Inequality of Opportunity in Mexico and its Regions: A Data-Driven Approach

Thibaut Plassot * Isidro Soloaga * Pedro Torres *

April 1, 2021

Abstract

This research proposes a first approximation of Inequality of Opportunity (IOp) in Mexico based on a concept of ex-post compensation, fully consistent with Roemer´s approach of Inequality of Opportunity. This framework considers the advantage reached by an individual to be determined by the circumstances at origin and by the effort exerted. Following Brunori and Neidhöfer (2020), we construct a data-driven procedure using regression trees to identify types based on circumstances. To identify degrees of effort, an algorithm estimates the distribution of outcome in each type based on coefficients of Bernstein polynomials. Our results underline the differences, in terms of opportunities, faced by individuals, based on the territory in which they grew up, the household context, and their sex. We find that the education and the wealth of parents are the principal circumstances that shape their trajectories. Importantly, territorial variables are significant determinants of IOp among the most disadvantaged at origin, but they hold less importance for the most advantaged. IOp is lower in urban areas, in the Northern region and in Mexico City compared to other regions. The Southern region and rural territories are the most unequal. We also estimate that the weight of IOp in terms of total inequality varies according if we adopt an ex-post or an ex-ante approach, which can lead to different conclusions.

JEL: C14; C81; D63

Keywords: Decision Trees; Inequality Of Opportunity; Mexico

* *Universidad Iberoamericana, Department of Economics, Prolongacion Paseo de la Reforma 880, Lomas de Santa Fe*, Mexico City, 01219.

Introduction

Research shows that people favor equality of opportunity, in the sense of leveling the playing field, over an equality of outcomes (Breen, 2010). In this sense, the literature on Inequality of Opportunity (herein referred to as IOp) focuses on two ethical principles: the compensation principle and the reward principle. The former sustains that inequalities due to circumstances are not acceptable and must be compensated, while the latter adds the notion of individual responsibility to egalitarian thought and advocates increase the recompense for any additional effort.

In this study, we follow a methodology suggested by Brunori and Neidhöfer (2020) to estimate ex-post IOp measures for Mexico, which are fully consistent with Roemer´s inequality of opportunity framework. The works of Roemer (1998a) contribute to the philosophical literature on equality and justice developed by, among others, Arneson (1989); Cohen (1989); Dworkin (1981a,b); Fleurbaey (1995) and Rawls (1971) covering the notions of choices, preferences, and circumstances (Kanbur and Wagstaff, 2014). The conceptual framework provided by Roemer considers that access to an advantage depends on circumstances and effort. While circumstances are, for example, the family background or the neighborhood environment, being elements "for which the society in question does not wish to hold individuals responsible"; effort contains the factors "for which the society does hold the individual responsible" (Roemer, 2004, p. 49). By aggregating individuals in types, defined as groups of people who share equal circumstances, equality of opportunity is achieved 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).

Importantly, Roemer´s approach considers that the preferences and levels of effort of individuals are influenced by circumstances, differentiating levels of effort from degrees of effort. Unlike levels of effort, degrees of effort are not influenced by circumstances at origin. Roemer proposes approximating degrees of effort by estimating quantiles of the distribution of outcome for each type. Two important assumptions in this method are that all circumstances (i.e., all types) are identified and are exogenous to the control of each individual, and that the effort exerted and the advantage achieved are positively associated.

To measure IOp, researchers generate counterfactual distributions. For example, Ferreira and Peragine (2015) remove fair inequality based on effort and apply an inequality indicator to the new distribution in order to measure any remaining unfair inequality. The compensation principle can be approached using two methods: ex-ante and ex-post. In the former, attention is focused on inequalities between individuals with different circumstances, the latter focuses on individuals with the same degree of effort. Even if both approaches seek to quantify inequalities due to circumstances, they are not comparable due to different methodological approaches (Fleurbaey and Peragine, 2013; Ramos and Van de Gaer, 2016; Roemer, 2002).

With regard to the reward principle, the two main definitions in the literature are the liberal reward and the utilitarian reward.[1] A wealth of literature underlines the tensions between the ex-post compensation and the different reward principles, whereas the ex-ante approach does not clash with these principles (Ferreira and Peragine, 2015; Fleurbaey and Peragine, 2013; Ramos and Van de Gaer, 2016). Nevertheless, the ex-ante approach does not allow for the direct consideration of the effort dimension, which represents the "fundamental ethical principle of Roemer's theory of equal opportunity, stating that individuals exerting the same effort should obtain the same outcome" (Brunori and Neidhöfer, 2020, p. 5). A majority of studies measure IOp from an ex-ante point of view, perhaps due to difficulties of obtaining data that explicitly measures effort.[2] Moreover, even in countries like Mexico where research into social mobility and inequality of opportunity has increased considerably in re-

[1]As mentioned by Ferreira and Peragine (2015), although liberal and utilitarian are the most used principles, other formulations have been proposed, among them, the inequality averse reward, the arithmetic average reward, and the minimal reward.

[2]Researchers look for data that could reflect effort, such as, for example, IQ tests Roemer (2002) or height at mid-life for ability Deary et al. (2005). For Mexico, the 2015 Social Mobility Survey (compiled by the *Espinosa Yglesias Center for Studies*) encompasses the measurement of cognitive and non-cognitive social-emotional abilities, in addition to preferences Campos Vázquez (2015). When effort is not directly measurable, Brunori and Neidhöfer (2020) propose a method to approximate degrees of effort as defined by Roemer, which is the approach we follow here.

cent decades, there is, to our knowledge, no existing measurement of ex-post IOp. All studies adopt an ex-ante approach (Monroy-Gómez-Franco and Corak, 2020; Plassot et al., 2019; Wendelspiess Chávez Juárez and Soloaga, 2013).

This study complements the existing literature and presents new measurements of IOp for Mexico at a sub-national level. We follow Brunori and Neidhöfer (2020) and construct a data-driven procedure for the identification of types of individuals based on their circumstances at age 14. We use regression trees to provide a graphical representation of the structure of opportunities in the society in question. To calculate ex-post IOp, we follow Brunori and Neidhöfer (2020) and focus on inequality between individuals with the same degree of effort. Through this approach, we measure degrees of effort using an algorithm that estimates the distribution of outcome in each type based on coefficients of Bernstein polynomials. To calculate ex-ante IOp, we follow Ferreira and Gignoux (2011) and focus on inequality between types. In both the ex-ante and the ex-post approaches, we use the Gini coefficient and Mean Log Deviation (MLD) as measurements of inequality. We replicate Cecchi et al. (2015) by using MLD to present the relative share of total inequality that can be attributed to IOp, as well as the proportion that can be attributed to differences in effort. All estimates are presented at a national level, encompassing urban and rural areas, Mexico City and five large regions.

Results show that the major circumstances that contribute to IOp are the education and the wealth of parents. Other circumstances, such as the sex of the respondent, the region they are from, and the rural/urban area of residence at age 14 are also determinants in the trajectories of individuals, although the relative importance of these variables varies in each territory. Importantly, territorial variables are significant factors in distinguishing types within the most disadvantaged at origin, but they are less important for the most advantaged. Our measures found that IOp and the number of types are higher in the south than those in other regions, while they are lower in the north and in Mexico City. IOp is higher among individuals who grew up in rural areas than in urban area, while the number of types is higher in urban areas. This is to say that more circumstances are taken into account when the structure of opportunities is measured for inhabitants of urban areas at age 14, which may reflect a less polarized society in urban areas over rural areas.[3] Using an ex-post approach, we estimate that IOp contributes to more than 50% of total inequality; whereas, with an ex-ante method, this figure stands at around 20%. Thus, our study provides evidence that the choice of the method used to approximate IOp has strong implications and can lead to different interpretations of the results.

In Section 1 we present the conceptual framework on which we have based our study. Section 2 presents the theoretical and methodological aspects of the measurement of IOp through both an ex-ante and an ex-post approach. Section 3 describes the source of information. Section 4 shows our results. Section 5 explores the robustness of our estimations. Section 6 presents conclusions.

1 Conceptual Framework

In this study, we consider the advantage achieved by a person (in this case a wealth index) as a function of a vector of circumstances and a variable that approximates effort. Society is divided into types, which are based on circumstances, and tranches, based on their degree of effort.

Circumstances are factors that are beyond personal responsibility, while effort represents those aspects for which individuals can be held responsible after taking her/his type into consideration. Consequently, the first step for research is to distinguish which factors belong to circumstances. For example, Roemer (2004) identified and ordered four channels of influence of the parents and the future advantage this may give their child: their network of social connections, their beliefs and skills, their ability, and, finally, their preferences and aspirations. Thus, there are four versions of Equality of Opportunity, ranging from the least to the most egalitarian, depending on whether the channels are considered to be part of the responsibility of the

[3] We believe this is one of the main advantages of our approach: it provides an opportunity to explore only statistically meaningful combinations of circumstances.

individual or not[4]. For example, a meritocratic approach holds individuals responsible for their abilities; in other words, it does not consider the genetic transmission of abilities as unfair (Roemer, 2004). In the same way, the approach employed by Dworkin (1981a,b) considers factors influenced by preferences as part of individual responsibility. On the contrary, Arneson (1989); Cohen (1989) or Roemer (1998a) affirm that preferences are also determined by circumstances at origin. Besides, Roemer differentiates the preferences influenced by the parents, which can be considered to be circumstances, from the autonomous preferences, which determine the individual´s effort and are orthogonal to (i.e., do not depend on) the characteristics of the parent.

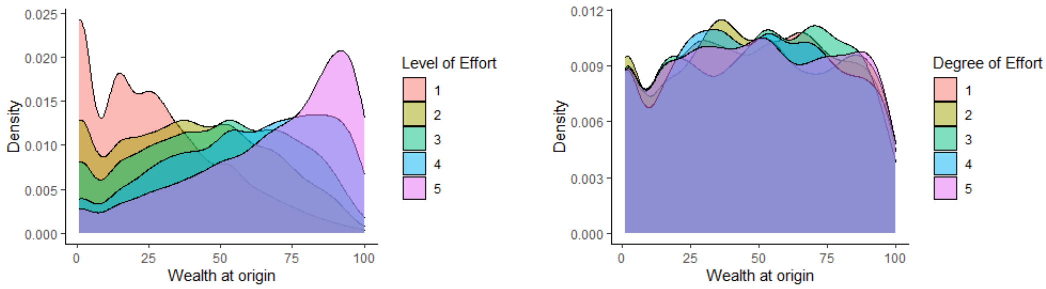
Roemer separates the levels of effort from the degrees of effort. The level of effort is measured generating quantiles covering the overall distribution of wealth at destination. The degree of effort is measured as the quantiles of wealth at destination after controlling for parental influence, and it is measured by generating quantiles for each type-specific distribution. We use the 2017 EMOVI survey in Mexico to show that level of effort and wealth at origin have a strong positive correlation (Figure 1 Panel (a)), whereas the degree of effort is somehow independent from wealth at origin (Figure 1 Panel (b)). In this research, we follow Roemer in calculating degrees of effort and considering that equality of opportunity is reached when all individuals who exert the same degree of effort have an equivalent outcome. This situation is represented in Figure 2,where higher degrees of effort equate to a higher outcome, and a situation in which degree of effort is orthogonal to circumstances.

To measure IOp, two approaches were predominantly developed, based on whether the compensation principle applies before or after observing effort: the ex-ante and ex-post compensation. While the first focuses on the inequality between-types, the second focuses on the inequality between individuals who exert the same degree of effort, known as within-effort (within-tranche) inequality. In both approaches, the

[4]Channels are ordered according to Roemer. Each version of IOp adds a channel of influence to the set of circumstances. For example, the first version only considers that the network of social connection is beyond the individual´s control, while the fourth version of IOp (most radical) considers all four channels to be beyond the individual's control.

Figure 1: Distribution of Wealth at origin according effort

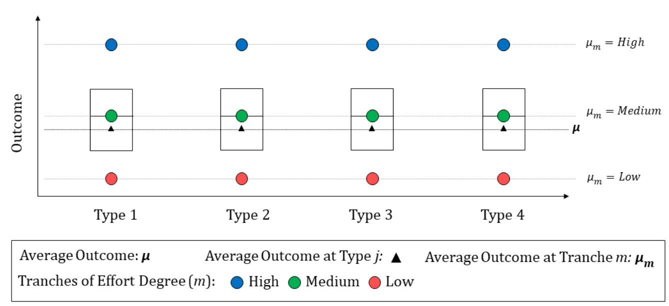
(a) Effort levels (b) Effort degrees



Notes: Estimations show the density of wealth at origin according to effort in Mexico using data from EMOVI 2017. Levels of effort are approximated by generating quantiles for the overall distribution, and degrees of effort are identified by generating quantiles for each type-specific distribution. Panel (a) shows a situation in which the average level of effort differs across types; Panel (b) shows a situation in which the degree of effort is similar across types.

inequality explained by circumstances can be calculated by contrasting the current distribution with a modified distribution. In the ex-ante approach, the modified distribution is obtained by removing the within-type inequality; for example, after assigning the mean of his/her type to each individual. In the ex-post approach, the modified distribution is obtained by removing the inequality between-tranches (between degrees of effort).

Figure 2: Counterfactual distribution



Notes: Figure shows a distribution in which all types present the same average outcome and in which individuals from the same tranche of effort have a similar outcome. In this situation, there is no more inequality due to circumstances, with the remaining inequality being due to effort.

2 Measurement of Inequality of Opportunity

By following Checchi and Peragine (2005), it is possible to define the outcome y of an individual i determined by a function g according to a vector of circumstances C beyond individual control, and a variable e that approximates the effort exerted by this individual.

$$y_{ijm} = g(C_j, e_m)$$

The outcome of a person i with Circumstances C_j and Effort e_m is defined by y_{ijm}. The vector C indexed by $j = 1, \ldots, k$ gives us the number k of types of persons with similar circumstances. The effort is portioned in n “tranches”, indexed by $m = 1, \ldots, n$. We can then construct a k x n dimensional matrix $|Y_{jm}|$ with k rows that represent the types, and n columns for the different tranches of effort.

$$|Y_{jm}| = \begin{bmatrix} & e_1 & e_2 & ... & e_n \\ C_1 & y_{11} & y_{12} & ... & y_{1n} \\ C_2 & y_{21} & y_{22} & ... & y_{2n} \\ ... & ... & ... & ... & ... \\ C_k & y_{k1} & y_{k2} & ... & y_{kn} \end{bmatrix}$$

2.1 IOp ex-ante (Types Approach)

As shown by Bourguignon et al. (2003); Checchi and Peragine (2005); Ferreira and Gignoux (2011); Monroy-Gómez-Franco and Corak (2020) and Soloaga and Wendelspiess (2010), it is possible to remove within-type inequality by creating a counterfactual distribution $|Y'_{BT}|$, where outcome y_{ijm} of an individual i in type j is replaced by the average outcome in its type

$$|Y'_{BT}| : y_{jm} = \mu_j, \forall j \in (1, k); \forall m \in (1, n)$$

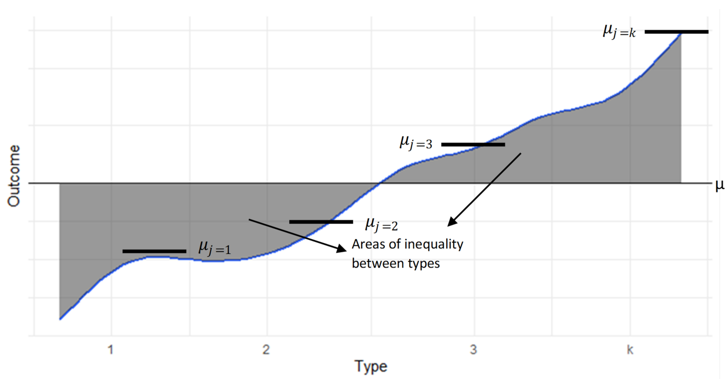
where μ_j is the mean outcome of individuals in type j, therefore.

$$|Y'_{BT}| = \begin{bmatrix} & e_1 & e_2 & ... & e_n \\ C_1 & \mu_1 & \mu_1 & ... & \mu_1 \\ C_2 & \mu_2 & \mu_2 & ... & \mu_2 \\ ... & ... & ... & ... & ... \\ C_k & \mu_k & \mu_k & ... & \mu_k \end{bmatrix}$$

If we apply this transformation to all individuals, the counterfactual distribution of outcome can be represented as in Figure 3. The remaining inequality is the between-types inequality due to circumstances. We can use an inequality measure (for example a Gini coefficient, Mean Log Deviation, or a Dissimilarity Index) on the new distribution as an indicator of IOp ex-ante such that

$$IOP_{EA} = I(Y'_{BT})^5$$

Figure 3: Inequality of Opportunity between-types



Notes: Figure shows a situation in which inequality due to effort has been removed, with the remaining inequality being due to circumstances (IOp ex-ante). In each type, the outcome of individuals is now equal to the mean of the group. The distance between the mean of each group with regard to the mean outcome of all types is a measure of IOp ex-ante.

[5] EA stands for ex-ante. BT stands for between types.

Additionally, Checchi and Peragine (2005) show how total inequality can be decomposed as the sum of the IOp ex-ante (between-types) and of the inequality due to effort (within-types). They use mean logarithmic deviation (MLD) as a measure to satisfy path independent decomposition (Foster and Shneyerov, 2000).

2.2 IOp ex-post (Tranches Approach)

IOp ex-post focuses on within-tranches of effort inequalities, considering between-tranches inequalities acceptable because they are the result of differences in effort. To remove the between-tranches inequalities, researchers propose creating a counterfactual distribution $|Y'_{WT}|$ where outcome y_{ijm} of an individual i in type j and tranche effort m is multiplied by the ratio between the population mean outcome μ and the average outcome of his/her tranche μ_m (Brunori and Neidhöfer, 2020; Checchi and Peragine, 2005)

$$|Y'_{WT}| : y_{jm} = y_{jm} * \frac{\mu}{\mu_m}; \forall j \in (1, ..., k); \forall m \in (1, ..., n)$$

where μ is the population mean outcome and μ_m is the average outcome in tranche m and WT stands for within-tranches. The distribution $|Y'_{WT}|$ is then

$$|Y'_{WT}| = \begin{array}{cc} & \begin{array}{cccc} e_1 & e_2 & ... & e_n \end{array} \\ \begin{array}{c} C_1 \\ C_2 \\ ... \\ C_k \end{array} & \begin{bmatrix} y_{11} * \frac{\mu}{\mu_1} & y_{12} * \frac{\mu}{\mu_2} & ... & y_{1n} * \frac{\mu}{\mu_n} \\ y_{21} * \frac{\mu}{\mu_1} & y_{22} * \frac{\mu}{\mu_2} & ... & y_{2n} * \frac{\mu}{\mu_n} \\ ... & ... & ... & ... \\ y_{k1} * \frac{\mu}{\mu_1} & y_{k2} * \frac{\mu}{\mu_2} & ... & y_{kn} * \frac{\mu}{\mu_n} \end{bmatrix} \end{array}$$

The outcome is re-scaled in such a way that all tranches have the same mean, allowing the elimination of between-tranches inequality while preserving the within-tranches differences. We can measure the IOp ex-post applying an inequality indicator to this new outcome distribution $|Y'_{WT}|$.

As in the ex-ante approach, we can decompose total inequality as the sum of the IOp ex-post (within-tranches) and of the inequality due to effort (between-tranches). The weight of each component can be quantified using a path-independent measure, such as MLD (Checchi and Peragine, 2005).

2.3 Identification of Types

We find some limitations in traditional approaches for selecting which variables to use for the definition of types. Generally speaking, types are determined through the interaction of all the possible combinations of circumstances (Plassot et al., 2019; Soloaga and Wendelspiess, 2010). For example, if two circumstances with two categories are considered - sex and rural/urban area of residence - then four types will be identified. In other words, all circumstances are relevant to all individuals, which is not necessarily true. Therefore, as the number of circumstances considered increases, the number of types rapidly increases, while the number of observations in each type decreases. To overcome these limitations, researchers have often re-codified or re-categorized the variables and limited the number of circumstances in order to maintain a relatively significant number of observations in each type and reduce variability.

Following Brunori et al. (2018) and Brunori and Neidhöfer (2020) we use conditional inference trees to identify types and relevant circumstances. According to Hothorn et al. (2006, p. 3), these are "efficiently applicable to regression problems where both response and covariates can be measured at arbitrary scales, including nominal, ordinal, discrete and continuous as well as censored and multivariate variables". Conditional inference trees present a non-arbitrary selection criterion. In the presence of a large set of circumstances, the tree will only consider those that have a statistically meaningful association with the outcome. The tree determines the interactions as it finds splitting points that have associative power with respect to the variance of the outcome. These same interactions allow us to identify intersections as certain interactions may be relevant for one type but not others. Lastly, regression trees provide easy and direct visualization to represent the structure of opportunities.

The algorithm stratifies observations into different regions of the prediction space: observations with certain characteristics are segmented and then partitioned again until a certain criterion is met. Once this criterion is met, we establish the output regions as $R_1 = \{Y|C_1\}, \ldots, R_j = \{Y|C_j\}$, where C_1 and C_j can take any given number of circumstances depending on the decision to keep growing or to stop the tree. The regions R_j are known as terminal nodes, and their predicted outcome is the average of all observations contained therein. The points along the tree where the predictor space is separated are referred to as internal nodes. Every split divides the tree into two new spaces known as branches.[6] It is worth mentioning that the algorithm allows to identify points where inequality is happening, providing visual representation, which is a plus in understanding the process.

2.4 Identification of Tranches or Effort Degrees

We estimate degrees of effort instead of levels of effort because "we cannot reasonably expect to see the same amount of effort exerted by an average member of a disadvantaged type as is exerted by an average member of an advantaged type" (Sherman, 2016, p. 214). To identify degrees of effort, we focus on the type-specific outcome distribution. The distribution can be approximated through a number of different methods. Here, we compare the Log-normal, the Kernel-Gaussian and the Bernstein polynomials. Appendix B.1 shows a comparison between these three approaches using an example from EMOVI 2017. While the Log-normal and Kernel assume an underlying distribution of the data, the Bernstein polynomials is much more flexible, as can be seen in Appendix B.1 Panel (c) [7]. Thus, following Brunori and Neidhöfer (2020), we proceed by using the distribution estimated through the Bernstein polynomials.

We regress a parametric approximation of the distribution using Bernstein poly-

[6] The splitting criteria is detailed in Appendix A.

[7] We also find that Log-normal and Kernel tend to misestimate empirical distributions when they move away from the assumed distribution, as can be seen in Appendix B.1 Panel (b).

nomials defined as a linear combination of Bernstein basis polynomials.

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

For every type k, the number of observations is split into a ten-fold cross-validation manner. For every fold, we estimate the shape of the type-specific outcome distribution with monotone increasing Bernstein polynomials of degree m. We use the coefficients to estimate the cumulative distribution and estimate the out of sample log-likelihood. The selected degree is the one that maximizes the out-of-sample log-likelihood (an example is given in Appendix B.2).

3 Data

We use data from the ESRU Survey on Social Mobility in Mexico (EMOVI) developed by the *Espinosa Yglesias Center for Studies* (CEEY). The 2017 survey is representative of the population between the ages of 25 and 64 in both rural/urban areas and for five large regions of the country and Mexico City.[8] The questionnaire allows us to compile information regarding the sociodemographic, territorial, household, and educational characteristics of respondents. This information is available for the current household of the respondent (in 2017) and for the one in which the informant lived at age 14.

Following Velez Grajales et al. (2018), we approximate wealth through an index that reflects the possession of some assets and access to house services. Table 1 presents the items selected to construct such an index. We first build a polychoric correlation matrix, then we perform a factor analysis to extract the first component as the index variable. We follow Brunori and Neidhöfer (2020) to build an outcome

[8]The Regions are: North (Baja California, Coahuila, Chihuahua, Monterrey, Sonora, Tamaulipas), North- West (Baja California Sur, Sinaloa, Zacatecas, Nayarit, Durango), Center-West (Aguascalientes, Colima, Jalisco, Michoacán, San Luis Potosí), South (Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz, Yucatán), and Mexico City.

variable at destination and at origin that is somehow free of lifecycle variation. To do this, we first regress the index variable for the age and age-squared of individuals then predict the expected outcome at each age. We then divide the index variable for each individual by the predicted outcome for the age of the respondent. Finally, we generate percentiles of wealth for an easier interpretation of the trees.

Table 1: Selected Items

	Household at age 14	Current Household
Electricity	X	X
Stove	X	X
Washing Machine	X	X
Fridge		X
TV	X	X
Cable TV		X
Landline	X	X
Cellular		X
Internet		X
Computer		X
Credit Card		X
Tubing Water	X	
Toilet inside	X	
Domestic Service	X	

Notes: Selected Items used to construct the Asset Index from EMOVI 2017.

In our analysis, we restrict the sample to individuals who were living with their parents (one or both of them) at the age of fourteen. The model we consider in this study uses the Percentile of Wealth as a dependent variable explained by a set of circumstances that are: Percentile of Wealth at Origin, Average Education Level of Parents[9], Rural or Urban area at Age 14, Region of Origin at Age of 14, and Sex. As a result of missing information in some variables used as circumstances, the final sample is composed of 15,091 observations[10] between 25 and 64 years old which

[9] In the event that the respondent was living with both parents at age of 14, then the parents´ education variable is the average number of years schooled by both parents, and in the event that he/she was living with only one parent, we focus on the years of schooling of this parent.

[10] The original sample contains 17,665 observations, from which 16,457 were living with one

represent 51.7 million people. Table 2 shows some descriptive statistics. The sample contains is made up of 53% of women, and 33% of respondents declared they were living in a rural area at the age of 14. The average education level of parents is between seven and eight years of schooling, which corresponds to unfinished junior highsecondary education. Concerning With regard to the territory, 11% of the sample was living in Mexico City at age 14, 15% in the Northern region, 14% in the Center-OccidentWest, 7% in the North-OccidentalWest, 24% in the South, and, finally, 28% in the Central region.

Table 2: Descriptive Statistics

	Mean	Std. Dev.	Min	Max	Obs.	Pop. Millions
Percentile of Wealth (Outcome)	50	28.8	1	100	15,091	51.7
Percentile of Wealth at Origin	50	29.6	1	100	15,091	51.7
Education Level of Parents*	7.9	5.9	0	22	15,091	51.7
Rural at Age 14**	0.33	0.4	0	1	15,091	51.7
Sex (Women=1)	0.53	0.5	0	1	15,091	51.7
Region at Age 14:						
North	15.0				2,326	7.78
North-West	7.2				1,911	3.73
Center-West	14.5				2,619	7.52
Center	27.6				2,212	14.29
South	24.3				3,567	12.57
Mexico City	11.3				2,456	5.85

Notes: Estimations using the EMOVI 2017. The mean and standard deviation are calculated using expansion factors of the survey.

*Years of schooling of parents calculated as the average years of schooling for father and mother for those individuals (living with both parents at age 14).

**Rural areas correspond to localities of less than 2,500 inhabitants.

or both parents at age of 14; however, due to missing information regarding region of origin (59 observations), education level of parents (857 observations) and wealth at origin (558 observations), the sample size was reduced to 15,091 observations.

4 Results

4.1 Structure of Opportunities in Mexico

The first step of the study is to identify types. As explained in Section 2 we resort on regression trees to classify respondents according to circumstances. we employ regression trees to classify respondents based on circumstances. Following the data from EMOVI 2017, Mexican society can be divided into 43 types according to the circumstances of individuals at the age of fourteen. Table 3 and the tree at the national level in Figure 4 represent the structure of opportunities in Mexico for individuals between the ages of 25 and 64. The tree is composed of 42 splitting points, of which the education level of parents splits criteria 12 times, wealth at origin is taken into account 11 times, region of origin is divided in 10 splitting points, rural or urban area at age 14 comprises 7 splitting points, and, finally, sex has 2 splitting points.

Table 3: Number of Circumstances used as a splitting point in each model

	Wealth at Origin	Parents Level of Education	Region	Rural/Urban Area	Sex	Total Branches	Number of Types
National	11	12	10	7	2	42	43
Urban	8	9	9		1	27	28
Rural	5	5	3		0	13	14
North	4	4		2	0	10	11
North-West	3	2		2	0	7	8
Center-West	4	3		2	1	10	11
Center	4	5		1	1	11	12
Mexico City	3	4		0	1	8	9
South	4	7		3	1	15	16

Notes: Estimations using EMOVI 2017. Table shows the number of times each circumstance is used as a splitting point in different trees.

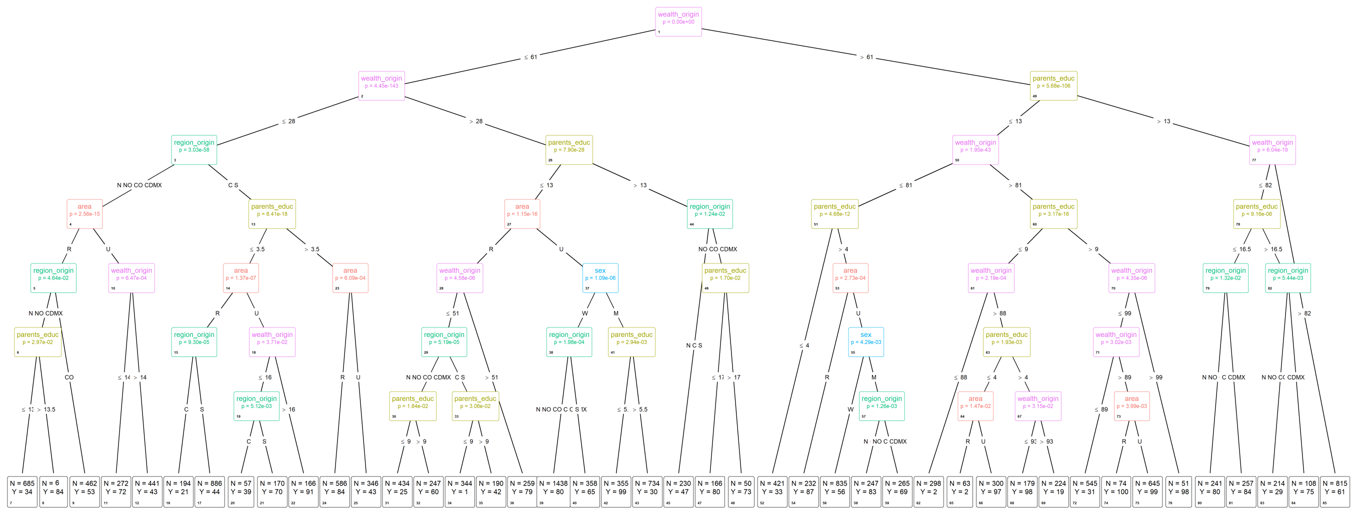
A first splitting point divides the society into two groups based on wealth at origin. This circumstance represents the principal determinant of future wealth. The tree splits individuals based on the threshold of the 61st percentile of wealth at origin. The more disadvantaged at age 14 are represented on the left side of the tree in Figure 4, while the more advantaged are represented on the right. For the most disadvantaged group, a second splitting point divides individuals below the 27th percentile at origin from those between the 27th and 61st percentile. For the most advantaged group (those above the 61st percentile at origin), a second splitting point divides individuals based on the education level of their parents. At this level, society is divided into four groups, and the variable used as the third splitting point varies for each group. For the two groups relatively most advantaged at origin, wealth at origin is the next most important circumstance in shaping the trajectory of individuals. In comparison, the education level of parents is taken into account for individuals between the 27th and 61st percentile at age 14. Finally, for individuals below the 27th percentile of wealth at age 14, the third splitting point is the region of residence at this same age.

We find that territorial variables (region and area of residence at age 14) have high significant importance for respondents living in the poorest households at the age of 14. For example, all of the 24 types on the left of the tree (most disadvantaged at origin) have at least one splitting point determined by the regional variable or the urban/rural variable. For the other half of the sample (most advantaged at origin), the rural/urban variable is considered in only 8 types, while the regional variable is found in 6 types. In the urban model, the region of origin represents a node for almost all types, reflecting more diversity than in rural areas with regard to the trajectories of individuals based on the territory of origin. The structure of opportunities for individuals who grew up in rural areas shows that the region of residence is significant for only the most disadvantaged at origin, with the first division being made between the South and the other regions. In the same way, among respondents who grew up in the South, the rural/urban variable is only significant for the most disadvantaged at origin.

Further inspection of the trees at the sub-national level can be found in Table 3. We can see that there are more types for respondents growing up in urban areas (28 types - see Appendix C.2) compared to rural areas (14 types - see Appendix C.1). We interpret this as an indication of a more polarized society in rural Mexico when compared to urban areas, given that more nodes means more diversity of outcomes.

At a regional level, the number of types is higher for those who grew up in the South (16 types - see Appendix C.8) and lower in the North-West (8 types - see Appendix C.4), and Mexico City (9 types - see Appendix C.7). In all trees, wealth at origin and the education level of parents are the principal factors used to determine types. In each territory the first splitting point is made according to wealth at origin. Within regions, the rural area of residence at age 14 is a determinant in all regions except for Mexico City. Finally, the sex variable is not a determinant to identify types in the two Northern regions and in the rural model.

Figure 4: Structure of Opportunities at the National level



Notes: Estimations using the EMOVI 2017. Figure shows the structure of the tree (structure of opportunities) at a national level. Variables in the tree are: Wealth at Origin (Purple), Parents Level of Education (Yellow), Sex (Light blue), Rural/Urban (Red) and Region at Origin (Green).

4.2 Inequality ex-post

To estimate IOp ex-post it is necessary to estimate degrees of effort. For our estimations using EMOVI 2017, levels of effort are not distributed equally between types. Appendix B.1 shows that the distribution of effort in a low-average outcome type tends to be skewed to the left (Panel (a)), while it tends to be skewed to the right for types with a high-average outcome (Panel (b)). We also have several types where the distribution seems to be much more even across individuals (Panel (c)). We make effort independent from circumstances when estimating degrees of effort as shown in Figures 1 and 2.

Table 4 shows that IOp (measured using the Gini coefficient) is much higher among those who grew up in rural areas (0.27) compared to those who grew up in urban territories (0.20). The same indicator for those who grew up in the South is 0.34 and reflects the highest IOp when comparing regions (Figure 5). The second most unequal region is the Center (0.26) with a relatively high level of IOp. Those that grew up in Mexico City (0.17) or the North (0.19) present the lowest IOp, while the North-West and the Center-West regions have an IOp indicator between these two previous groups (0.21).

Table 4: IOp ex-post measures

	IOp ex-post		Inequality between-tranches of effort		Total Inequality	
	Gini	MLD	Gini	MLD	Gini	MLD
National	0.26	0.14 (56%)	0.24	0.11 (44%)	0.32	0.26
Urban	0.20	0.09 (50%)	0.22	0.09 (50%)	0.26	0.18
Rural	0.27	0.15 (40%)	0.33	0.22 (60%)	0.39	0.36
North	0.19	0.07 (42%)	0.23	0.10 (58%)	0.27	0.18
North-Occident	0.21	0.09 (41%)	0.26	0.13 (59%)	0.31	0.22
Center-Occident	0.21	0.10 (48%)	0.24	0.11 (52%)	0.29	0.20
Center	0.26	0.13 (52%)	0.25	0.12 (48%)	0.32	0.25
Mexico City	0.17	0.06 (49%)	0.19	0.07 (51%)	0.22	0.13
South	0.34	0.22 (54%)	0.31	0.18 (46%)	0.40	0.40

Notes: Estimations using EMOVI 2017. The Table represents the estimations obtained using an ex-post approach for IOp. Inequality of Opportunity and Inequality due to effort are calculated through the Gini coefficient and Mean Log Deviation (MLD). Given that MLD permits path independent decomposability, we present the weight as a percentage for each component of total inequality: IOp ex-post and inequality due to effort. Some totals do not tally given the rounding up of decimals.

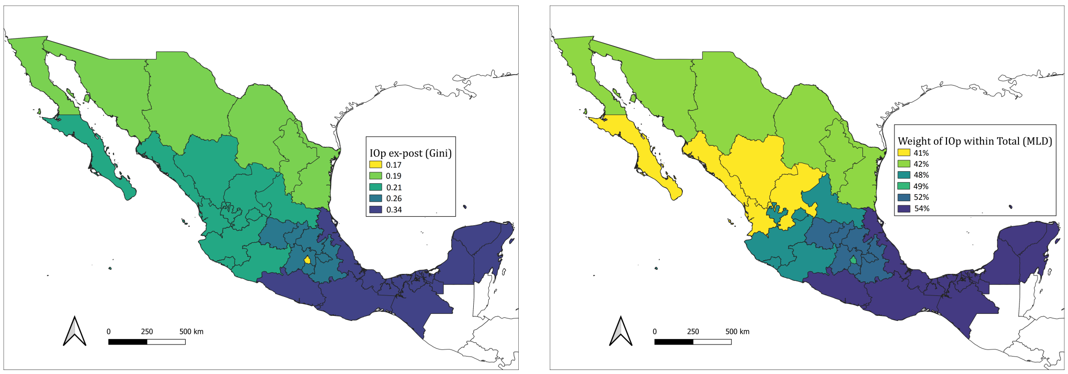
We now calculate the IOp and inequality due to effort using Mean Log Deviation (MLD) instead of the Gini coefficient as observed in Checchi and Peragine (2005). MLD allows decomposable measures (Shorrocks, 1980), where the sum of IOp and inequality due to effort represents total inequality. This type of indicator satisfies path independent decomposability (Foster and Shneyerov, 2000) and permits both absolute and relative measures[11]. At a national level, IOp represents 56% of total inequality, while the remaining inequality is due to effort, as can be seen in Table 4. The contribution of IOp is higher in urban areas (50%) compared to rural ones (40%), where the inequality between-tranches of effort has a more important weighting. At a regional level, we distinguish a first group formed by the North and the North-West

[11] Ferreira and Gignoux (2011) insist that both measures are complementary and of interest for the analysis.

where the weight of IOp (42% and 41%, respectively) is lower than inequality due to effort. A second group comprises Mexico City and the two Central regions, where the share of IOp is around 50%. Finally, in the South, 54% of inequality is due to IOp.

Figure 5: Regional dissagregation of ex-post IOP

(a) IOp ex-post (Gini) (b) Weight of IOp in total inequality(MLD)



Notes: Figure shows the ex-post IOp measured through the Gini coefficient (a) and the relative importance of ex-post IOp by MLD (b) in different regions in Mexico. It shows how the Northern region and Mexico City have the lowest IOp, while the Southern region has the highest. Total IOp weight in total inequality is higher in the South and the Center than in other regions.

4.3 Inequality ex-ante

An ex-ante IOp indicator is observed through the construction of the trees: if more than one type is determined, then equality of opportunity is not reached. To measure IOp through an ex-ante approach, we focus on the between-type inequality. There is wide-ranging literature arguing that the clash between the compensation principle and the reward principle is a consequence of the incompatibility between ex-ante and ex-post approaches (Checchi and Peragine, 2005; Cecchi et al., 2015; Fleurbaey and Peragine, 2013; Fleurbaey et al., 2017; Ooghe et al., 2003). Although we do not dwell on that discussion in this article, we would like to present the results obtained through both approaches, as observed in Aaberge (2011) or Checchi and Peragine (2005).

We calculate IOp based on the types identified through the trees. We follow Ferreira and Gignoux (2011) who propose using MLD as an inequality indicator. Absolute values of IOp with an ex-ante approach are much lower than measures obtained with an ex-post method. Nevertheless, by using both methods, we can observe the same differences across territories at origin: Rural areas and the Southern region are more unequal, while the Northern region and Mexico City are less unequal. Looking at relative measures in Table 5, we can see that each approach leads to different conclusions. At a national level and through an ex-post approach, the weight of IOp within total inequality is around 56%, whereas with an ex-ante approach this weight only represents 19%.

Table 5: IOp ex-ante Measures

	IOp ex-ante		Inequality Within-Type		Total Inequality Wealth	
	Gini	MLD	Gini	MLD	Gini	MLD
National	0.17	0.05 (19%)	0.30	0.21 (81%)	0.31	0.26
Urban	0.13	0.03 (15%)	0.25	0.15 (85%)	0.26	0.17
Rural	0.16	0.04 (12%)	0.37	0.32 (88%)	0.38	0.35
North	0.11	0.02 (10%)	0.26	0.16 (90%)	0.26	0.16
North-West	0.13	0.03 (12%)	0.29	0.20 (88%)	0.30	0.22
Center-West	0.14	0.03 (16%)	0.27	0.17 (84%)	0.29	0.20
Center	0.16	0.04 (16%)	0.30	0.21 (84%)	0.32	0.25
Mexico City	0.10	0.01 (11%)	0.22	0.12 (89%)	0.22	0.13
South	0.21	0.07 (17%)	0.38	0.33 (83%)	0.39	0.40

Notes: Estimations using the EMOVI 2017. The Table represents the estimations obtained using an ex-ante approach for IOp. Inequality of Opportunity and Inequality due to effort are calculated using the Gini coefficient and Mean Log Deviation (MLD). Given that MLD permits path independent decomposability, we present the weight of each component of total inequality: IOp and Inequality due to effort. Some totals do not tally as a result of the rounding up of decimals.

5 Robustness Checks

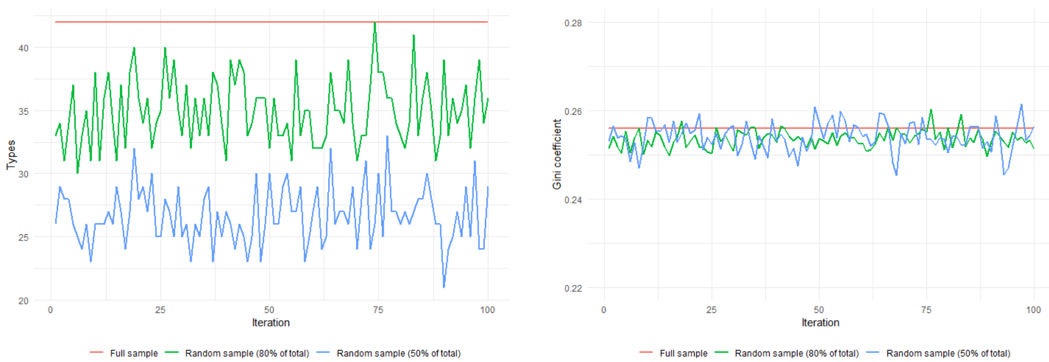
As a robustness check of the sensitivity of estimations to sample size and the set of circumstances considered, we use an approach similar to a bootstrap, called "bagging". The sample is split into n different training sets and a tree is estimated for every split (meaning we estimate n trees). We then proceed to average out the predictions. The method to estimate types and IOp is repeated 100 times, and, in each iteration, two random sub-samples are chosen: one with 80% of the full sample (12,073 observations) and a second with 50% of the full sample (7,546 observations).

A first result is that as the size of the sample increases, so does the number of types. By using the full sample, there are 43 types at the national level, whereas this number decreases to 35 types on average over the 100 random samples of 12,073 observations and to 27 types when using 7,546 observations (Figure 6 Panel (a)). The number of types also varies in each iteration from 30 to 42 types with 80% of the full sample and from 21 to 33 types when using 50%. With regard to IOp estimations, we do not observe such volatility across the iterations. As can be seen in Figure 6 Panel (b), which represents IOp ex-post estimations using the Gini coefficient, the measure varies between 0.250 and 0.260 when using 80% of the full sample. The same result is found when comparing IOp ex-ante estimations or when using MLD instead of Gini. If variations are smaller, they can have stronger implications in the interpretation of results in the case of wider differences in the sample size.

Table 6 shows how the number of types estimated using the trees decreases as we reduce the set of circumstances. The difference in the number of types is much more important when the omitted variable has an important weighting in the formation of the nodes (e.g., education level of parents), compared to variables with a lower importance, such as sex. Furthermore, IOp measures decline slightly as the number of circumstances decreases. As a consequence, the weighting of the IOp within the total increases as we add circumstances to the model

Figure 6: Results of 100 iterations

(a) Number of Types (b) Gini Coefficient



Notes: Estimations using the EMOVI 2017. Figure shows the estimations obtained over the 100 iterations with different subsamples (Full sample is red; 80% is green; 50% is blue). Panel (a) shows the number of types estimated in each iteration, and Panel (b) shows the Gini coefficient.

Table 6: Change in results when circumstances vary

		(a)	(b)	(c)	(d)
		Full Circumstances	(a) Without Sex	(b) Without Education Level of Parents	(c) Without Area at age 14
	Number of types	43	43	30	26
Gini	IOp expost	0.256	0.256	0.247	0.244
	Between-Tranches	0.244	0.244	0.251	0.252
	Within-Tranches	0.298	0.298	0.300	0.300
	IOp exante	0.172	0.172	0.163	0.161
	Total Inequality	0.310	0.315	0.315	0.315
MLD	IOp expost	0.149	0.144	0.135	0.133
	Between-Tranches	0.106	0.114	0.123	0.126
	Within-Tranches	0.203	0.209	0.213	0.215
	IOp exante	0.052	0.050	0.045	0.044
	Total Inequality	0.255	0.258	0.258	0.258

Notes: Estimations using the EMOVI 2017. The first model (a) considers all the circumstances identified in the previous parts of the article: Sex, Education Level of Parents, Wealth at Origin, Region at Age 14, and Area of Residence at age 14. The second model (b) considers all these circumstances except the sex variable. The third model (c) considers all the circumstances except Sex and Education Level of Parents. Finally, the fourth model (d) only considers Sex, Wealth at Origin, and Region at Age 14 as circumstances. We can observe how the IOp and the number of types vary in each model.

6 Discussion

Literature in Mexico has advanced in identifying the circumstances that are determinants in shaping the trajectory of individuals, particularly when focusing on the weighting of the territorial and households environment at age 14, in addition to highlighting gender differences. The most recent step forward has been to provide indicators at a regional level and for urban and rural territories. Studies have shown the differences among regions (Delajara, 2018) and between states (Vélez-Grajales, 2017) in the indicators of intergenerational social mobility in wealth. Concerning IOp, indicators with an ex-ante approach have been presented for different regions (Monroy-Gómez-Franco and Corak, 2020; Plassot et al., 2019) but there are no studies with an ex-post approach that follow Roemers' approach of IOp. Finally, researchers have estimated the weight of ex-ante IOp within total inequality (Monroy-Gómez-Franco and Corak, 2020; Velez Grajales et al., 2018). The present study completes this literature by providing IOp estimations for territories at a sub-national level, and it is the first of its kind to provide ex-post IOp measures for Mexico. It is also the first attempt to determine types and represent the structure of opportunity of Mexican society through conditional inference trees. Furthermore, we identify effort through Bernstein polynomials that allow the data to express the underlying distribution of effort in each type as proposed by Brunori and Neidhöfer (2020).

We find that, by 2017, Mexican society can be divided into 43 types based on circumstances at origin. The wealth and the education level of parents at origin are the principal determinants of IOp, while the variables of region, rural/urban area of residence and sex are determining factors only in some groups or territories. One of the main results is that territorial variables (region and rural/urban area of residence at age of 14) are significant circumstances mainly for those individuals most disadvantaged at the origin. IOp is higher in rural areas than in urban ones, and the number of types is higher in urban areas. The Southern region is the region that presents the most types and highest IOp, whereas Mexico City and the Northern region have the lowest number of types and lower IOp indicators. The Center is the second most unequal region in terms of opportunity, followed by the Center-West and the North-West, respectively. These findings confirm those published by Plassot et al. (2019) using an ex-ante IOp indicator, or Delajara (2018), who found lower absolute upward mobility and higher intergenerational association in the South compared to the higher mobility and lower association in the North.

We then present ex-ante and ex-post IOp measures and found that the ranking of regions according to the level of IOp is the same using both approaches. Comparing the weight of IOp within total inequality, we conclude that the choice of the method has strong implications in terms of the interpretation of results. Using an ex-ante approach above 80% of total inequality can be explained by differences in effort, while the remaining percentage depends on circumstances. The ex-post method displays a different picture, one in which about half of the total inequality is explained by differences in effort. While other studies with an ex-ante approach have found a contribution of IOp to total inequality between 30% and 48%[12] in Mexico (CEEY, 2019) we must underline that these studies include a wider range of circumstances variables than in this study. As mentioned by Ferreira and Gignoux (2011), the more circumstances that are added, the higher the share of IOp within total inequality. At a territorial level, we found that the weight of inequality due to circumstances within the total inequality is more important in urban areas than in rural ones. Finally, the weight of IOp within the total inequality is lower in the Northern regions and higher in the South.

In terms of methodology, the first conclusion is that the use of conditional inference trees as employed by Brunori and Neidhöfer (2020) offers a straightforward measurement of IOp in Mexico. The trees permit graphical representations of the opportunities provided by society at any given time, and this format can be easily communicated to the public. Using regression (and classification) trees, the importance of each circumstance differs based on the type but also the territory observed.

[12] Velez Grajales et al. (2018) estimate that the lower bound of this contribution is around 30%. Monroy-Gómez-Franco and Corak (2020) include variables about neighborhood characteristics at origin (in addition to personal and family characteristics and territorial circumstances), finding that around 48% of inequality in outcomes is due to IOp.

Therefore, researchers can identify which dimensions are the principal barriers to the Equality of Opportunity in each territory or population group.

To conclude, we present some suggestions for further research. Firstly, it would be relevant to estimate types through a random forest or different unsupervised learning algorithms, such as K-means or Gaussian Mixture (and Bayesian Gaussian Mixture). Random Forests allow both observations and variables to change in each tree, providing more robust estimations. Given the richness of dimensions contemplated in the EMOVI surveys, more variables can be added as circumstances (e.g., household size, neighborhood characteristics, social transfers, migration, or private schooling) to represent the overall structure of opportunities in Mexico. With regard to effort, it would be recommendable to use other sources to directly estimate effort and compare it to the IOp measures of this study, and, for example, with the 2015 survey on social mobility, designed by *El Colegio de México (Colmex)* and including questions about ability, skills, and effort.

References

Aaberge, Rolf Mogstad, M. . P. V.: 2011, Measuring long-term inequality of opportunity, *Journal of Public Economics* **95**(3-4), 193–204.

Ald, F.: 2017, The Bernstein Mechanism Differential Privacy with the Bernstein Mechanism, pp. 1–6.

Arneson, R. J.: 1989, *Equality of Opportunity for Welfare*.

Bourguignon, F., Ferreira, F. H. and Menéndez, M.: 2003, Inequality of Outcomes and Inequality of Opportunities in Brazil, *William Davidson Institute Working Paper No.* (630), 1–43.

Breen, R.: 2010, Social mobility and equality of opportunity Geary Lecture, spring 2010, *Economic and Social Review* **41**(4), 413–428.

Brunori, P., Hufe, P. and Mahler, D.: 2018, The roots of inequality : estimating inequality of opportunity from regression trees, *World Bank Policy Research Working Paper* .

Brunori, P. and Neidhöfer, G.: 2020, The Evolution of Inequality of Opportunity in Germany: A Machine Learning Approach.

Campos Vázquez, R.: 2015, Promoviendo la Movilidad Social en México: Informe de Movilidad Social 2015, p. 67.

Cecchi, D., Peragine, V. and Serlenga, L.: 2015, Income inequality and opportunity inequality in europe: Recent trends and explaining factors, (6th).

CEEY: 2019, Hacia la igualdad regional de oportunidades, *Technical report*.

Checchi, D. and Peragine, V.: 2005, Regional disparities and inequality of opportunity: the case of Italy, *IZA Discussion Paper* (1874).

Chetty, R., Hendren, N., Kline, P. and Saez, E.: 2014, Where is the land of opportunity? The geography of intergenerational mobility in the United States, *Quarterly Journal of Economics* **129**(4), 1553–1623.

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

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

Corak, M.: 2017, Divided Landscapes of Economic Opportunity : The Canadian Geography of Intergenerational Income Mobility, *HCEO Working Paper Series* (2017), 1–58.

Cortés, F.: 2013, Medio siglo de desigualdad en el ingreso en méxico, *Economía UNAM* **10**, 12–34.

Deary, I. J., Taylor, M. D., Hart, C. L., Wilson, V., Smith, G. D., Blane, D. and Starr, J. M.: 2005, Intergenerational social mobility and mid-life status attainment: Influences of childhood intelligence, childhood social factors, and education, *Intelligence* **33**(5), 455–472.

Delajara, M. Graña, D.: 2018, La movilidad social intergeneracional en méxico y sus regiones es el resultado de regresiones rango-rango, *Sobre México Temas de Economía* .

Dworkin, R.: 1981a, What is equality? part 1: Equality of welfare, *Philosophy Public Affairs* **10**(3), 185–246.

Dworkin, R.: 1981b, What is equality? part 2: Equality of resources, *Philosophy Public Affairs* **10**(4), 283–345.

Ferreira, F. H. G. and Peragine, V.: 2015, Equality of opportunity: Theory and evidence, *Microeconomics: Welfare Economics Collective Decision-Making eJournal* .

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

Fleurbaey, M.: 1995, Three Solutions for the Compensation Problem, *Journal of Economic Theory* **65**(2), 505–521.

Fleurbaey, M. and Peragine, V.: 2013, Ex ante versus ex post equality of opportunity, *Organizations Markets: Motivation Incentives eJournal* .

Fleurbaey, M., Peragine, V. and Ramos, X.: 2017, Ex post inequality of opportunity comparisons, *Social Choice and Welfare* **49**(3-4), 577–603.

Foster, J. E. and Shneyerov, A. A.: 2000, Path Independent Inequality Measures, *Journal of Economic Theory* **91**(2), 199–222.

Guan, Z.: 2016a, Efficient and robust density estimation using Bernstein type polynomials, *Journal of Nonparametric Statistics* **28**(2), 250–271.

Guan, Z.: 2016b, Efficient and robust density estimation using Bernstein type polynomials, *Journal of Nonparametric Statistics* **28**(2), 250–271.

Hothorn, T.: 2020, Package ‘ basefun ’.

Hothorn, T., Hornik, K., Strobl, C. and Zeileis, A.: 2017, party: A laboratory for recursive partitioning. R package version 1.2-3.

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

Jencks, C. and Tach, L.: 2011, Would Equal Opportunity Mean More Mobility?, *SSRN Electronic Journal* (May).

Kanbur, R. and Wagstaff, A.: 2014, How useful is inequality of opportunity as a policy construct ?, (6980).

Leblanc, A.: 2012, On estimating distribution functions using Bernstein polynomials, *Annals of the Institute of Statistical Mathematics* **64**(5), 919–943.

Monroy-Gómez-Franco, L. and Corak, M.: 2020, A Land of Unequal Chances: Social Mobility Across Mexican Regions, (10).

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

Ooghe, E., Schokkaert, E. and Van de Gaer, D.: 2003, Equality of opportunity versus equality of opportunity sets, *Social Choice and Welfare* **28**.

Paes de Barros, Ricardo, F. F. H. G. M. J. R. and Saavedra Chanduvi, J.: 2008, *Measuring Inequality of Opportunities in Latin America and the Caribbean*.

Peragine, V. and Fleurbaey, M.: 2009, Ex ante versus ex post equality of opportunity Marc Fleurbaey Ex ante versus ex post equality of opportunity, *Society for the Study of Economic Inequality* .

Piketty, T.: 2019, *Capital and Ideology*, E´ditions du Seuil, Paris.

Plassot, T., Rubio, G. and Soloaga, I.: 2019, Movilidad social intergeneracional y desigualdad de oportunidades en México. Educación y activos : un enfoque territorial, (07).

Plomin, R. and Nesselroade, J.: 1990, Behavioral genetics and personality change., *Journal of personality* **58 1**, 191–220.

Ramos, X. and Gaer, D. V. D.: 2017, Is Inequality of Opportunity Robust to the Measurement Approach ? Is Inequality of Opportunity Robust to the Measurement Approach?, (11157).

Ramos, X. and Van de Gaer, D.: 2016, Approaches To Inequality of Opportunity: Principles, Measures and Evidence, *Journal of Economic Surveys* **30**(5), 855–883.

Rawls, J.: 1971, *A Theory Of Justice*, Oxford University Press.

Roemer, J.: 1998a, Equal opportunity for health.

Roemer, J.: 1998b, Equality of Opportunity — John E. Roemer | Harvard University Press.

Roemer, J.: 2002, Equality of opportunity: A progress report., *Soc Choice Welfare* (19), 455–471.

Roemer, J. E.: 2003, Defending equality of opportunity, *The Monist* **86**(2), 261–282.

Roemer, J. E.: 2004, *Equal opportunity and intergenerational mobility: going beyond intergenerational income transition matrices*, Cambridge University Press, p. 48–57.

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

Sammons, P., Elliot, K., Sylva, K., Melhuish, E., Siraj-Blatchford, I. and Taggart, B.: 2004, The impact of pre-school on young children's cognitive attainments at entry to reception, *British Educational Research Journal* **30**(5), 691–712.

Sammons, P., Sylva, K., Melhuish, E., Siraj, I., Taggart, B. and Hunt, S.: 2008, Influences on children's attainment and progress in key stage 2: cognitive outcomes in year 6.

Sherman, J. A.: 2016, *Renewing liberalism.*

Shorrocks, A.: 1980, A class of additively decomposable inequality measures, *Econometrica* **48**, 613–25.

Soloaga, I. and Wendelspiess, F.: 2010, *Desigualdad de Oportunidades: aplicaciones al caso de México*, Vol. 2.

Sylva, K., Melhuish, E., Sammons, P., Siraj, I. and Taggart, B.: 2004, The effective provision of pre-school education (eppe) project: Findings from pre-school to end of key stage 1.

Torche, F.: 2015, Diferencias de género en la movilidad social de méxico, *México, ¿El Motor Inmóvil?; Centro de Estudios Espinosa Yglesias* .

Toro, H.: 2015, Cohortes laborales y origen socioeconómico como determinantes del logro ocupacional, *México, ¿El Motor Inmóvil?; Centro de Estudios Espinosa Yglesias* .

Velez Grajales, R., Monroy-Gómez-Franco, L. and Yalonetzky, G.: 2018, Inequality of opportunity in mexico.

Vélez-Grajales, Roberto Stabridis, O. . M.-C. E.: 2017, Still looking for the land of opportunity. the case of mexico.

Wendelspiess Chávez Juárez, F. and Soloaga, I.: 2013, Scale vs. Translation Invariant Measures of Inequality of Opportunity when the Outcome is Binary, *SSRN Electronic Journal* .

Appendices

A Conditional Inference Trees

In this work we use conditional inference trees, first proposed by Hothorn et al. (2006). These trees are based on the concept of recursive binary partitioning.

Suppose we have a response variable Y and a set of explanatory variables C (we use C instead of X as we talk about circumstances in this specific example). We presume that both Y and C are taken from some sample space. Let us assume that the conditional distribution of Y given C is such as

$$D(Y|C) = D(Y|C_1, ..., C_m) = D(Y|f(C_1, ..., C_m)$$

The algorithm estimates the regression relationship on a learning sample

$$L_n = \{(Y_i, X_{1i}, ..., X_{mi}); i = 1, ..., n\}$$

proceeding as follows

1. **Variable selection:** Test the null hypothesis of independence between the outcome variable and all circumstances $C_j = 1, \ldots, k$ in each node w

$$H_0^j = D(Y_i|C_{ji}) = D(Y_i) \rightarrow H_0 = \bigcap_{j=1}^{m} H_0^j$$

the global null hypothesis $H_0 0$ is tested on multiple linear statistics of the way

$$T_j(L_n, w) = vec\left(\sum_{i=1}^{n} w_i g_j(C_{ji}) h(Y_i, (Y_1, ..., Y_n))^T\right) \in \mathbb{R}^{p_j, q}$$

where g is a non-random transformation of the covariates C_j and h depends on the response variable Y. The distribution of $T_j(L_n, w)$ depends on the joint distribution of Y and C and can be tested through *permutation tests*.

2. **Splitting Criteria:** If the global H_0 cannot be rejected, it is necessary to measure the association of each covariate C_j, $j = 1, ..., m$ by H_0^j to the response variable Y, the one with the lowest adjusted p-value is set as the first splitting variable. 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 discrepancy between two samples is measured by means of a linear statistic given by

$$T_{j*}^S(L_n; w) = vec(\sum_{i=1}^{n} w_i I(C_{ji} \in S) h(Y_i(Y_i, ..., Y_n))^T) \in \mathbb{R}^q$$

   where the split S^* with a test statistic maximized over all possible subsets S is established as

$$S^* = \underset{S}{argmax} c(t_{j*}^S, \mu_{j*}^S, \Sigma_{j*}^S)$$

   the splitting point with the lowest p-value is selected and two branches are generated that correspond to two subsamples.

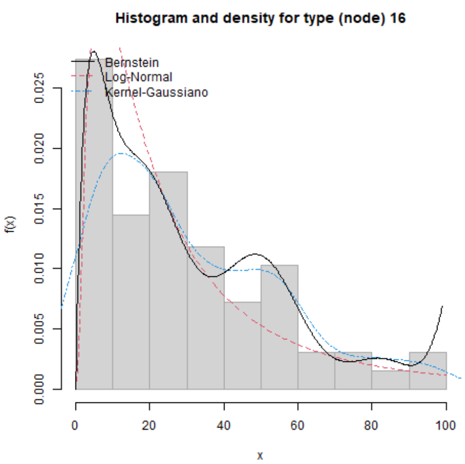
3. **Stopping Criteria:** This procedure is repeated on all possible subsamples until the global null hypothesis in 1 is reached, at which point types (terminal nodes) are identified.

The value of α is pre-specified at 0.05 to get results at the 95% significance level. For further information regarding specifics of the algorithm see Hothorn et al. (2006).

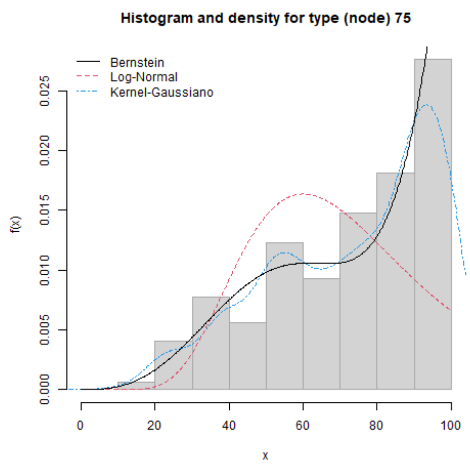
B Effort Distribution

Figure B.1: Histogram and Density of Type specific Distributions

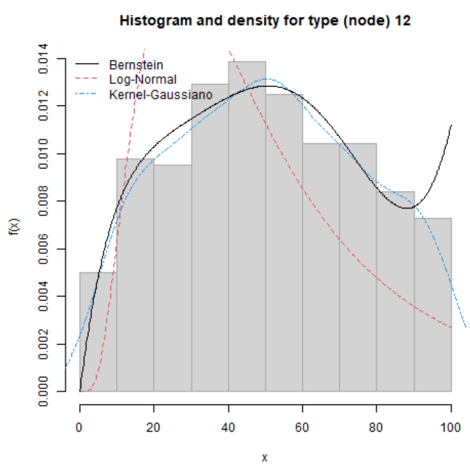
(a) Low average outcome



(b) High average outcome

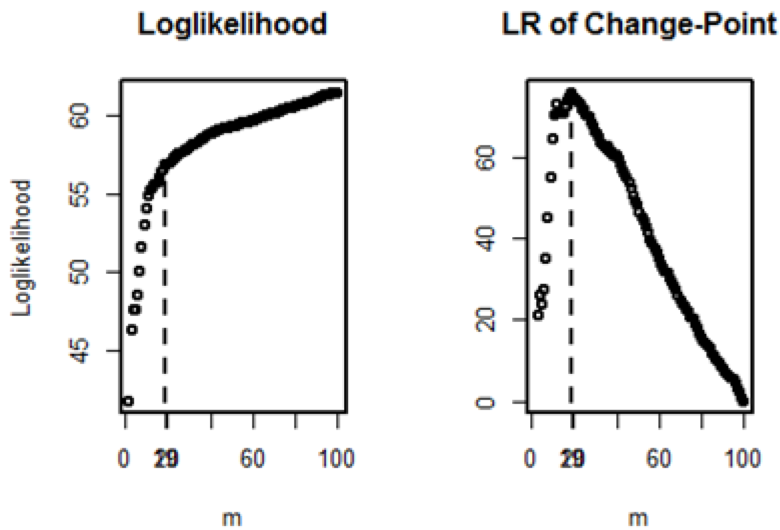


(c) Average outcome



Notes: Estimations showing the histogram and density approximation through different methodologies (red: Log-normal; blue: Kernel-Gaussian; black: Bernstein polynomials) of wealth at destination (x) for different types in Mexico using data from the EMOVI 2017. Panel (a) shows a type where the average outcome is low; (b) shows a type where average outcome is high; and (c) a type where outcome is at the average level. We use Bernstein polynomials for further estimations.

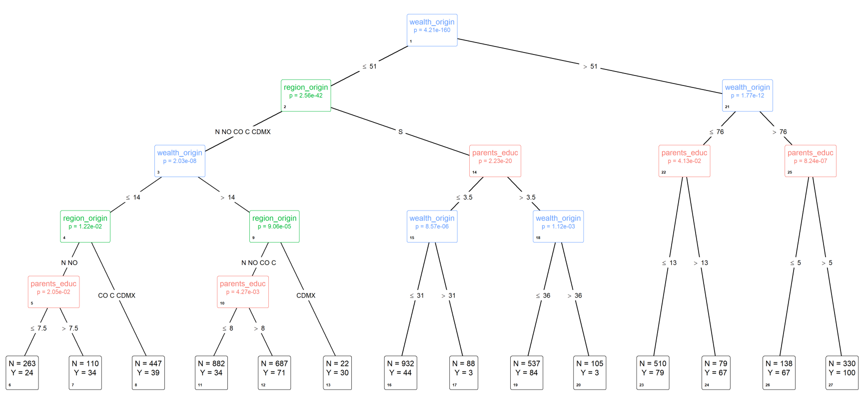
Figure B.2: Example of the Selection of m that Maximizes the Out-of-sample Loglikelihood



NB: Estimation using the *mable* function in R as proposed by Guan (2016b) and data from the EMOVI 2017.

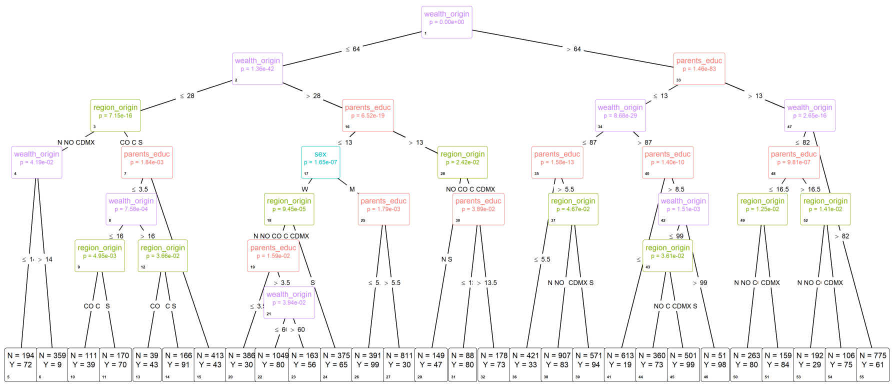
C Trees

Figure C.1: Rural



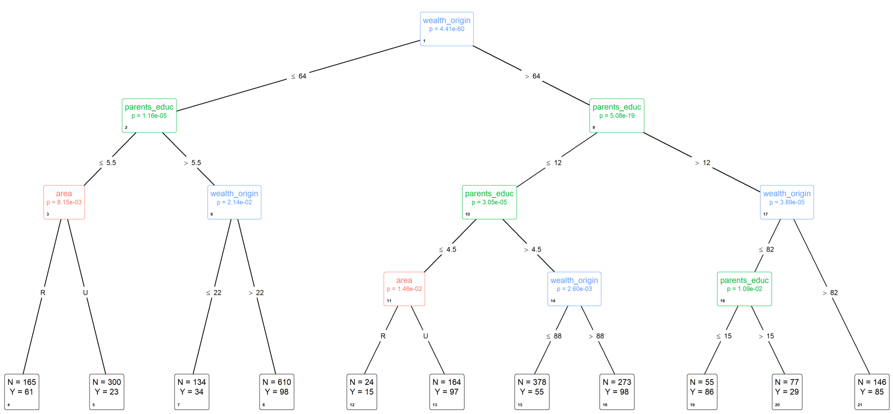
Notes: Figure shows the structure of the tree (structure of opportunities) for Rural areas. Variables in the tree are: Wealth at Origin (Blue), Education Level of Parents (Red) and Region at Origin (Green).

Figure C.2: Urban



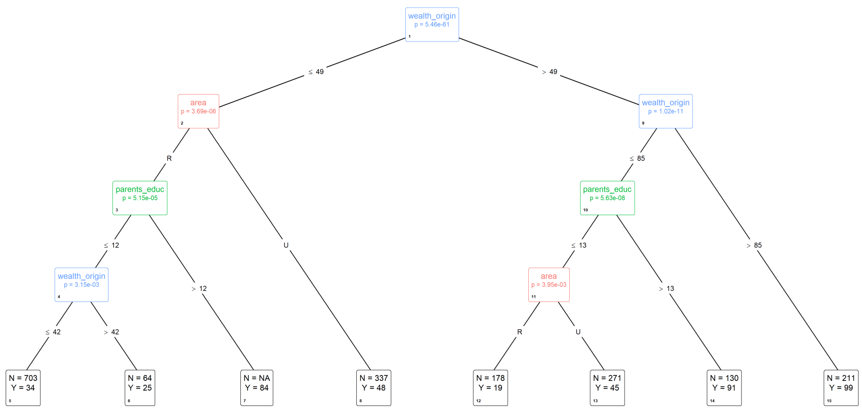
Notes: Figure shows the structure of the tree (structure of opportunities) for Urban areas. Variables in the tree are: Wealth at Origin (Purple), Education Level of Parents (Red), Region at Origin (Light green) and Sex (Light blue).

Figure C.3: North



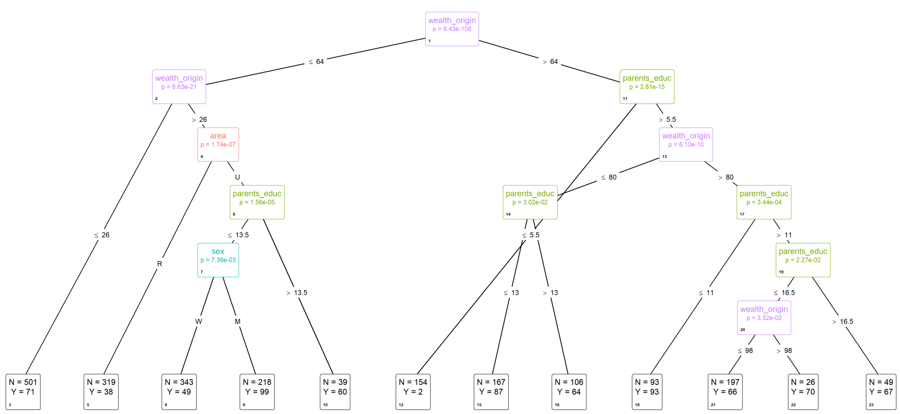
Notes: Figure shows the structure of the tree (structure of opportunities) for the Northern region. Variables in the tree are: Wealth at Origin (Blue), Education Level of Parents (Green) and Region at Origin (Red).

Figure C.4: North-West



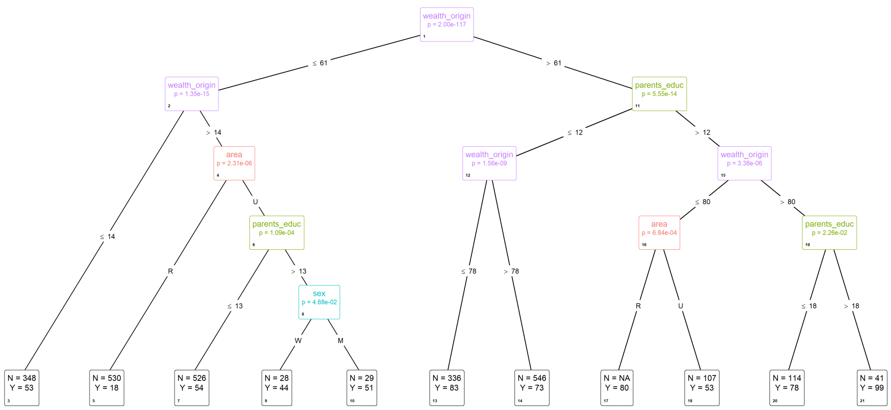
Notes: Figure shows the structure of the tree (structure of opportunities) for the North-Western region. Variables in the tree are: Wealth at Origin (Blue), Education Level of Parents (Light green), and Area (Red).

Figure C.5: Center



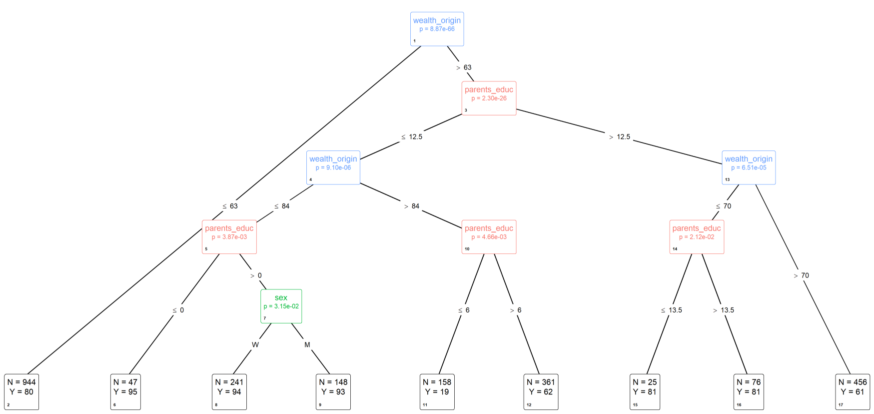
Notes: Figure shows the structure of the tree (structure of opportunities) for the Central region. Variables in the tree are: Wealth at Origin (Purple), Sex (Light blue), Education Level of Parents (Green) and Region at Origin (Red).

Figure C.6: Center-Occident



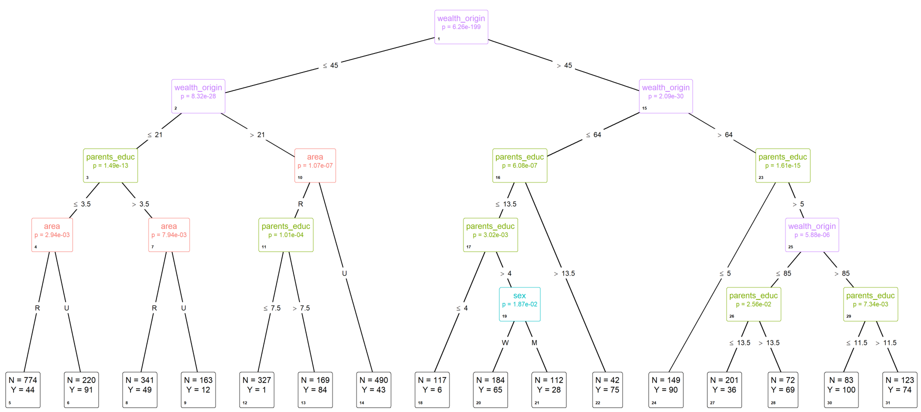
Notes: Figure shows the structure of the tree (structure of opportunities) for the Center-Western region. Variables in the tree are: Wealth at Origin (Purple), Education Level of Parents (Light green), Region at Origin (Red) and Sex (Light blue).

Figure C.7: Mexico City



Notes: Figure shows the structure of the tree (structure of opportunities) for Mexico City. Variables in the tree are: Wealth at Origin (Light blue), Sex (Green) and Education Level of Parents (Red).

Figure C.8: South



Notes: Figure shows the structure of the tree (structure of opportunities) for the Southern region. Variables in the tree are: Wealth at Origin (Purple), Education Level of Parents (Light green),Region at Origin (Red) and Sex (Light blue).