variables, which limit proper comparisons across years. We present in Appendix A.1 a synthesis of this literature and the IOp estimations, the source of data and the circumstances used by each author. As we can see there, estimations are very different according to the choice of the inequality indicator, the outcome variable, as well as in the circumstances chosen. Moreover, they vary also in the way other factors like the size and composition of the sample, the method (parametric or non-parametric), or the approach chosen (ex-ante or ex-post) are implemented.

In other countries, researchers have studied trends in IOp over time. For example, Bussolo et al. (2019) compare IOp measures and their weight within total inequality on the long term for five European countries; Brzezinski (2020) analyzes the change in IOp between 2004 and 2010 for twenty-three European countries; Aaberge et al. (2010) estimate IOp in Norway from 1967 to 2006 with both an ex-ante and ex-post approach. The studies with a temporal approach permit to better identify over time the weight of each circumstance within IOp (Suárez Álvarez & López Menéndez, 2018; Brunori & Neidhofer, 2021), the linkages between macroeconomics dimensions like growth and IOp (Carranza, 2021; Bradbury, 2016), or also how the changes in socio-economic characteristics[2] of some territories can be correlated with variations in IOp over time (Souza et al. 2017). Finally, authors underline the difficulty to compare studies done with different methodologies and variables and emphasize

[1] Literature generally use the Gini index, Mean Log Deviation, Dissimilarity Index or the R-squared as measures of absolute and relative inequality.
[2] Variables estimated at the regional level like the average years of schooling, the rate of informality in the economy, per capita GDP, health or educational expenditure.

the need to standardize measures across countries and over time (Breen & Jonsson, 2005; Shavit and Blossfeld,1993).

In Mexico some studies analyze inequality across time, addressing the influence on inequality that changes in social policies implemented by the different governments and the macroeconomic conjuncture (Cortes, 2013; Campos et al., 2012) may have. Concerning IOp and social mobility we can only mention the works of Torche (2020) that uses a cohort analysis to study the IOp in educational attainment, and Solis (2012) who describes changes in social mobility in the occupational status. Nevertheless, to the best of our knowledge, there are no studies that focus on the evolution of IOp in the wealth dimension.

We contribute to this literature by presenting comparable measures of total inequality and IOp for years 2006, 2011 and 2017 in Mexico. We use the same set of circumstances and construct percentiles based on wealth indexes with the same assets for each year, thus making our results comparable across time. Following Roemer's theory and the works of Brunori et al. (2021), Brunori & Neidhöfer (2020), Plassot et al. (2022) and Salas-Rojo & Rodríguez (2022) we resort to a data driven approach identifying types through Conditional Inference Trees and Random Forests. We then estimate effort by constructing quantiles on each type-specific distribution using Bernstein polynomials. We calculate IOp using an ex-ante approach with weak criterion as usually estimated for Mexico (Juárez, 2015; Grajales et al., 2018; Monroy-Gómez-Franco et al., 2021; and Monroy-Gómez-Franco & Vélez-Grajales, 2021). Finally, we provide an ex-post approach that considers the effort dimension as originally proposed by Roemer and as calculated by Brunori & Neidhöfer (2020) and Plassot et al. (2022).

We find that IOp as well as total inequality increases sharply between 2006 and 2011, most probably as a consequence of the 2008 economic crisis. On the contrary, both indicators decreased between 2011 and 2017 but remained slightly higher than the level of 2006. For the whole period, the share of IOp on total inequality has slowly decreased using an ex-ante method but remains stable using an ex-post method, which reflects differences in the evolution of IOp according to the definition adopted[3]. The percentile of household´s wealth at age 14 is the principal circumstance in shaping opportunities, contributing with more than 50% of total IOp. Parents´ education is the other principal contributor to IOp: father´s as well as mother´s schooling are statistically significant and their contribution to IOp is about 20% each one. Mother´s education takes more importance over the years considered, both in its contribution to IOp as well as in the construction of the trees. Although other circumstances like gender, going to a private school, or whether the parents were speaking an indigenous language or not, were also relevant for the construction of types, their individual contribution to IOp was lower than 6%.

Besides this introductory Section 1, in what follows Section 2 discusses the conceptual framework, Section 3 presents the specifics of our methodology for measuring both ex-ante and ex-post IOp as well as the weight decomposition of circumstances within IOp, Section 4 describes our data, Section 5 shows our results, and finally, Section 6 provides a discussion of the results.

2. Conceptual Framework

According to Roemer's theory, it is possible to define the outcome *y* of an individual *i* by the effort *e* that he/she exerts belonging to tranche *m* and his/her circumstances *C* (which are similar to all individuals of type *j*) through a function *f* additively separable between circumstances *C* and effort *e* as

$$y_i = f\left(C_j, e_m\right) \; \forall j \in [1, k] \;\&\; \forall m \in [1, n] \tag{1}$$

where *k* represents the number of types in a given society and *n* the number of tranches from which it is possible to represent each individual's outcome in a *k* x *n* dimensional matrix $|Y|$.

[3] Checchi et al. (2010) demonstrated how some policies at early age are more correlated with a variation of ex-ante IOp, while other policies, like redistributive and fiscal actions, have a major impact on reducing ex-post IOp.

It is then possible to estimate IOp from 2 different points of view[4]: *i)* ex-ante and *ii)* ex-post. The first one is the types-approach, which focuses on inequalities between social groups (between-types) and states an ex-ante compensation criterion. In this approach the inequality within-type is considered fair, so the weak criterion method proposes to create a counterfactual distribution $|Y'_{BT}|$ where the outcome y_i of each individual is replaced by π_i the average of its type (Checchi & Peragine, 2010; Ferreira & Gignoux, 2011).

$$|Y'_{BT}|: \hat{y}_i = \pi_i \tag{2}$$

Once we apply this transformation, by comparing the original inequality with that coming after replacing y_i by $\hat{y}_i$, we obtain the inequality between-types, i.e., that inequality which is explained only by the circumstances before observing the effort. Policies focusing on reducing differences between groups at an early age (in the educational or health dimension for example) can influence this indicator

The second one is the tranche-approach, which focuses on inequalities between individuals with the same degree of effort and uses the ex-post compensation criterion. We consider that this approach is closer to the initial proposition of Roemer (1998, 2002) that states that IOp is observed when individuals in the same tranche of effort have different probabilities to achieve an advantage. For this goal, researchers create a counterfactual distribution $|Y'_{WT}|$ by removing the between-tranches inequality and focusing on within-tranche inequality. To rescale the outcome, each individual's outcome is divided by the average score of his/her tranche so that all tranches have the same mean but maintaining the variance within tranches.

$$|Y'_{WT}|: \breve{y}_i = \frac{y_i}{\pi_m} \tag{3}$$

[4] Each approach has different hypothesis and are therefore incompatible in the sense that the ex-ante method concentrates on inequality between social groups with same circumstances, and the ex-post on inequality between individuals with the same degree of effort (Checchi & Peragine, 2010; Ferreira & Peragine, 2015; Fleurbaey, 2008; Fleurbaey & Peragine, 2013).

Once we apply this transformation, by comparing the original inequality with that coming after replacing y_i by $\check{y}_i$, we obtain the inequality within-tranches, i.e., the inequality which is explained only by the circumstances after observing the effort.

A measure of inequality $I(\cdot)$ must be applied to the new distributions to obtain both ex-ante and ex-post IOp.

$$IOp = I(Y'), \forall Y' \in [Y'_{BT}, Y'_{WT}] \tag{4}$$

In this work we use the Mean Log Deviation (MLD) because it is decomposable as well as the Gini coefficient (Gini) because it is widely used in the literature and allows for comparisons. Finally, we also present the R^2 as an ex-ante measure.

3. Empirical Strategy

a. Types

Function f in Equation 1 can be estimated within the general framework of Supervised Machine Learning models from existing observations to make out-of-sample predictions of the dependent variable y. Two specific algorithms within this general framework have direct applications for Roemer's theory: regression trees and random forests. These algorithms have several comparative advantages over more traditional methods such as the linear regression which also fall within this general framework (Brunori et al., 2019; Salas-Rojo & Rodríguez, 2022). A first advantage is that regression trees and random forests allow for missing values (further discussion of these procedure is given later, see Appendix B.1). They are also more flexible in considering non-linearities in the predictor variables, they present a non-arbitrary way of selecting relevant variables and can find intersectionality without having to rescale or transform any of the variables. Moreover, trees have an easy-to-interpret visual representation[5]. We use a specific set of trees called conditional inference trees as proposed by Hothorn et al. (2006) and as used by Brunori et al. (2021), Brunori & Neidhöfer (2020), Plassot et al. (2022) and Salas-Rojo & Rodríguez (2022).

[5] Although Random Forests lose this interpretability, they provide more accurate predictions. For a more in depth discussion see Brunori et al. (2019, 2021) and Appendix B1.

The algorithm proceeds as follows. In a first step the algorithm selects the relevant variables by testing the null hypothesis of independence between the outcome variable and all circumstances $C_j = 1,...,k$ in each node w.

$$H_0^j = D\left(Y_i \middle| C_{ji}\right) = D(Y_i) \rightarrow H_0 = \bigcap_{j=1}^{m} H_0^j \tag{5}$$

The global null hypothesis H_0 is tested on multiple linear statistics where the joint distribution of Y and C can be tested through permutation tests. If the global H_0 cannot be rejected, it is necessary to measure the association of each covariate $C_j, j = 1, \ldots, m$ by H_0^j to the response variable Y, the one with the lowest adjusted p-value is set as the first splitting variable. We specify the value of α at 0.05 to get results at the 95% confidence level. The splitting point divides the sample into two groups according to the values taken by the variable; for dichotomous variables, the sample is divided between the two categories and the threshold is obvious. For other types of variables, we need a non-arbitrary way of determining the threshold required to split the sample. The algorithm identifies each possible binary partition inducing a two-sample statistic where, for all possible subsets S of the sample space C_j, the discrepancy between the two is measured. The splitting point with the lowest p-value is selected and two branches that correspond to two subsamples are generated. This procedure is repeated on all possible subsamples until the global null hypothesis in 1 is reached, at which point types (terminal nodes) are identified. Importantly, when a value is missing for a circumstance of a specific observation, the algorithm sets the weight of each node w to zero for the computation. The algorithm then looks for a split by changing the weight that is similar to its result when w is zero.

Tree-based methods tend to have lower predictive accuracy when compared to other regression methods, as shown in Appendix B.1 and by Brunori et al. (2021). This can be the case because they may leave out a certain variable due to high correlation during step 1 of the algorithm. Conditional Random Forests improve upon Trees by building many of these trees and average across them all to make predictions. Trees are constructed through the same procedure as above, however, limiting circumstances to be considered at each splitting point to a subset so that $C' \subset C$ are used. These modifications, jointly with the estimation of N trees, allow for the correction of tree-based method's limitations. We prune each Forest by running them 200 times varying the number of trees to be computed. At every run we change the value (ranging from 1 to 200) and select the one where the decrease in the Mean Squared Error (MSE) loses significance (see Appendix B2). For further information regarding specifics of the algorithm see Hothorn et al. (2006).

b. Tranches (Effort)

The tranches, or degree of effort are estimated independently for each type. We identify quintiles over each type-specific distribution and approximate these by using Bernstein polynomials as

$$B_m(t,a,b) = \sum_{j=0}^{m} \beta_i b_{j,m}(t,a,b) \tag{6}$$

where the number of observations in each type is divided in a ten-fold cross-validation manner. The type-specific distribution is then estimated for every fold using Equation (6). The cumulative distribution is estimated using the coefficients *t, a* and *b,* in order to estimate the out of sample log-likelihood, choosing the polynomial degree *m* as the one that maximizes this log-likelihood, as proposed by Guan (2016) and used in Plassot et al. (2022).

c. Assessing the relative importance of each variable

To estimate the relative importance of each variable within IOp, we follow a more traditional procedure within economics and use the Shapley decomposition on the R^2 as used among others by Monroy-Gómez-Franco et al. (2021). The decomposition of the weight of each circumstance within R^2 is based on the additive property of Shapley values. This indicates that each prediction $\hat{y}_i$ can be decomposed into the sum of the attribution of each circumstance $\emptyset_i^{(c)}$and the predicted average $\emptyset_0$.

$$\hat{y}_i = \emptyset_0 + \sum_{c=1}^{C} \emptyset_i^{(c)} \tag{7}$$

We calculate the importance of each circumstance by estimating the R^2_{base} with all the circumstances, then estimating the Shapley prediction by removing one circumstance at a time and generating a matrix with dimensions N x C, where N stands for the total number of observations in the sample. For each of the circumstances we estimate a modified R^2_{shp}, the difference between $R^2_{base} - R^2_{shp}$ and normalize them to calculate percentages as proposed by Liang (2021). We present an alternative for the R^2 given that we are not relying on linear models, as shown by Redell (2019).

$$R^2_{adjusted} = \frac{var_{\hat{y}}}{var_{\hat{y}} + var_{(y-\hat{y})}} \tag{8}$$

This alternative metric has two properties that are desirable in IOp computation. Firstly, it maintains the properties of traditional R^2 and is limited to values between 0 and 1[6], giving a direct percentual interpretation of IOp. Secondly, it can be decomposed making measurable the weight of each circumstance in IOp.

4. Data

The source of information is the ESRU Surveys on Social Mobility in Mexico (EMOVI) that were done in 2006, 2011 and 2017 by the Espinosa Yglesias Center for Studies (CEEY). The Surveys are statistically representative of the 25 to 64 years old Mexican population. To focus on the influence of parents, we eliminate observations for individuals that at age 14 were not living with at least one of their parents. The samples used contain 6,796 observations in 2006, 10,196 in 2011, and 16,457 in 2017.

The surveys permit to assess the circumstances of the respondents at age 14 through retrospective questions on household's assets and parents' schooling levels. As one of the main focuses of this paper is to be able to compare across surveys, we selected the set of circumstances that were available in the three years. As it is well known in the IOp literature, since these variables are only a subset of the whole circumstances that could affect individuals' trajectories, all measures are lower-bound estimates. Circumstances chosen[7] are: i) parent´s years of schooling; ii) wealth level in the household at age 14 (measured through

[6] It is 0 when the explained variance $var_{\hat{y}}$ is zero, and 1 when the unexplained variance $var_{(y-\hat{y})}$ is zero.

[7] Our set of circumstances is very similar to that of Monroy et al. (2021), although these authors also consider circumstances like the skin tone, the residence in an urban or rural territory at age 14, and whether the father was an agricultural worker. The set is also comparable to the one used by Wendelspiess Chávez Juárez (2015), although the author also includes a variable indicating whether the parents of the respondent owned a house (Appendix A.1)

an asset index); iii) respondent´s sex; iv) whether the respondent's parents speak an indigenous language, and v) whether the person went to a private school at least one year while attending school.

Following Velez-Grajales et al. (2019) or Monroy et al. (2021), we construct an asset index for the current household of the respondent (outcome variable) and another asset index for when the respondent was 14 years old (circumstance variable). As indicated above, to construct indexes comparable across years, we identified the assets information available in the three surveys (Table 1). We conducted Multiple Correspondence Analysis (MCA), which is appropriate for dichotomic variables, to rank respondents according to their asset's level. The first factor or dimension represents the combination of variables explaining the higher share of total variance. The proportion of the variance explained for the household of the respondent at age 14 by the first dimension is 94% in 2006, 91% in 2011, and 94% in 2017. Concerning the assets of the current household of the respondents this proportion is lower (83% in 2006, 85% in 2011, and 88% in 2017). We extract the first dimension and standardize the outcome. Finally, we generate percentiles of wealth for this index (see detailed information in Appendixes C1 and C2).

Table 1: Assets used in the construction of our wealth indexes

	At age 14	At the time of the Survey
Electricity	X	X
Stove	X	X
Washing machine	X	X
Fridge		X
TV	X	X
Cable TV		X
Landline phone	X	X
Cellular phone		X
Internet		X
Computer		X
Bank Acount	X	X
Credit Card		X
Tubing water	X	
Toilet inside	X	
Domestic service	X	

Notes: Own, based on ESRU-EMOVI surveys for 2006, 2011 and 2017.

For year 2006, the survey is mostly composed by women (87% of respondents), whereas this percentage is 53% for 2011 and 2017 (Table 2). The 2017 survey includes a higher rate of respondents from rural areas (33%) than in 2011 (19%) and 2006 (17%). Finally, the 2006 survey includes a lower rate of respondents whose parents speak an indigenous language (8%) in comparison with that for the other years, whereas the 2011 survey includes a larger share of respondents that went to private schools.

Table 2: Descriptive Statistics for the years 2006, 2011, and 2017

	2006	2011	2017
Observations (Original Database)	8,520	11,001	17,665
Observations (Working Database)	6,796	10,196	16,457
Men	13%	47%	47%
Women	87%	53%	53%
Mother's years of school	2.88	3.83	4.48
Father's years of schooling	3.51	4.04	4.84
Parents speaking indigenous language	8%	15%	12%
Respondent went to private school	4%	10%	4%

Note: Own, based on ESRU-EMOVI surveys for 2006, 2011, and 2017.The Working Database is the database after eliminating respondents that were not living with at least one of his/her parents.

Both wealth indexes have a similar distribution (see Table 3), being centered between the 45th and 50th percentile. We see a little decrease of the 1st quantile, the median, the 3rd quantile, and the mean in 2011. These same statistics are very similar but slightly higher for 2006 than for 2017.

Table 3: Descriptive Statistics for Wealth

		Min	25% (1st Q)	50%	75% (3rd Q)	Max	Mean
Wealth at age 14							
	2006	1	31	56	77	100	53
	2011	1	22	44	69	100	45
	2017	1	26	51	77	100	50
Wealth at the time of the study							
	2006	1	31	53	76	100	53
	2011	1	20	43	68	100	45
	2017	1	28	52	76	100	51

Notes: Table shows basic descriptive statistics for both wealth indexes for the years 2006, 2011 and 2017.

5. Results

a. Structure of Trees

A first indicator of ex-ante IOp is given by the construction of trees. Trees are a representation of the structure of opportunities in terms of wealth according to groups with different circumstances called types. Using the full sample for each year we can observe the increase in the number of types from 23 in 2006, to 32 in 2011, and to 45 in 2017 (see Appendixes D1, D2 and D3). This increment in the number of types is essentially due to the number of observations in each sample. As indicated in the literature, although the estimations of inequality indicators are relatively stable, the number of types increases with sample size (Brunori & Neidhöfer, 2020; Plassot et al., 2022).

The variables used as splitting points and the order in which they are selected allow a first approximation of the relative importance of each circumstance to define types and then determine the counterfactual outcome. For each year, the first variable used to construct the tree is the wealth percentile at origin (at age 14), dividing the sample into two groups. For the first group, which represents the most disadvantaged at age 14 (left of the trees in Appendix D), the wealth at origin is the variable that determines the second splitting point for the three years considered in the analysis. For the second group that concentrates the most advantaged individuals (right of the tree in Appendix D) the second most important variable for 2006 is father's education, and for 2011 and 2017 it is mother's education. These three variables are the circumstances most used in the construction of the trees. Mother´s education takes relatively more importance after 2006, while father's education becomes somehow of less importance: mother´s education is determinant for 56% of the types in 2006, 97% in 2011 and 96% in 2017 while father´s education is determinant for 96% of the types in 2006, 66% in 2011 and 84% in 2017.

The variable that reflects whether the parents of the respondent speak an indigenous language is only determinant for the most disadvantaged in terms of wealth at age 14 and is generally associated with lower outcomes for the three years. The variable that captures whether the respondent attended a private school is only determinant for children of the wealthiest households and in the middle of the outcome distribution at age 14, in particular for year 2006, where this variable sets apart children born in a household ranked in a percentile higher than the 87th of the wealth distribution and had a father with high levels of education. The sex of the respondent has relatively low importance to define types, which may reflect the fact that the outcome variable is an asset index calculated at the household level.

b. Inequality and IOp

Analysing the Gini coefficient and MLD, we observe a sharp increase of total inequality between 2006 and 2011 (from 0.29 to 0.35 using Gini; from 0.20 to 0.31 using MLD) followed by a reduction between 2011 and 2017 (from 0.35 to 0.32 using Gini; from 0.31 to 0.26 using MLD). Nonetheless, the level of inequality in 2017 is still higher to the one of 2006 as can be seen in Table 4. These results are probably a consequence of the 2008 economic crisis which exacerbated inequalities (Campos et al., 2014).

Table 4: Total Inequality

	Gini	MLD
2006	0.29	0.20
2011	0.35	0.31
2017	0.32	0.26

Notes: Table shows absolute values of total inequality in Mexico for 2006, 2011, and 2017 using both the Gini coefficient and Mean Log Deviation (MLD).

IOp indicators follow the same trend: they increase between 2006 and 2011 and then decrease in 2017 (see Table 5). Using an ex-ante approach, we observe that IOp levels were very similar in 2006 and 2017 (0.164 and 0.168 using Gini; 0.047 and 0.048 using MLD). Nevertheless, the share of IOp within total inequality decreases, going from 22.8% to 18% when using MLD or from 33.2% to 27.4% using R^2. Just as seen with total inequality, in 2011 there is an increase of the absolute value in IOp compared to 2006, however, the relative importance within total inequality is very similar in these two years.

Table 5: IOp ex-ante

	Absolute		Relative		
	Gini	MLD	Gini	MLD	R2
2006	0,164	0,047	56,5%	22,8%	33,2%
2011	0,203	0,069	57,1%	22,3%	34,2%
2017	0,168	0,048	53,0%	18,0%	27,4%

Notes: Table shows ex-ante IOp measures, both in absolute values and as percentages of total inequality for 2006, 2011, and 2017. Gini coefficient is not perfectly decomposable and should therefore not be interpreted as a direct measure of the share of IOp within total inequality.

Using an ex-post approach, the increase and decrease in absolute IOp levels appear to be proportional to the observed increases and decreases in total inequality, and the share of IOp remains stable across time at around 57% when using MLD (see Table 6). We observe a slight increase from 2006 to 2011 and a decrease from 2011 to 2011. As found by Plassot et al. (2022) the share of IOp is higher with an ex-post than with an ex-ante approach.

Table 6: IOp ex-post

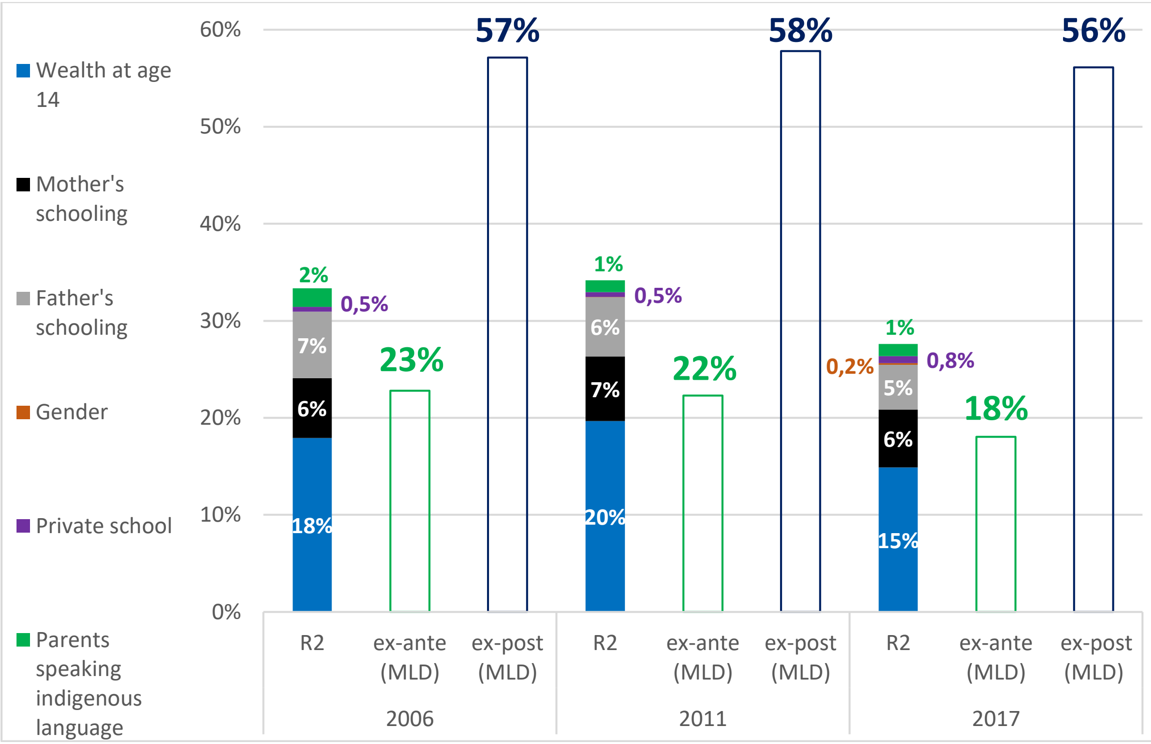
	ex-post			
	Absolute		Relative	
	Gini	MLD	Gini	MLD
2006	0,230	0,117	79,2%	57,1%
2011	0,298	0,180	83,6%	57,8%
2017	0,256	0,148	81,0%	56,1%

Notes: Table shows ex-post IOp measures, both in absolute values and as percentages of total inequality for 2006, 2011, and 2017. Gini coefficient is not perfectly decomposable and should therefore not be interpreted as a direct measure of the share of IOp within total inequality.

c. Shapley Decomposition

Finally, we decompose the weight of each circumstance within the ex-ante IOp using the R^2. Results are very similar to the ones shown by the variables used as splitting points in each tree as discussed in section 5.a. Wealth levels at age 14 are the principal contributors to IOp for the three years and explains more than half of IOp. The weight of this variable increases between 2006 and 2011 from 54% to 58%, before decreasing back to 54% in 2017. Parent´s schooling levels (mother or father) are the second and third biggest contributors to IOp with around 20% each one. It is interesting to observe that in 2006 the weight of the contribution of mother´s schooling is slightly lower to that of the father´s (19% and 21% of IOp) but higher since 2011, and weighs 22% in 2017, whereas the contribution of the education of the father is 17% for the same year. It is difficult to pinpoint the source of these marginal changes in these influences and to determine if the differences in the composition of the sample of the EMOVI across years can lead to biased interpretation. The ability of at least one parent to speak an indigenous language accounts for less than 6% of IOp and having attended to private school for less than 3%. Finally, the sex variable weighs less than 1% of total households´ wealth inequality. Both, ex-ante and ex-post results using MLD, are statistically different in each year at the 95% confidence level.

Figure 1: Weight of total IOp and of each Circumstance



Notes: Figure shows relative measures of ex-ante IOp using R^2 and MLD; ex-post measures using MLD; as well as the weight decomposition of circumstances over R^2 for the years 2006, 2011 and 2017. Both, ex-ante and ex-post results using MLD, are statistically different in each year at the 95% confidence level.

6. Discussion

This research is the first to our knowledge that presents ex-ante and ex-post IOp measures comparable for Mexico for different years using the same methodology and circumstances. Our results suggest a change in inequality across time. First, there is a strong increase of total inequality and IOp from 2006 to 2011, which most probably is due to the 2008 economic crisis, which, among other things, affected negatively formal employment and increased unemployment in Mexico (Freije et al., 2013). The decrease in inequality and IOp between 2011 and 2017 indicates a general post-crisis recovery.

When using an ex-ante approach with a weak criterion, we find that the share of IOp within total inequality stays at similar levels for 2006 and 2011, around 22% and 34% when using MLD and R^2 respectively, however, we do see a decrease in its share at the post-crisis period from 2011 to 2017 (going from 22% to 18% using MLD and from 34% to 28% using R^2). On the other hand, using an ex-post approach, we find that the share of IOp within total inequality remains stable at around 57% for the three years. In other words, the share of inequality derived from differences between social groups has slowly decreased over the period while the share of inequality derived from differences between individuals with the same effort remains stable. The former result can be a consequence of policies striving to equalize (ex-ante) opportunities and to reduce differences in the effective access and quality of education and health between different groups and territories, while the policies striving to reduce (ex-post) inequality through the fiscal and redistributive policies (or any action after observing effort in the adult life), were less effective (or absent) over the period[8].

The share of IOp explained by the percentile of wealth at age 14 contributes to more than half of IOp. The weight of the mother´s education on IOp (around 20%) is comparable to that of father´s for the three years, however, the first one gains a bit of importance since 2011 and its contribution to IOp is relatively higher for 2017. These results are consistent with those found by Monroy-Gómez-Franco et al. (2021) and Plassot et al. (2019), where the principal factor that contributes to IOp is wealth at the age of 14. Monroy-Gómez-Franco et al. (2021) found also that the education of both parents is the second most important circumstance in determining wealth, while Plassot et al. (2019) found these to be territorial variables. The low contribution of gender within total inequality is consistent with both works and can probably be explained by the fact that assets are measured at the household level and are therefore similar both for men and women when the respondent lives with his/her spouse/husband. Whether or not the parents speak an indigenous language present a low

[8] During the period considered, coverage of social public expenditure, particularly for public health and housing conditions, increased. Taxes and subsidies remained about the same.

contribution to IOp, but is significant in determining some types, especially the ones that are more disadvantaged at the age of 14, probably due to the fact that these lie more often in rural areas as shown by (Plassot et al., 2022). The private school variable is significant to differentiate types for respondents born in wealthier households, suggesting the possibility of social stratification through social networks.

A limitation of this study is that, to be able to make comparisons across years, we only consider a smaller set of circumstances, when other studies demonstrated the importance of other dimensions like the size of the city or the region (Delajara & Graña, 2018; Monroy-Gómez-Franco & Vélez-Grajales, 2020) or the skin tone (Monroy-Gómez-Franco et al., 2021; Monroy-Gómez-Franco & Vélez-Grajales, 2020). In this sense our estimations are lower-bound measures with the advantage of being comparable over time. Future surveys in Mexico with regional level and and/or rural/urban representativeness will permit to better understand the relation between policies at the national and regional levels and their efficiency on reducing or increasing inequality over time.

References

Aaberge, R., Mogstad, M., & Peragine, V. (2010) : Measuring longterm inequality of opportunity, *Discussion Papers, No. 620, Statistics Norway, Research Department, Oslo*

Arneson, R. J. (1989). Equality and equal opportunity for welfare. *Philosophical Studies*, *56*(1), 77–93. https://doi.org/10.1007/BF00646210

Bradbury, K., & Triest, R.K. (2016). Inequality of Opportunity and Aggregate Economic Performance. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, vol. 2, no. 2, 2016, pp. 178–201. *JSTOR*, https://doi.org/10.7758/rsf.2016.2.2.08.

Breen, R., & Jonsson, J. O. (2005). Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility. *Annual Review of Sociology*, 31, 223-243. https://doi.org/10.1146/annurev.soc.31.041304.122232

Brzezinski, M. (2020). The evolution of inequality of opportunity in Europe, in *Applied Economics Letters*, 27:4, 262-266, DOI: 10.1080/13504851.2019.1613493

Brunori, P., Hufe, P., & Mahler, D. (2021). The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees and Forests. *IZA Institute of Labor Economics*. https://doi.org/10.2139/ssrn.3917304

Brunori, P., & Neidhöfer, G. (2020). The Evolution of Inequality of Opportunity in Germany: A Machine Learning Approach. *ZEW Discussion Papers*. https://doi.org/10.2139/ssrn.3570385

Brunori, P., Peragine, V., & Serlenga, L. (2019). Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare*, *52*(4), 635–661. https://doi.org/10.1007/S00355-018-1165-X/FIGURES/4

Bussolo, M., Checchi,D., & Peragine,V. (2019). Long-Term Evolution of Inequality of Opportunity. *Policy Research Working Paper; No. 8700. World Bank, Washington, DC*. © World Bank. https://openknowledge.worldbank.org/handle/10986/31168 License: CC BY 3.0 IGO."

Campos, R., Esquivel, G., & Lustig, N. (2012). The rise and fall of income inequality in Mexico, 1989-2010. *UNU-WIDER Working Paper*, (10), January 2012.

Campos Vázquez, R., Lustig, N., & Santillán, A. (2014). A methodological note on the measurement of labor income in Mexico. *Estudios Económicos*, *29*(1), 107-123. https://doi.org/10.24201/ee.v29i1.72

Checchi, D., Peragine, V., & Serlenga, L. (2010). Fair and unfair income inequalities in Europe. *IZA Discussion Paper No. 5025*. Bonn, Germany.

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *Journal of Economic Inequality*, *8*(4), 429–450. https://doi.org/10.1007/s10888-009-9118-3

Cohen, G. A. (1989). On the Currency of Egalitarian Justice. *Ethics*, *99*(4), 906–944. https://doi.org/10.1086/293126

Cortés, F. (2013). Medio siglo de desigualdad en el ingreso en México. In *Economía UNAM*, 10(29):12–34, 2013.

Delajara, M., & Graña, D. (2018). La movilidad social intergeneracional en M ~ exico y sus regiones es el resultado de regresiones rango-rango. *Sobre México Temas de Economía*,

1(1), 22–37.
Ferreira, F. H. G., & Gignoux, J. (2011). The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of Income and Wealth*, *57*(4), 622–657. https://doi.org/10.1111/j.1475-4991.2011.00467.x
Ferreira, F. H. G., & Peragine, V. (2015). Equality of Opportunity: Theory and Evidence. *IZA Discussion Paper*. https://doi.org/10.2139/ssrn.2598934
Fleurbaey, M. (1995). Three solutions for the compensation problem. *Journal of Economic Theory*, *65*(2), 505–521. https://doi.org/10.1006/jeth.1995.1018
Fleurbaey, M. (2008). Fairness, responsibility, and welfare. In *Oxfor University Press*. Oxford University Press.
Fleurbaey, M., & Peragine, V. (2013). Ex Ante Versus Ex Post Equality of Opportunity. *Economica*, *80*(317), 118–130. https://doi.org/10.1111/j.1468-0335.2012.00941.x
Grajales, R. V., Monroy-Gómez-Franco, L. A., & Yalonetzky, G. (2018). Inequality of Opportunity in Mexico. *Journal of Income Distribution*, 134–158. https://jid.journals.yorku.ca/index.php/jid/article/view/40429
Guan, Z. (2016). Efficient and robust density estimation using Bernstein type polynomials. *Journal of Nonparametric Statistics*, *28*(2), 250–271. https://doi.org/10.1080/10485252.2016.1163349
Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674. https://doi.org/10.1198/106186006X133933
Hufe, P., Peichl, A. & Weishaar, D. (2022). Lower and upper bound estimates of inequality of opportunity for emerging economies. *Soc Choice Welf* **58,** 395–427 (2022). https://doi.org/10.1007/s00355-021-01362-7
Liang, J. (2021). *Package "ShapleyValue": Shapley Value Regression for Relative Importance of Attributes Version 0.2.0*. CRAN. https://cran.r-project.org/web/packages/ShapleyValue/ShapleyValue.pdf
Monroy-Gómez-Franco, L., & Corak, M. (2019). A Land of Unequal Chances: Social Mobility Across Mexican Regions. *Eighth Meeting of the Society for the Study of Economic Inequality* (ECINEQ). Paris School of Economics, July 3-5, 2019. http://www.ecineq.org/ecineq_paris19/papers_EcineqPSE/paper_446.pdf
Monroy-Gómez-Franco, L., & Vélez-Grajales, R. (2020). Skin Tone Differences in Social Mobility in Mexico: Are We Forgetting Regional Variance? *Journal of Economics, Race, and Policy*. https://doi.org/10.1007/s41996-020-00062-1
Monroy-Gómez-Franco, L., & Vélez-Grajales, R. (2021). Skin Tone Differences in Social Mobility in Mexico: Are We Forgetting Regional Variance? *Journal of Economics, Race, and Policy*, *4*(4), 257–274. https://doi.org/10.1007/s41996-020-00062-1
Monroy-Gómez-Franco, L., Vélez-Grajales, R., & Yalonetzky, G. (2021). Layers of Inequality: Unequal Opportunities and Skin Color in Mexico. *Review of Black Political Economy*. https://doi.org/10.1177/00346446211044149
Plassot, T., Rubio, G., & Soloaga, I. (2019). *Movilidad social intergeneracional y desigualdad de oportunidades en México. Educación y activos: un enfoque territorial*. https://ceey.org.mx/wp-content/uploads/2020/03/07-Plassot-Rubio-y-Soloaga-2019.pdf
Plassot, T., Soloaga, I., & Torres, P. (2022). Inequality of Opportunity in Mexico and its Regions: A Data-Driven Approach. *The Journal of Development Studies*, 1–17. https://doi.org/10.1080/00220388.2022.2055465

Rawls, J. (1971). A Theory Of Justice. In *Oxfor University Press*. https://doi.org/10.4324/9781315097176

Redell, N. (2019). *Shapley Decomposition of R-Squared in Machine Learning Models*. https://arxiv.org/abs/1908.09718v1

Roemer, J. E. (1998). Equality of Opportunity. In *Cambridge, MA: Harvard*. Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674004221

Roemer, J. E. (2002). Equality of opportunity: A progress report. *Social Choice and Welfare*, *19*(2), 455–471. https://doi.org/10.1007/s003550100123

Roemer, J. E. (2004). Equal opportunity and intergenerational mobility: Going beyond intergenerational income transition matrices. *Generational Income Mobility in North America and Europe* (pp. 48–57). Cambridge University Press. https://doi.org/10.1017/CBO9780511492549.004

Salas-Rojo, P., & Rodríguez, J. G. (2022). Inheritances and wealth inequality: a machine learning approach. *Journal of Economic Inequality*. https://doi.org/10.1007/s10888-022-09528-8

Shavit, Y., & Blossfeld, H. P. (1994). Persistent Inequality: Changing Educational Attainment in Thirteen Countries. *British Journal of Educational Studies* 42 (4):413-415.

Solís, P. (2012). Social mobility in Mexico Trends, Recent Findings and Research Challenges. *Trace. 62. 7-20*. 10.22134/trace.62.2012.454.

Souza, W., Annegues, A., & Oliveira, V. (2017). Thoughts on the inequality of opportunities: New evidence. *CEPAL Review*. 2017. 103-121. 10.18356/7b2d8869-en.

Suárez Álvarez, A., & López Menéndez, A.J. (2018). Assessing Changes over Time in Inequality of Opportunity: The Case of Spain, *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, Springer, vol. 139(3), pages 989-1014, October.

Torche, F. (2020). Change in Intergenerational Mobility in Mexico: A Synthetic Cohort Analysis», *Papers 2020_06*, Centro de Estudios Espinosa Yglesias.

Vélez Grajales, R., Monroy-Gómez-Franco, L. A., & Yalonetzky, G. (2019). Inequality of Opportunity in Mexico. *Journal of Income Distribution*, *27*(3-4), 134–158. https://doi.org/10.25071/1874-6322.40429

Wendelspiess Chávez Juárez, F., & Soloaga, I. (2013). Scale vs. Translation Invariant Measures of Inequality of Opportunity when the Outcome is Binary. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2226822

Wendelspiess Chávez Juárez, F. (2015). Measuring Inequality of Opportunity with Latent Variables. *Journal of Human Development and Capabilities*, 16:1, 106-121, DOI: 10.1080/19452829.2014.907247

Appendix

A. Literature on Inequality of Opportunity in wealth in Mexico

Data	Source	Advantage	Circumstances		Absolut			Relative		
					Gini	MLD	DI	R2	Gini	MLD
EMOVI 2017	Monroy-Gómez-Franco, L., & Corak, M. (2019)	Asset Index	Indigenous status, skin tone, maximum parental education attainment, household wealth index of the origin household, characteristics of the neighborhood, urban or rural community of origin, region of origin and sex of the respondent.	**Lb**				35%		
				Ub				45%		
EMOVI 2017	Plassot, T., Rubio, G., & Soloaga, I. (2019)	Asset Index	Maximum parental education attainment, wealth index of the household at age 14, urban or rural area at age 14, region at age 14, age, living with both parents at age 14, private school, migration, household´s size at age 14, and sex of the respondent.	**Lb**			0,18			
				Ub			0,43			
EMOVI 2017	Plassot, T., Soloaga, I., & Torres, P. (2022)	Asset Index	Parents speaking indigenous language, skin tone, average parental education attainment, wealth index of the household at age 14, urban or rural at age 14, region at age 14 and sex of the respondent.	**Lb**	0,17	0,05				20%
				Ub	0,26	0,15				56%
MMSI 2016	Monroy-Gómez-Franco, L., Vélez-Grajales, R., & Yalonetzky, G. (2021)	Asset Index	Parents speaking indigenous language, skin tone, maximum parental education attainment, household wealth index of the origin household, urban or rural community of origin, father was agricultural worker and sex of the respondent.	**Lb**				4%		
				Ub				42%		
EMOVI 2011	Vélez Grajales, R., Monroy-Gómez-Franco, L. A., & Yalonetzky, G. (2019)	Asset Index	Parents speaking Indigenous language, father and mother education attainment, household wealth index of the origin household, urban or rural community of origin, father was agricultural worker and sex of the respondent.	**Lb**		0,06		29%		28%
				Ub		0,08		37%		36%
ENCEL 1999-2009	Hufe, P., Peichl, A. & Weishaar, D. (2022)	Individual and household income	Speaking Indigenous language, age and sex of the respondent.	**Lb**	0,11	0,03			35%	13%
				Ub	0,14	0,03			44%	17%
MXFLS 2009	EqualChances.org	Household income	Parental education, parental occupation and origin (e. i. Race, ethnic origin, area of birth).	**Lb**	0,11				19%	
				Ub	0,16				29%	
EMOVI 2006	Wendelspiess Chávez Juárez, F. (2015)	Log income and Asset Index	Indigenous self-classification, father and mother education attainment, household of origin wealth index, parents own a house, and sex of the respondent.	**Lb**				17%		
				Ub				36%		

Note: Lb refers to Lower-bound, Ub refers to Upper-bound. DI refers to Dissimilarity Index, MLD for Mean Log Deviation. All estimates of IOp adopt an ex-ante compensation criterion, except for Plassot, Torres, and Soloaga (2022) where the upper-bound estimates are the result of ex-post compensation criterion. The relative measures correspond to the quotient of IOp on total inequality.

B. Selection & Training of the Model

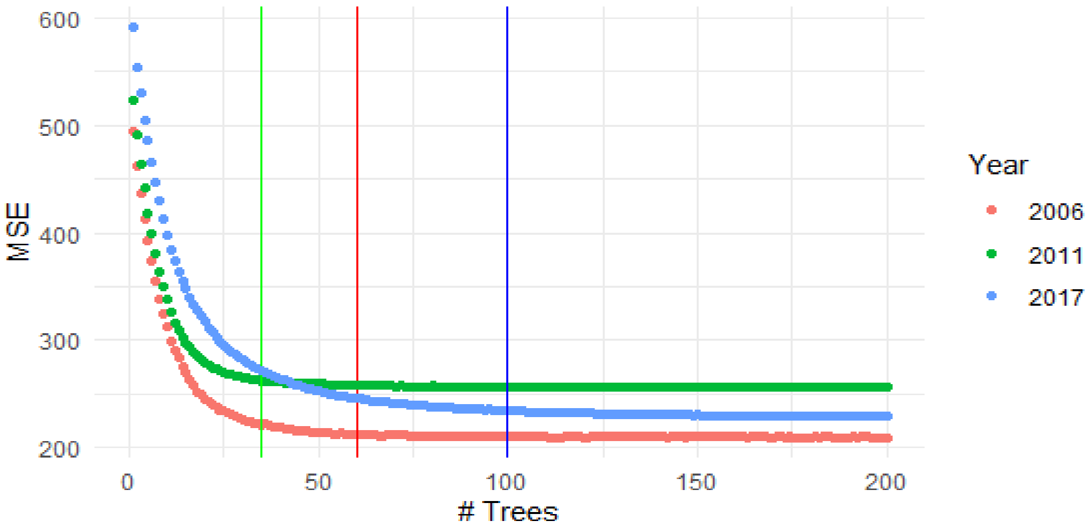
B1. Selection Metrics of the Model

To show the relative performance of each method, we compare the results of using three different methods (OLS, Conditional Inference Trees, and Conditional Inference Forests) for the estimation of the counterfactual outcome for year 2017. Conditional Inference Trees and Forests have two main advantages over OLS estimations, the first one being its direct application in the definition of types as defined in Roemer's (1998) theory and the second one being the number of observations that are used due to the procedure of surrogate splitting (missing values) as explained in Section 2 above. Furthermore, as discussed by Brunori et al. (2019, 2021) we find Conditional Inference Forests to be more accurate than both OLS and Conditional Inference Trees.

	Observations	MSE	MAE	RMSE	R2
OLS	11,870	561.25	19.68	23.69	0.31
C.I. Tree	16,457	588.72	20.15	24.263	0.27
C.I. Forest	16,457	541.24	19.39	23.265	0.33

Notes: Table show some accuracy results as a measure of performance of different algorithms for the year 2017. The first row shows the results for Ordinary Least Squares, the second one for Conditional Inference Trees and the third one for Conditional Inference Random Forests.

B2. Decrease of MSE by number of Trees



Notes: The figure shows the relationship of the mean squared error and the number of trees in each forest for years 2006, 2011 and 2017.

C. Wealth Indexes

C1. Multiple Correspondence Analysis, Inertia explained by each factor

		Wealth at age of 14			Actual Wealth		
	Dimension	Principal Inertia	Percent	Cumulative	Principal Inertia	Percent	Cumulative
2017	1	0,132	93,6	93,6	0,056	87,6	87,6
	2	0,001	0,9	94,6	0,002	3,6	91,2
	Total	0,141	100		0,064	100	
2011	1	0,102	90,5	90,5	0,049	84,5	84,5
	2	0,002	2,1	92,6	0,003	5,4	90,0
	Total	0,113	100		0,058	100	
2006	1	0,169	94,0	94,0	0,057	83,6	83,6
	2	0,001	0,7	94,7	0,005	7,2	90,8
	Total	0,180	100		0,068	100	

Notes: Estimations using the EMOVI. Principal Inertia is the inertia explained by the axis or dimension k. Percent is the percentage of the total inertia explained by the dimension k. We only represent the first two dimensions.

C2.Multiple Correspondence Analysis: Eigenvalues and proportion of the variance accounted for by the factors

	2006 (Actual Wealth)					
	Overall			1st Dimmension		
	Mass	Quality	% Inertia	Mass	Quality	% Inertia
Stove						
No	0.005	0.897	0.047	2.712	0.629	0.035
Yes	0.079	0.897	0.003	-0.165	0.629	0.002
Electricity						
No	0.001	0.969	0.016	3.143	0.552	0.011
Yes	0.082	0.969	0.000	-0.042	0.552	0.000
Fridge						
No	0.010	0.867	0.071	2.395	0.697	0.059
Yes	0.073	0.867	0.010	-0.340	0.697	0.008
Washing machine						
No	0.022	0.904	0.068	1.733	0.830	0.067
Yes	0.061	0.904	0.025	-0.637	0.830	0.025
Phone						
No	0.040	0.957	0.047	1.157	0.946	0.053
Yes	0.044	0.957	0.042	-1.045	0.946	0.048

	2006 (Wealth at age 14)					
	Overall			1st Dimmension		
	Mass	Quality	% Inertia	Mass	Quality	% Inertia
Tubing water						
No	0.051	0.930	0.082	1.254	0.927	0.081
Yes	0.060	0.930	0.071	-1.081	0.927	0.070
Toilet inside the house						
No	0.062	0.941	0.069	1.062	0.940	0.069
Yes	0.050	0.941	0.086	-1.318	0.940	0.086
Electricity						
No	0.034	0.947	0.084	1.586	0.941	0.084
Yes	0.078	0.947	0.037	-0.687	0.941	0.037
Stove						
No	0.054	0.933	0.081	1.213	0.931	0.080
Yes	0.057	0.933	0.078	-1.166	0.931	0.077
Washing machine						
No	0.086	0.956	0.027	0.570	0.953	0.028
Yes	0.025	0.956	0.093	-1.937	0.953	0.095

Cellular phone							**Domestic service**						
No	0.047	0.970	0.039	0.978	0.968	0.045	No	0.107	0.969	0.001	0.111	0.902	0.001
Yes	0.036	0.970	0.051	-1.278	0.968	0.059	Yes	0.005	0.969	0.032	-2.603	0.902	0.031
TV							**TV**						
No	0.006	0.966	0.031	2.032	0.711	0.026	No	0.062	0.962	0.060	0.991	0.962	0.061
Yes	0.077	0.966	0.003	-0.168	0.711	0.002	Yes	0.049	0.962	0.077	-1.269	0.962	0.078
Cable TV							**Phone**						
No	0.066	0.938	0.022	0.609	0.914	0.024	No	0.094	0.952	0.014	0.384	0.940	0.014
Yes	0.018	0.938	0.084	-2.282	0.914	0.091	Yes	0.017	0.952	0.079	-2.183	0.940	0.079
Computer							**Bank account**						
No	0.069	0.859	0.025	0.604	0.826	0.025	No	0.105	1.011	0.002	0.122	0.948	0.002
Yes	0.014	0.859	0.121	-2.881	0.826	0.120	Yes	0.006	1.011	0.026	-2.063	0.948	0.026
Internet													
No	0.075	0.858	0.012	0.400	0.808	0.012							
Yes	0.008	0.858	0.117	-3.749	0.808	0.113							
Bank account													
No	0.071	0.923	0.011	0.402	0.882	0.012							
Yes	0.012	0.923	0.066	-2.414	0.882	0.069							
Credit card													
No	0.076	0.913	0.008	0.337	0.863	0.009							
Yes	0.008	0.913	0.081	-3.274	0.863	0.083							

	2011 (Actual Wealth)							2011 (Wealth at age 14)					
	Overall			1st Dimmension				Overall			1st Dimmension		
	Mass	Quality	% Inertia	Mass	Quality	% Inertia		Mass	Quality	% Inertia	Mass	Quality	% Inertia
Stove							**Tubing water**						
No	0.006	0.882	0.055	2.800	0.689	0.045	No	0.037	0.916	0.106	1.704	0.910	0.107
Yes	0.078	0.882	0.004	-0.205	0.689	0.003	Yes	0.074	0.916	0.053	-0.844	0.910	0.053
Electricity							**Toilet inside house**						
No	0.002	0.959	0.015	2.397	0.588	0.010	No	0.058	0.951	0.068	1.114	0.951	0.072
Yes	0.082	0.959	0.000	-0.052	0.588	0.000	Yes	0.053	0.951	0.074	-1.205	0.951	0.078
Fridge							**Electricity**						
No	0.011	0.887	0.072	2.470	0.778	0.066	No	0.022	0.913	0.111	2.249	0.901	0.110
Yes	0.073	0.887	0.011	-0.368	0.778	0.010	Yes	0.089	0.913	0.027	-0.548	0.901	0.027
Washing machine							**Stove**						
No	0.026	0.929	0.070	1.694	0.905	0.075	No	0.042	0.922	0.103	1.575	0.920	0.105
Yes	0.057	0.929	0.032	-0.767	0.905	0.034	Yes	0.069	0.922	0.064	-0.970	0.920	0.065
Phone							**Washing machine**						
No	0.057	0.939	0.031	0.774	0.925	0.034	No	0.082	0.943	0.030	0.618	0.934	0.031
Yes	0.026	0.939	0.070	-1.718	0.925	0.076	Yes	0.029	0.943	0.087	-1.770	0.934	0.090
Cellular phone							**Domestic service**						
No	0.036	0.975	0.049	1.246	0.974	0.057	No	0.108	0.913	0.001	0.071	0.663	0.001
Yes	0.047	0.975	0.038	-0.968	0.974	0.044	Yes	0.003	0.913	0.024	-2.255	0.663	0.017
TV							**TV**						

No	0.003	0.896	0.040	3.086	0.629	0.030
Yes	0.080	0.896	0.002	-0.121	0.629	0.001
Cable TV						
No	0.064	0.939	0.024	0.635	0.919	0.026
Yes	0.019	0.939	0.079	-2.107	0.919	0.086
Computer						
No	0.060	0.843	0.048	0.888	0.824	0.047
Yes	0.024	0.843	0.123	-2.257	0.824	0.120
Internet						
No	0.066	0.828	0.035	0.711	0.799	0.034
Yes	0.017	0.828	0.138	-2.768	0.799	0.130
Bank account						
No	0.082	1.003	0.000	0.042	0.903	0.000
Yes	0.001	1.003	0.011	-3.373	0.903	0.012
Credit card						
No	0.075	0.990	0.005	0.278	0.952	0.006
Yes	0.008	0.990	0.049	-2.636	0.952	0.055

No	0.038	0.917	0.108	1.699	0.913	0.109
Yes	0.073	0.917	0.055	-0.871	0.913	0.056
Phone						
No	0.099	0.935	0.008	0.283	0.879	0.008
Yes	0.013	0.935	0.064	-2.223	0.879	0.062
Bank account						
No	0.110	0.895	0.000	0.035	0.557	0.000
Yes	0.001	0.895	0.017	-2.749	0.557	0.010

	2017 (Actual Wealth)					
	Overall			1st Dimmension		
	Mass	Quality	% Inertia	Mass	Quality	% Inertia
Stove						
No	0.005	0.888	0.062	3.318	0.772	0.055
Yes	0.078	0.888	0.004	-0.210	0.772	0.003
Electricity						
No	0.001	0.993	0.010	2.818	0.693	0.008
Yes	0.082	0.993	0.000	-0.036	0.693	0.000
Fridge						
No	0.006	0.879	0.073	3.254	0.786	0.065
Yes	0.077	0.879	0.006	-0.260	0.786	0.005
Washing machine						
No	0.017	0.939	0.076	2.151	0.913	0.079
Yes	0.066	0.939	0.019	-0.551	0.913	0.020
Phone						
No	0.051	0.907	0.044	0.933	0.893	0.044
Yes	0.032	0.907	0.069	-1.477	0.893	0.070
Cellular phone						
No	0.012	0.983	0.053	2.163	0.963	0.058
Yes	0.071	0.983	0.009	-0.378	0.963	0.010
TV						
No	0.012	1.003	0.043	1.961	0.973	0.048
Yes	0.071	1.003	0.007	-0.342	0.973	0.008
Cable TV						
No	0.040	0.972	0.047	1.146	0.969	0.052
Yes	0.044	0.972	0.043	-1.038	0.969	0.047

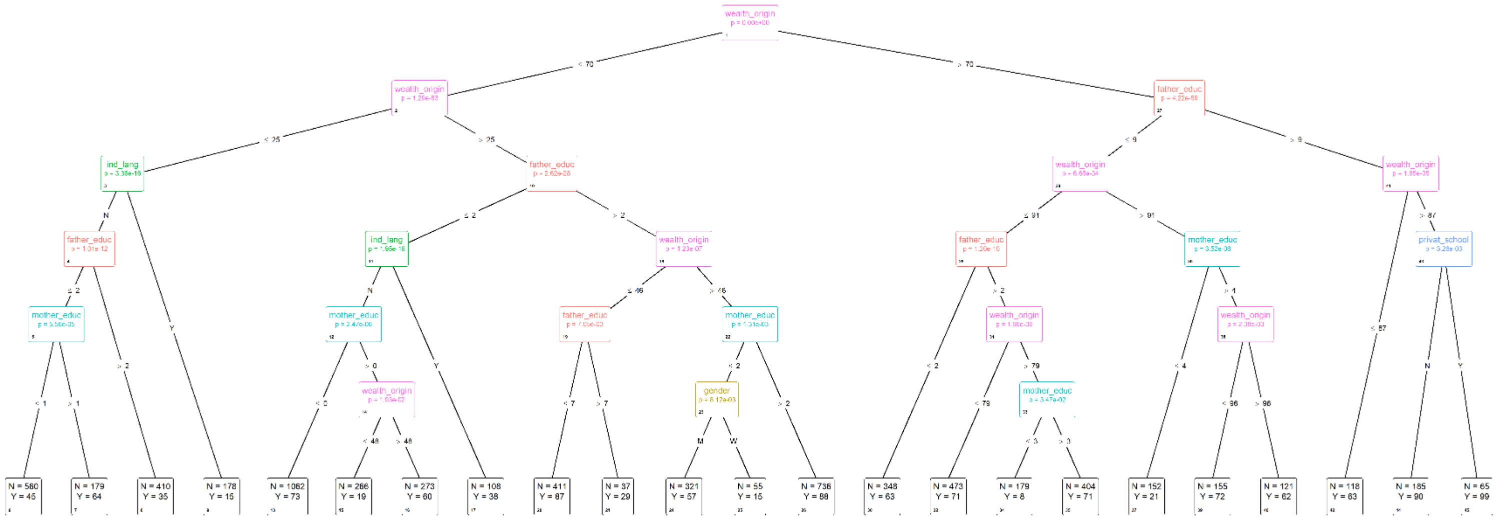
	2017 (Wealth at age 14)					
	Overall			1st Dimmension		
	Mass	Quality	% Inertia	Mass	Quality	% Inertia
Tubing water						
No	0.038	0.935	0.094	1.559	0.932	0.093
Yes	0.073	0.935	0.050	-0.826	0.932	0.050
Toilet inside house						
No	0.048	0.948	0.080	1.294	0.947	0.081
Yes	0.063	0.948	0.062	-0.998	0.947	0.062
Electricity						
No	0.018	0.950	0.089	2.208	0.939	0.090
Yes	0.093	0.950	0.018	-0.438	0.939	0.018
Stove						
No	0.040	0.937	0.092	1.511	0.934	0.092
Yes	0.071	0.937	0.053	-0.860	0.934	0.052
Washing machine						
No	0.074	0.955	0.043	0.766	0.953	0.044
Yes	0.037	0.955	0.087	-1.551	0.953	0.088
Domestic service						
No	0.104	0.975	0.002	0.133	0.892	0.002
Yes	0.007	0.975	0.028	-1.917	0.892	0.027
TV						
No	0.041	0.940	0.089	1.471	0.937	0.089
Yes	0.070	0.940	0.052	-0.863	0.937	0.052
Phone						
No	0.088	0.951	0.021	0.493	0.938	0.021
Yes	0.023	0.951	0.080	-1.860	0.938	0.081

Computer							**Bank account**						
No	0.056	0.892	0.043	0.875	0.873	0.043	No	0.097	0.946	0.007	0.272	0.903	0.007
Yes	0.027	0.892	0.088	-1.788	0.873	0.088	Yes	0.014	0.946	0.053	-1.937	0.903	0.051
Internet													
No	0.047	0.862	0.070	1.194	0.849	0.067							
Yes	0.036	0.862	0.092	-1.572	0.849	0.089							
Bank account													
No	0.064	0.898	0.016	0.496	0.862	0.016							
Yes	0.019	0.898	0.055	-1.682	0.862	0.054							
Credit card													
No	0.070	0.892	0.011	0.391	0.845	0.011							
Yes	0.013	0.892	0.061	-2.143	0.845	0.059							

Notes: Estimations using the EMOVI. "Quality" is a measure of the representativeness of each item by the components, values near 1 reflect more representation. "Mass" corresponds to the proportion or weight of each category. It can be represented as the frequency of the category of a variable divided by the sum of the frequencies of all the categories of all variables. The sum of the mass for all categories of all variables equals 1. Finally, "% Inertia" is the percentage of the total inertia explained by the category of each variable. "Coord" refers to the coordinate of the category on the dimension k. "Sqcorr" is the square correlation of the category with the dimension k. "Contribution" is the proportion of the inertia on the dimension k that is explained by this category.

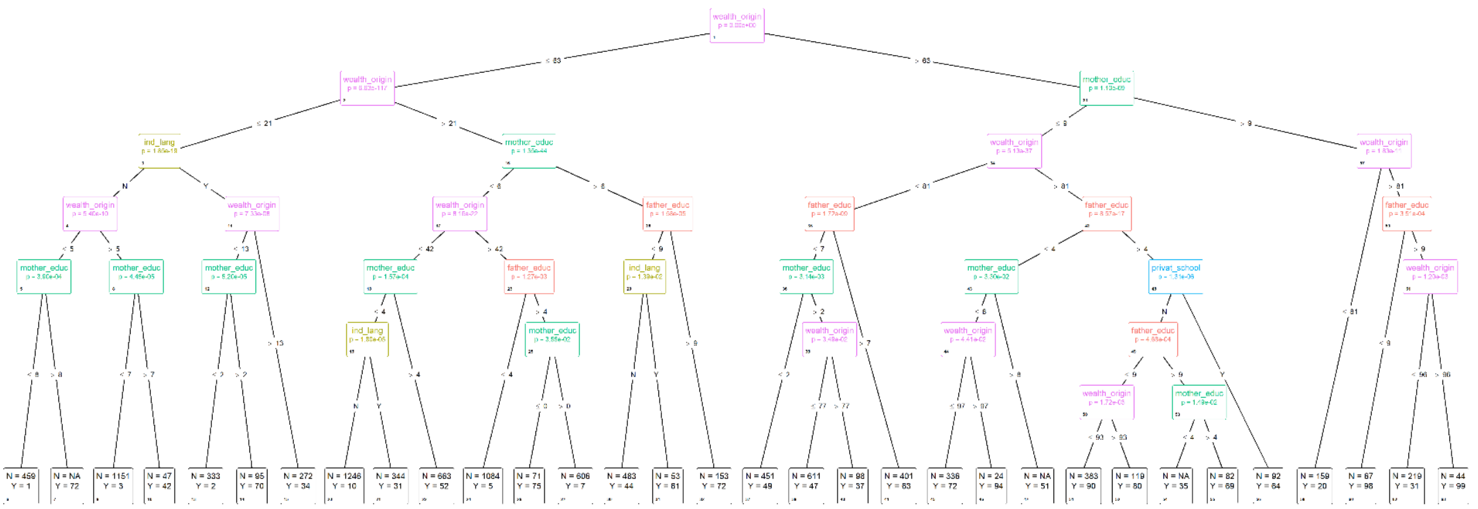
D. Trees

D1. 2006



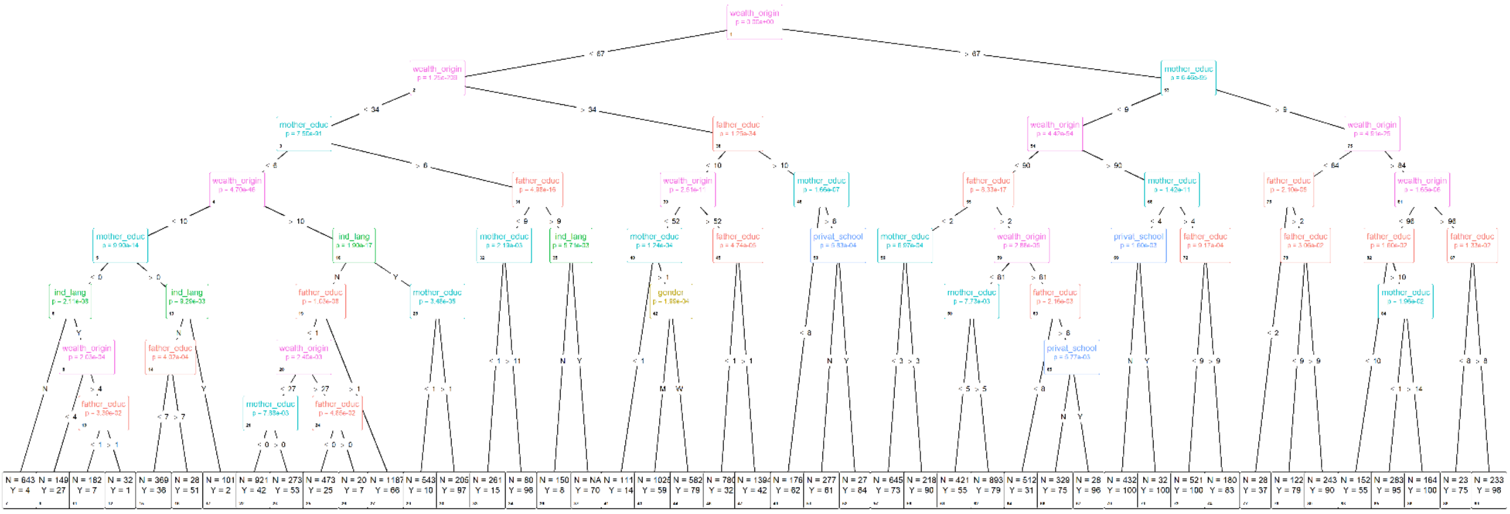
Notes: Figure shows the structure of the tree for 2006. Colors are: i) Wealth at Origin (pink); ii) Ability of parents to speak an indigenous language (green); iii) Father's years of schooling (red); iv) Mother's years of schooling (light blue); v) Respondent's sex (yellow); vi) Private School (blue).

D2. 2011



Notes: Figure shows the structure of the tree for 2011. Colors are: i) Wealth at Origin (pink); ii) Ability of parents to speak an indigenous language (yellow); iii) Father's years of schooling (red); iv) Mother's years of schooling (green); v) Private School (light blue).

D3. 2017



Notes: Figure shows the structure of the tree for 2017. Colors are: i) Wealth at Origin (pink); ii) Ability of parents to speak an indigenous language (green); iii) Father's years of schooling (red); iv) Mother's years of schooling (light blue); v) Respondent's sex (yellow); vi) Private School (blue).