NOVEMBER 2023
BxWP2023-12

BORDEAUX ECONOMICS WORKING PAPERS
CAHIERS D'ECONOMIE DE BORDEAUX

Mapping Inequality of Opportunity in France and its Regions: A Data-Driven Analysis of Income Inequality from Fiscal Administrative Data

Thibaut PLASSOT
Univ. Bordeaux, CNRS, BSE, UMR 6060, F-33600 Pessac, France

Michael SICSIC
Institution: CRED, University Paris II Panthéon-Assas, Paris, France

Bordeaux Sciences Economiques
Bordeaux School of Economics

BSE UMR CNRS 6060
Université de Bordeaux
Avenue Léon Duguit, Bât. H
33608 Pessac – France
Tel : +33 (0)5.56.84.25.75
http://bse.u-bordeaux.fr/

Abstract

This paper provides measures of Inequality of Opportunity (IOp) in France and its regions using an original database that synthesizes wide administrative information. The data directly links young adults' incomes in 2019 to those of their parents in 2010, and to other circumstances like gender, family capital, household type, living area, occupation status, education and migratory status of parents. The contributions of the article are threefold: first, it calculates IOp in France based on fiscal administrative information for the first time; second, it provides measures of both ex-ante and ex-post IOp, and with different robustness checks; and third, it computes IOp for each region in France and identifies bottlenecks to equality of opportunity. We show that ex-post IOp accounts for almost half of the total inequality, while ex-ante IOp represents a much smaller proportion. Moreover, we find that IOp measured with relative rank is smaller than IOp measured with absolute income. We also provide an extensive survey of the international literature and show that France is characterized by a moderate level of IOp compared to other countries. Lastly, we highlight the heterogeneity at the subnational level by identifying four groups of regions according to their inequality profiles.

Keywords: Inequality of Opportunity, Regional Analysis, Conditional Inference Trees, Administrative data, France.

JEL: C14, D31, D63, O15, P25.

To cite this paper: **PLASSOT, Thibaut and SICSIC, Michael,** (2023), **Mapping Inequality of Opportunity in France and its Regions: A Data-Driven Analysis of Income Inequality from Fiscal Administrative Data**, Bordeaux Economics Working Papers, BxWP2023-12

https://ideas.repec.org/p/grt/bdxewp/2023-12.html

Mapping Inequality of Opportunity in France and its Regions: A Data-Driven Analysis of Income Inequality from Fiscal Administrative Data

Thibaut PLASSOT[1], Michaël SICSIC[2]

Abstract

This paper provides measures of Inequality of Opportunity (IOp) in France and its regions using an original database that synthesizes wide administrative information. The data directly links young adults' incomes in 2019 to those of their parents in 2010, and to other circumstances like gender, family capital, household type, living area, occupation status, education and migratory status of parents. The contributions of the article are threefold: first, it calculates IOp in France based on fiscal administrative information for the first time; second, it provides measures of both ex-ante and ex-post IOp, and with different robustness checks; and third, it computes IOp for each region in France and identifies bottlenecks to equality of opportunity. We show that ex-post IOp accounts for almost half of the total inequality, while ex-ante IOp represents a much smaller proportion. Moreover, we find that IOp measured with relative rank is smaller than IOp measured with absolute income. We also provide an extensive survey of the international literature and show that France is characterized by a moderate level of IOp compared to other countries (when estimates are available). Lastly, we highlight the heterogeneity at the subnational level by identifying four groups of regions according to their inequality profiles.

Keywords: Inequality of Opportunity, Regional Analysis, Conditional Inference Trees, Administrative data, France.

JEL Classification: C14, D31, D63, O15, P25

---

[1] Institution: Université de Bordeaux – CNRS, Bordeaux, France ; Email: thiplass@gmail.com. Corresponding author

[2] Institution: CRED, Université Paris II Panthéon-Assas, Paris, France ; Email: sicsic.michael@gmail.com

This study has received financial support from the French State in the framework of the Investments for the Future programme IdEx université de Bordeaux / GPR HOPE

I. Introduction

The theoretical framework developed by Roemer (1998) about Inequality of Opportunity (IOp henceforth, or EOp for Equality of Opportunity) allows to go beyond a simple analysis of the inequality of outcomes, separating the drivers of inequality into responsibility and non-responsibility factors. IOp refers to differences in life chances that are determined by circumstances beyond individuals' control, such as their socioeconomic background, race, gender, and geographic location. This approach postulates that individuals' outcomes are a function of both their circumstances, for which society cannot hold them responsible, and their degree of effort, for which they are held responsible. Crucially, the degree of effort must be purged of the influence of circumstances.

The study of IOp is indispensable to address the thematic of social justice and wellbeing and is closely linked to ethical and normative questions. Identifying the principal circumstances that shape opportunities and addressing these sources of inequality is a key policy priority. Measuring IOp also serves as a useful benchmark to evaluate the effectiveness on IOp of policy interventions such as fiscal and redistributive policies or programs aimed at reducing disparities in the effective access to essential services like education or healthcare. Finally, inequality and IOp indicators allow comparisons across time and territories.

The empirical literature has approached this concept from two distinct but complementary perspectives: the ex-ante approach, which examines inequalities before observing effort, and the ex-post approach, which analyzes inequalities after observing the effort or responsibility factor. The former concentrates on differences between "types" defined as groups of persons with similar circumstances, and the latter on differences within "tranches" of effort defined as groups of persons with similar degree of effort. According to Roemer, EOp is achieved when individuals who exert the same degree of effort obtain comparable outcomes, so that the impact of the circumstances on the outcome is insignificant.

France is characterized by a relatively low level of inequality compared to other countries and especially OECD countries. Indeed, France is in the top third of countries with the lowest disposable income inequality (OECD 2023)[3], and this trend is mainly linked to a low

[3] See also the Gini index of the disposable income of the World Bank, which indicates that France has much lower inequality than the world average, but also lower inequality than the United States, the United Kingdom, Italy and Spain. The inequality level is nevertheless slightly higher than in counties like Sweden, Norway, Finland, or Denmark.

level of pre-redistribution inequality (Bozio et al. 2020).[4] Nevertheless, little is known about IOp in France and in comparison with other countries: is the relatively low inequality of standard of living in France mainly composed of IOp? Which circumstances have more weight in the IOp? Are there regional differences in the absolute and relative levels of IOp?

We answer these questions in this paper by estimating measures of IOp in France at the national and regional level. We use a recent and original panel database that synthesizes wide administrative information with many income and socio-demographic variables (Robert-Bobée and Gualbert 2021), and we concentrate on two cohorts of young adults around 28 years old,[5] representing 36,402 parent-child pairs. The database is built on Sicsic (2023)'s work, which calculates for the first time the intergenerational social mobility of income in France by comparing directly the income (from labor and unemployment) of young adults to that of their parents when they were living in the same tax household (around 18 years old). Literature on IOp in France is scarce, and we present for the first time measures based on income using recent data from fiscal records and a data-driven methodology. We implement the algorithm proposed by Brunori and Neidhöfer (2021) to estimate types and effort degrees, and then we calculate IOp with two approaches, ex-ante and ex-post (and with two different indicators: the Gini index and Mean Log Deviation or MLD henceforth). Each approach provides unique and complementary information that enriches the understanding of the issue, captures distinct aspects of the phenomenon and is associated with different institutional frameworks (Fleurbaey and Peragine 2013, Checchi et al. 2016).

Until recently, there was no statistical source that could directly link parents' incomes to those of their children in France, unlike the United States and Scandinavian countries where such matching has been possible for a longer period of time (see section III for an extensive literature review). In this work we use administrative data, which permit more accurate measures given that they are based on income and come from fiscal records, rather than using income declared in surveys, asset indexes, or parents' education as seen in other studies. The earnings variables are reported by employers and are especially reliable as they are controlled by the fiscal administration with frequent audits. Moreover, it includes very high incomes, unlike surveys. Our database also contains a proxy of the family wealth when individuals lived with their parents (at age 18), by using parents' capital income in the analysis. The database

[4] Also, the top 10% pre-tax national income share is one of the lowest (31% in 2021) in comparison with other Western countries (World Inequality Database).
[5] These cohorts represent 4% of the French population.

also includes variables on the area of residence, the occupation status and education of the parents, the type of household (single-parents, both parents), and the migratory status of the parents, all from the Census data.

The main contributions of this article are threefold. First, we calculate IOp in France linking parents' incomes directly to those of their children according to an original dataset based on fiscal administrative information. We find that the principal circumstance in shaping the future income of individuals is the standard of living of their parents, while the gender and the type of housing at age 18 are also strongly statistically significant in determining types. Variables reflecting the territorial environment of individuals at age 18, such as the region of residence and the size of the municipality, also play an important role in IOp. Second, we provide measures of both ex-ante and ex-post IOp for France (while usually only one of the two indicators is estimated in the literature), and compare it with other countries. We show that ex-post IOp accounts for almost half of the total inequality (0.08 out of 0.18, with MLD indicator), while ex-ante IOp represents a much smaller proportion, and that France is characterized by a moderate IOp in international comparison, and an lower-middle level compared to European countries. We apply a number of robustness checks and sensibility of measures according to the chosen form of income variables (continuous or relative ranks), the approach, or the indicators used. We find for instance that IOp measured with relative positions in the income distribution is smaller than IOp measured with absolute income. Third, we compute IOp for each region, which allows to identify the bottlenecks to EOp in each region and for each type of individual. We observe four groups of regions: (i) the northern regions have high income IOp and inequality, (ii) in the South, Occitanie and Provence-Alpes-Côte d'Azur also present substantial IOp, particularly ex-post IOp, (iii) three Western regions and Centre-Val de Loire are characterized by very low absolute IOp, (iv) Île-de-France and Auvergne-Rhône-Alpes display similar indicators aligned with the national average.

The principal constraint of the study is that we only focus on two cohorts of individuals who were respectively 27 and 28 years old in 2019 (born in 1991 and 1992), and whose income is different from their permanent income. However, our results suffer only slightly from a life-cycle bias, as by the age of 28, almost all children are in the labor market (only 1% of individuals are in initial education). Chetty et al. (2014) show that intergenerational mobility is only slightly altered if the income of the children is measured at age 28 or 35.[6] Moreover, focusing on only two generations of almost the same age and with parents with approximatively the same age limits the biases and facilitates the treatment of age effects, usually complex.

Our paper relies on different strands of the literature (see section 3 for an in-depth literature survey). A first strand defined the theory of IOp conceptually (Roemer 1998; Fleurbaey 2008; Fleurbaey and Peragine 2013; Trannoy 2016; Roemer and Trannoy 2016). A second strand estimated IOp empirically: Checchi et al. (2016), Brunori et al. (2023a and 2023b), and Hufe et al. (2022) performed cross-country comparisons, while Carpentier and Sapata (2013) measured IOp for France. Our paper is also linked to papers that estimate intergenerational mobility with administrative data: Chetty et al. (2014) in the US and also Corak and Heisz (1999), Helsø (2021), Heidrich (2017), and Acciari et al. (2022) in Canada, Denmark, Sweden, and Italy. In France, Lefranc and Trannoy (2006) and Kenedi and Sirugue (2021) estimate intergenerational mobility statistics imputing parent's income, and Sicsic (2023) estimate rank-rank correlations and transition matrices directly linking parents and their offspring. Lastly, this paper is related to literature at a sub-national level. Chetty et al. (2014), Acciari et al. (2022), Kenedi et Sirugue (2021) and Sicsic (2023) estimate intergenerational mobility at the sub-national level, and Betthäuser (2021), Galvis and Meisel Roca (2014), Plassot et al. (2022), or Carpantier and Sapata (2013) present IOp estimations at a regional level, in different countries.

The paper is organized as follows: Section 2 synthetizes the conceptual framework about IOp and its place in the literature about social justice. Section 3 details the existing literature about ex-ante and ex-post IOp and highlights the main results, particularly in France. Section 4 contains the description of the data and some descriptive statistics on the sample. Section 5 presents the method used to calculate IOp and the algorithm used to estimate types and effort degrees. Section 6 presents the results of the study, at the national and regional level. Finally Section 7 discusses the results and highlights the contributions of this research.

[6] For instance, rank-rank correlation is only 5% higher at age 40 than at 28, according to the authors. Note also that Chetty et al. (2016) studied the influence of location on earnings, by measuring earnings for children at age 26, a younger age than our study.

II. Conceptual framework

Literature background

Theories of social justice have evolved beyond the sole analysis of inequality of outcomes and have increasingly incorporated the notion of personal responsibility into the egalitarian and non-welfarist approach. In the realm of political philosophy, Rawls (1971) and then Dworkin (1981a, 1981b) have respectively emphasized the importance of equality of primary goods and resources, and recommend focusing on inequality caused by factors for which individuals should not be held responsible. Both authors also acknowledge that individuals are responsible for their preferences and choices in their life plans. Cohen (1989) further contributed to this debate and affirms that preferences are partially influenced by circumstances and resources.

Contrary to Rawls' approach, Sen (1979) and Roemer (1992) do not consider individuals as fully responsible for their preferences. These authors study the diversity of society in terms of social class, aspirations or abilities. Sen (1985) describes capabilities as the freedom of choice (or opportunities) to achieve certain available functionings (i.e., what individuals manage to do - beings and doings). Sen's Capability Approach (CA) recognizes that the choice of the combination of functionings belongs to the personal responsibility of individuals. The CA is grounded on the ex-ante approach of IOp when observing differences between opportunity sets before considering the choices, and does not consider the exercise of freedom. In this sense, this approach differs from the ex-post IOp that focuses on inequality after observing individual's responsibility (Trannoy 2016). In addition, Roemer considers that in the CA, the responsibility factors are not defined sufficiently clearly, and the method does not indicate objectives to reach, nor an index of functioning (Roemer 1996). If Sen focuses on capabilities which represent all the possible ways of life (functionings), the Roemerian approach limits the analysis to the opportunities faced by individuals towards a specific and concrete objective like income or education (Igersheim 2006).

The framework of IOp provides a concrete mechanism, or algorithm, to first draw the line between effort and circumstances, and then orientate the distribution of resources (Roemer and Trannoy 2015; Gaertner and Schokkaert 2012). IOp as developed by Roemer (1998) is concerned with the extent to which circumstances beyond an individual's responsibility determine life trajectories. There are two central principles in the IOp literature: compensation and reward. The former endorses the idea that inequalities caused by factors for which the individual cannot be held responsible must be removed. The latter principle refers to the extent to which inequality caused by factors of personal responsibility must be preserved. Different reward principles have been proposed, for example utilitarian (Roemer, 1998) and natural rewards (Fleurbaey 1994, 1995; Bossert 1995). The literature has well documented the tension between the compensation and reward principles (Trannoy 2016), underlining that this tension extends into a conflict between the ex-ante and ex-post approach (Fleurbaey and Peragine 2013). We address this question at the end of the section, after explaining these notions conceptually.

Conceptual framework

IOp theory considers that the outcome y_i of an individual i is a function of circumstances C beyond an individual's control and of the degree of effort e defined as the factors for which society can hold individuals responsible.

$$y_i = g(C_j, e_m) \qquad (1)$$

Population can be partitioned into k "types" of individuals who share the same circumstances, and C is indexed by $j=1,\ldots,k$. The population is also partitioned into n "tranches" of individuals with the same degree of effort and e is indexed by $m=1,\ldots,n$. The total population can be represented in a $k \times n$ dimensional matrix.

The concept of IOp has been analyzed using two approaches: ex-ante and ex-post. The ex-ante IOp or type approach refers to the unequal distribution of opportunities before observing individuals' effort. Attention is focused on the impact of circumstances in shaping individuals' life chances. Ex-ante IOp analyzes differences between types or social groups. As proposed by Van de Gaer (1993), a counterfactual distribution is generated by replacing the outcome of individuals y_{jm} by the average outcome of their type μ_j, in order to remove any within-type inequality which is considered as fair inequality according to the reward principle.

$$|Y'_{BT}|: y_{jm} = \mu_j \qquad (2)$$

Ex-post IOp underscores the unequal outcomes that individuals achieve after exerting effort. It focuses on differences between individuals with the same degree of effort or within-tranche inequality. When the effort is not directly observed, we can approximate the degree of effort by generating quantiles on each type-specific distribution. Consequently, the identification of types of persons with similar circumstances is necessary in both approaches, and the identification of effort is required only in the ex-post approach. For ex-post IOp, we also construct a counterfactual distribution to remove the between-tranche inequality and focus on the within-tranche inequality. To do this we multiply the outcome of each individual by the quotient of the average outcome of the population divided by the average outcome of the tranche he is in (Checchi and Peragine 2010).

$$|Y'_{WT}|: y_{jm} = y_{jm} \frac{\mu}{\mu_m} \quad (3)$$

Once the counterfactual distributions have been constructed, scholars apply an indicator of inequality to measure IOp. When the indicator allows decomposition (such as MLD), it is possible to estimate the share of fair and unfair inequalities.

Fleurbaey and Peragine (2013) have demonstrated the tension between the two approaches and more specifically the clash between the compensation principle and various reward principles. We do not enter deeper in this discussion in this paper and assume that each approach brings complementary information. Checchi et al. (2016) conclude that policies aimed at reducing ex-ante IOp may include investing in early childhood education and care, or improving the quality of public schools and health systems in marginalized areas. In this sense, policies that address ex-ante IOp include affirmative action programs that counteract historical and systemic discrimination against certain groups, and anti-poverty programs, such as cash transfers and food assistance, which aim to reduce the impact of poverty on children's development and future life chances. At the same time, policies to reduce ex-post IOp may focus on actions to ensure that individuals have the tools and resources necessary to succeed and achieve their full potential, regardless of their background or initial conditions (Checchi et al. 2016). Some examples of policies tackling ex-post IOp are social spending to improve the access to higher education and vocational training, to promote entrepreneurship and innovation, and to provide social protections such as unemployment insurance, pensions and disability benefits.

III. Previous results of the literature on Inequality of Opportunity

Although the international literature on the topic is extensive, it remains at the same time relatively scant and inadequate to enable robust comparisons across time and territories due to the variety of methods, approaches, and variables employed by each author. To enable cross-country comparisons, we have synthesized the key findings from the literature on IOp in Tables A1 and A2 (Appendix), focusing mainly on France and European countries, specifying the type of dependent variable, and the year for which the data is available. The table presents absolute and relative estimations based on the use of the Gini or MLD indicator, and either an ex-ante or ex-post approach.

Literature on IOp in France has demonstrated that ex-ante IOp slightly decreased between the 1970s and the 2000s according to Lefranc et al. (2009), who analyzed the inequality in income depending on social origins. For the same period of time, Bussolo et al. (2023) confirm this trend for France, and also present cross-country comparisons for different years. These authors show that France in 2005 had a lower absolute IOp than Germany, Italy and Great Britain; nevertheless, the relative weight of IOp was lower in Great Britain than in France (13% versus 21%). Another study by Lefranc et al. (2008) calculating ex-ante IOp in various European countries in the 1990s confirms a very low IOp in Sweden, Norway and West Germany, an higher IOp in the United States and Italy, while France, Belgium or Great Britain have an IOp between these countries. The authors also note that France, West Germany and the Netherlands have similar levels of total inequality but in France the weight of IOp is stronger. Other European studies like Suárez Álvarez and López Menéndez (2021) or Brzezinski (2018) integrate a dynamic dimension to study over a short period of time (respectively 2004-2010, and 2005-2011) how the economic crisis affected the inequality structure in each country. The former also provided evidence to validate some convergence between European countries in terms of inequality and IOp.

A seminal contribution in the field was made by Checchi et al. (2016), who present ex-ante and ex-post IOp measures in absolute and relative terms for 25 European countries, using survey data from the European Union statistics on income and living conditions (EU-SILC). The principal results show a positive relation between the total income inequality observed in a country and the share of IOp (whether ex-ante or ex-post) within the total. In addition to focusing on absolute levels of IOp, they also present relative measures of IOp. These two dimensions allow to identify three groups of countries with different profiles in terms of IOp: the centrally-planned countries (high total inequality and intermediate IOp); the continental countries including France (intermediate total inequality and high IOp); and egalitarian countries (low inequality and IOp). Lastly, we only found two studies using an ex-post approach to estimate IOp in France (Checchi et al. 2016; Carpantier and Sapata 2013).

All this research was carried out using household surveys which contain self-reported incomes and retrospective information declared by the respondents, and are therefore subject to biases. There is a paucity of research on income polarization in France based on administrative data, indeed the majority of research on IOp in Europe uses the Household Budget Survey (BDF) of the INSEE, or the EU-SILC (among others, Checchi et al. 2016; Brzezinski 2015; Suárez Álvarez and López Menéndez 2018; Brunori and Neidhöfer 2021).

Literature on IOp at the sub-national level remains thin. Some exceptions are the work of Betthäuser (2021) for the regions of European countries, Galvis and Meisel Roca (2014) for Colombia, or Plassot et al. (2022) for Mexico. For France, Carpantier and Sapata (2013) used the EU-SILC data from 2005 to present ex-post IOp in France and its regions. They found significant differences across regions and a positive correlation between total inequality and IOp, but on a relatively small sample. It is worth mentioning that there is a wide literature on total inequality in France with a territorial approach reflecting significant differences between the region Ile-de-France and other regions (Combes et al. 2011), and between metropolitan and overseas departments (Govind, 2020). A work by Bonnet et al. (2021) analyses the change in spatial income inequality across the departments of metropolitan France since 1922: they found a very substantial decline in inequality between regions over the past hundred years. Finally, the field of social mobility, has developed a more comprehensive approach to territorial analysis. For example, Sirugue (2020) demonstrated how spatial segregation impacts social mobility in France between 2010 and 2016. Also, Sicsic (2023) present indicators of social mobility (absolute upward and downward mobility, rank-rank coefficients) for regions and departments of France using fiscal information on income.

The aforementioned elements underscore the need to enhance the literature on inequality through comparable measures and indicators that can be used for cross-study comparisons. It is important to introduce measures based on administrative data and to provide estimations at the sub-national level.

IV. Data

Our main source of information is the Permanent Demographic Sample (EDP), a panel of individuals built by INSEE that collects administrative information. EDP includes data from population census, annual declarations of social data (DADS), the electoral register, and, since 2015, fiscal and social data (Fidéli and Filosofi) (see Robert-Bobée and Gualbert 2021 for a comprehensive description of the data). The fiscal data in the EDP database cover all incomes from 2010 to 2019, meaning that we can directly compare the incomes of a sample of young adults in 2019 to those of their parents in 2010 when parents and children lived in the same tax household, nine years earlier. We focus on 36,402 parent-child pairs, and use the income of individuals aged 27 and 28 in 2019 who are not studying. At this age, almost all young adults are in the labor market, as only 1% of individuals are in initial education at age 28 (Bernard, 2021). Additionally, Chetty et al. (2014), Kenedi and Sirugue (2021) and Sicsic (2023) confirm that results in terms of social mobility (rank-rank correlation) for young adults only suffer very slightly from the life cycle bias.

By combining different sources of data (especially fiscal records and Census), we identify the circumstances through the sociodemographic, economic, and territorial characteristics in the environment of the child in 2010, and his/her income as a young adult in 2019. We describe the variables selected for the analysis below, and descriptive statistics are presented in Table 1. First, our outcome or dependent variable is the individual income from work and unemployment observed in 2019 for young adults. The distribution of this variable according to the region of residence in 2010 can be visualized in Figure 1. In this figure we observe some regional differences in terms of income, the most notable being between the region Ile-de-France and the other regions.

Table 1: Summary statistics of the sample

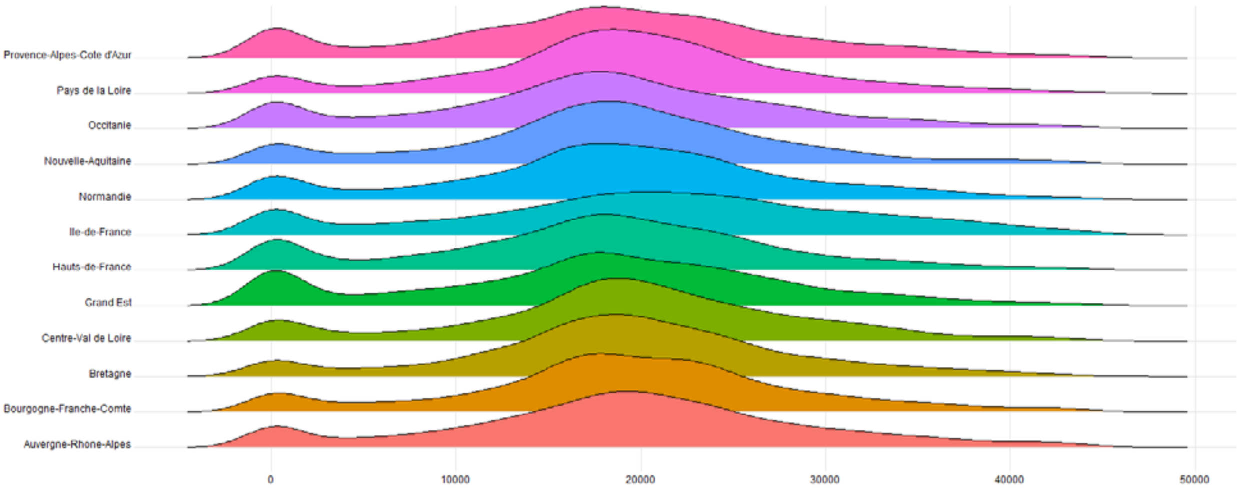
absolute outcome		**Gender**	
Mean (SD)	20000 (12500)	Men	18372 (50.5%)
Median [Min, Max]	19300 [0, 288000]	Women	18030 (49.5%)
relative outcome (percentiles)		**Housing type**	
Mean (SD)	51.1 (28.9)	Landlord	25755 (70.8%)
Median [Min, Max]	51.0 [1.00, 100]	Social Housing	4687 (12.9%)
Parents Standard of Living		Other	5922 (16.3%)
Mean (SD)	23400 (19000)	Missing	38.0 (0.1%)
Median [Min, Max]	19800 [-58100, 978000]	**Household type**	
Missing	38.0 (0.1%)	Complex HH	18090 (49.7%)
Parents Standard of Living (percentiles)		Couple with 1 or 2 children	9491 (26.1%)
Mean (SD)	51.8 (28.8)	Couple with 3 children or more	6936 (19.1%)
Median [Min, Max]	52.0 [1.00, 100]	Single-parent HH	1885 (5.2%)
Capital income		**Parental Education**	
Mean (SD)	6150 (39200)	Without	3070 (8.4%)
Median [Min, Max]	1180 [-206000, 4840000]	Vocational	8404 (23.1%)
Missing	38.0 (0.1%)	Bac	2828 (7.8%)
Capital income (percentiles)		Bac+	5431 (14.9%)
Mean (SD)	51.7 (28.7)	Missing	16669 (45.8%)
Median [Min, Max]	52.0 [1.00, 100]		

Table 1: Summary statistics of the sample (continued)

Occupational		**Mobility**	
Farmers	519 (1.4%)	Immobile	14041 (38.6%)
SEBO	1476 (4.1%)	Mobile	7179 (19.7%)
Executives	3849 (10.6%)	Missing	15182 (41.7%)
Intermediate	4891 (13.4%)	**Regions**	
Employees	2805 (7.7%)	Auvergne-Rhone-Alpes	4390 (12.1%)
Manuals	5772 (15.9%)	Bourgogne-Franche-Comte	1621 (4.5%)
Other	421 (1.2%)	Bretagne	2054 (5.6%)
Missing	16669 (45.8%)	Centre-Val de Loire	1453 (4.0%)
Immigrant status		Grand Est	3323 (9.1%)
Non Immigrant	17841 (49.0%)	Hauts-de-France	3931 (10.8%)
Immigrant	1954 (5.4%)	Normandie	2012 (5.5%)
Missing	16607 (45.6%)	Nouvelle-Aquitaine	2927 (8.0%)
Municipality Size		Occitanie	2934 (8.1%)
Outside city's attraction	2220 (6.1%)	Pays de la Loire	2290 (6.3%)
less than 50,00 inhab	4306 (11.8%)	Provence-Alpes-Cote-d'Azur	2497 (6.9%)
between 50,000 and 200,000 inhab	6801 (18.7%)	Ile-de-France	6918 (19.0%)
between 200000 and 700,000 inhab	8161 (22.4%)	Missing	52.0 (0.1%)
700,000 inhab or + (without Paris)	7449 (20.5%)		
Paris Area	7465 (20.5%)		

Notes: The total sample contains 36,402 observations. The same statistics for each region can be found in Table A3. Coverage: Metropolitan France, individuals born between 1989 and 1992, included in their parents' tax return in 2010, 2011 or 2012 and who have positive or zero income in 2019. Source: Insee-DGFiP-Cnaf-Cnav-CCMSA, échantillon démographique permanent 2020.

Figure 1: Distribution of the income in 2019 according to the region of residence in 2010



Note: To facilitate the visualization, the x-scale of the boxplots has been fixed at 50,000 euros, and does not represent extreme values. Source: Insee-DGFiP-Cnaf-Cnav-CCMSA, échantillon démographique permanent 2020.

Secondly, we describe the variables considered as circumstances. We include the standard of living of the tax household (that of the parents) to which the young person was attached in 2010 (at age 18). It corresponds to disposable income (i.e., income from work, unemployment, pension, and capital, minus taxes, to which monetary social benefits are added) divided by an equivalence scale. It also takes into account alimony payments made in the case of separation, and in the event that the young person was working, these revenues are subtracted from the tax household. These incomes include all taxable incomes, which can be a proxy of the wealth of the family.[7] Another variable of interest is the capital income in 2010. It is important to note that the three variables described above are used in a first (baseline) and second model in a continuous form (absolute incomes), then, in a third model, we generate percentiles (calculated by cohorts) to obtain relative positions.

We also control for the sex of the young adult, for the housing status (i.e., whether the household is an owner, social sector tenant or private sector tenant), and the household type in 2010 (i.e., either a couple with one or two children, a couple with three or more children, a single-parent family, or complex households). Regarding parents, the models take into account the level of education, as well as the socio-occupational category[8] of the reference parent (the one with the higher income). The rest of the circumstance variables are of geographical type. We include a variable reflecting whether the reference parent (with higher income) emigrated from another country or not, and we include the region[9] of residence at age 18. Importantly, we exclude the Corsica region from the study due to an insufficient number of observations. Finally, we include a communal level variable reflecting the type of territory in terms of the area of attraction of cities. For this last variable (municipality size), we distinguish six communal categories: Outside urban attraction areas, less than 50 thousand inhabitants, between 50 thousand and 200 thousand inhabitants, between 200 thousand and 700 thousand inhabitants, more than 700 thousand inhabitants (excluding Paris), and Paris.

As observed in Table 1, there are three variables, namely parental education, occupation, and immigrant status, where 46% of the observations have missing information. Additionally, two other variables, household type and municipality size, have approximately 20% of observations with missing information. It is worth mentioning that our methodology can handle missing data by considering the information of surrogate variables. As a robustness check we also calculate the IOp using a sample of 19,708 observations without any missing data. The descriptive statistics for this sample can be found in Table A3 (Appendix). We note that the average income of young adults, the standard of living and capital of parents is almost the same

[7] Note that incomes imputed to owner-occupiers are not included.

[8] We consider seven categories: farmers, self-employed, executives, intermediate professions, employees, workers, and others.

[9] In total 12 regions after excluding Corsica: Auvergne-Rhône-Alpes, Bourgogne-Franche-Comte, Bretagne, Centre-Val de Loire, Grand Est, Hauts-de-France, Ile-de-France, Normandie, Nouvelle-Aquitaine, Occitanie, Pays de la Loire, Provence-Alpes-Côte d'Azur.

in both samples, but the observations with missing values are characterized by a high rate of single-parents, lower incomes, and living in Ile-de-France at age 18.

Finally, note that while the literature generally considers as factors beyond the individual's responsibility all the circumstances at the age of 14 or before, we use the information when the subjects were 18 years old and living with their parents.

V. Method to identify types and degree of effort

Roemer's contribution in 1998 has been translated into different definitions of IOp. As we mentioned above, in both the ex-ante and ex-post approaches we need to identify types, and only in the ex-post we identify degrees of effort. For both approaches, we follow Brunori and Neidhöfer (2021)'s algorithm. One assumption is that the outcome of individuals is positively associated with the degree of effort. To identify degrees of effort we concentrate on the part of the effort that is not linked with circumstances (i.e., type).

Type identification

We identify the types who share similar circumstances using a data-driven approach as proposed by Brunori and Neidhöfer (2021). This recursive partition method extends the works of Li Donni et al. (2015) who use a latent class model to partition the population into types, and Hothorn et al. (2006) who use conditional inference trees to overcome overfitting and selection bias problems present in other decision tree methods. Conditional inference trees allow to predict the response variable based on covariates. The method uses diverse test procedures to determine the association between the input variables (here circumstances) and the outcome (response variable). One of the main advantages is that the procedure only selects the relevant circumstances, identifies the thresholds for partitioning, and considers the interactions between covariates.

The recursive binary splitting algorithm first tests the global null hypothesis of independence between the response variable and the covariates. It is noteworthy that if no statistically significant association is found for any circumstance, it indicates ex-ante equality of opportunity. In such a case, there is only one type in the society and no between-type inequality is perceived. In the opposite case, we detect the covariate with the lowest p-value and this variable is chosen as the first splitting point to divide the sample into two first groups. Secondly, the threshold or splitting point chosen within the variable to divide the population into two groups depends on the type of variable. When this is binary, each category forms a group; when the variable is continuous or categorical, the algorithm will select the better cutoff after testing for the discrepancy between all the possible subsets and choosing the partition with the lowest p-value.

Thirdly, the algorithm repeats the steps, choosing for each of the new groups the other variable with lowest p-value. The algorithm ceases its iteration once any statistically significant association between the response variable and the covariates can no longer be discerned. The types identified are referred to as terminal nodes, and the inequality between types is an estimation of ex-ante IOp. As mentioned earlier, our data present a high number of observations with missing values for certain variables. To address this, conditional inference trees operate surrogate splits. The procedure begins by assigning a weight of zero to observations with missing values for a specific variable and then searches for an initial split. Subsequently, it identifies surrogate variables that exhibit similar patterns to the variable with missing values and looks for the most accurate surrogate split that replicates the initial split. More details on this methodology can be found in Hotorn et al. (2006), or Brunori and Neidhöfer (2021).

Effort identification

The effort dimension is required to estimate ex-post IOp. When the effort cannot be observe or directly measured, the identification of the types is a preliminary step to approximate effort. It is further assumed that the outcome increases as a monotonic function of the effort. Importantly, Roemer differentiates the level from the degree of effort, considering that the level of effort is partly influenced by circumstances. Roemer (2002) argues that "*we should somehow adjust for the fact that those efforts are drawn from distributions which are different*." To address this concern and following Roemer's approach, we define quintiles on each type-specific distribution of the outcome.

To estimate the shape of the outcome distribution in each type, a sufficient number of observations is required in each type. Following Brunori and Neidhöfer (2021), who identify effort based on a procedure proposed by Hothorn (2018), we use a linear combination of Bernstein basis polynomials (Equation 4) and a ten-fold cross-validation approach to approximate the shape of the outcome distribution in each type.

$$B_m(t,a,b) = \sum_{i=0}^{m} \beta_i b_{,mj}(t,a,b) \quad (4)$$

Equation 5 presents the Bernstein basis polynomial of degree *m*. We search for the most appropriate degree for a positive outcome variable $t \epsilon [a,b]$.

$$b_{j,m}(t,a,b) = \frac{1}{(b-a)^m}\binom{m}{j}(t-a)^j(b-t)^{m-j}, \quad \forall j = 1, \dots, m \quad (5)$$

The process first divides the population of each type into ten samples or folds (f = 1,…,10). Then it repeats the procedure for different orders from one to ten as follows. For each order, an iterative process for every fold from one to ten is realized, the f-th fold corresponds to the test sample and is removed, while the remaining sample becomes the training sample. For each fold, we employ on the training sample monotone increasing Bernstein polynomials of degree m to model the shape of the outcome distribution specific to the type. Then, we predict the cumulative distribution of the type on the test sample and by extension, the out-of-sample log-likelihood. For each order from 1 to 10 we calculate the sum of the test sample log-likelihood. We finally identify the order that maximizes the out-of-sample log-likelihood.

VI. Results

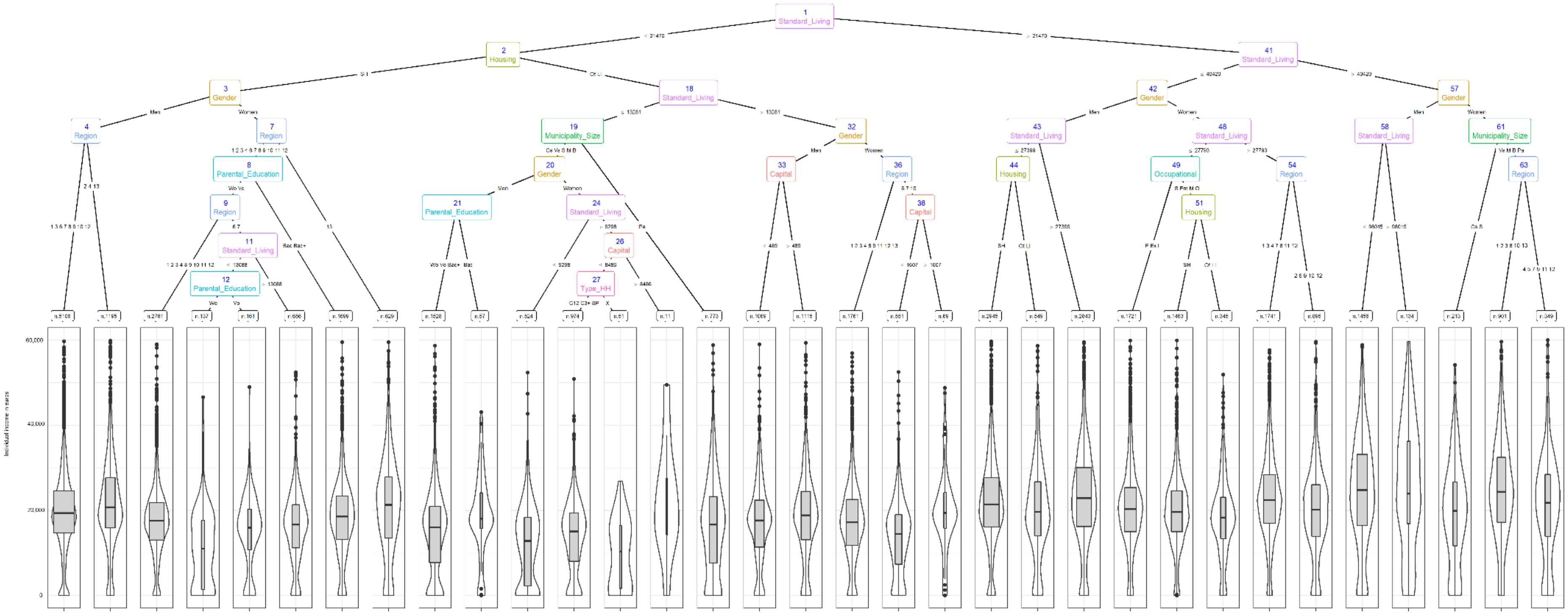
National

Identification of types through conditional inference trees

First, we describe the structure of the tree derived from the first (baseline) model (Figure 2). The tree consists of 33 types distributed across eight levels of depth.[10] The first split is determined by the parents' standard of living, the variable with the highest statistical significance or predictive accuracy. This first split distinguishes households with a standard of living exceeding 21,470 euros per year, positioned on the right side of the tree, from those below this threshold, located on the left side. On the left side, the housing type becomes the second most significant variable, and dissociates social housing residents from landlords and other residency types. On the right side, the second most important variable is once again the standard of living, with a threshold of 40,420 euros. Gender emerges as a significant splitting point at the second depth level and is used in six splitting points throughout the tree. Types composed of women tend to have lower average incomes compared to men with similar circumstances.

[10] The level of depth of a conditional inference tree refers to the number of levels from the root node to the terminal nodes that represent types. Each level of the tree represents a decision or split based on a particular variable or condition. A deeper tree would have more levels, while a shallower tree would have fewer levels. The depth of a conditional inference tree is determined during the model training process and the choice is data-driven. The optimal depth must be balanced to avoid overfitting when growing an overly complex tree, and avoid underfitting when the tree is too simplistic.

Figure 2: Conditional inference tree for France, Model 1



Note: To facilitate the visualization, the y-scale of the boxplots has been fixed at 60,000 euros, and does not represent extreme values. The circumstances considered are Standard of Living (purple), Housing (light green), Gender (yellow), Region (blue), Municipality Size (green), Capital (red), Occupational (green), Type of Household (pink), Parental Education (light blue). Regions are numerated from 1 to 13 and correspond respectively to Auvergne-Rhone-Alpes (1), Bourgogne-Franche-Comte (2), Bretagne (3), Centre-Val de Loire (4), Corse (5), Grand Est (6), Hauts-de-France (7), Normandie (8), Nouvelle-Aquitaine (9), Occitanie (10), Pays de la Loire (11), Provence-Alpes-Cote d'Azur (12), and Ile-de-France (13). The Corse region is excluded from this analysis. For the categories of the Housing variable SH corresponds to Social Housing, Ll to Landlords, and Ot to other forms of ousing. For the categories of the Parental Education variable Bac corresponds to the Baccalauréat[11] level, Bac+ to a level superior to the Baccalauréat, Vo corresponds to Vocational, and Wo to a level inferior to the Baccalauréat. For the categories of the Municipality Size variable, Cs corresponds to Outside urban attraction areas, Vs to Areas of less than 50,000 inhabitants, S to Areas between 50,000 and 200,000 inhabitants, M to Areas between 200,000 and 700,000 inhabitants, B to Areas with 700,000 inhabitants or more (without Paris), and Pa to Paris Area. For the categories of the Type of household variable C12 corresponds to households with one or two children, C3+ to households with three children or more, SP to Single persons, and W to Complex households. For the categories of the Occupational variable F corresponds to Farmers, Ex to Executives, I to Intermediate professions, S to self-employed, Em to Employees, and O to others. Source and coverage: see table 1.

[11] The Baccalaureat is a secondary education diploma awarded to students who have successfully completed their high school studies.

The region where individuals lived at 18 years old serves as a significant variable in six splitting points, since the fourth depth level. However, the groups of regions identified differ depending on the circumstances, and there is no consistent division. Another important variable is the size of the municipality, which becomes decisive at the fourth level. For individuals with a low standard of living at age 18, it separates the Paris area from other types of municipalities, while for those with a high standard of living, it distinguishes small areas from other municipality types. Parental education is only significant on the left side of the tree, indicating its relevance for individuals who were most disadvantaged at age 18. The parents' capital is considered in three nodes, and the type of household is used in one node, distinguishing complex households and single parents from couples with children. Lastly, the occupational variable is chosen in one node, separating farmers, executives, and intermediate professions from other occupations, with the former having higher average incomes than the latter.

IOp estimations

We describe here our baseline model using absolute income values, while the other models are used to compare results in the robustness checks section. The ex-post IOp estimated using the Gini index attains 0.24, while the total inequality has a value of 0.32 (Table 2). Using the MLD, we found an IOp of 0.08 out of 0.18 for the total inequality, representing 46% of the total. The ex-ante IOp estimates were lower, with a Gini index of 0.09 and an MLD of 0.01, representing 8% of the total inequality. Our results align with existing literature showing that ex-post are generally higher than ex-ante estimations (Checchi et al. 2016; Fleurbaey and Peragine, 2013; Plassot et al. 2022). Thus, ex-post IOp accounts for almost half of the total inequality in France, while ex-ante IOp represents a much smaller proportion.

Table 2: IOp estimations for France

	Gini			MLD				
	Total Inequality	ex-post	ex-ante	Total Inequality	ex-post	ex-ante	% ex-post	% ex-ante
Model 1 (Baseline)	0.32	0.24	0.09	0.18	0.08	0.01	46%	8%

Note: The Model 1 calculates IOp considering absolute incomes in 2019. MLD stands for Mean Log Deviation.
Coverage: Metropolitan France, individuals born between 1989 and 1992, included in their parents' tax return in 2010, 2011 or 2012 and who have positive or zero income in 2019.
Source: Insee-DGFiP-Cnaf-Cnav-CCMSA, échantillon démographique permanent 2020.

Comparison with other estimates

To start with, we compare the IOp estimates for France with those of other countries, particularly ex-ante IOp which is the main indicator in the literature. Table A1 in Appendix presents an in-depth review of IOp estimation in the literature. Our results show that ex-ante IOp measured by the Gini index or MLD indicator is lower than in Italy, Spain, Hungary, and several non-European Union countries such as South Africa, Ethiopia, Mexico, Argentina, or Chile, but higher than in Finland and Sweden[12]. The ex-ante IOp value for France is very close to that found in Germany, Belgium, and the UK, depending on the publication. France is thus characterized by a modeate ex-IOp level compared to European countries, a moderate IOp in international comparison (when estimates are available). Ex-post IOp is more difficult to compare with other countries due to the scarcity and variability of estimates.

We then compare our study's findings with previous research on IOp in France, acknowledging that differences in data, methodology, and the choice of circumstances exist (Table A1). In our first model, we found that total inequality measured using the Gini coefficient is 0.24, which is similar to the values reported by Checchi et al. (2016) and Suárez Álvarez and López Menéndez (2021). Meanwhile, total inequality measured with the MLD is 0.18, slightly higher than Checchi et al. (2016) and Suárez Álvarez and López Menéndez (2021), but much lower than Bussolo et al. (2020). The ex-post IOp calculated in our first model aligns with the work of Carpentier and Sapata (2013) who found a value of 0.22 for the ex-post inequality of opportunity (IOp). Similarly, our estimate of 0.09 for the ex-ante IOp using the Gini coefficient is in line with the findings of Brunori et al. (2023a) and Suárez Álvarez and López Menéndez (2021), who reported values of 0.09 and 0.08, respectively. When using MLD, our ex-post IOp estimate is 0.08, which is close to Checchi et al. (2016)'s IOp of 0.05. Our ex-ante IOp estimate of 0.01 using the MLD is consistent with Suárez Álvarez and López Menéndez (2021) and Checchi et al. (2016), who reported similar values of 0.01 and 0.02, respectively. However, our results differ from Bussolo et al. (2020), who found a value of 0.07 for ex-ante IOp in France in 2005.

The relative weight of ex-ante IOp is 8% in our first model and is in line with Suárez Álvarez and López Menéndez (2021), who found a comparable proportion of 10%. Our measure is slightly lower than the estimate of 13% reported by Checchi et al. (2016), and much lower than the estimate of 21% measured by Bussolo et al. (2020). When adopting an ex-post approach, IOp accounts for 46% of total inequality, which is much higher than the 28% reported by Checchi et al. (2016). These differences may be due to a number of methodological and data-related differences. We believe that our administrative data, which directly uses the

[12] Our comparisons here are mainly with Checchi et al. (2016) and Brunori et al. (2023a).

parents' standard of living, gives a more accurate estimate of the IOp, but another source of difference could be that we observe the circumstances of an individual at age 18, while it is generally at age 14 in the other studies.

Regional analysis

Identification of types through conditional inference trees for each region

The trees for each region are presented in Figures A1 to A12. The number of types differs in each region, from three types in Centre-Val de Loire, to a maximum of twelve types in Hauts-de-France (HFr). Other regions present a high number of types, in particular Normandie (eleven), Grand Est (nine) or Bourgogne Franche-Comté (eight), while the other regions are characterized by five to seven types.

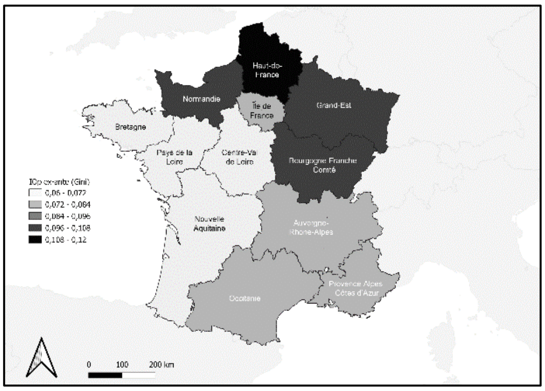
The variable chosen for the first splitting point is the standard of living in eight regions, and the housing type in four regions. Gender is used in all the regions in the first or second level of depth, and the capital variable is significant in five regions. The occupational variable is significant in three regions, while parental education and type of household are only used in two regions. Finally, the municipality size is only determinant in HFr, and the immigrant characteristic of the parents is significant in Auvergne-Rhône-Alpes (ARA), where children of immigrants have -other circumstances being equal- a lower average income.

IOp estimations

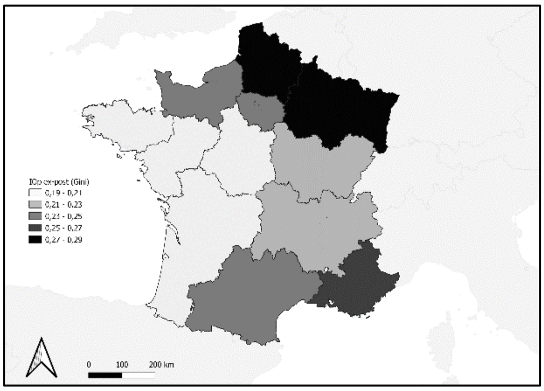
The results reveal heterogeneities between regions or groups of regions (Table A4). For instance, Bretagne (Bre) and PLo demonstrate the lowest ex-post IOp, as measured by the Gini index, with a value of 0.19, whereas Grand Est (GE) and HFr exhibit the highest IOp, with respective values of 0.29 and 0.28 (Figure 3). Regardless of the approach (ex-ante or ex-post), or inequality indicator used, regions can be categorized into groups.

Figure 3: Regional IOp estimations using Gini and MLD

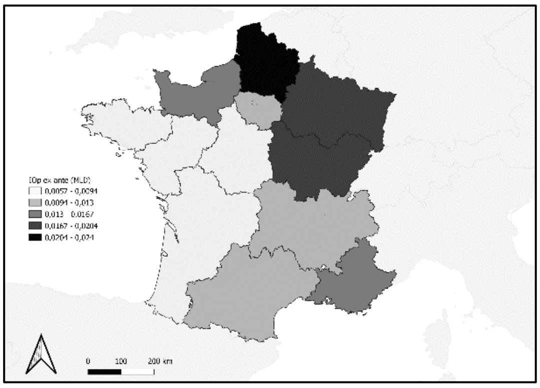
a) ex-ante IOP (Gini)



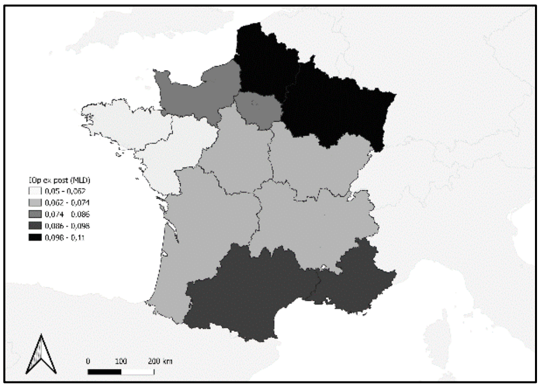
b) ex-post IOP (Gini)



c) ex-ante IOP (MLD)



d) ex-post IOP (MLD)



Note: The names of the regions are only specified in the first figure.
Source: Insee-DGFiP-Cnaf-Cnav-CCMSA, échantillon démographique permanent 2020.

Firstly, the northern regions, including HFr, GE, Normandie (Nor), and Bourgogne-Franche-Comté (BFc), are characterized by high levels of IOp. Notably, HFr and GE stand out for their significantly higher absolute levels of IOp compared to other regions. Furthermore, these two regions also exhibit the highest total inequality (Figure A13 and A14), with IOp accounting for a larger share of the total (Figure A15 and A16). It should be highlighted that BFc presents a relatively high level of ex-ante IOp, but ex-post IOp is marginally below the national average. A second group composed by Occitanie (Occ) and Pro, two regions in the south of the country, demonstrate substantial absolute IOp, particularly in terms of ex-post IOp, and a strong total inequality. The share of IOp within total is lower in these two regions than in the first group but higher than the national average.

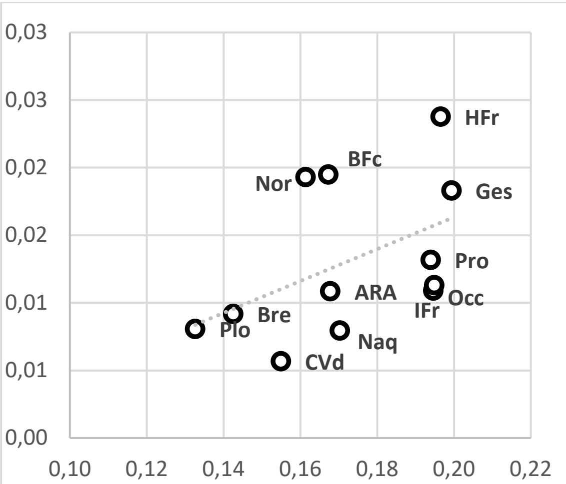
A third group encompasses the western regions, including Bre, PLo, Centre-Val de Loire (CVL), and Nouvelle Aquitaine (NAq). These regions exhibit very low levels of absolute

IOp compared to other regions. Notably, Bre and PLo are the two regions with the lowest IOp, and they also demonstrate the lowest total inequality, with a minimal share of IOp within the total. Interestingly, NAq presents relatively low levels of IOp, but the level of total inequality is close to the national average, resulting in a very low share of IOp in this region. Lastly, the fourth group includes the regions of Ile de France (IFr) and ARA, which exhibit similar indicators across all measures, with absolute and relative IOp levels closely aligned with the national average.

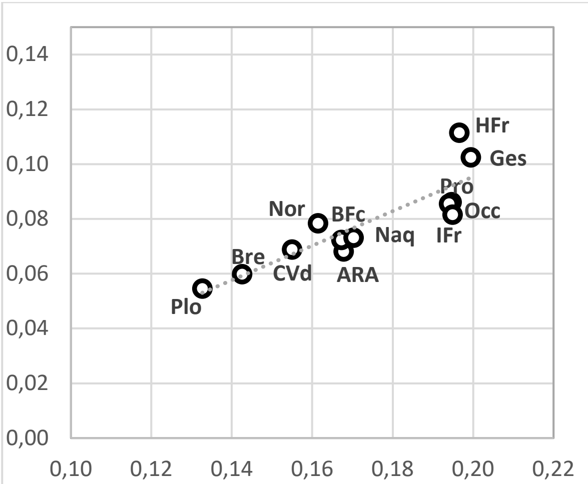
Additionally, when comparing regions, we observe a positive relation between total inequality and IOp measured with MLD (Figure 4). This is consistent with Corak (2013) who found a correlation between higher inequality and lower social mobility (relation also called the Great Gatsby Curve), when comparing OECD countries. A comparable relation can also be observed with a Gini indicator, or between the share of IOp and absolute IOp, or the share of IOp and Total inequality.

Figure 4: Total Inequality and IOp using MLD

a) Total Inequality and ex-ante IOp (MLD)



b) Total Inequality and ex-post IOp (MLD)



Note: the x-axis represents the total inequality, and the y-axis represents absolute IOp measures. Labels used for each region correspond to: Auvergne-Rhone-Alpes (ARA), Bourgogne-Franche-Comte (BFc), Bretagne (Bre), Centre-Val de Loire (CVd), Grand Est (Ges), Hauts-de-France (HFr), Normandie (Nor), Nouvelle-Aquitaine (Naq), Occitanie (Occ), Pays de la Loire (Plo), Provence-Alpes-Cote d'Azur (Pro), and Ile-de-France (IFr). Source: Insee-DGFiP-Cnaf-Cnav-CCMSA, échantillon démographique permanent 2020.

Checchi et al. (2016) underline how estimations and rankings can be sensitive to the approach, the method, or the indicator used. With our data we found a strong correlation between Gini and MLD estimations at the regional level (Figures A17 and A18), and between ex-ante and ex-post measures (Figures A19 and A20). Despite this strong correlation, the ranking of certain regions can vary slightly according to the approach. Indeed, social policies in the domain of education and health will have more impact on ex-ante than on ex-post IOp, while policies addressing labor market institutions, and fiscal actions are more linked to ex-post IOp (Checchi et al. 2016). By examining both ex-ante and ex-post indicators, we observe different profiles of regions and the policy implications of these territorial differences are clear: policymakers must consider regional characteristics when designing policy interventions.

Robustness checks

As a robustness check we run two alternative models. The second model excludes observations with missing values in at least one variable. The tree for this model is almost the same than in the first model, while ex ante IOp is the same and ex-post IOp is to some extent a little lower (Table 3 and Figure A21). The ranking of certain regions according to IOp differs a little in comparison with other models.[13]

Table 3: IOp estimations: robustness checks

	Gini			MLD				
	Total Inequality	ex-post	ex-ante	Total Inequality	ex-post	ex-ante	% ex-post	% ex-ante
Baseline	0.32	0.24	0.09	0.18	0.08	0.01	46%	8%
Model 2	0.30	0.22	0.09	0.17	0.07	0.01	44%	8%
Model 3	0.33	0.18	0.09	0.27	0.08	0.01	28%	5%

Note: The Model 2 excludes observations with missing values, and the Model 3 focuses on relative positions by generating percentiles on the distribution of the income variable. Coverage: Metropolitan France, individuals born between 1989 and 1992, included in their parents' tax return in 2010, 2011 or 2012 and who have positive or zero income in 2019. Source: Insee-DGFiP-Cnaf-Cnav-CCMSA, échantillon démographique permanent 2020.

In the third model, instead of observing some variables in absolute terms,[14] we consider relative positions within the distribution using percentiles. The principal differences between our first and third model are the following. In the former, we obtain 34 types, in the latter we calculate 44 types (Figure A22). Regarding absolute IOp values, results are close in the two models, nonetheless when using the Gini index we estimate an ex-post IOp value of 0.24 in the

[13] Ile-de-France and Provence-Alpes-Cote d'Azur are overrepresented within the sample with missing values, and observations with missing information are characterized by a low level of wealth at age 18, and an important proportion of single-parent households.

[14] The variables rescaled as percentiles are: income of young adults, standard of living of parents, and the capital owned by the parents.

first model, and a value of 0.18 in the third. This result is due to the sensitiveness of the Gini index to extreme values. Indeed, taking absolute incomes rather than percentiles ranks leads to great differences in absolute incomes between individuals with the same degree of effort.

When comparing the results for the MLD, the value for total inequality is 0.27 in the third model compared to 0.18 in the first model, but the IOp absolute values are stable. Consequently, the relative weight of IOp is lower in the third model compared to the first, for example 28% instead of 46% in an ex-post approach and 5% instead of 9% in the ex-ante approach. The share of ex-post IOp using the third model is consistent with the proportion of 28% reported by Checchi et al. (2016) using the same indicator. These variations can also be attributed to the choice of using absolute or relative positions within the distribution. Besides, MLD incorporates the natural logarithm of the variable which compresses values towards zero and is less sensitive to extreme values. The linearization of the variable when using percentiles leads to a modification of the distribution and to higher total inequality between segments of the distribution, but IOp remains stable.

The major contribution of our article is to use administrative data that include very high income, and in the sense we think useful to provide measures in absolute terms with income and capital variables in a continuous form. Still, variables based on absolute terms may not be directly comparable across different populations or time periods, as they are influenced by changes in the overall level of the variable, and include some structural changes. In contrast, percentiles are more comparable as they focus on the relative position. Recognizing the significance of both approaches, we deem it crucial to present both measures.

VII. Conclusion and discussion

This article provides valuable insights into national and sub-national inequalities in France, and highlights the importance for policy adaptations of considering social and territorial diversity to address these inequalities. By using administrative information and data-driven methods, we were able to calculate both ex-ante and ex-post IOp at the national and sub-national level. Our method permitted regional comparisons and identified the principal circumstances beyond an individual’s control at age 18 that shape the life path of young adults observed at age 28 or 29.

Following Roemer’s conceptual framework, we use conditional inference trees to identify types or groups of young adults with the same circumstances at age 18. The visualization of the trees allows to identify in each region, and for different profiles of individuals, which variables are the principal contributors to IOp. The principal circumstance in shaping the future income of individuals is the standard of living of their parents. Gender and the type of housing at age 18 are also strongly statistically significant in determining types. Variables reflecting the territorial environment of individuals at age 18 like the region of residence and the size of the municipality demonstrate an important influence on IOp.

After identifying the types according to circumstances beyond individual responsibility, we calculate the ex-ante IOp as the between-type inequality. We also approximate personal responsibility and calculate ex-post IOp as the within-tranche inequality. When comparing our findings with existing literature by various authors, our estimations indicate that France exhibits a moderate IOp level compared to other European countries. We also find that IOp measured with relative rank is smaller than IOp measured with absolute income, which is an important result for comparing studies that measure IOp differently. Besides, our estimations enable regional analysis, in particular the classification of regions based on inequality indicators. We observe four groups of regions, first, the northern regions have high income IOp and inequality, with Hauts-de-France and Grand Est standing out with significantly higher levels compared to other regions. These regions also exhibit the highest share of IOp within total inequality. Secondly, in the South, Occitanie and Provence-Alpes-Côte d'Azur also present substantial IOp, particularly ex-post IOp. A third group is formed by three western regions and Centre-Val de Loire, characterized by very low absolute IOp and inequality compared with other regions. Finally, Île-de-France and Auvergne-Rhône-Alpes display similar indicators aligned with the national average.

Although there are limitations related to the relatively young age of individuals, we believe that our findings have important implications for policymakers and future research in this area. Our work enhances comprehension of regional inequality dynamics and the underlying factors that contribute to them. For instance, it sheds light on notable disparities in income between men and women, between regions of residence at age 18, and differences between children whose parents have dissimilar economic capital or educational levels, assuming all other circumstances being equal. Finally, this work insists on the significance of considering territorial diversity to tackle inequalities. We demonstrate how the importance of factors beyond individual responsibility vary across regions.

References

Acciari, P., Polo, A.,Violante, G.L.: ‘And Yet It Moves’: Intergenerational Mobility in Italy. American Economic Journal: Applied Economics, 14(3), (2022).

Bernard, J.: Les jeunes ni en emploi, ni en études, ni en formation : jusqu’à 21 ans, moins nombreux parmi les femmes que parmi les hommes, INSEE focus n°229, mars (2021)

Betthäuser, B. A., Kaiser, C., Trinh, N. A.: Regional Variation in Inequality of Educational Opportunity across Europe. Socius, 7 (2021). doi: 10.1177/23780231211019890

Bonnet, F., d’Albis, H., Sotura, A.: Income Inequality across French Departments over the Last 100 Years. Economie et Statistique / Economics and Statistics, 526-527, 49-69 (2021). doi: 10.24187/ecostat.2021.526d.2052

Bossert, W.: Redistribution mechanisms based on individual characteristics. Mathematical Social Sciences, 29, 1-17 (1995). doi: 10.1016/0165-4896(94)00760-6

Bozio, A., Garbinti, B., Goupille-Lebret, J., Guillot, M., Piketty, T.: Predistribution vs. Redistribution: Evidence from France and the U.S. CEPR DP15415 (2020)

Brunori, P., Neidhöfer, G.: The Evolution of Inequality of Opportunity in Germany: A Machine Learning Approach. Review of Income and Wealth, 67(4), 900-927 (2021). doi: 10.1111/roiw.12502

Brunori, P., Hufe, P., Mahler, D.: The roots of inequality: estimating inequality of opportunity from regression trees and forests. Scandinavian Journal of Economics (2023a). https://doi.org/10.1111/sjoe.12530

Brunori P, Ferreira FH, Neidhöfer G. Inequality of opportunity and intergenerational persistence in Latin America. WIDER Working Paper 2023/39. Helsinki: UNU-WIDER. (2023b). https://doi.org/10.35188/UNU-WIDER/2023/347-5

Brzezinski, M.: Income inequality and the Great Recession in Central and Eastern Europe. Economic Systems, 42(2), 219-247 (2018). doi: 10.1016/j.ecosys.2017.07.003

Bussolo, M., Checchi, D., Peragine, V.: Long-term evolution of inequality of opportunity: Educated parents still matter. Journal of Economic Inequality (2023). doi: 10.1007/s10888-022-09562-6

Carpantier, J-F. Sapata, C.: An Ex-Post View of Inequality of Opportunity in France and its Regions. Journal of Labor Research, 34(3), 281-311 (2013)

Checchi, D., Peragine, V., Serlenga, L.: Inequality of Opportunity in Europe: Is There a Role for Institutions? In: Cappellari, L., W. Polachek, S., Tatsiramos, K. (eds.) Inequality: Causes and Consequences, vol. 43, pp. 1-44, Emerald Group Publishing Limited (2016). doi: 10.1108/S0147-912120160000043008

Chetty, R., Hendren, N., Kline, P., Saez, E.: Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. The Quarterly Journal of Economics, 129(4), 1553-1623 (2014). doi: 10.1093/qje/qju022

Chetty, R., Hendren, N., Katz, L.: The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. American Economic Review, 106(4), 855-902 (2016). doi: 10.1257/aer.20150572

Cohen, G. A.: On the currency of egalitarian justice. Ethics, 99(4), p.p. 906-944 (1989). doi: 10.1086/293126

Combes, P-P., Lafourcade, M., Thisse, J., Toutain, J-C.: The rise and fall of spatial inequalities in France: A long-run perspective. Explorations in Economic History, 48(2), 243-271 (2011). doi: 10.1016/j.eeh.2010.12.004

Corak, M, Heisz A.: The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data. Journal of Human Resources, 34, no. 3, 504–533 (1999). https://doi.org/10.2307/146378

Corak, M.: Income inequality, equality of opportunity, and intergenerational mobility. Journal of Economic Perspectives, 27(3), 79-102 (2013). doi: 10.1257/jep.27.3.79.

Dworkin, R.: What is equality. Part 1: Equality of welfare. Philosophy and Public Affairs, 10, 185-246 (1981a)

Dworkin, R.: What is equality. Part 2: Equality of resources. Philosophy and Public Affairs, 10, 283-345 (1981b)

Fleurbaey, M.: On fair compensation. Theory Decision, 36, 277-307 (1994). doi: 10.1007/BF01079932

Fleurbaey, M.: Equality and responsibility. European Economic Review, 39(3-4), 683-689 (1995). doi: 10.1016/0014-2921(94)00075-B

Fleurbaey, M. Peragine, V.: Ex-ante versus ex-post equality of opportunity. Economica, 80, 118-130 (2013). doi: 10.1111/j.1468-0335.2012.00941.x

Fleurbaey, M.: Fairness, Responsibility and Welfare, Oxford: Oxford University Press (2008)

Gaertner, W. Schokkaert, E.: Empirical social choice: Introduction. In: Gaertner, W., Schokkaert, E. (eds.) Empirical social choice: Questionnaire experiments, pp. 1-8. Springer (2012). doi: 10.1007/s11127-014-0164-4

Galvis-Aponte, L. A., Meisel-Roca, A.: Aspectos regionales de la movilidad social y la igualdad de oportunidades en Colombia. Revista de Economía del Rosario, 17(2), 257-297 (2014). doi: 10.12804/rev.econ.rosario.17.02.2014.03

Govind, Y.: Post-colonial Trends of Income Inequality: Evidence from the Overseas Departments of France. PSE Working Papers, halshs-03022303, HAL (2020).

Heidrich, S.: Intergenerational Mobility in Sweden: A Regional Perspective. Journal of Population Economics, 30(4), 1241–1280. (2017). https://doi.org/10.1007/s00148-017-0648-x

Helsø, A.-L.: Intergenerational Income Mobility in Denmark and the United States. Scand. J. of Economics, 123: 508-531. (2021). https://doi.org/10.1111/sjoe.12420

Hothorn, T., Hornik, K., Zeileis, A.: Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics, 15(3), 651-674 (2006). doi: 10.1198/106186006X133933

Hothorn, T.: basefun: Infrastructure for Computing with Basis Functions (2018)

Igersheim, H.: A.K. Sen and J.E. Roemer: une même approche de la responsabilité ?, Post-Print hal-00279378, HAL (2006).

Kenedi, G., Sirugue, L.: The Anatomy of Intergenerational Income Mobility in France and its Spatial Variations. PSE working paper (2021)

Lefranc, A., Trannoy, A.: Intergenerational Earnings Mobility in France: Is France More Mobile than the US? *Annales d'Économie et de Statistique*, 78, 57-77. (2005), https://doi.org/10.2307/20079128

Lefranc, A., Pistolesi, N., Trannoy, A.: Inequality of opportunities vs. inequality of outcomes: are western societies all alike? Review of Income and Wealth, 54(4), 513-543 (2008). doi: 10.1111/j.1475-4991.2008.00289.x

Lefranc, A., Pistolesi, N., Trannoy, A.: 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 (2009). doi: 10.1016/j.jpubeco.2009.07.008

Li Donni, P., Rodríguez, J. G., Rosa Dias, P.: Empirical definition of social types in the analysis of inequality of opportunity: a latent classes approach. Social Choice and Welfare, 44(3), 673-701 (2015). doi: 10.1007/s00355-014-0851-6

OECD.: Income inequality indicator (2023). doi: 10.1787/459aa7f1-en (Accessed on 17 March 2023)

Plassot, T., Soloaga, I., Torres, P.: Inequality of Opportunity in Mexico and its Regions: A Data-Driven Approach. The Journal of Development Studies, 58:9, 1857-1873 (2022). doi: 10.1080/00220388.2022.2055465

Rawls, J.: A Theory of Justice. Harvard University Press (1971)

Robert-Bobée, I. Gualbert, N.: L'échantillon démographique permanent : en 50 ans, l'EDP à bien grandi ! Courrier des Statistiques, 6 (2021)

Roemer, J.E.: A Pragmatic Theory of Responsibility for the Egalitarian Planner, Papers 391, California Davis - Institute of Governmental Affairs (1992)

Roemer, J.: Theories of distributive justice. Harvard University Press (1996)

Roemer, J.: Equality of opportunity. Harvard University Press (1998)

Roemer, J.: Equality of opportunity: A progress report. Social Choice and Welfare, 19, issue 2, p. 455-471 (2002). doi: 10.1007/s003550100123

Roemer, J., Trannoy, A.: Equality of Opportunity. In: Atkinson, A. B., Bourguignon, F. (eds.) Handbook of Income Distribution (Vol. 2, pp. 217-300). Elsevier (2015)

Sen, A.: Equality of What? In: McMurrin (ed.) Tanner Lectures on Human Values, Cambridge: Cambridge University Press, pp. 197–220 (1979).

Sen A.: Commodities and Capabilities. Amsterdam: North-Holland (1985)

Sicsic M.: Who Climbs Up the Income Ladder? An Analysis of Intergenerational Income Mobility in France. Economie et Statistique / Economics and Statistics, 540, 3–20 (2023). doi: 10.24187/ecostat.2023.540.2100

Sirugue, M.: Inequality of opportunity in France: The role of spatial segregation in the ethnic gap (Master's thesis). Paris School of Economics, Paris, France (2020).

Suárez Álvarez, A., López Menéndez, A. J.: Dynamics of inequality and opportunities within European countries. Bulletin of Economic Research, 73, 555-579 (2021). doi: 10.1111/boer.12266

Trannoy, A.: Equality of opportunity: A progress report. Revue d'économie politique, 126(5), 621-651 (2016)

Van de Gaer, D.: Equality of Opportunity and Investment in Human Capital, Ph.D. Dissertation, Katholieke Universiteit Leuven, Belgium (1993)

Appendix: Additional Tables and Figures

Table A1: IOp estimations in the literature

				Gini		MLD			
Authors	**Advantage**	**Country**	**Year**	**ex-post**	**ex-ante**	**ex-post**	**ex-ante**	**% ex-post**	**% ex-ante**
Carpentier and Sapata (2013)	(fair) Income	France	2005	0,22					
Fleurbaey et al. (2015)	Income	Germany	2009			0,09	0,05	33%	18%
Brunori et al. (2023a)	Income	France	2011		0,09				
		Germany			0,07				
		Belgium			0,09				
		Italy			0,11				
		Spain			0,13				
		UK			0,07				
		Finland			0,02				
		Sweden			0,02				
		Hungary			0,113				
Brunori and Neidhofer (2021)	Deviation of income from its expected value	Germany	2016	0,1					
Bussolo et al. (2020)	Income	France	2005				0,07		21%
		Germany	2013				0,08		21%
		Italy	2014				0,14		38%
		UK	2014				0,04		13%
Checchi et al. (2016)	Income	France	2005			0,05	0,02	28%	13%
		Belgium				0,06	0,02	40%	17%
		Italy				0,06	0,03	32%	14%
		Spain				0,07	0,04	34%	20%
		United Kingdom				0,08	0,04	41%	20%
		Germany				0,06	0,03	31%	18%
		Finland				0,03	0,01	21%	10%
		Sweden				0,03	0,01	24%	11%
		Hungary				0,06	0,02	30%	10%

		Czech Republic				0,07	0,02	30%	10%
		Austria				0,07	0,04	39%	21%
		Estonia				0,08	0,03	32%	11%
Suárez Álvarez and López Menéndez (2021)	Income	France	2011		0,08		0,01		10%
		Italy			0,11		0,02		12%
		UK			0,09		0,01		7%
		Belgium			0,08		0,01		11%
		Finland			0,05		0,004		4%
		Hungary			0,12		0,02		16%
Brunori et al. (2023b)	Income (2014)	Argentina	2014	0,12	0,17	0,05	0,05	17%	17%
		Brazil	2014	0,22	0,30	0,14	0,15	30%	31%
		Chile	2015	0,16	0,26	0,11	0,10	23%	21%
		Ecuador	2014	0,15	0,21	0,08	0,07	19%	17%
		Peru	2014	0,17	0,23	0,10	0,08	29%	23%
Vélez Grajales et. (2018)	Asset Index	Mexico	2011						39%
	Income						0.09		40%
Plassot et al. (2022)	Asset Index	Mexico	2017	0,26	0,17	0,15	0,05	56%	20%
Pistolesi (2007)	Earnings	USA	1968–2001					29%	23%
Hufe et al. (2022)	Income	Argentina	2015		0,13		0,03		34%
		China	2014		0,22		0,08		46%
		Ethiopia	2009		0,35		0,21		56%
		Indonesia	2013		0,25		0,1		52%
		Mexico	2009		0,11		0,02		35%
		Russia	2017		0,18		0,05		37%
		South Africa	2017		0,29		0,13		51%

Note: The rows correspond to author's estimations found in the literature. UK stands for United Kingdom. The data used in each survey is specified in Table A2.

Table A2: Sources of information for the different IOp estimations presented in Table 1

Authors	**Source of Information**
Carpentier and Sapata (2013)	EU statistics on income and living conditions (EU-SILC)
Brunori et al. (2023a)	EU statistics on income and living conditions (EU-SILC) Household Budget Survey (HBS), German Socio-economic Panel (SOEP)
Bussolo et al. (2020)	Survey on Household Incomes and Wealth (SHIW) Understanding Society-Household Longitudinal Survey (UKHLS)
Cecchi et al. (2016)	EU statistics on income and living conditions (EU-SILC)
Suárez Álvarez and López Menéndez (2021)	EU statistics on income and living conditions (EU-SILC)
Brunori and Neidhofer (2021)	German Socio-Economic Panel (SOEP)
Fleurbaey et al. (2015)	German Socio-Economic Panel7 (SOEP) Encuesta Nacional sobre la Estructura Social Pesquisa Nacional por Amostra de Domicílios
Brunori et al. (2023b)	Encuesta de Caracterización Socioeconómica Nacional Encuesta de Condiciones de Vida Encuesta Nacional de Hogares
Vélez Grajales et. (2018)	ESRU Survey on Social Mobility in Mexico
Plassot et al. (2022)	ESRU Survey on Social Mobility in Mexico
Pistolesi (2007)	Michigan Panel Study of Income Dynamics Encuesta Permanente de Hogares China Health and Nutrition Survey Ethiopia Rural Household Survey
Hufe et al. (2022)	Indonesian Family Life Survey (IFLS) Encuesta Evaluation de los Hogares (ENCEL) Russia Longitudinal Monitoring Survey (RLMS) National Income Dynamics Study (NIDS)

Note: The source of information in each row corresponds to the respective row of the Table A1 for each author.

Table A3: Descriptive Statistics

	France	France (without missings)	Auvergne-Rhone-Alpes	Bourgogne-Franche-Comte	Bretagne
	(N=36402)	(N=19708)	(N=4390)	(N=1621)	(N=2054)
Outcome					
Mean (SD)	20000 (12500)	20100 (12100)	20600 (12400)	20900 (13600)	20200 (11600)
Median [Min, Max]	19300 [0, 288000]	19400 [0, 258000]	19800 [0, 204000]	19700 [0, 258000]	19500 [0, 171000]
Parents Standard Living					
Mean (SD)	23400 (19000)	23500 (16900)	24000 (17100)	21400 (12900)	22400 (14000)
Median [Min, Max]	19800 [-58100, 978000]	20300 [-58100, 717000]	20500 [-39400, 480000]	19300 [-34500, 153000]	19400 [-10900, 184000]
Missing	38.0 (0.1%)		0 (0%)	0 (0%)	0 (0%)
Capital income					
Mean (SD)	6150 (39200)	5940 (27800)	6890 (35200)	4240 (16900)	6300 (24600)
Median [Min, Max]	1180 [-206000, 4840000]	1280 [-206000, 1760000]	1490 [-64100, 1770000]	972 [-154000, 308000]	1520 [-56500, 547000]
Missing	38.0 (0.1%)		0 (0%)	0 (0%)	0 (0%)
Outcome percentile					
Mean (SD)	51.1 (28.9)	51.8 (28.3)	52.5 (28.7)	52.8 (27.9)	51.7 (26.9)
Median [Min, Max]	51.0 [1.00, 100]	52.0 [1.00, 100]	53.0 [1.00, 100]	54.0 [1.00, 100]	52.0 [1.00, 100]
Parents Standard Living percentile					
Mean (SD)	51.8 (28.8)	53.3 (27.8)	53.7 (28.4)	48.7 (27.5)	50.8 (26.8)
Median [Min, Max]	52.0 [1.00, 100]	55.0 [1.00, 100]	55.0 [1.00, 100]	49.0 [1.00, 100]	51.0 [1.00, 100]
Capital income percentile					
Mean (SD)	51.7 (28.7)	54.1 (27.4)	56.7 (27.7)	48.2 (27.7)	57.6 (26.3)
Median [Min, Max]	52.0 [1.00, 100]	55.0 [1.00, 100]	59.0 [1.00, 100]	46.0 [1.00, 100]	60.0 [1.00, 100]
Gender					
Men	18372 (50.5%)	10074 (51.1%)	2245 (51.1%)	840 (51.8%)	1061 (51.7%)
Women	18030 (49.5%)	9634 (48.9%)	2145 (48.9%)	781 (48.2%)	993 (48.3%)
Housing type					
SH	25755 (70.8%)	15377 (78.0%)	3207 (73.1%)	1214 (74.9%)	1663 (81.0%)
Ot	4687 (12.9%)	1799 (9.1%)	508 (11.6%)	154 (9.5%)	146 (7.1%)
Ll	5922 (16.3%)	2532 (12.8%)	675 (15.4%)	253 (15.6%)	245 (11.9%)
Missing	38.0 (0.1%)		0 (0%)	0 (0%)	0 (0%)
Household type					
C12	18090 (49.7%)	11768 (59.7%)	2260 (51.5%)	859 (53.0%)	1004 (48.9%)
C3+	9491 (26.1%)	5981 (30.3%)	1135 (25.9%)	410 (25.3%)	619 (30.1%)
SP	6936 (19.1%)	1186 (6.0%)	839 (19.1%)	293 (18.1%)	368 (17.9%)
X	1885 (5.2%)	773 (3.9%)	156 (3.6%)	59.0 (3.6%)	63.0 (3.1%)
Parental Education					
Wo	3070 (8.4%)	3063 (15.5%)	339 (7.7%)	171 (10.5%)	121 (5.9%)
Vo	8404 (23.1%)	8398 (42.6%)	1013 (23.1%)	449 (27.7%)	596 (29.0%)
Bac	2828 (7.8%)	2824 (14.3%)	392 (8.9%)	120 (7.4%)	195 (9.5%)
Bac+	5431 (14.9%)	5423 (27.5%)	717 (16.3%)	196 (12.1%)	293 (14.3%)

Missing	16669 (45.8%)		1929 (43.9%)	685 (42.3%)	849 (41.3%)
Occupational					
F	519 (1.4%)	518 (2.6%)	59.0 (1.3%)	44.0 (2.7%)	68.0 (3.3%)
S	1476 (4.1%)	1474 (7.5%)	226 (5.1%)	66.0 (4.1%)	106 (5.2%)
Ex	3849 (10.6%)	3846 (19.5%)	478 (10.9%)	133 (8.2%)	210 (10.2%)
I	4891 (13.4%)	4881 (24.8%)	663 (15.1%)	234 (14.4%)	300 (14.6%)
Em	2805 (7.7%)	2804 (14.2%)	313 (7.1%)	132 (8.1%)	162 (7.9%)
M	5772 (15.9%)	5765 (29.3%)	667 (15.2%)	313 (19.3%)	336 (16.4%)
O	421 (1.2%)	420 (2.1%)	55.0 (1.3%)	14.0 (0.9%)	23.0 (1.1%)
Missing	16669 (45.8%)		1929 (43.9%)	685 (42.3%)	849 (41.3%)
Immigrant status					
Non Immigrant	17841 (49.0%)	17765 (90.1%)	2202 (50.2%)	870 (53.7%)	1183 (57.6%)
Immigrant	1954 (5.4%)	1943 (9.9%)	267 (6.1%)	69.0 (4.3%)	24.0 (1.2%)
Missing	16607 (45.6%)		1921 (43.8%)	682 (42.1%)	847 (41.2%)
Municipality_Size					
Cs	2220 (6.1%)	1443 (7.3%)	341 (7.8%)	154 (9.5%)	274 (13.3%)
Vs	4306 (11.8%)	2548 (12.9%)	423 (9.6%)	344 (21.2%)	291 (14.2%)
S	6801 (18.7%)	3945 (20.0%)	731 (16.7%)	714 (44.0%)	362 (17.6%)
M	8161 (22.4%)	4437 (22.5%)	1052 (24.0%)	402 (24.8%)	658 (32.0%)
B	7449 (20.5%)	3945 (20.0%)	1843 (42.0%)	0 (0%)	469 (22.8%)
Pa	7465 (20.5%)	3390 (17.2%)	0 (0%)	7.00 (0.4%)	0 (0%)
Regions					
Auvergne-Rhone-Alpes	4390 (12.1%)	2461 (12.5%)			
Bourgogne-Franche-Comte	1621 (4.5%)	936 (4.7%)			
Bretagne	2054 (5.6%)	1204 (6.1%)			
Centre-Val de Loire	1453 (4.0%)	869 (4.4%)			
Grand Est	3323 (9.1%)	1982 (10.1%)			
Hauts-de-France	3931 (10.8%)	2268 (11.5%)			
Ile-de-France	6918 (19.0%)	3043 (15.4%)			
Normandie	2012 (5.5%)	1190 (6.0%)			
Nouvelle-Aquitaine	2927 (8.0%)	1649 (8.4%)			
Occitanie	2934 (8.1%)	1608 (8.2%)			
Pays de la Loire	2290 (6.3%)	1359 (6.9%)			
Provence-Alpes-Cote d'Azur	2497 (6.9%)	1139 (5.8%)			
Missing	52.0 (0.1%)				

Table A3: Descriptive Statistics (continued)

	Centre-Val de Loire (N=1453)	**Grand Est** (N=3323)	**Hauts-de-France** (N=3931)	**Ile-de-France** (N=6918)	**Normandie** (N=2012)
Outcome					
Mean (SD)	19800 (10500)	18900 (12700)	18400 (12200)	22300 (14600)	19600 (11300)

Median [Min, Max]	19600 [0, 99900]	18500 [0, 132000]	18200 [0, 215000]	21500 [0, 288000]	19100 [0, 92800]
Parents Standard Living					
Mean (SD)	22400 (13500)	22200 (13700)	20100 (12800)	28800 (27600)	21000 (14100)
Median [Min, Max]	19600 [-23800, 221000]	19600 [-7970, 242000]	17500 [-10400, 266000]	23400 [-28600, 717000]	18600 [-11100, 363000]
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Capital income					
Mean (SD)	5160 (23300)	5660 (24000)	3930 (18900)	8630 (39000)	3790 (15600)
Median [Min, Max]	1180 [-35500, 691000]	1290 [-56000, 847000]	764 [-69100, 699000]	1410 [-127000, 1350000]	939 [-42800, 401000]
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Outcome percentile					
Mean (SD)	51.9 (27.3)	48.2 (29.7)	47.1 (28.3)	56.4 (30.7)	50.4 (28.0)
Median [Min, Max]	53.0 [1.00, 100]	47.0 [1.00, 100]	45.0 [1.00, 100]	61.0 [1.00, 100]	50.0 [1.00, 100]
Parents Standard Living percentile					
Mean (SD)	51.3 (27.2)	50.5 (27.9)	44.0 (28.1)	60.0 (30.0)	47.2 (27.4)
Median [Min, Max]	51.0 [1.00, 100]	51.0 [1.00, 100]	41.0 [1.00, 100]	66.0 [1.00, 100]	47.0 [1.00, 100]
Capital income percentile					
Mean (SD)	52.0 (27.0)	53.3 (27.6)	42.5 (28.3)	55.0 (29.3)	46.7 (27.7)
Median [Min, Max]	52.0 [1.00, 100]	55.0 [1.00, 100]	38.0 [1.00, 100]	58.0 [1.00, 100]	45.0 [1.00, 100]
Gender					
Men	759 (52.2%)	1688 (50.8%)	2009 (51.1%)	3343 (48.3%)	1009 (50.1%)
Women	694 (47.8%)	1635 (49.2%)	1922 (48.9%)	3575 (51.7%)	1003 (49.9%)
Housing type					
SH	1089 (74.9%)	2410 (72.5%)	2721 (69.2%)	4291 (62.0%)	1409 (70.0%)
Ot	164 (11.3%)	404 (12.2%)	601 (15.3%)	1424 (20.6%)	326 (16.2%)
Ll	200 (13.8%)	509 (15.3%)	609 (15.5%)	1203 (17.4%)	277 (13.8%)
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Household type					
C12	771 (53.1%)	1737 (52.3%)	1807 (46.0%)	2998 (43.3%)	1067 (53.0%)
C3+	368 (25.3%)	870 (26.2%)	1216 (30.9%)	1876 (27.1%)	521 (25.9%)
SP	248 (17.1%)	582 (17.5%)	692 (17.6%)	1517 (21.9%)	347 (17.2%)
X	66.0 (4.5%)	134 (4.0%)	216 (5.5%)	527 (7.6%)	77.0 (3.8%)
Parental Education					
Wo	131 (9.0%)	347 (10.4%)	469 (11.9%)	488 (7.1%)	224 (11.1%)
Vo	429 (29.5%)	923 (27.8%)	1035 (26.3%)	932 (13.5%)	546 (27.1%)
Bac	119 (8.2%)	246 (7.4%)	277 (7.0%)	415 (6.0%)	169 (8.4%)
Bac+	190 (13.1%)	466 (14.0%)	487 (12.4%)	1208 (17.5%)	251 (12.5%)
Missing	584 (40.2%)	1341 (40.4%)	1663 (42.3%)	3875 (56.0%)	822 (40.9%)
Occupational					
F	31.0 (2.1%)	49.0 (1.5%)	42.0 (1.1%)	10.0 (0.1%)	37.0 (1.8%)
S	75.0 (5.2%)	117 (3.5%)	117 (3.0%)	222 (3.2%)	85.0 (4.2%)
Ex	131 (9.0%)	318 (9.6%)	294 (7.5%)	1011 (14.6%)	161 (8.0%)
I	213 (14.7%)	480 (14.4%)	516 (13.1%)	714 (10.3%)	323 (16.1%)

Em	112 (7.7%)	276 (8.3%)	335 (8.5%)	405 (5.9%)	158 (7.9%)
M	285 (19.6%)	701 (21.1%)	919 (23.4%)	604 (8.7%)	411 (20.4%)
O	22.0 (1.5%)	41.0 (1.2%)	45.0 (1.1%)	77.0 (1.1%)	15.0 (0.7%)
Missing	584 (40.2%)	1341 (40.4%)	1663 (42.3%)	3875 (56.0%)	822 (40.9%)
Immigrant status					
Non Immigrant	781 (53.8%)	1795 (54.0%)	2166 (55.1%)	2330 (33.7%)	1126 (56.0%)
Immigrant	89.0 (6.1%)	194 (5.8%)	109 (2.8%)	727 (10.5%)	64.0 (3.2%)
Missing	583 (40.1%)	1334 (40.1%)	1656 (42.1%)	3861 (55.8%)	822 (40.9%)
Municipality_Size					
Cs	150 (10.3%)	281 (8.5%)	144 (3.7%)	0 (0%)	173 (8.6%)
Vs	266 (18.3%)	542 (16.3%)	405 (10.3%)	0 (0%)	432 (21.5%)
S	391 (26.9%)	712 (21.4%)	1199 (30.5%)	0 (0%)	445 (22.1%)
M	538 (37.0%)	1129 (34.0%)	804 (20.5%)	0 (0%)	879 (43.7%)
B	0 (0%)	659 (19.8%)	1030 (26.2%)	0 (0%)	0 (0%)
Pa	108 (7.4%)	0 (0%)	349 (8.9%)	6918 (100%)	83.0 (4.1%)

Table A3: Descriptive Statistics (continued)

	Nouvelle-Aquitaine (N=2927)	**Occitanie** (N=2934)	**Pays de la Loire** (N=2290)	**Provence-Alpes-Cote d'Azur** (N=2497)
Outcome				
Mean (SD)	19200 (10600)	18800 (12300)	19900 (10300)	19100 (12300)
Median [Min, Max]	18800 [0, 87700]	18200 [0, 222000]	19300 [0, 95900]	18600 [0, 191000]
Parents Standard Living				
Mean (SD)	22000 (13900)	22500 (19200)	22400 (25400)	22600 (14900)
Median [Min, Max]	19500 [-31300, 181000]	19200 [-58100, 642000]	19400 [-24100, 978000]	19800 [-25700, 273000]
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Capital income				
Mean (SD)	5530 (22900)	6240 (41200)	7230 (107000)	5350 (18400)
Median [Min, Max]	1230 [-182000, 521000]	1210 [-206000, 1760000]	1130 [-134000, 4840000]	896 [-102000, 353000]
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Outcome percentile				
Mean (SD)	49.2 (27.1)	47.9 (28.6)	51.4 (26.6)	48.7 (29.3)
Median [Min, Max]	48.0 [1.00, 100]	45.0 [1.00, 100]	52.0 [1.00, 100]	47.0 [1.00, 100]
Parents Standard Living percentile				
Mean (SD)	50.0 (28.0)	49.7 (28.9)	50.4 (26.5)	50.9 (29.0)
Median [Min, Max]	50.0 [1.00, 100]	49.0 [1.00, 100]	50.0 [1.00, 100]	51.0 [1.00, 100]
Capital income percentile				
Mean (SD)	52.7 (27.7)	51.5 (29.8)	51.2 (27.3)	46.6 (30.4)
Median [Min, Max]	54.0 [1.00, 100]	53.0 [1.00, 100]	51.0 [1.00, 100]	43.0 [1.00, 100]
Gender				

Men	1492 (51.0%)	1470 (50.1%)	1185 (51.7%)	1240 (49.7%)
Women	1435 (49.0%)	1464 (49.9%)	1105 (48.3%)	1257 (50.3%)
Housing type				
SH	2160 (73.8%)	2143 (73.0%)	1818 (79.4%)	1620 (64.9%)
Ot	244 (8.3%)	222 (7.6%)	204 (8.9%)	288 (11.5%)
Ll	523 (17.9%)	569 (19.4%)	268 (11.7%)	589 (23.6%)
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Household type				
C12	1601 (54.7%)	1542 (52.6%)	1179 (51.5%)	1257 (50.3%)
C3+	611 (20.9%)	613 (20.9%)	712 (31.1%)	537 (21.5%)
SP	562 (19.2%)	610 (20.8%)	335 (14.6%)	540 (21.6%)
X	153 (5.2%)	169 (5.8%)	64.0 (2.8%)	163 (6.5%)
Parental Education				
Wo	210 (7.2%)	186 (6.3%)	194 (8.5%)	184 (7.4%)
Vo	724 (24.7%)	649 (22.1%)	673 (29.4%)	430 (17.2%)
Bac	272 (9.3%)	268 (9.1%)	175 (7.6%)	177 (7.1%)
Bac+	443 (15.1%)	505 (17.2%)	319 (13.9%)	348 (13.9%)
Missing	1278 (43.7%)	1326 (45.2%)	929 (40.6%)	1358 (54.4%)
Occupational				
F	54.0 (1.8%)	53.0 (1.8%)	60.0 (2.6%)	12.0 (0.5%)
S	140 (4.8%)	138 (4.7%)	75.0 (3.3%)	107 (4.3%)
Ex	294 (10.0%)	345 (11.8%)	224 (9.8%)	247 (9.9%)
I	400 (13.7%)	409 (13.9%)	344 (15.0%)	285 (11.4%)
Em	266 (9.1%)	265 (9.0%)	188 (8.2%)	193 (7.7%)
M	461 (15.8%)	359 (12.2%)	444 (19.4%)	265 (10.6%)
O	34.0 (1.2%)	39.0 (1.3%)	26.0 (1.1%)	30.0 (1.2%)
Missing	1278 (43.7%)	1326 (45.2%)	929 (40.6%)	1358 (54.4%)
Immigrant status				
Non Immigrant	1549 (52.9%)	1467 (50.0%)	1328 (58.0%)	1024 (41.0%)
Immigrant	104 (3.6%)	148 (5.0%)	32.0 (1.4%)	124 (5.0%)
Missing	1274 (43.5%)	1319 (45.0%)	930 (40.6%)	1349 (54.0%)
Municipality_Size				
Cs	324 (11.1%)	170 (5.8%)	120 (5.2%)	89.0 (3.6%)
Vs	459 (15.7%)	453 (15.4%)	447 (19.5%)	244 (9.8%)
S	732 (25.0%)	739 (25.2%)	446 (19.5%)	330 (13.2%)
M	682 (23.3%)	427 (14.6%)	622 (27.2%)	916 (36.7%)
B	730 (24.9%)	1145 (39.0%)	655 (28.6%)	918 (36.8%)
Pa	0 (0%)	0 (0%)	0 (0%)	0 (0%)

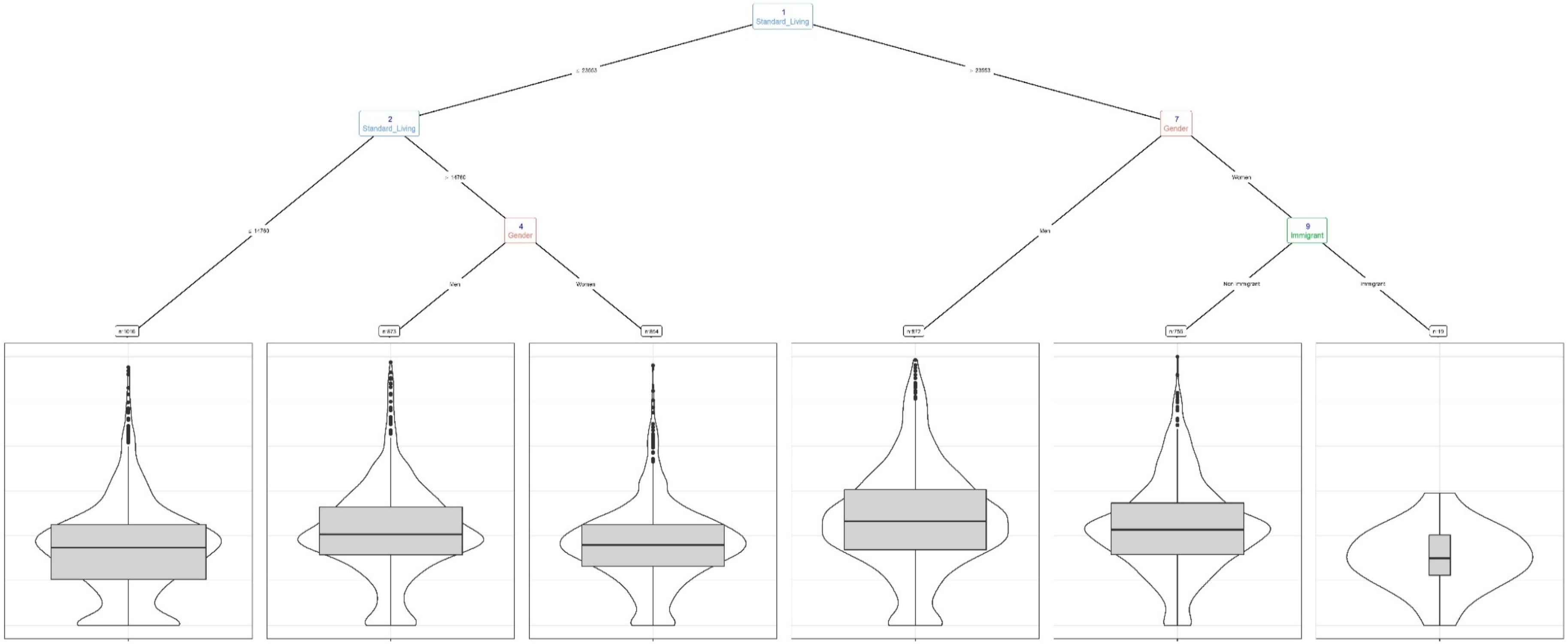
Note: The third column corresponds to the sample without missing values as used in the Model 2. Coverage: Metropolitan France, individuals born between 1989 and 1992, included in their parents' tax return in 2010, 2011 or 2012 and who have positive or zero income in 2019. Source: Insee-DGFiP-Cnaf-Cnav-CCMSA, échantillon démographique permanent 2020.

Table A4: Number of types and IOp estimations at the national and regional level, Model 1

	Number of types	Gini			Mean Log Deviation		
		ex-post	ex-ante	Total	ex-post	ex-ante	Total
National	**34**	**0,241**	**0,094**	**0,319**	**0,08**	**0,01**	**0,18**
Auvergne-Rhône-Alpes	6	0,217	0,082	0,310	0,07	0,01	0,17
Bourgogne-Franche-Comte	8	0,217	0,098	0,303	0,07	0,02	0,17
Bretagne	6	0,189	0,067	0,279	0,06	0,01	0,14
Centre-Val de Loire	3	0,203	0,057	0,280	0,07	0,01	0,15
Grand Est	9	0,294	0,099	0,351	0,10	0,02	0,20
Hauts-de-France	12	0,284	0,119	0,330	0,11	0,02	0,20
Normandie	11	0,237	0,107	0,305	0,08	0,02	0,16
Nouvelle-Aquitaine	7	0,201	0,063	0,291	0,07	0,01	0,17
Occitanie	5	0,245	0,080	0,328	0,09	0,01	0,19
Pays de la Loire	7	0,187	0,067	0,272	0,05	0,01	0,13
Provence-Alpes-Côte d'Azur	7	0,257	0,083	0,336	0,09	0,01	0,19
Ile-de-France	7	0,245	0,083	0,334	0,08	0,01	0,19

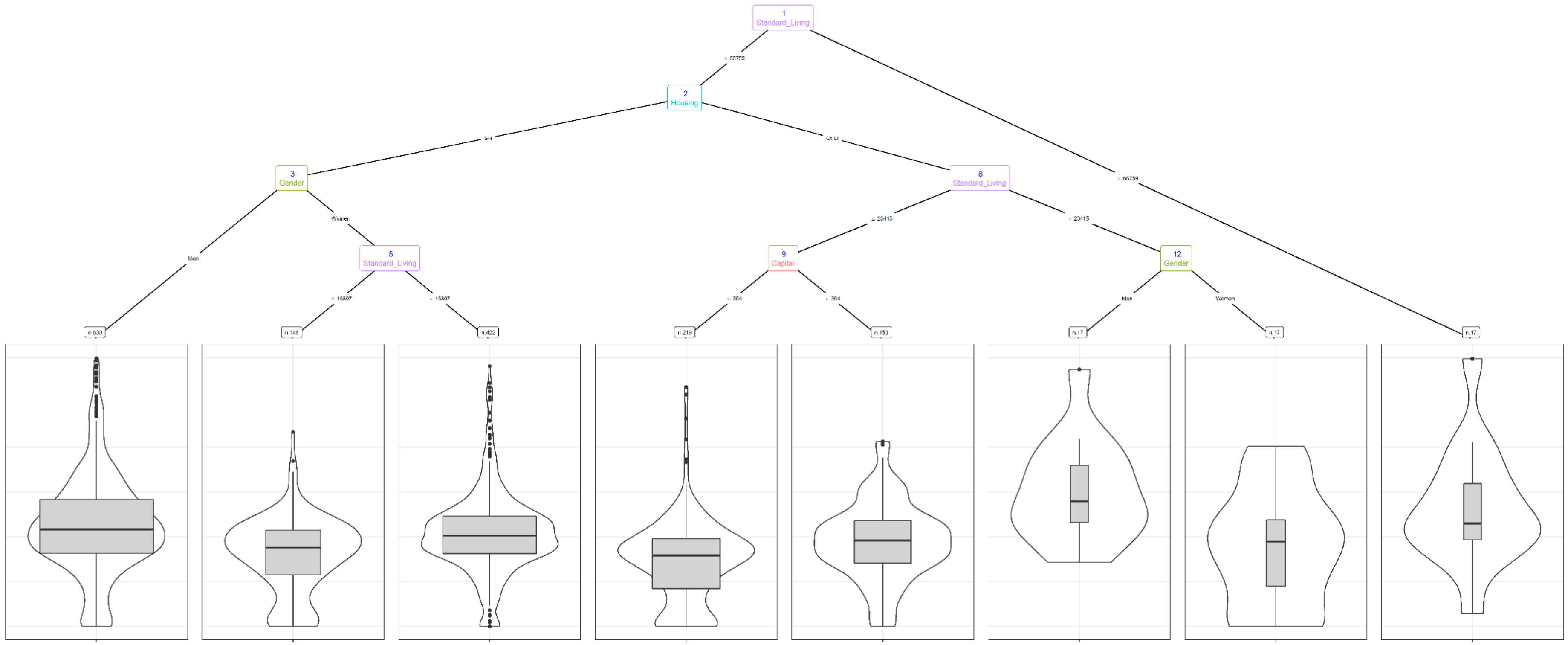
Note: Total corresponds to total income inequality. Coverage: Metropolitan France, individuals born between 1989 and 1992, included in their parents' tax return in 2010, 2011 or 2012 and who have positive or zero income in 2019. Source: Insee-DGFiP-Cnaf-Cnav-CCMSA, échantillon démographique permanent 2020.

Figure A1: Conditional inference tree, region Alpes-Côte d'Azur



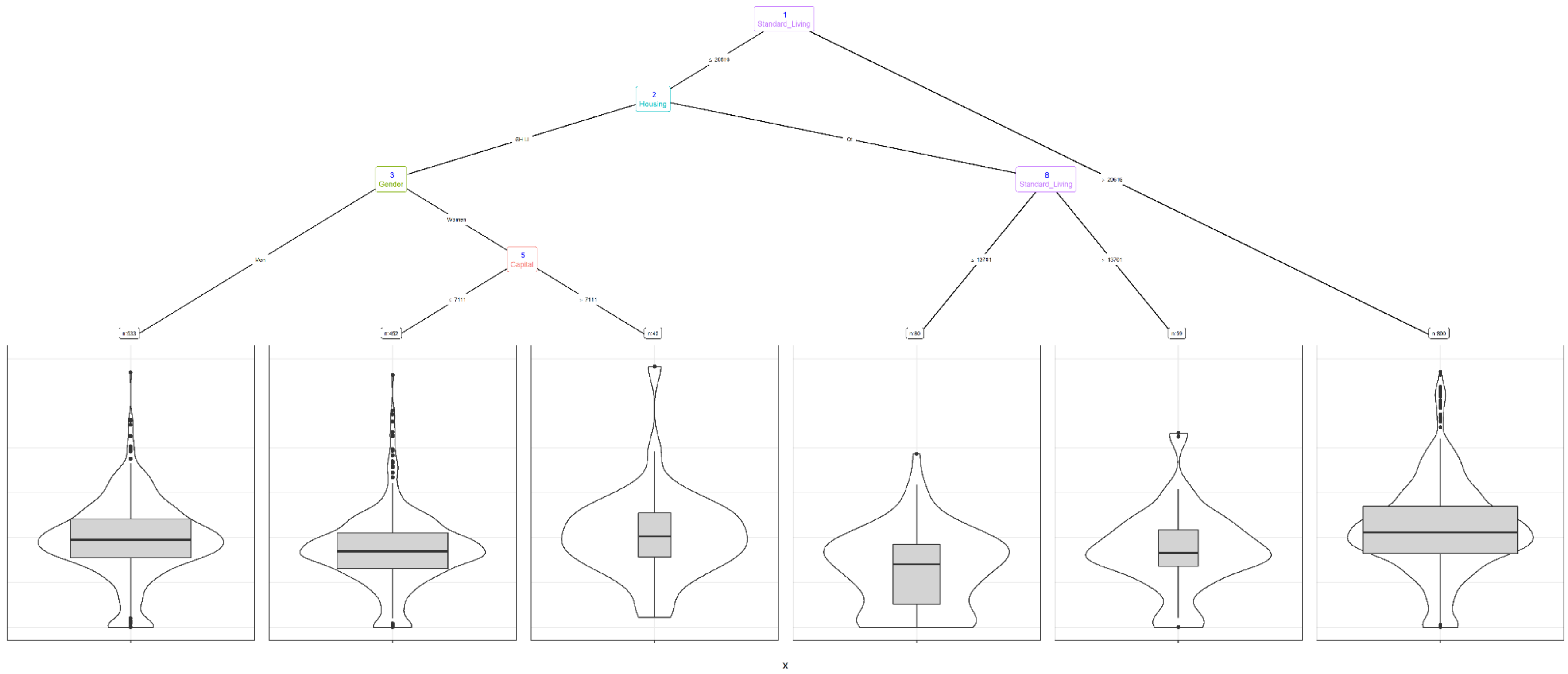
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. Source and coverage: see Table A4.

Figure A2: Conditional inference tree, region Bourgogne-Franche Comté



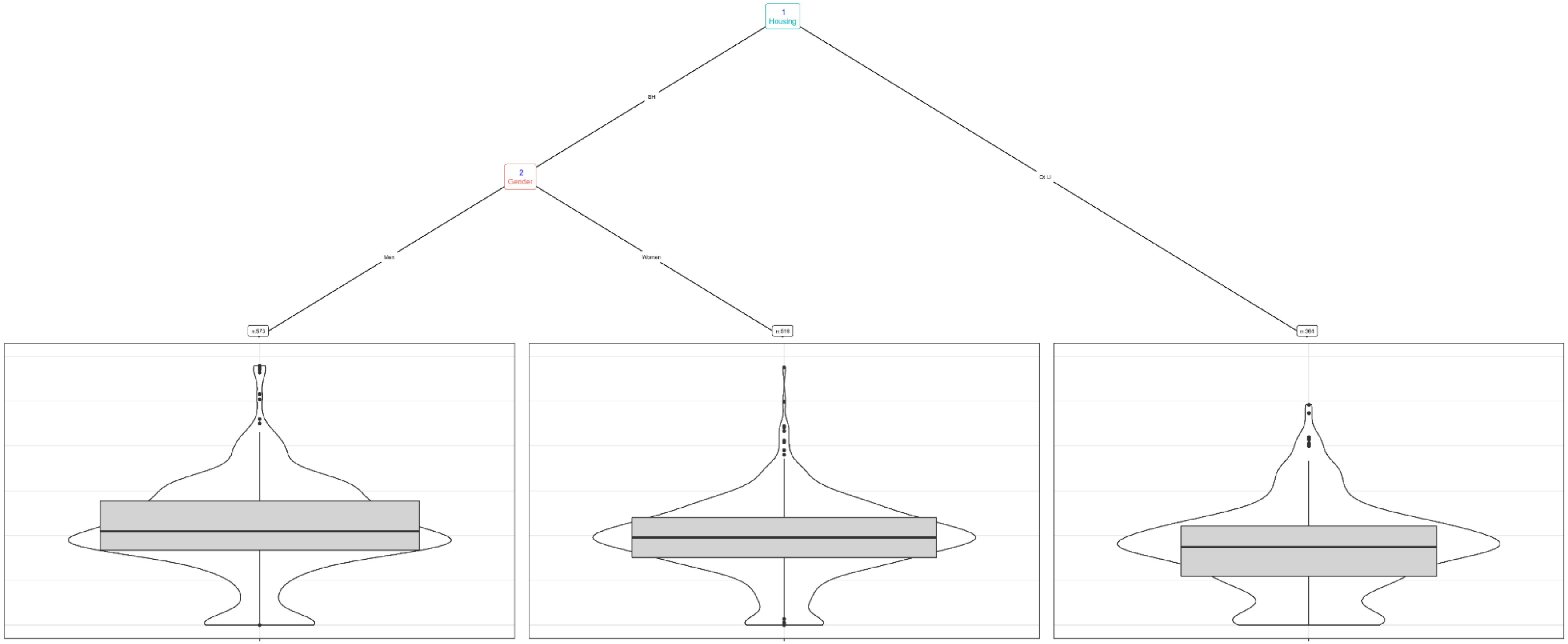
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. Source and coverage: see Table A4.

Figure A3: Conditional inference tree, region Bretagne



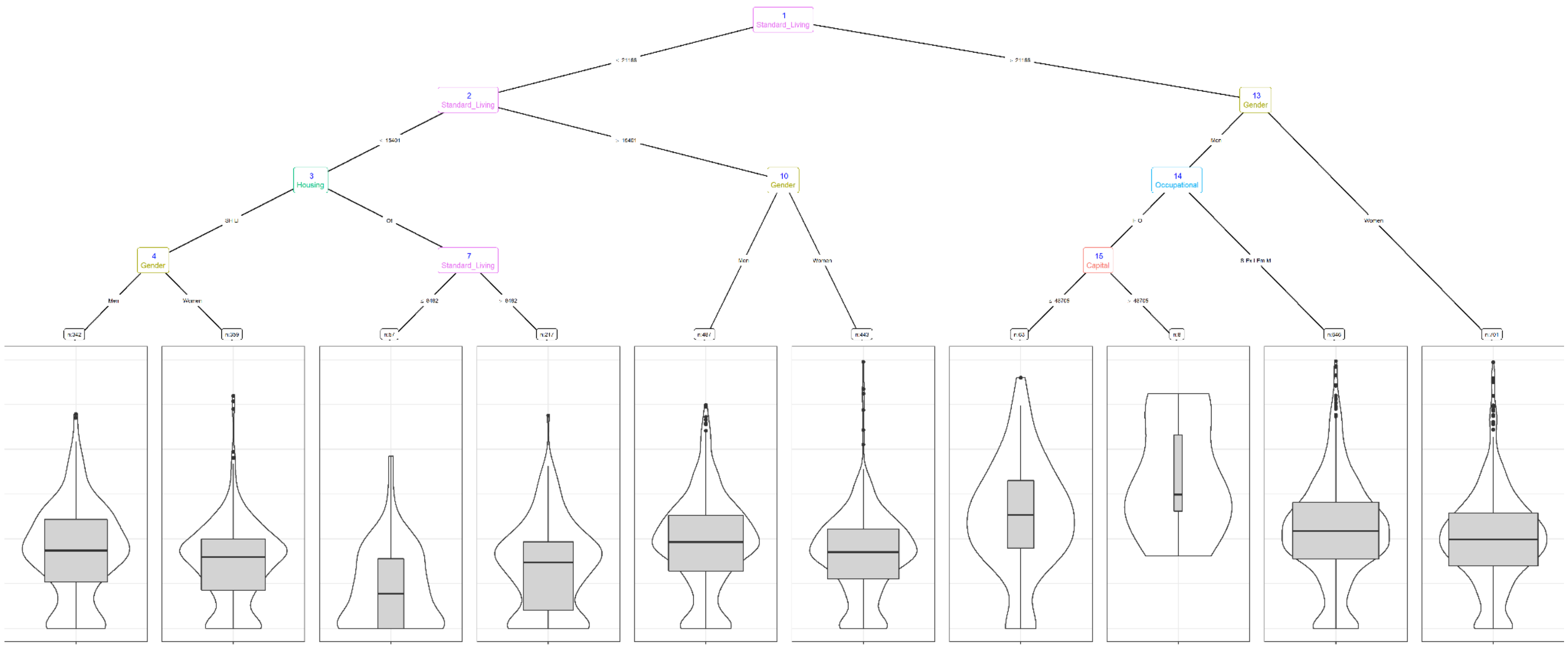
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. Source and coverage: see Table A4.

Figure A4: Conditional inference tree, region Centre - Val de Loire



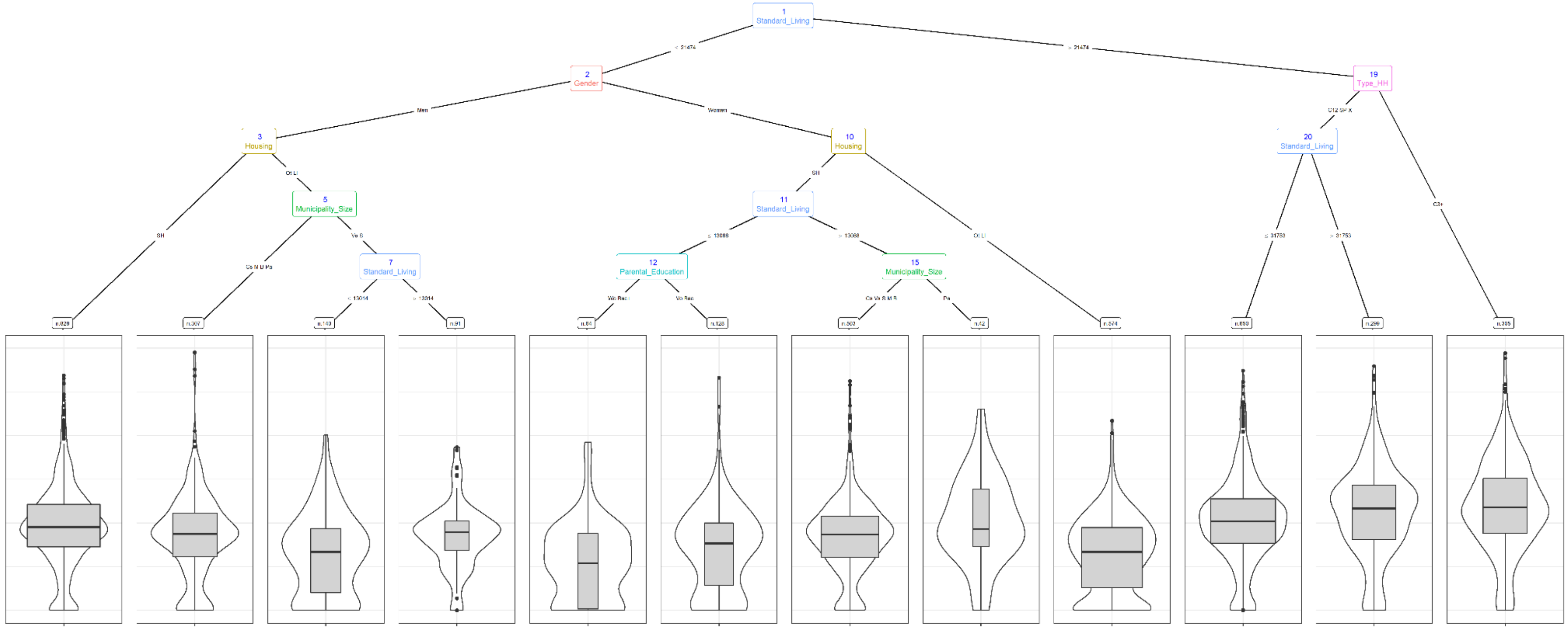
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. Source and coverage: see Table A4.

Figure A5: Conditional inference tree, region Grand Est



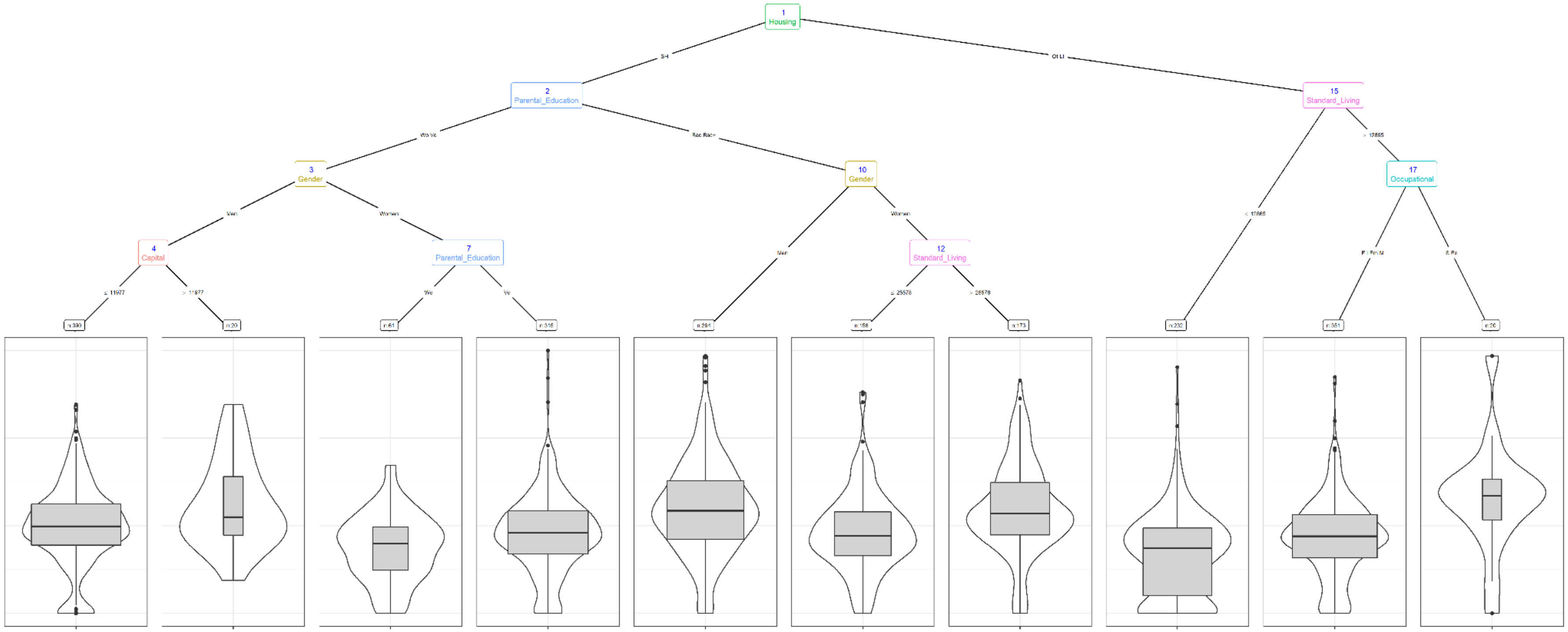
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. Source and coverage: see Table A4.

Figure A6: Conditional inference tree, region Hauts-de-France



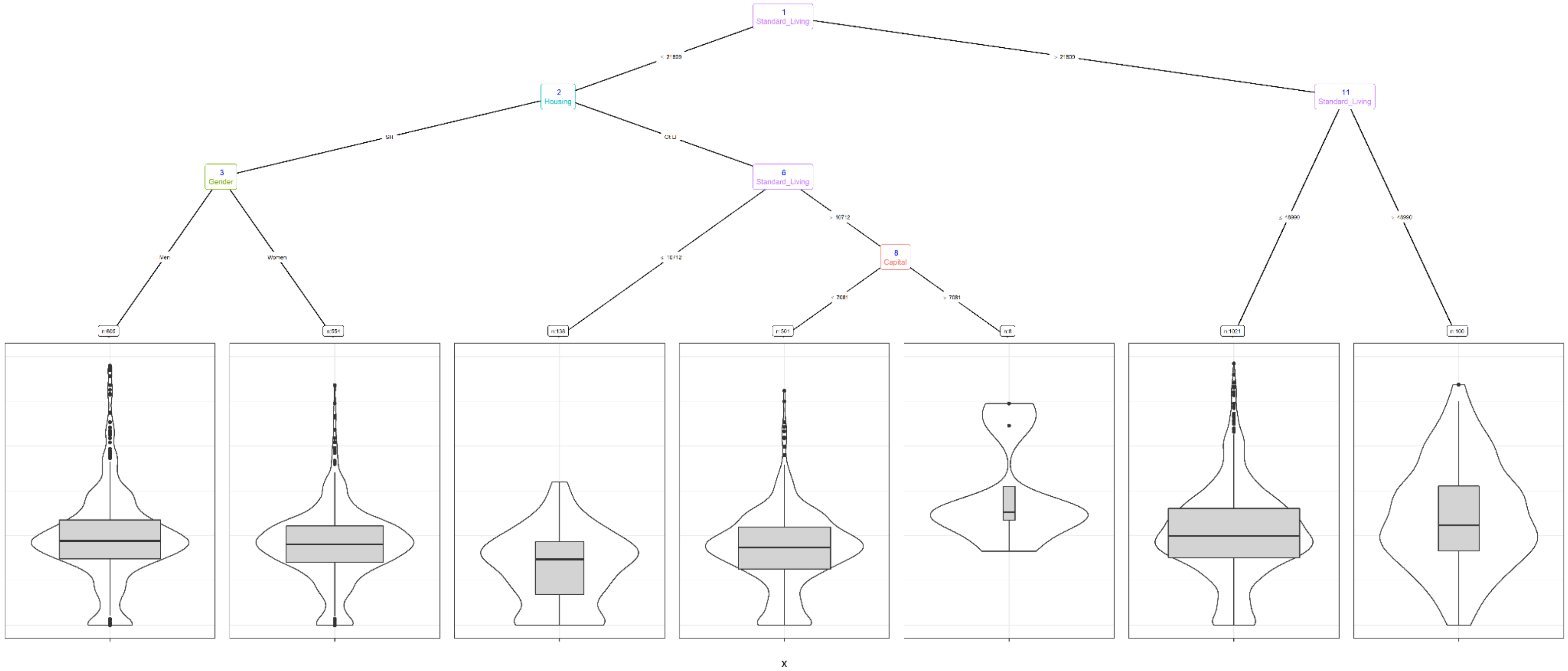
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros.

Figure A7: Conditional inference tree, region Normandie



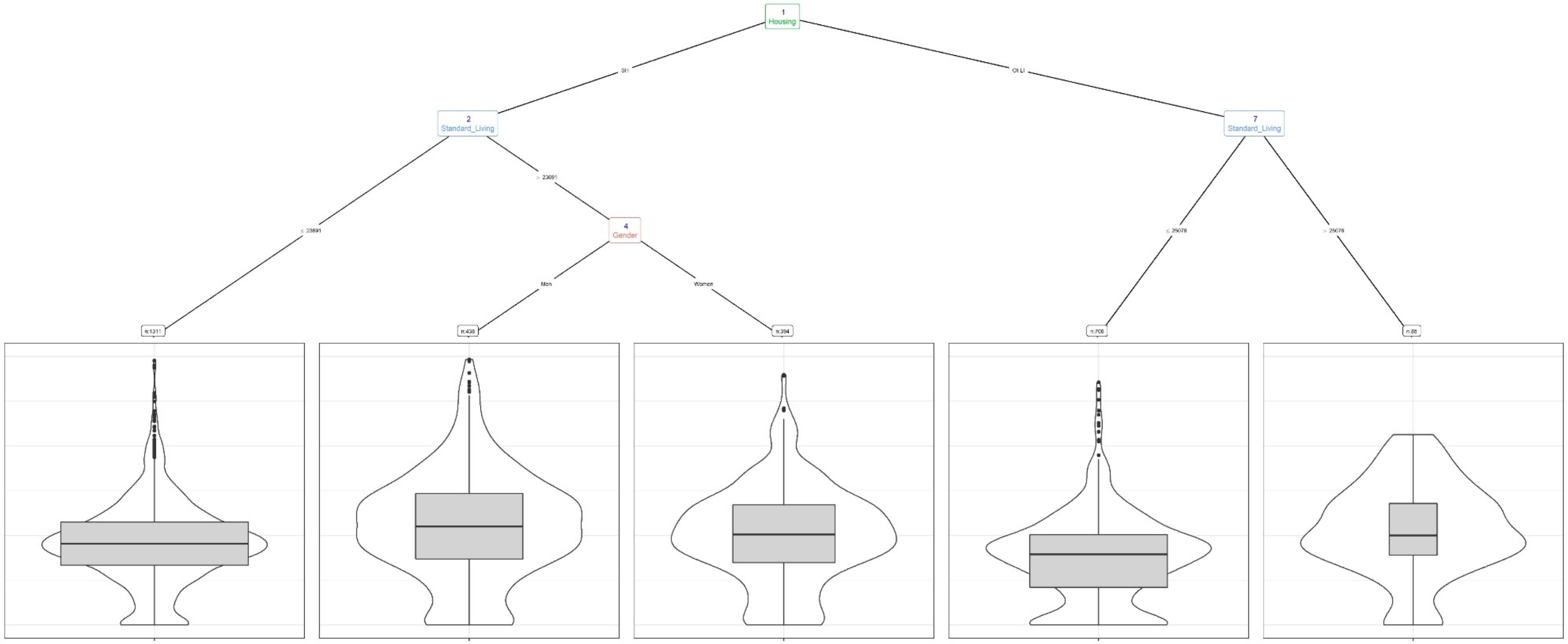
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros.

Figure A8: Conditional inference tree, region Nouvelle Aquitaine



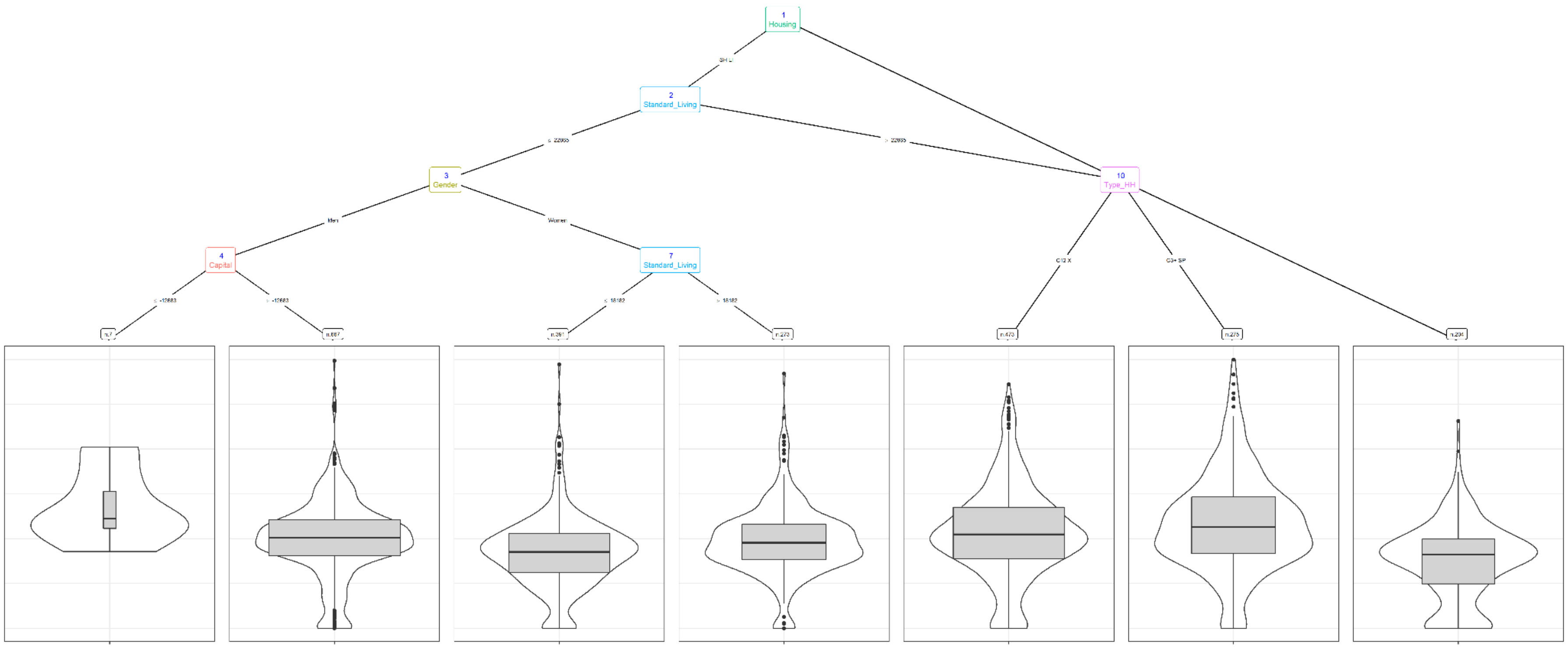
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. Source and coverage: see Table A4.

Figure A9: Conditional inference tree, region Occitanie



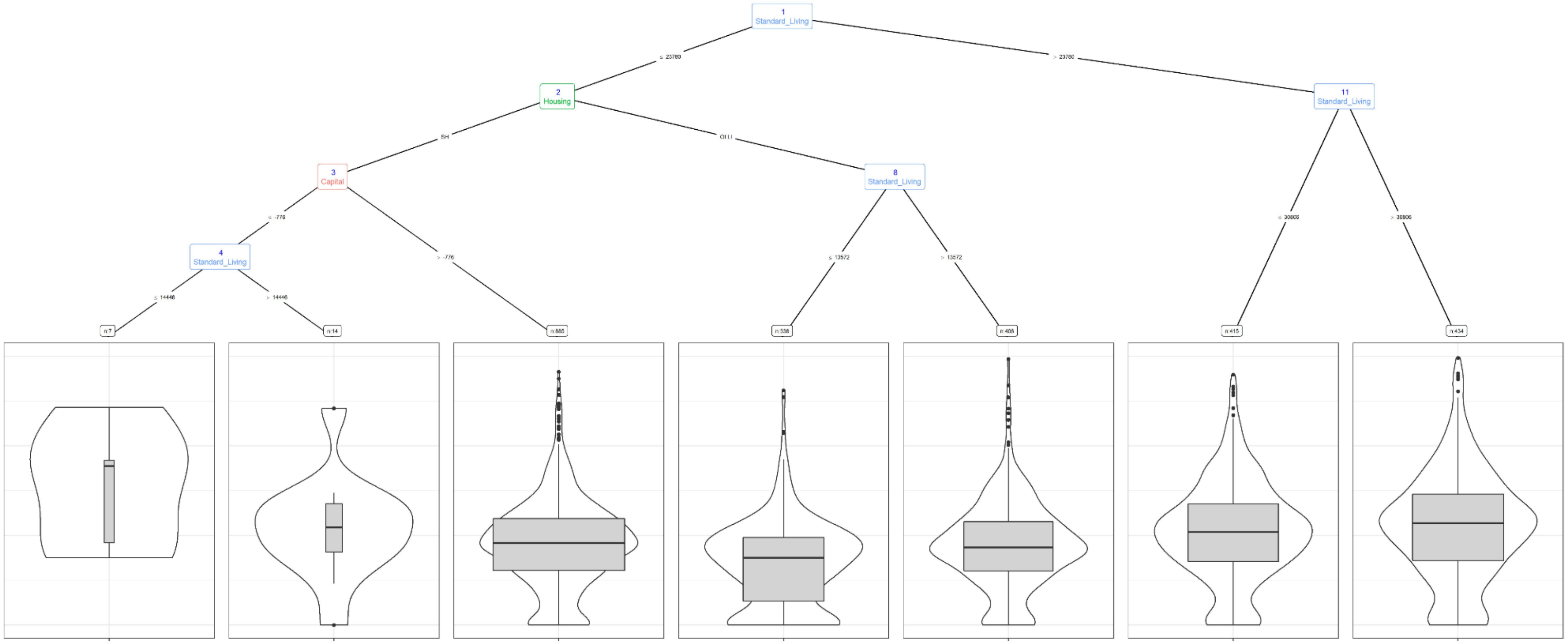
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. Source and coverage: see Table A4.

Figure A10: Conditional inference tree, region Pays de la Loire



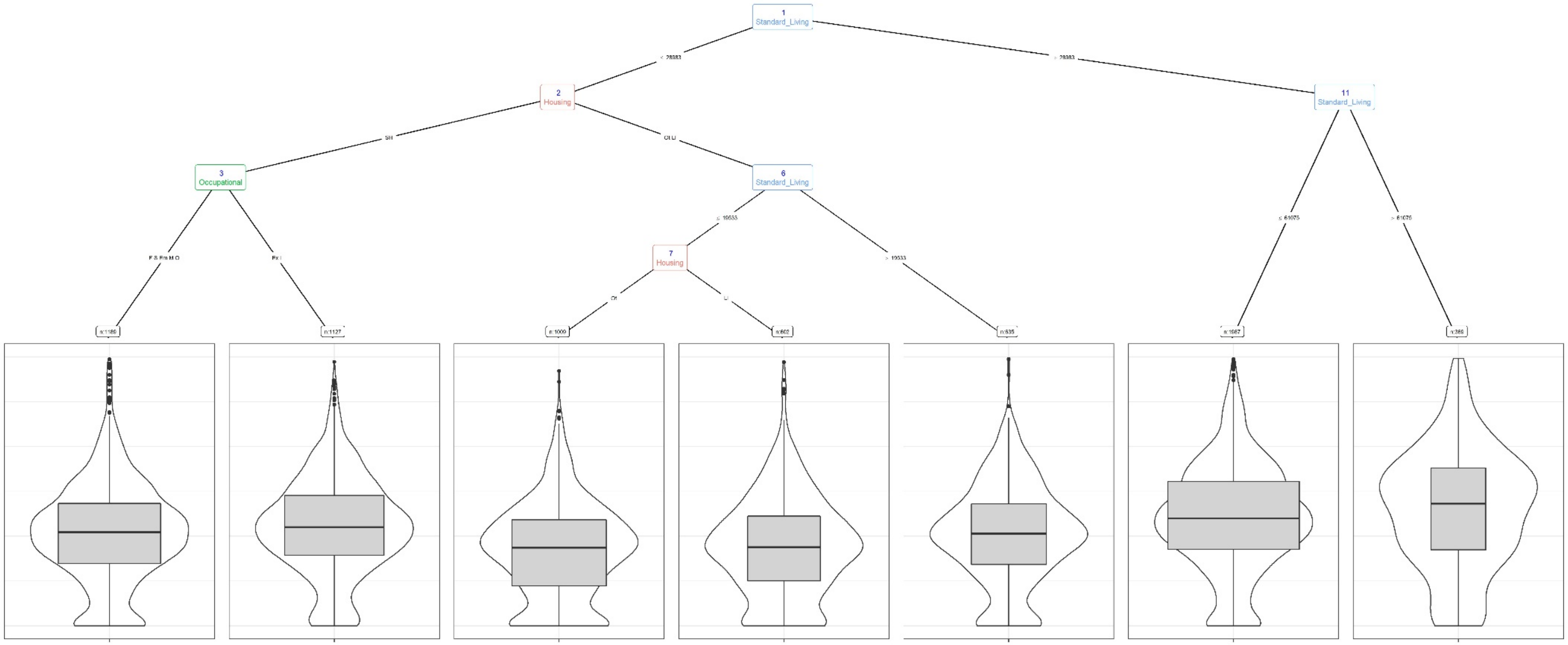
Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. The first split separates Landlords and social housing residents from other types of residency. Source and coverage: see Table A4.

Figure A11: Conditional inference tree, region Provence-Alpes-Côte d'Azur



Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. Source and coverage: see Table A4.

Figure A12: Conditional inference tree, region Ile de France



Note: The y-axis represents a range of income per capita from 0 euros to 60,000 euros, with each horizontal line denoting an interval of 10,000 euros. Source and coverage: see Table A4.

Figure A13: Total Inequality (Gini)

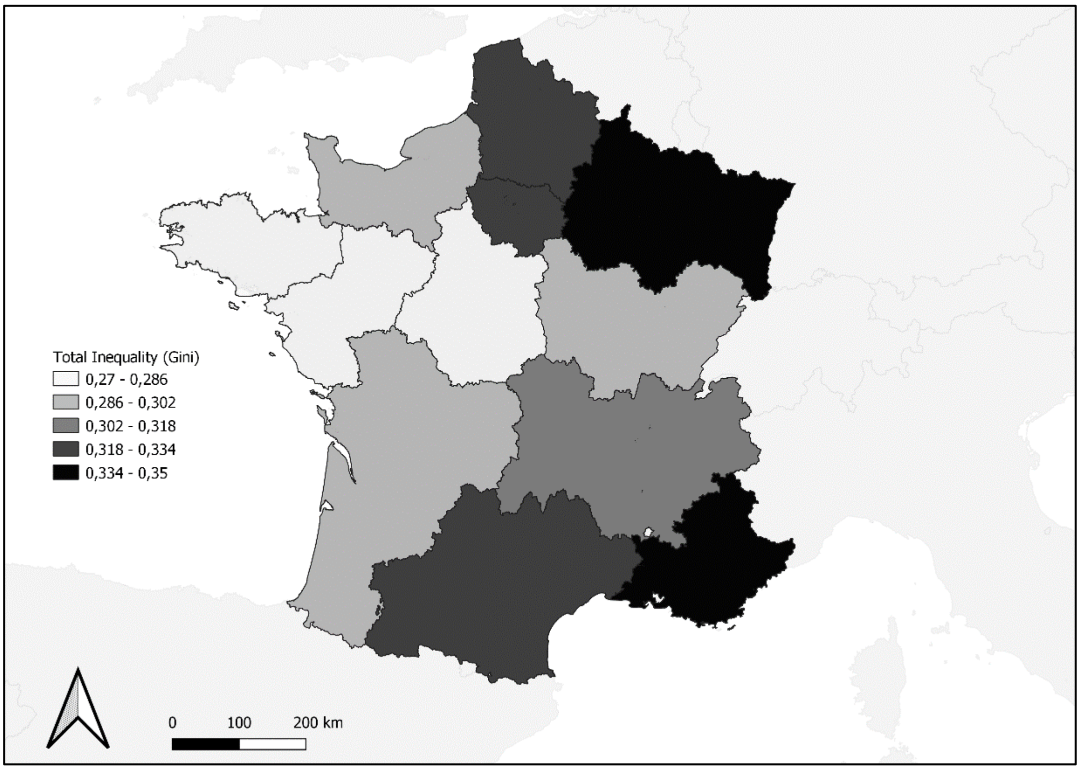


Figure A14: Total Inequality (MLD)

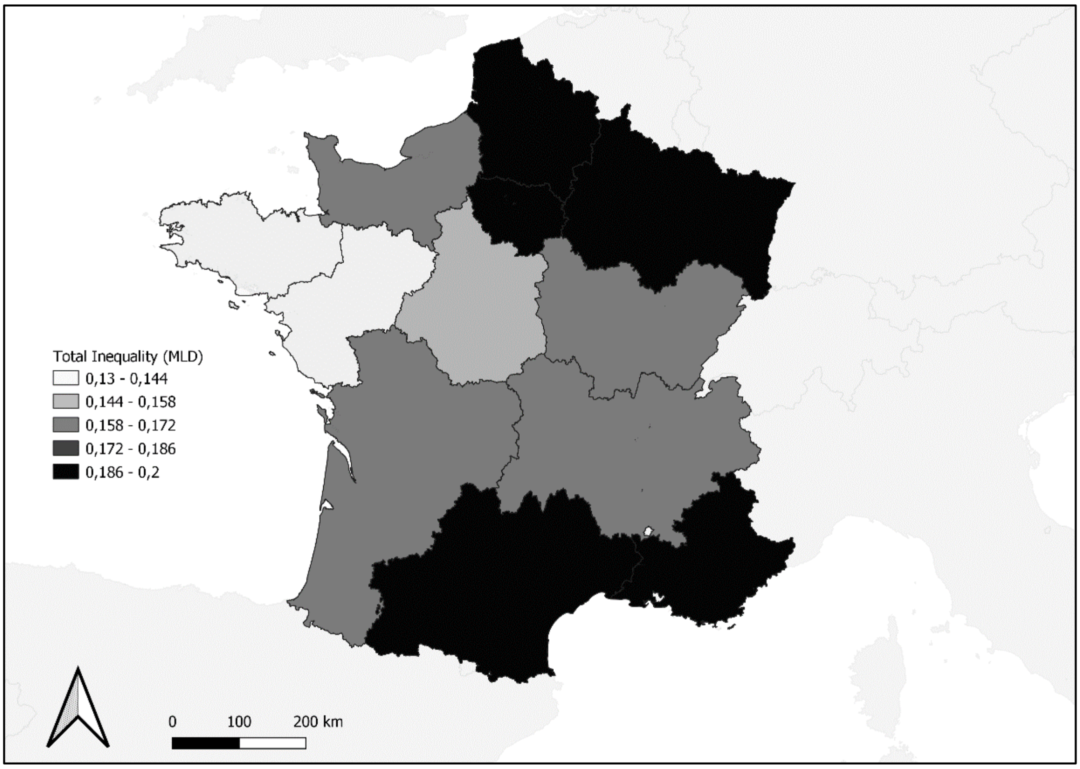


Figure A15: Share of ex-ante IOp

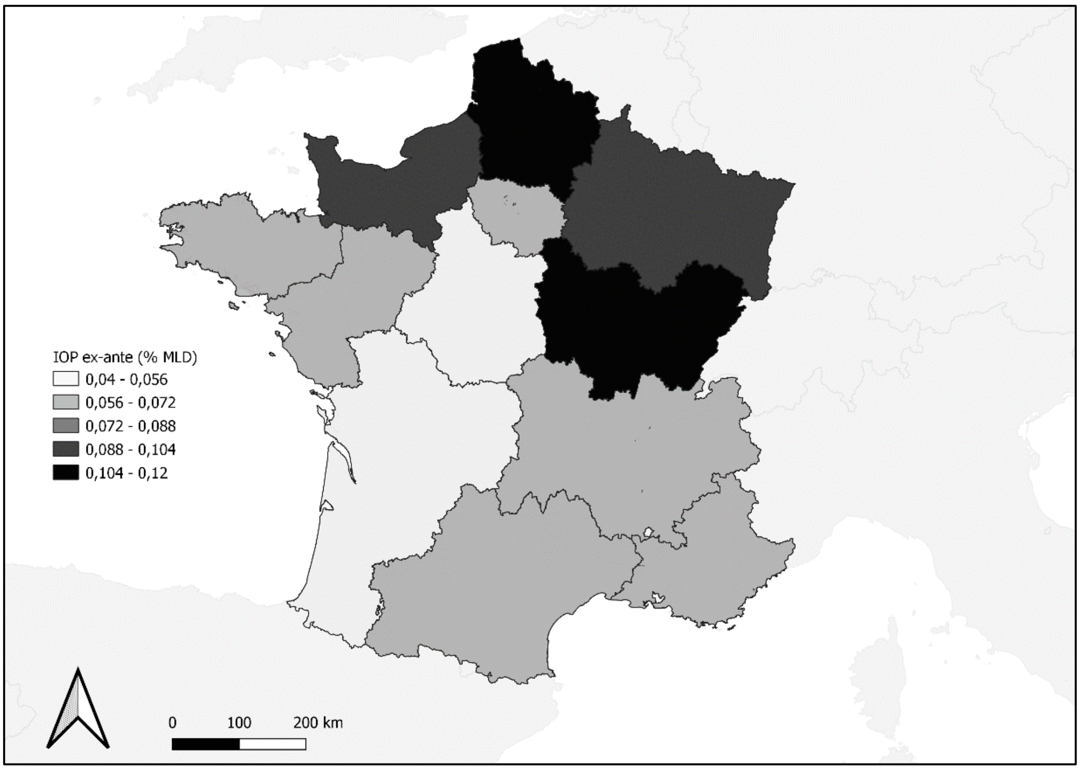


Figure A16: Share of ex-post IOp

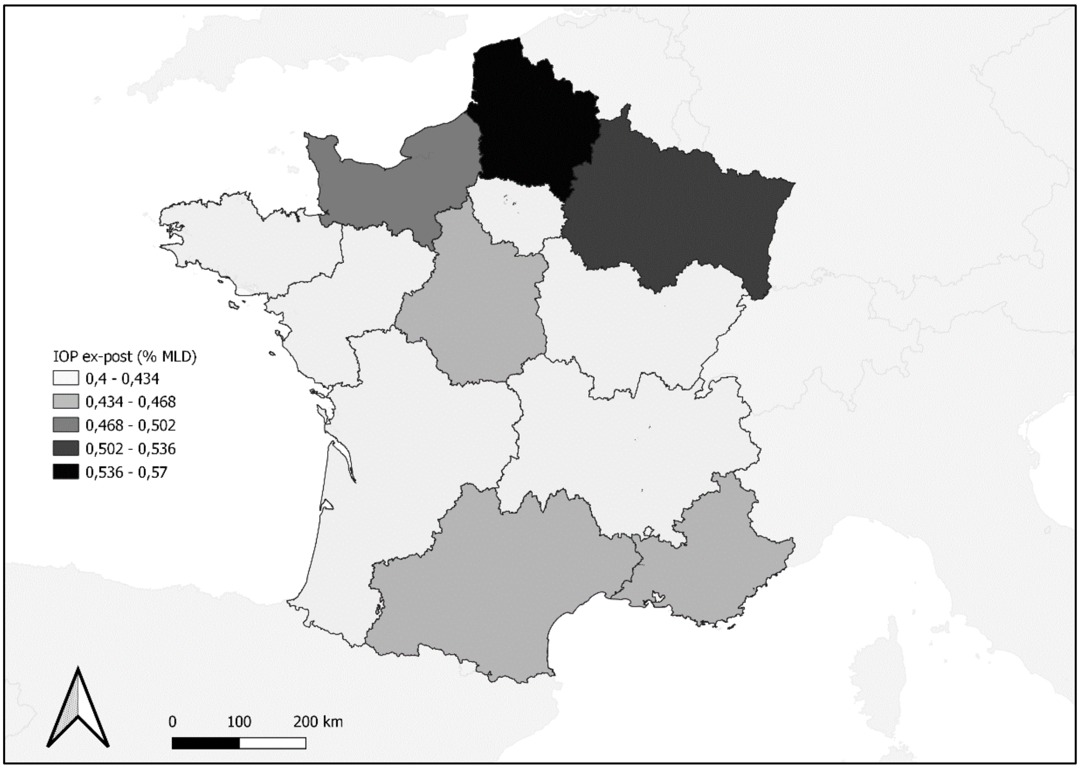
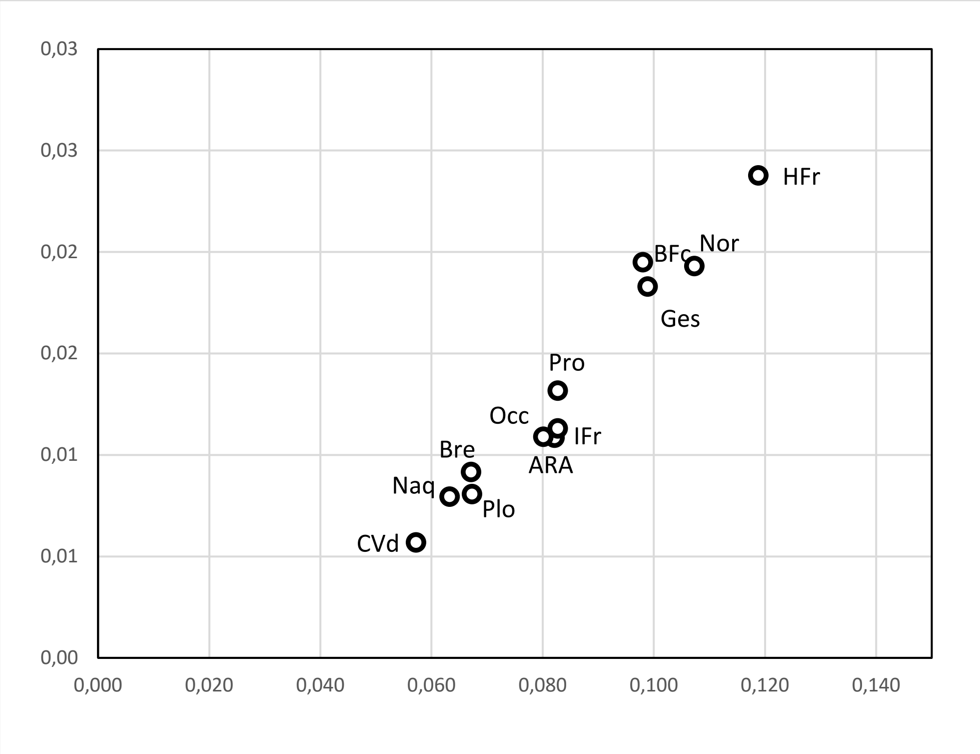
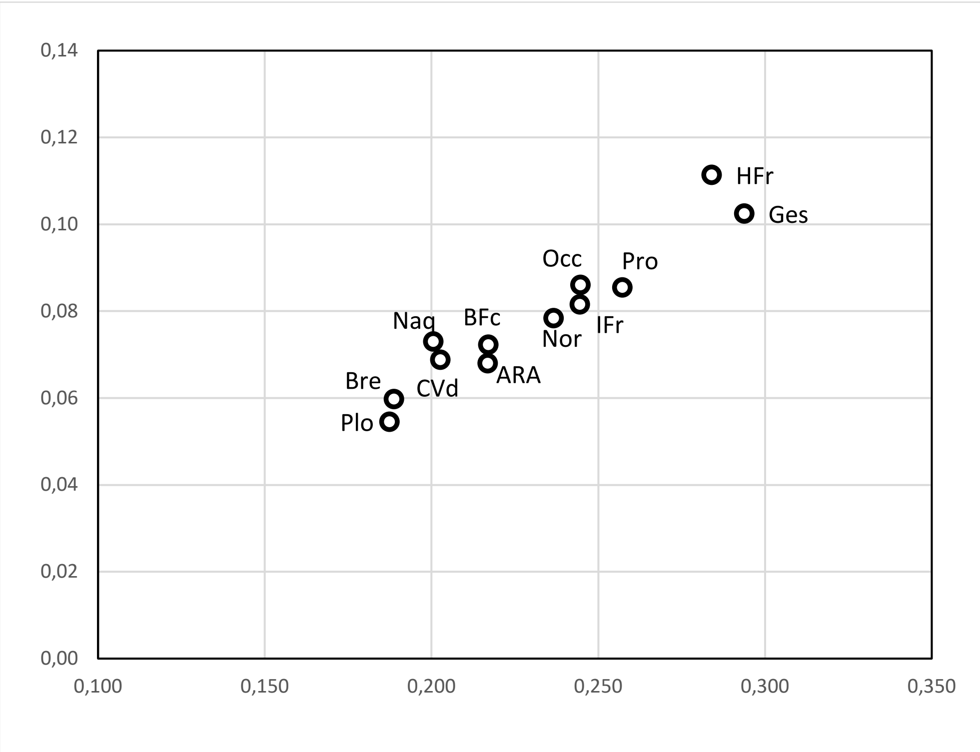


Figure A17: ex-ante IOp : Gini vs MLD



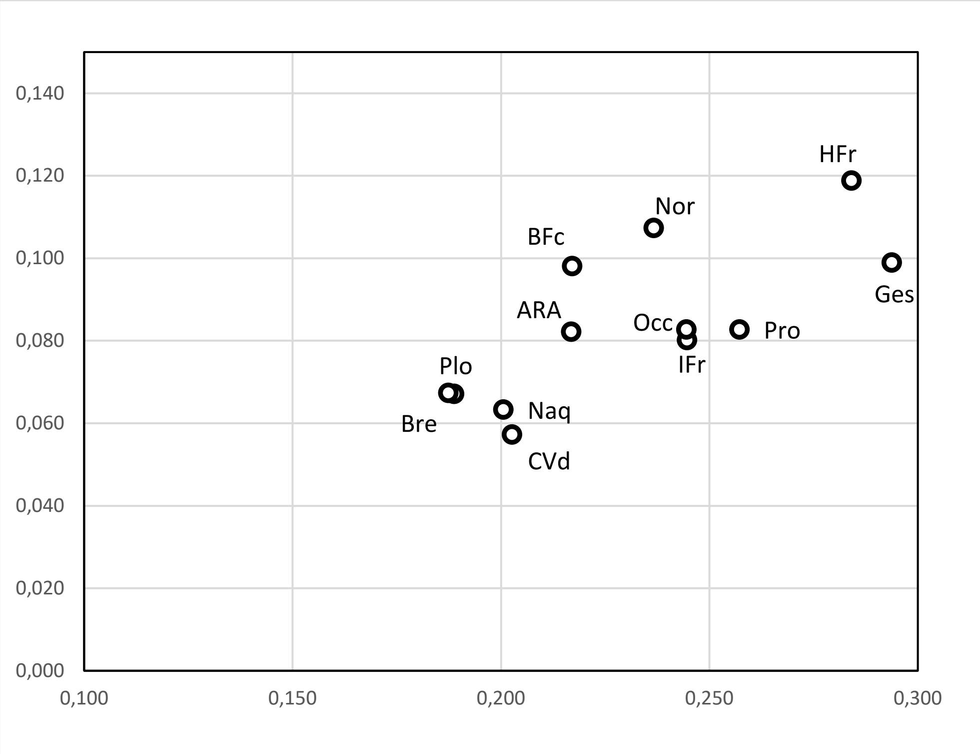
Note: the x-axis represents the Gini measures, and the y-axis represents the MLD measures. Labels used for each region are specified in the note of Figure 4. Source and coverage: see Table A4.

Figure A18: ex-post IOp : Gini vs MLD



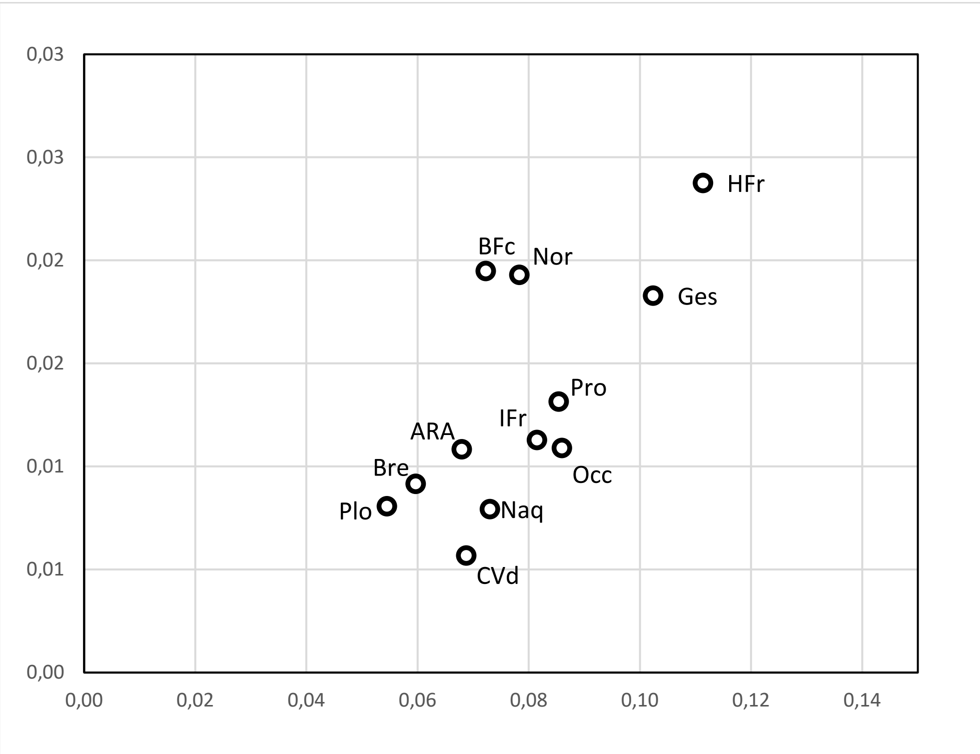
Note: the x-axis represents the Gini measures, and the y-axis represents the MLD measures. Labels used for each region are specified in the note of Figure 4.

Figure A19: ex-ante IOp vs ex-post IOP (Gini)



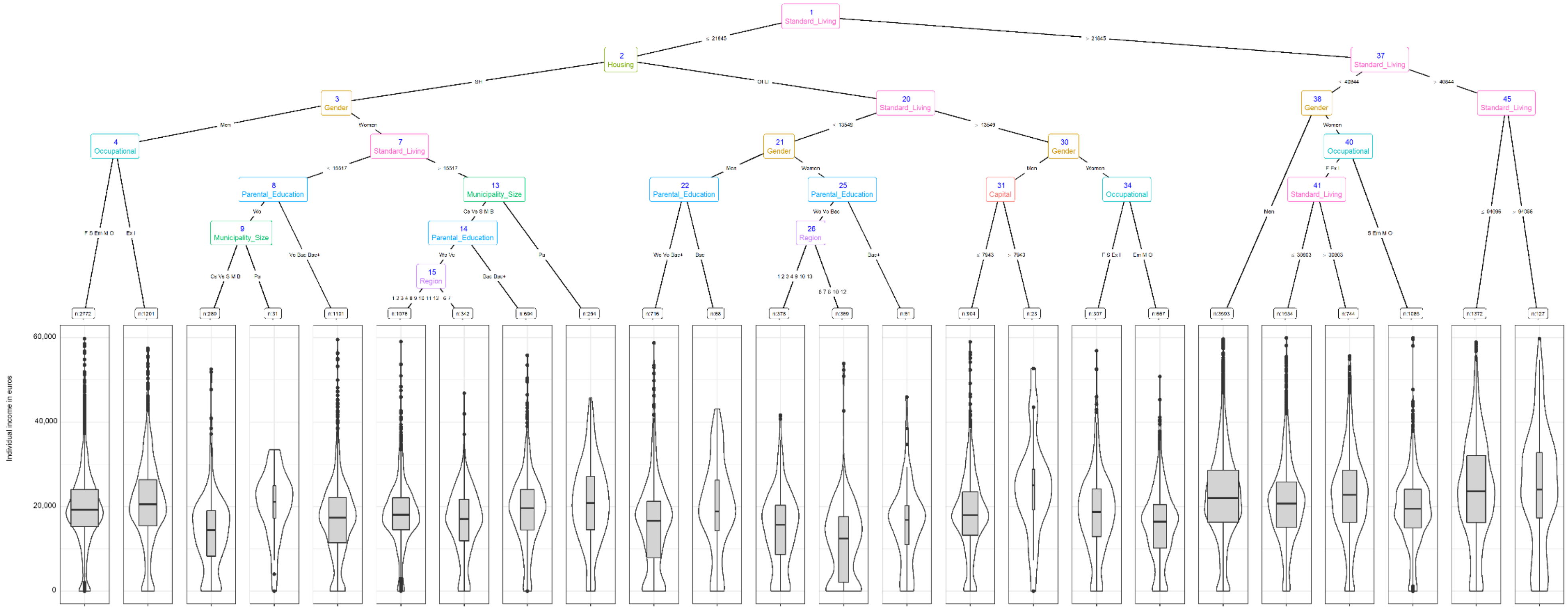
Note: the x-axis represents ex-post IOp measures, and the y-axis represents the ex-ante IOp. Labels used for each region are specified in the note of Figure 4.

Figure A20: ex-ante IOp vs ex-post IOP (Gini)



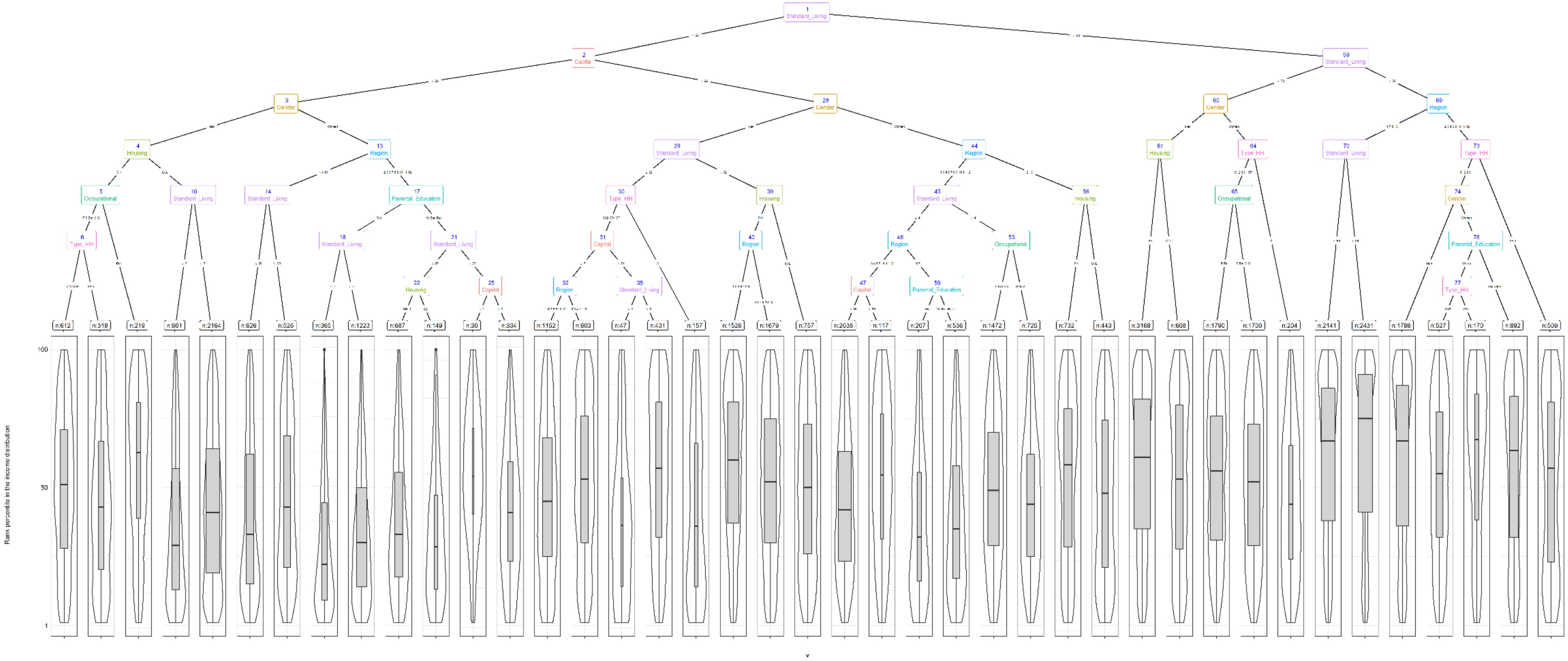
Note: the x-axis represents ex-post IOp measures, and the y-axis represents the ex-ante IOp. Labels used for each region are specified in the note of Figure 4.

Figure A21: Conditional inference tree, Model 2 excluding observations with missing values



Note: To facilitate the visualization, the y-scale of the boxplots has been fixed at 60,000 euros, and does not represent extreme values. Source and coverage: see Table A4.

Figure A22: Conditional inference tree, Model 3 with rank percentiles



BSE UMR CNRS 6060

Université de
Bordeaux
Avenue Léon
Duguit, Bât.
H 33608
Pessac,
France

Tel : +33 (0)5.56.84.25.75
http://bse.u-bordeaux.fr/

Derniers numéros – Last issues

Ernest MIGUELEZ is the scientific coordinators of the Bordeaux Economics Working Papers. The layout and distribution are provided by Cyril MESMER.