Playing the birth lottery in South America and Europe*

Annaelena Valentini[a], Paolo Brunori[b,c], Francisco Ferreira[b], Pedro Salas-Rojo[b]

[a]University of Siena [b]III - London School of Economics [c]University of Florence

January 27, 2023

Abstract

This paper presents an analysis of inequality of opportunity at a continental level in South America and Europe. The concept of inequality of opportunity refers to the share of inequality that can be predicted by circumstances beyond individual control, such as sex, socioeconomic background, and ethnicity. Previous studies have estimated inequality of opportunity at the country level and have often neglected the role of the country of birth. However, as suggested by Milanovic (2015), country of birth may be a key factor to understand inequality of opportunity. Our study builds upon this intuition. We build upon Milanovic's analysis of global inequality of opportunity and contribute to the literature in two ways: first, we take into account migration as a key variable instead of assuming it is a negligible factor, and second, we expand the set of circumstances considered to include sex, socioeconomic background, and ethnicity. Our study utilizes a large set of survey data to analyze inequality of opportunity in Latin America and Europe. Our findings reveal that, consistent with Milanovic, region and country of birth are the single most important sources of inequality of opportunity in both continents.

Keywords: inequality of opportunity, place of birth, migration, income distribution, Europe, Latin America.

JEL codes: D31,J60,O52,O54

*We are grateful to Torsten Hothorn and Achim Zeileis for the kind and useful support.

1 Introduction

We do not get to choose the country, the environment, the family, or the genetic endowment we get born with. Still, the results of the *birth lottery* largely determine the choices and outcomes we get as we live and grow. Overall, this is what *Inequality of Opportunity (IOp)* refers to: how circumstances, factors beyond individual's control, such as place of birth, gender, race, and parental background, are predictive of valuable outcomes. Empirical applications of this idea have traditionally focused on a specific country (Lefranc et al. (2009); Checchi and Peragine (2010)) and cross-country comparisons (Ferreira and Gignoux (2011); Brunori et al. (2023)). Here we move one step forward and turn our attention to supranational inequalities of opportunities, notably, we consider the continent as a whole and we assess the relationship between the continental income distribution and country of origin, along with other circumstances related to familiar background. The two empirical cases of interest are the European and Latin America continents.

Social justice and equality of opportunity are integral to the European project. The equal opportunity is enshrined in the EU's Treaties which state that the EU must work to eliminate discrimination and promote equality between individuals on the grounds of sex, race, ethnic origin, religion or belief, disability, age, or sexual orientation. This community of nations has, indeed, often expressed two lofty aims: first, that the life chances available to European citizens should not be overly determined by their family background or other accidents of birth.[1] Second, that disparities across European regions should be gradually eliminated.[2] Yet, there is abundant evidence that, after a period of convergence in which poorer countries have been catching up with the rest of the continent (1980-2008), country of birth continues to shape the life outcomes of European residents (Marrero and Rodríguez, 2012; Milanovic, 2015; Pina and Sicari, 2021). Using unique data on the national origin of parents, this paper sheds light on two main questions. First, how important is the lottery of birth in contemporary Europe: what share of the continent's overall income inequality is accounted for by circumstances at birth, such as parental education, occupation, and country of birth? Second, we ask which is the most relevant factor, between parental background or country of birth, in shaping inequality of

[1] The aim of realizing equal opportunities is

[2] As stated in Article 158 of the Treaty establishing the European Community: "In order to promote its overall harmonious development, the Community shall develop and pursue its actions leading to the strengthening of its economic and social cohesion. In particular, the Community shall aim at reducing disparities between the levels of development of the various regions and the backwardness of the least favored regions or islands, including rural areas."

opportunity in Europe.

The second empirical case we analyze in this paper is the Latin American continent. Estimating the share of income inequality explained by birth circumstances which include the individual's place of birth, appears to be highly relevant in this context. It is widely known that it is one of the most unequal regions in the world but what is even more striking are the enormous economic disparities existing between the different regions within Latin America. The peculiar context and the availability of survey data which report more detailed information on the place of birth of the respondent, namely the specific region of birth, allow us to estimate the role of the birthplace in shaping inequality of opportunity in the continent and how this interact with other birth circumstances.

In 2015, Branko Milanovic published a study in which he aimed to estimate global inequality of opportunity. In order to achieve this goal, he employed a two-pronged approach: first, he assumed that migration was a negligible factor, and second, he limited the number of controlled variables. Although this allowed for satisfactory coverage of the global population, as he recognizes, using country of residence as a controlled variable may not be entirely accurate, as individuals have the ability to migrate. Furthermore, Milanovic's study did not take into account key variables, such as sex, ethnicity, and socioeconomic background, which are commonly utilized in the measurement of inequality of opportunity. Despite the fact that the analysis was based on only two circumstances, country of residence and its inequality, it estimated that approximately half of the total inequality in global income is due to inequality of opportunity. This finding is noteworthy as it suggests that a significant proportion of income disparities can be attributed to factors such as access to education, healthcare, and other opportunities rather than inherent differences in ability or productivity.

The study conducted by Milanovic serves as the starting point for our research. Building upon his approach, we aim to improve upon it in two ways. Firstly, we take into account migration as a key variable in understanding global income inequality of opportunity. Additionally, we expand the set of circumstances considered in the analysis to include typical ascriptive characteristics such as sex, ethnicity, and socioeconomic background. The trade-off for a more comprehensive approach is a reduction in coverage. In what follow we will restrict the analysis to the level of inequality at the continental level considering two continents: South America and Europe.

We depart from Milanovic's approach also in terms of estimation methodology. Following Brunori et al. (2023) we adopt a data-driven approach based on the Conditional Inference Trees

(CITs), a Machine Learning algorithm introduced by Hothorn et al. (2006).

CITs divide the population into non-overlapping subgroups by partitioning the regressors' space. Then individuals' prediction is obtained by taking the average of the subgroup they belong to. CITs are particularly suitable to deal with highly non-linear data-generating processes and, in our specific case, they exactly implement the partition in types postulated by the theory of equality of opportunity (van de Gaer, 1993; Roemer, 1998). Moreover, CITs search for all possible combinations of regressors that can best explain the dependent variable variability without overfitting the data. This translates into the minimization of the risk of upward and downward bias when estimating inequality of opportunity (Ferreira and Gignoux (2011); Brunori et al. (2019)). Finally, in order to improve the robustness of our conclusions, we bag regression trees drawing a large number of samples from the original observed distribution. This allows us to make inferences about our estimates by taking into account two sources of uncertainty: the fact that we do not observe the entire population, and the fact that each tree represents only one possible partition. We recognize that the "true" partition of types is inherently unknowable.

For the analysis of inequality of opportunity in Europe the data comes from the European Union Statistics on Income and Living Conditions (EU-SILC), a widely-used survey run in the European Union that gathers information about several income sources and socioeconomic factors related to living conditions. In particular, we focus on household disposable income. The survey is delivered on an annual basis, but only three waves (2005, 2011, 2019) have specific modules collecting information about circumstances. Still, we use the 2019 wave because it is the one providing exact information on parents' countries of birth, which can be used to infer the country of birth of the respondent when born outside the country of residence. This allows us to get a dataset of 372,072 observations, representative of the European population, covering 24 European countries and up to 174 countries of origin. The remaining circumstances we use in our analysis are sex, parental occupation, education and main labor activity, and a variable indicating the status of a second-generation immigrant. Then, after accounting for equivalences of scales, the age of the household, and differences in purchasing power, we estimate the share of overall income inequality in Europe that can be predicted with the above-mentioned circumstances.

According to our estimates, the portion of total income inequality in the European continent that can be attributed to circumstances amounts to 68%. The country of birth appears to have the most relevant role in shaping inequality of opportunity in the continent, accounting for more than two-thirds of the estimated income IOp.

As regards South America, we use nine household surveys fielded between 2008 and 2014, that contain information about household income, sex, father and mother's education and occupation, ethnicity, and area of birth. The countries covered are Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guatemala, Panama, and Peru. The surveys used are listed in A1 in Appendix A; all the datasets were obtained from the SEDLAC harmonized database maintained by CEDLAS at the University of La Plata (Argentina). At this stage we do not report the results concerning the analysis of inequality of opportunity in Latin America, which is still in progress.

2 Inequality of Opportunity

Consider a population of discrete individuals indexed by $i \in (1, \ldots, N)$ and a variable y characterizing our economic outcome of interest, income, whose distribution is a function of the set of circumstances faced by the individual, C_i, and the amount of effort exerted, e_i, such that $y_i = f(C_i, e_i)$. Circumstances are defined as a finite discrete vector of J elements and are assumed to be exogenous because they cannot be affected by individual choices. The population can be divided into M exhaustive and mutually exclusive groups, called types, such that all individuals belonging to the same type T_j share the same circumstances. Define the set of type partitions as Π.

Simultaneously, consider the individual effort to be a continuous variable that depends on both, personal decisions and circumstances, such that individual income can be rewritten as $y_i = f(C_i, e_i(C_i))$.

As income is a continuous variable, following the ex-ante approach in Van de Gaer (1993), we could say that circumstances have no role in the final income distribution if:

$$\bar{y}_j = \bar{y}_s \quad \forall \quad j, s | T_j, T_s \in \Pi \tag{1}$$

According to (1) there is Equality of Opportunity if the expected income is the same across all types. Therefore, any difference in the outcome between two types is a violation of the equal opportunity principle. Then to measure the extent of this violation, that is to estimate the level of inequality of opportunity, simply mean to estimated between-type inequality.

In practice this is done by constructing a counterfactual smoothed distribution, $\hat{y}_i$, where each observation is assigned with the average outcome of her type. Applying a suitable inequality measure I($\cdot$) - such as the Gini index, the mean logarithmic deviation, or the variance - to this counterfactual distribution gives us the inequality attributed to the set of observed circumstances (absolute IOp).

$$\text{Absolute IOp: } I(\hat{y}_i) \tag{2}$$

Dividing this measure over total inequality reveals the share of overall inequality systematically correlated with observable circumstances (relative IOp):

$$\text{Relative IOp: } \frac{I(\hat{y}_i)}{I(y_i)} \tag{3}$$

2.1 Estimation

Until quite recently, most IOp analyses have followed the so-called parametric approach (Bourguignon et al. (2007), Ferreira and Gignoux (2011)). Instead of trying to calculate the average outcome for all possible types, a simple OLS regression method is employed to determine the level of correlation between the circumstances and the outcome.:

$$ln(y_i) = \alpha + \psi C_i + \epsilon_i \tag{4}$$

The prediction of such regression was then used as smoothed counterfactual to estimate IOp:

$$\hat{y}_i = exp[\hat{\alpha} + \hat{\psi} C_i] \tag{5}$$

Note that this regression-based method is equivalent to the estimation of all types' mean under the assumption that the covariance between circumstances and outcome is linear and additive (Brunori et al, 2023).

Such formulation has repeatedly been used in the empirical literature and has been considered to deliver lower bound estimates of IOp because of the problem of partial observability of relevant circumstances. As shown by Ferreira and Gignoux (2011), the explained variance of model (4) can only (weakley) increase if an additional circumstance is added to the list of regressors.

Still, recent contributions have shown that estimates based on this approach could be unreliable (Brunori et al. (2019)). First, the traditional interpretation of the parametric IOp estimation as a lower bound has been questioned. If two types have the same average outcome in the population but different average outcome in the sample observed, the income dispersion across types will be spuriously attributed to IOp, introducing an upward bias. Second, the assumption of a linear and additive relationship between corcumstances and outcome limits the ability of the model to properly fit the data, introducing a further bias with opposite sign. Overall, the uncertainty regarding the true partition in types has compromised the trust in parametric IOp estimates.

As a solution, novel contributions have proposed data-driven approaches, that take information on the data to generate statistically meaningful and non-arbitrary type partitions (Li Donni et al., 2015; Brunori et al., 2023). The remainder of the section explains how one specific regression tree algorithm, the Conditional Inference Trees, is perfectly suited to estimate ex-ante IOp.

2.1.1 Conditional Inference Regression Trees

A regression tree is a Machine Learning algorithm trained to predict an output variable out-of-sample by splitting the observed sample into non-overlapping subgroups, based on a partition of the regressors' space. The final prediction is typically the average realization of the output variable by subgroups. The binary splitting used to grow trees makes them particularly suitable to fit non-linear data-generating processes. From the wide variety of available regression trees, the Conditional Inference Trees (CIT: Hothorn et al. (2006), Hothorn and Zeileis (2008), Hothorn and Zeileis (2015)), have recently been proposed by Brunori et al. (2023) to estimate ex-ante IOp. These trees are schemed as follows:

1. First, CIT performs a test on the global null hypothesis of independence between the outcome (y) and all circumstances (c), at some $1 - \alpha$ level of confidence:

   $H_0^C = D(y|C) = D(y)$

   For each circumstance, the algorithm stores the p-statistic, which needs to be adjusted to avoid type I errors. We apply the Bonferroni correction, quite common for multiple-hypothesis testing $p_{adj} = 1 - (1 - p)^C$.

2. The algorithm then selects the circumstance with the lowest p-value, i.e., the one with the strongest association with the dependent variable y. If $p_{adj} > \alpha$, the algorithm stops. Otherwise, it continues by setting the selected circumstance as a splitting variable.

3. The following step consists in choosing the splitting point, the value used to divide the sample in two subsamples. When the variable is binary this step is trivial, as there is only one way to be split. But, when the variable is continuous, the algorithm needs to check all potential partitions. Thus, let us consider:

   $y_z = y_i : C_i < x$

   $y_{-z} = y_i : C_i \geq x$

   where x defines each possible value in the continuous variable, and z the resulting subsamples. For each x, the algorithm performs a difference-in-means t-test and obtains a p-value.

4. The value x producing the smallest p-value is then selected as splitting value.

5. The whole process is repeated for each resulting subsample until the null hypothesis of independence cannot be rejected for the critical $(1 - \alpha)$.

The recursive binary splitting process leads to a statistically meaningful type partition and takes into account potential non-linearities and interactions across variables. Still, CIT programming bears a technical shortcoming. When non-ordered categorical circumstances are used as regressors, the tree tests across all possible groupings of its categories. This implies a factorial number of combinations, that grows exponentially with the number of categories in the regressor. To make the problem computationally finite, CIT limit the number of possible unordered categories to 30 (that already delivers an astonishing number of $2.65 * 10^{32}$ different combinations). As further explained in the data Section (below), we employ a circumstance, the individual's country of birth, that can take up to 174 possible values. Including this variable as being unordered categorical would pose an unsolvable limit for the computation of CIT. As a solution, we modify the algorithm such that the country of birth can be included in the analysis.

3 Continental Inequality of Opportunity in Europe

3.1 Data

In this study, we evaluate the inequality of opportunity at a continental level in two very different contexts. In regards to Europe, we have ideal information for comparing the phenomenon across different countries because the data is complete and fully comparable. Additionally, we also have information about the country of birth, which allows us to take into account migration when analyzing the phenomenon. Analyzing inequality of opportunity in South America is much more challenging because the data does not come from a single source. Each country has its own survey and the circumstances observed are not always fully comparable. On the other hand, data from South American countries is richer in terms of information. First, it records the region of birth rather than the country of birth and second, it contains information about ethnicity.

For the analysis of European inequality of opportunity we use the European Union Statistics on Income and Living Conditions (EU-SILC) 2019, a widely-used survey run in the European Union that gathers information about several income sources and socioeconomic factors mostly related to living conditions. The survey is collected on a yearly frequency in 31 countries conforming the EU plus some others.[3]. We use the 2019 wave that, for the working-age population, collects information on factors related to childhood and adolescence, as well as the country of origin of the parents, which is used to proxy the individual's country of origin (see below).

While most IOp applications consist of within or between-countries comparisons, in this paper we explore, for the first time, continental IOp. This implies that our estimates of European IOp are based on a pool of 24 different European countries, such as they are treated as coming from the very same region. Note that the original EU-SILC survey data does not proportionally represent the population structure in Europe, e.g. Germany has a population about 16 times larger than Norway but only twice the sample size in the EU-SILC. To fix this problem we stratify the sample to take into account both the share of resident in each European country, and in each country, the share of first generation immigrants from the most frequent countries of birth. As a result, and keeping individuals in working age (between 20 and 75 years old) our final sample representing Europe income distribution is formed by 372,072 observations.

The outcome considered is annual disposable household income (*variable HY020*), adjusted by Purchasing Power Parity conversion factors such that all monetary quantities are fully comparable in international US 2019 dollars. We also account for different household composition by dividing incomes over the squared root of the household size. Table A2 in the Appendix shows, by country, the summary statistics of incomes after these adjustments. Finally, to account for life cycle income dynamics, we adjust the dependent variable erasing the age effect on income.

EU-SILC framework does not provide detailed information on the respondent's country of birth.

[3]Besides 27 countries belonging to the European Union back in 2011 we also have data from Croatia (who entered the EU in 2014), Norway and Switzerland (who do not belong to the EU).

Still, it includes a variable stating whether she was born in the country of residence, in another European country, or outside the EU. For individuals born in the country of residence, we make no further arrangements. However, to get a thinner idea about the country of origin for those born abroad, we use the information on the parental country of birth included in the 2019 wave. For those who report having been born outside the country of residence, we assign the country of birth of their parents. As a result, we are left with 174 different places of birth. As concerns the status of the second-generation immigrant, we create a dummy variable which takes value 1 when the individual is born in the country of residence but his/her parents either had not the citizenship or, in case the information on the citizenship is missing, were not born in that same country.

In our analysis, we also take into consideration other factors such as the respondent's sex, their parents' education level, their parents' occupation and their parents' main source of income. The complete list of all the circumstances and their respective categories that we use in our analysis can be found in Table A3. All remaining descriptive statistics are available upon request.

3.2 Results

We estimate the continental CIT by selecting 99% as confidence level, we further tune the algorithm by imposing the minimum number of observation in terminal nodes to 100. CIT with depth 6 produces a partition in 52 types. The structure of the tree is very complex to keep the structure readable, we split the tree in the *root* tree (Figure 1), which shows the splits up to the nodes defined only by *country of birth*, and 8 sub-trees that grow below the terminal nodes of the root tree (nodes 6, 7, 8, 12, 13, 14, 23, 24). In Figure 1 we show all the splits up to the ones that are based on a circumstance different from country of birth, such as parental education, sex, etc. The fundamental importance of the country in which one is born in shaping opportunities is then evident: the tree, indeed, grows deep using only this key variable. Every final nodes of the root tree gives rise to a sub-tree of varying depths and different splitting variables. All the sub-trees, growing below the terminal nodes of the tree in Figure 1, can be found in the Appendix B. The following information may be useful to browse and interpret the trees. The number in parenthesis in each node simply indicates the node. The rectangular boxes are the Terminal nodes of the tree, representing the final types in which the sample is partitioned. The first number in the final nodes shows the share of population belonging to that group while the second number provides the predicted income of the group of individuals belonging to that node where incomes are re-scaled with EU mean = 1.

An overview of the root tree which splits the sample in groups of individuals only defined by their birth country graphically suggests that being born in poor/rich countries highly correlates with the future individual expected level. Indeed, the mean expected income of the individuals belonging to the sub-samples in the final nodes in Figure 1 increases from left to the right of the root tree and the countries grouped in the nodes are approximately ranked from the poorest to the richest. Denmark, Switzerland, Norway, US, Taiwan, Luxembourg, among others, appear altogether in the right nodes. We observe individuals born in UK, France, Japan, among others, in the group showing a mean expected income which is 144% of the mean European income while those born in countries like Italy, India, Germany etc. reporting a mean expected income which is around the continental average. In the left nodes, instead, we find individuals born from poorer countries like Albania, Slovak Republic, Kyrgyzstan, among others, having an expected mean income which is around 50% of the EU average.

It is curious to notice that all the final nodes of the root tree give rise to sub-trees apart from one exception. Node 13 does not show any further split in the subsequent levels of the tree (sub tree 13 in Figure 2). This suggests that the sub-sample of individuals born in Switzerland, Denmark, Malta, Saudi Arabia, Taiwan, and the US, constitutes a type representing 2.5% of the entire sample. It has an high expected income, which is 175% of the European mean income, and there is no other factor among familiar circumstances, sex, and immigration status that makes a significant difference in terms of expected income. Among various circumstances of birth, the only fact of being born in one of those countries, gives you an expected income which is almost double the European average.

Looking at the terminal nodes 77, 78, 79, 80 of the sub tree 23 in Figure 2, we notice that the expected income of the individuals born in the 14 countries shown on the branch leading to node 23 in Figure 1, is between 46% and 59% of the European average, with individuals whose mother has a high educational level doing slightly better in terms of expected income. Overall, the four types corresponding to the terminal nodes 77, 78, 79, 80 represent around 12% of the European population, as shown by the share reported in each rectangular box. On the contrary, the tree predicts that the most advantaged individuals living in Europe, represent around 3% of the total population and report expected income that almost doubles the mean (look at the terminal nodes of the subtree 14 in Figure 2), are those born in countries like Singapore (SG), Luxembourg (LU) or Norway (NO).

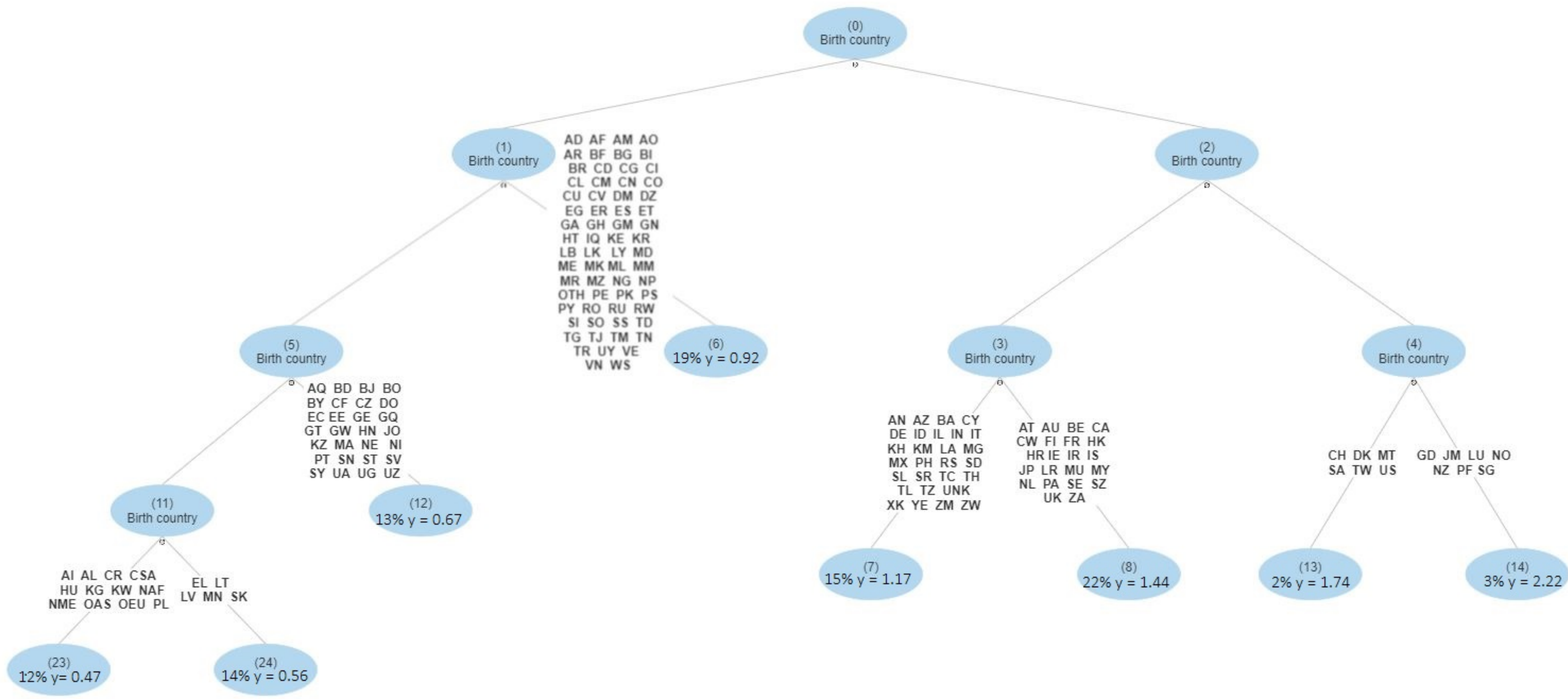

Figure 1: European opportunity tree. The tree is shown up to the nodes in which the circumstance *country of birth* is used as a splitting variable. The first number indicates the share of the population belonging to that sub-sample while the second number is the predicted income of the group. A sub-tree for each of the terminal nodes (6,7,8,12,13,14,23,24) of this tree is found in Appendix.

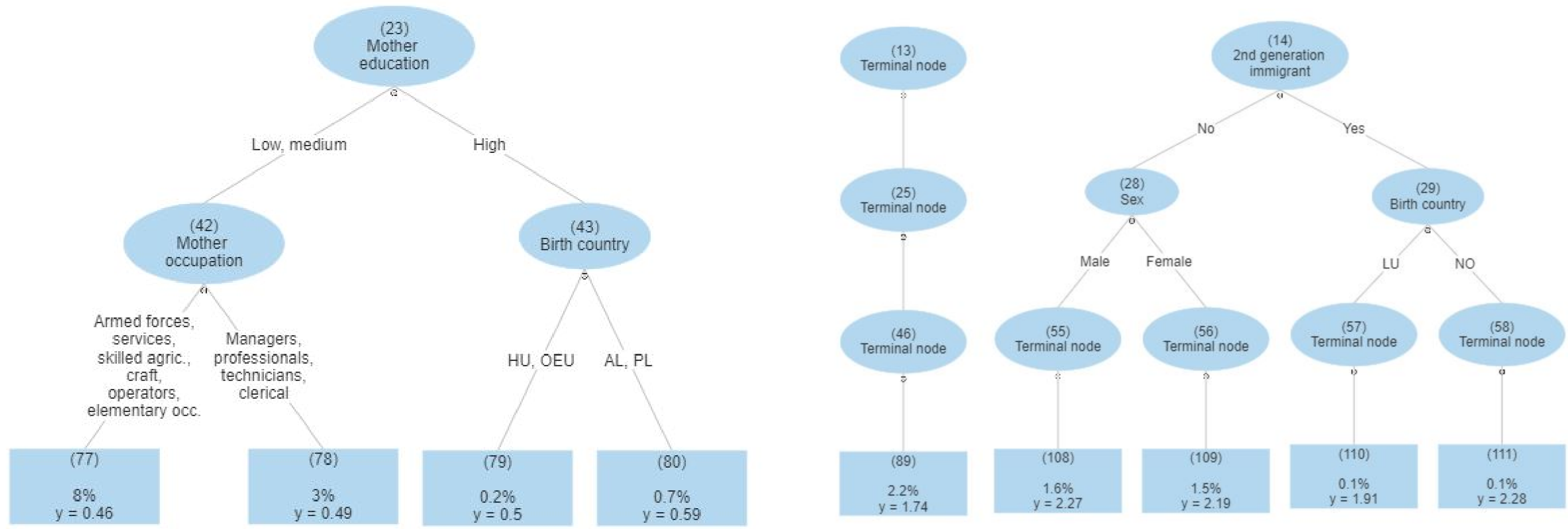

Figure 2: Sub trees 23, 13, and 14. The most advantaged and disadvantaged types.

Finally we use the CIT estimated to estimate the level of inequality of opportunity at the European level and to assess the relative importance of different circumstances.

Table 1 shows that the overall income inequality in Europe amounts to 0.41 Gini points, while the Gini absolute IOp, estimated on the smoothed distribution $\hat{y}$ after grouping observations with the tree (see equation 2), comes to 0.28 points. This implies that our six circumstances (the country of birth, second-generation immigrant status, sex, parental education, occupation, and the main activity status) account for more than two-thirds of total income inequality (68%). It is interesting to notice that the estimate of the overall European IOp is more than double the highest levels of within-country IOp, around 0.13 Gini points, found in Bulgaria (BG), Portugal (PT) and Luxembourg (LU) by Brunori et al. (2023).

Table 2 focuses on the relative contribution of each circumstance on the estimated income IOp. We compute it by re-estimating IOp after removing every circumstance, one by one, comparing this new value with the baseline shown in Table 1. We show two results. First, the difference -in Gini points- between both IOp estimates. Second, to simplify the interpretation, we set to 100 the most important variable (the one with the largest difference with respect to the baseline estimation, in this case, country of birth), and index the rest accordingly.

Among the six variables we account for, we find the country of birth of the respondent to be the most relevant circumstance in shaping IOp in Europe, accounting for 75% of total IOp. Indeed, the remaining circumstances are found to have a negligible contribution.

	Gini	Mean Logarithmic Deviation
Income inequality	0.41	0.33
IOp	0.28	0.14

Table 1: Income inequality and income inequality of opportunity in Europe

	IOp (Gini)	Relative importance
IOp without country of birth	0.07	100
IOp without second generation immigrant status	0.283	0.47
IOp without sex	0.283	0.47
IOp without parental education, occupation and main activity status	0.282	0.94
IOp without parental education	0.284	0
IOp without parental occupation	0.284	0
IOp without parental activity status	0.284	0

Table 2: Circumstances relative importance in income IOp in Europe

4 Continental Inequality of Opportunity in South America

[TBA]

5 Conclusions

[TBA]

References

Bourguignon, F., Ferreira, F. H., and Menéndez, M. (2007). Inequality of opportunity in brazil. *Review of income and Wealth*, 53(4):585–618.

Brunori, P., Hufe, P., and Mahler, D. G. (2023). The roots of inequality: Estimating inequality of opportunity from regression trees. *The Scandinavian Journal of Economics, forthcoming.*

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

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

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.

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.

Hothorn, T. and Zeileis, A. (2008). Generalized maximally selected statistics. *Biometrics*, 64(4):1263–1269.

Hothorn, T. and Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in r. *The Journal of Machine Learning Research*, 16(1):3905–3909.

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

Marrero, G. A. and Rodríguez, J. G. (2012). Inequality of opportunity in europe. *Review of Income and Wealth*, 58(4):597–621.

Milanovic, B. (2015). Global inequality of opportunity: How much of our income is determined by where we live? *Review of Economics and Statistics*, 97(2):452–460.

Pina, A. and Sicari, P. (2021). Enhancing regional convergence in the european union. *OECD Economics Department Working Papers*, No. 1696(3).

A Table Appendix

Country	Survey name	Survey wave	Acronym
Argentina	Encuesta Nacional sobre la Estructura Social	2014	ENES
Bolivia	Encuesta de Hogares	2008	EH
Brazil	Pesquisa Nacional por Amostra de Domicílios	2014	PNAD
Chile	Encuesta de Caracterización Socioeconómica Nacional	2015	CASEN
Colombia	Encuesta de Caracterización Socioeconómica Nacional	2010	CASEN
Ecuador	Encuesta de Condiciones de Vida	2014	ECV
Guatemala	Encuesta Nacional sobre Condiciones de Vida	2011	ENCOVI
Panama	Encuesta de Niveles de Vida	2008	ENV
Peru	Encuesta Nacional de Hogares	2015	ENHAO

Table A1: Latin America household surveys

Country	N	Avg eq. income	Gini	MLD
AT	8878	38794.00	0.28	0.18
BE	10875	35673.16	0.26	0.14
CY	7912	27482.08	0.32	0.17
CZ	13566	14301.47	0.24	0.10
DE	20204	33637.91	0.28	0.18
DK	8624	47751.54	0.24	0.14
EE	10200	16588.85	0.29	0.18
EL	27190	12307.00	0.31	0.22
ES	28112	23538.96	0.32	0.24
FI	16164	41864.33	0.27	0.13
FR	18076	35197.05	0.30	0.17
HU	10809	8284.38	0.28	0.16
IE	6848	40804.14	0.30	0.15
IT	30595	28333.29	0.32	0.27
LT	8158	12664.48	0.34	0.24
LU	7575	57393.73	0.31	0.20
LV	7787	11976.74	0.35	0.27
NL	21343	37270.64	0.27	0.16
NO	10252	60988.60	0.25	0.14
PL	36477	10074.81	0.29	0.17
PT	24015	15244.12	0.33	0.20
SE	8894	38871.46	0.25	0.16
SI	18381	21466.00	0.24	0.10
SK	11137	11787.00	0.23	0.10

Table A2: Summary statistics by country

Circumstance	Categories
Respondent's country of birth:	174 values (classification of Countries using SCL Geo codes)
Respondent's sex:	• Male • Female
Father/mother education:	• Low level (less than primary, primary education, or lower secondary education) • Medium level (upper secondary education and post-secondary non-tertiary education) • High level (short-cycle tertiary education, bachelor's or equivalent level, master's or equivalent level, doctoral or equivalent level)
Father/mother's main occupation (based on International Standard Classification of Occupations, published by the International Labour Office ISCO-08):	• Armed Forces Occupations • Managers • Professionals • Technicians and Associate Professionals • Clerical Support Workers • Services and Sales Workers • Skilled Agricultural, Forestry and Fishery Workers • Craft and Related Trades Workers • Plant and Machine Operators and Assemblers • Elementary occupations
Father/mother activity status:	• Employed working full-time • Employed working part-time • Self-employed or helping family business • Unemployed/Looking for a job • In retirement • Permanently disabled and/or unfit to work • Fulfilling domestic tasks and care responsibilities • Other inactive
Second generation immigrant status:	Binary. The respondent is considered to be a second-generation immigrant when he/she reports being born in the country of residence from parents who were both born in another country or who did not have citizenship of the country of residence (when the respondent was 14 years old).

Table A3: List of circumstances

B Figure Appendix

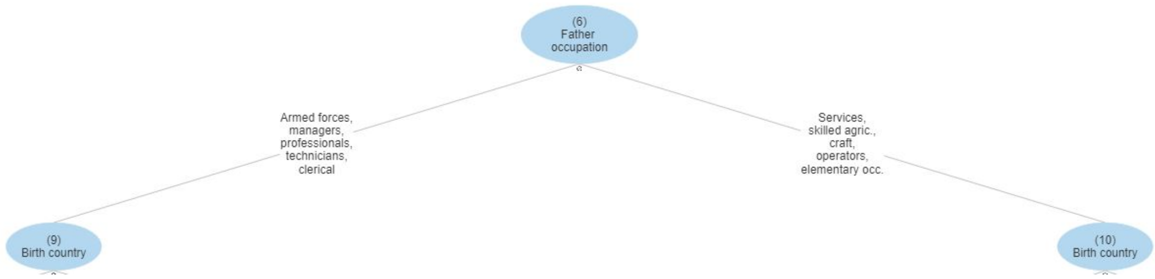

Figure A1: Sub tree 6. One single split is shown for reasons of readability. The subsequent sub-trees starting from nodes 9 and 10 are shown, respectively, in Figure A2 and Figure A3 .

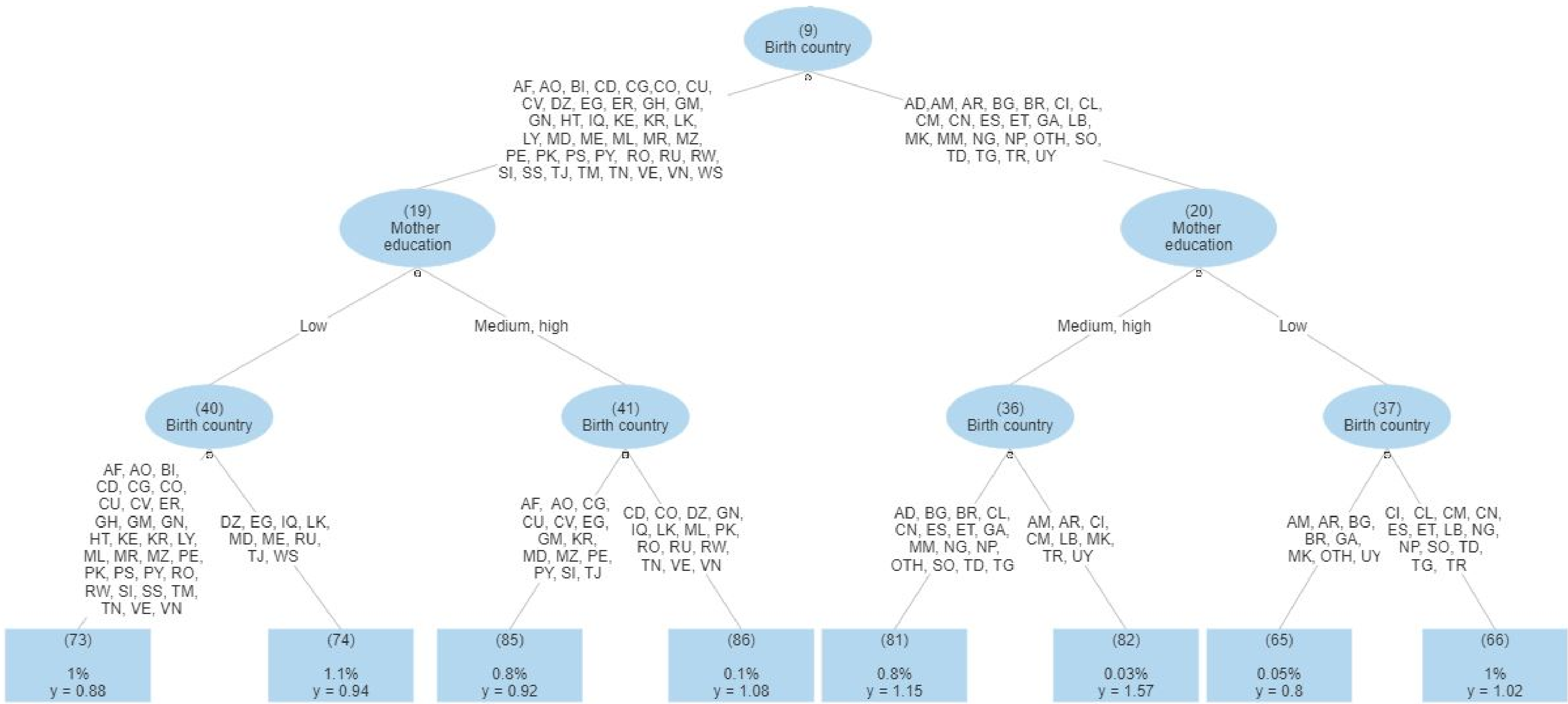

Figure A2: Sub tree 6.9

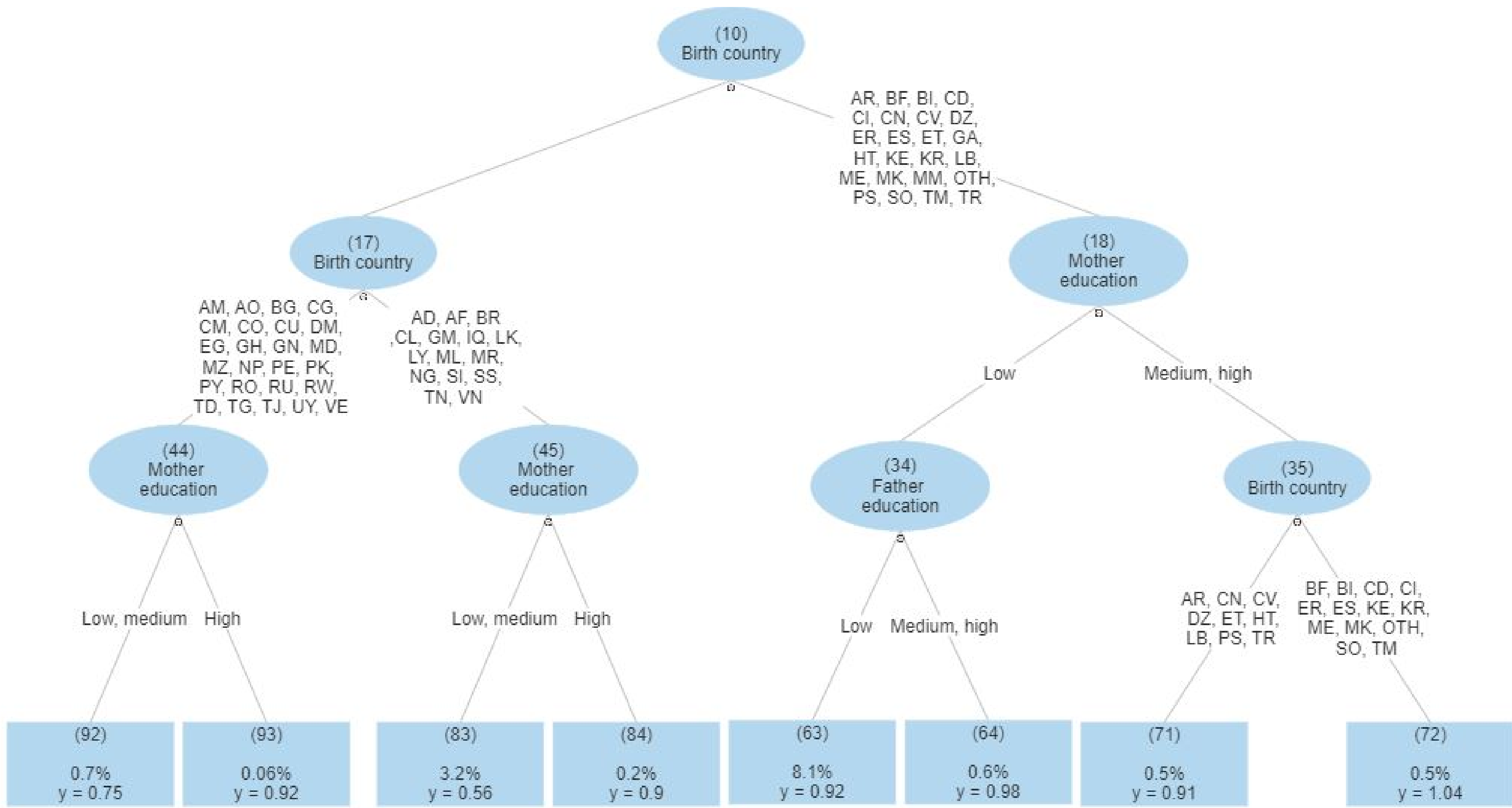

Figure A3: Sub tree 6.10

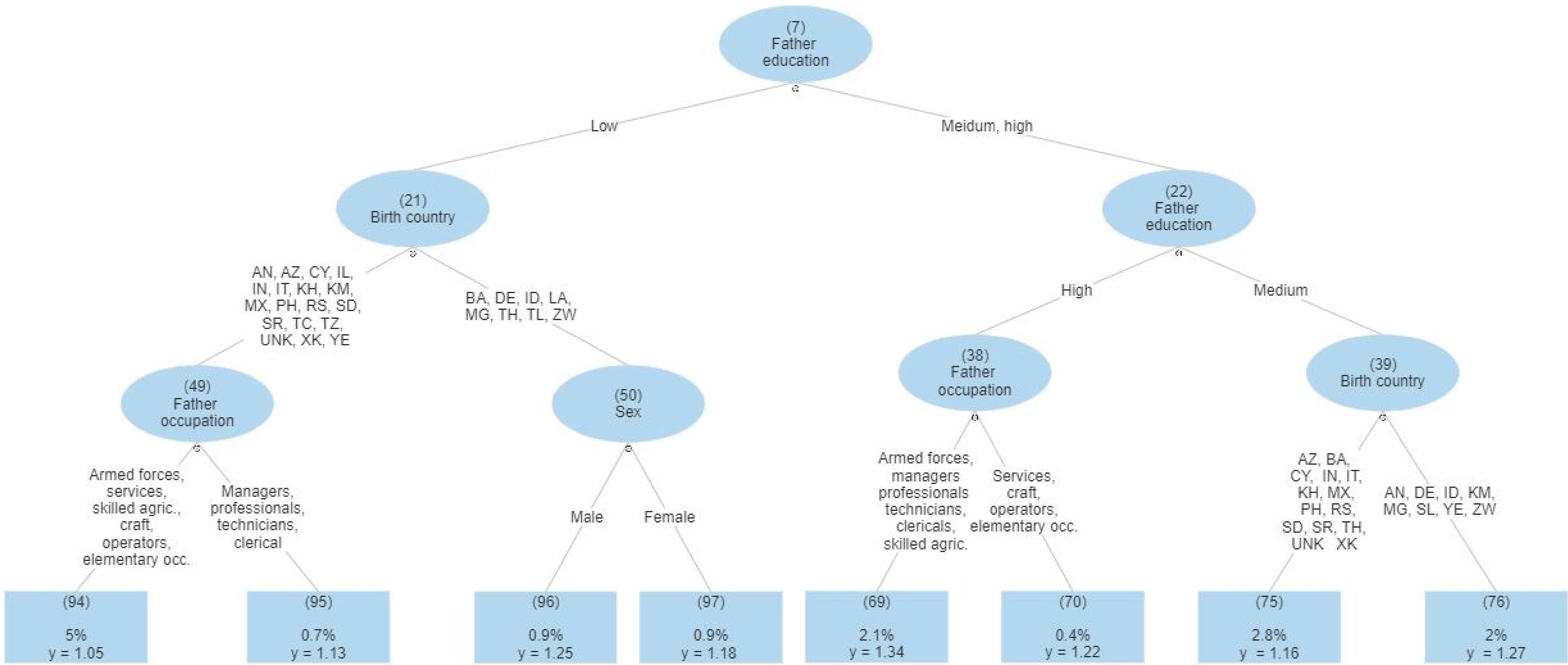

Figure A4: Sub tree 7

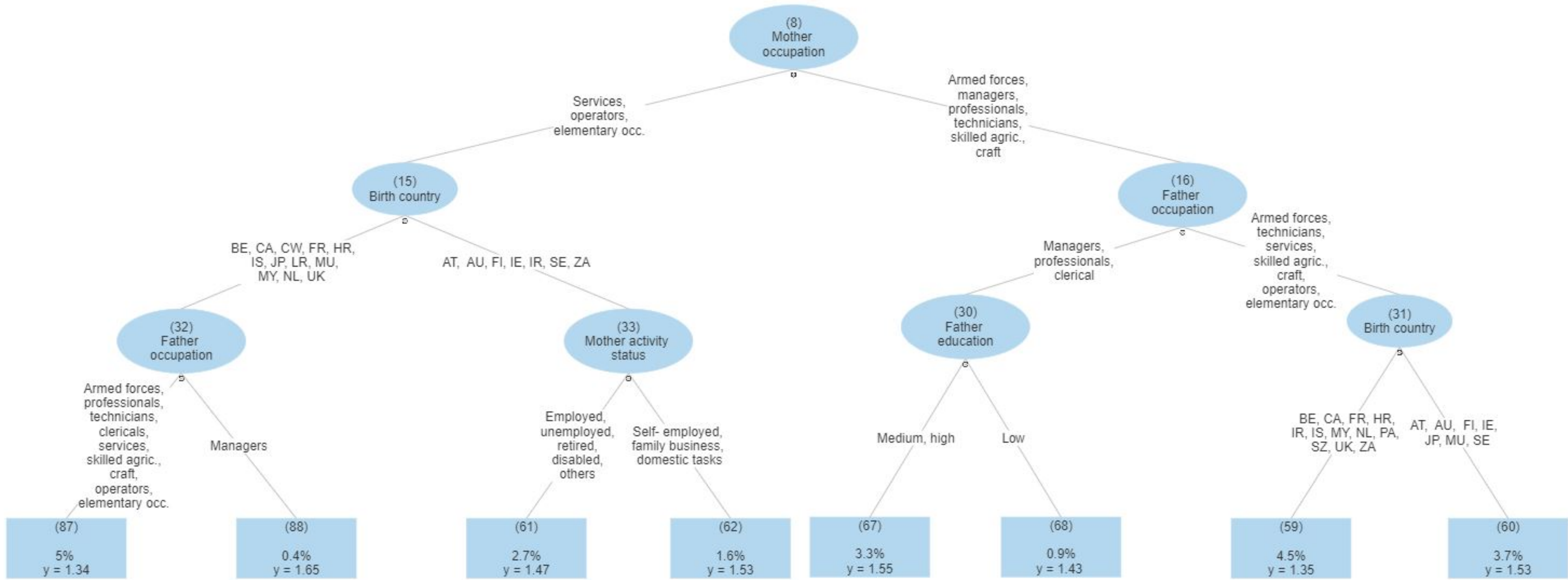

Figure A5: Sub tree 8

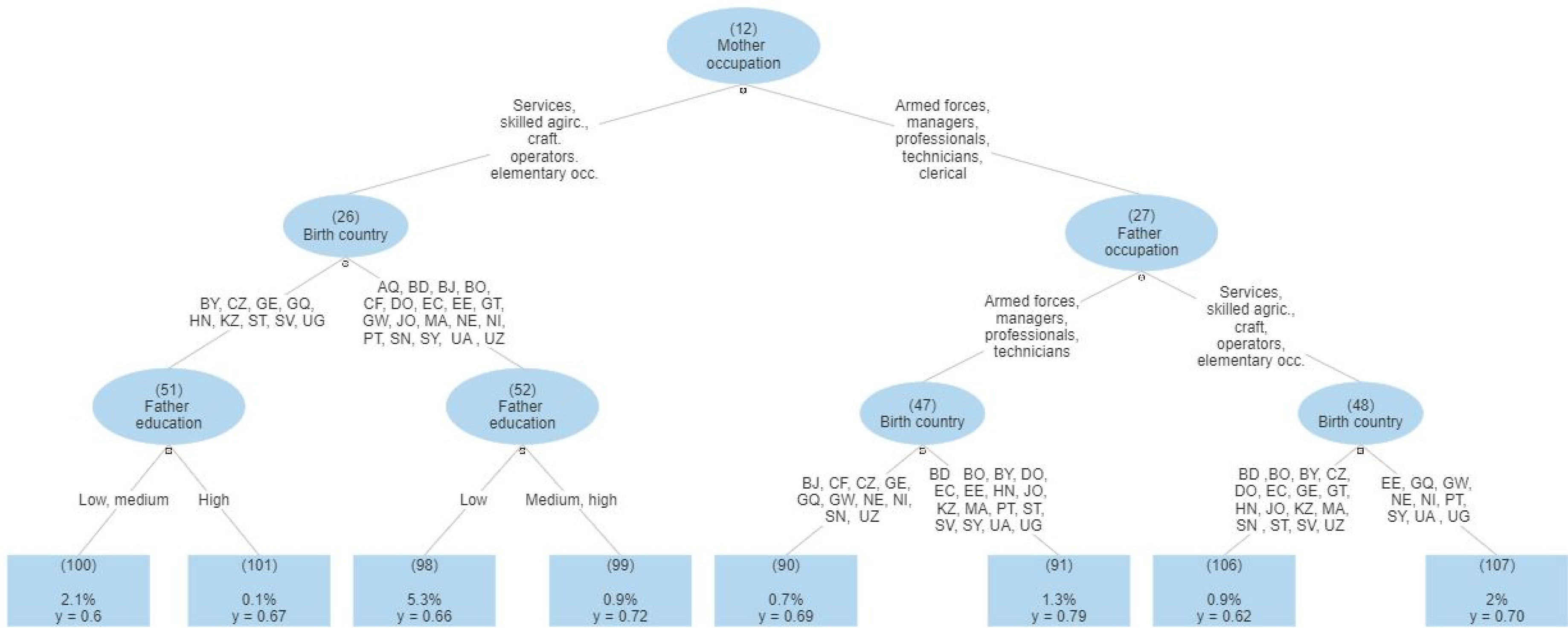

Figure A6: Sub tree 12

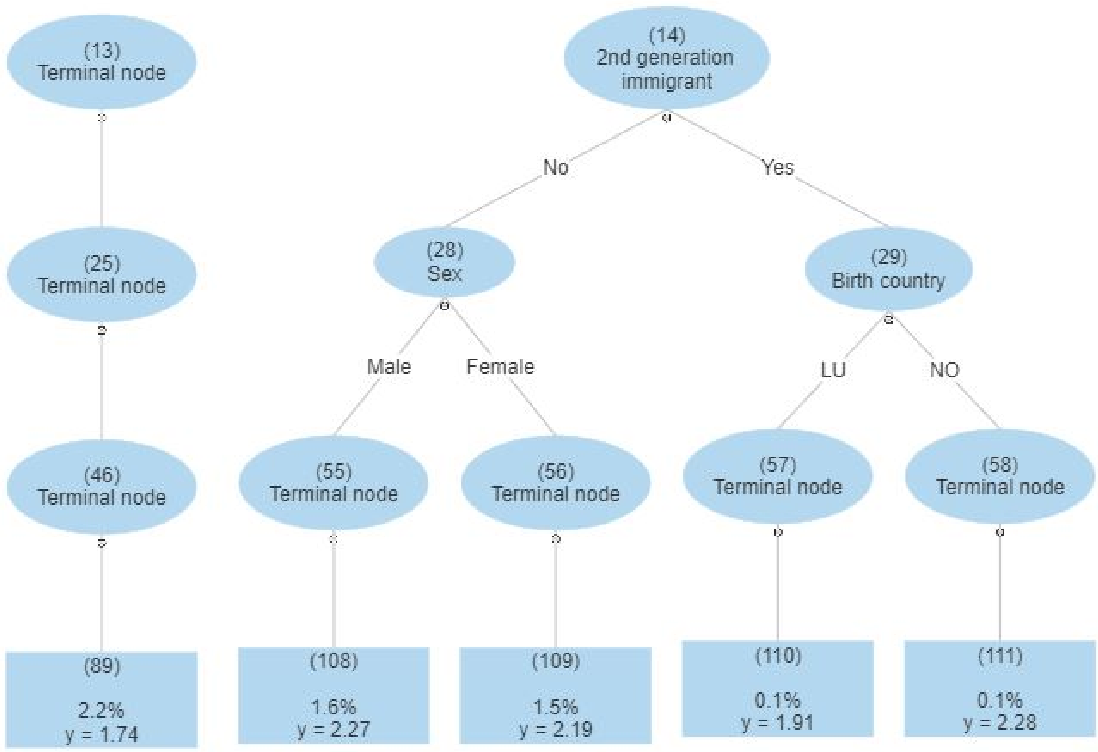

Figure A7: Sub trees 13 and 14

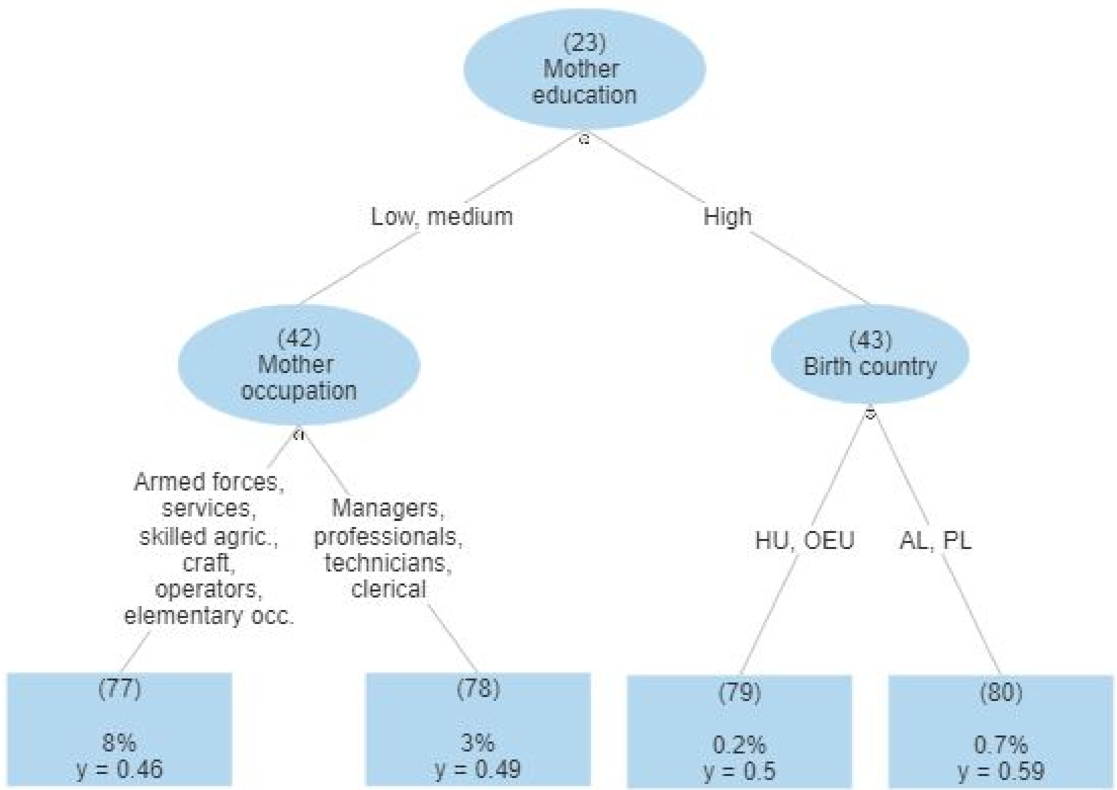

Figure A8: Sub tree 23

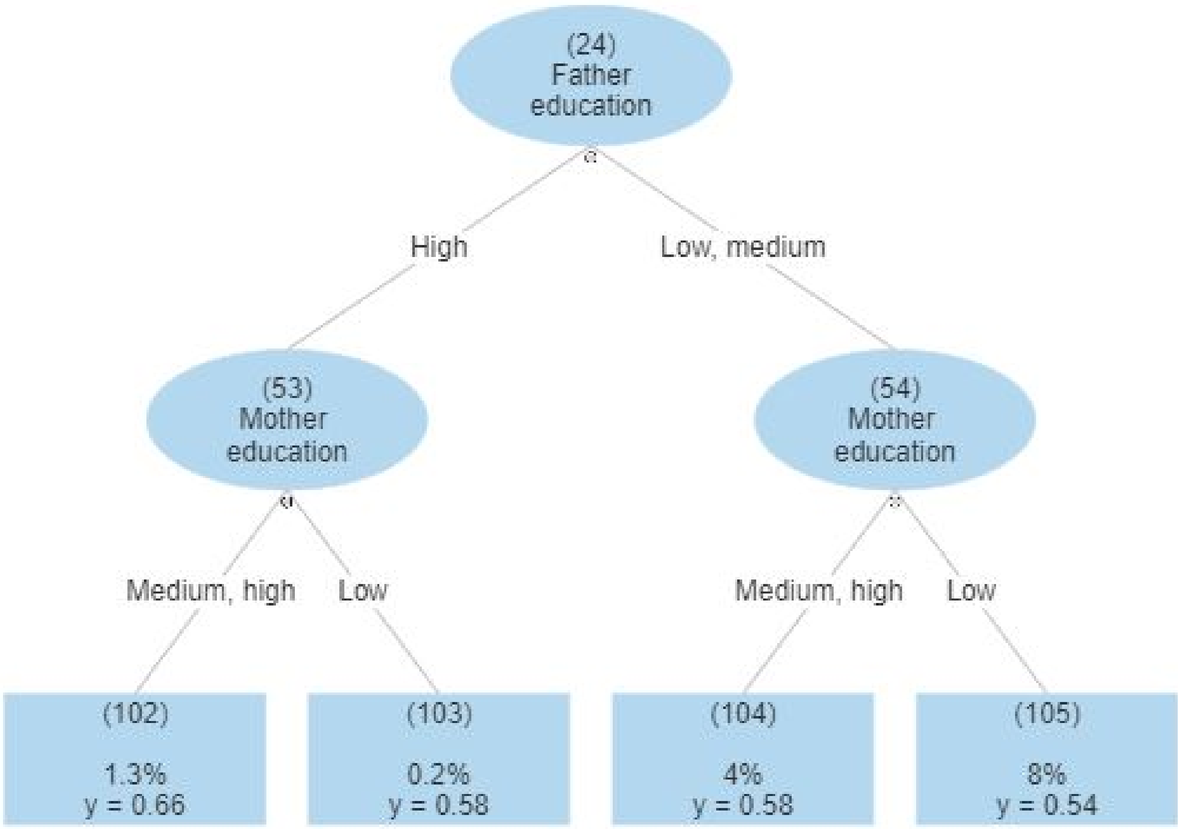

Figure A9: Sub tree 24