LATVIJAS BANKA
EIROSISTĒMA

ISBN 978-9934-578-66-3

LUDMILA FADEJEVA
KRISTA KALNBĒRZIŅA

CHILDHOOD CIRCUMSTANCES DEFINING THE INEQUALITY OF OPPORTUNITY IN EUROPE: WHAT ARE THE TREES TELLING US?

WORKING PAPER
4 / 2024

Latvijas Banka
K. Valdemāra iela 2A, Riga, LV-1050
Tel.: +371 67022300 info@bank.lv
http://www.bank.lv https://www.macroeconomics.lv

Childhood circumstances defining the inequality of opportunity in Europe: what are the trees telling us?

Ludmila Fadejeva*, Krista Kalnbērziņa†

October 15, 2024

Abstract

In this paper, we use boosted forests method to identify the main childhood circumstances associated with inequality of opportunity in Europe. The five main factors that influence income are the education of parents, financial situation of the household, gender, country of birth, and degree of urbanisation. The ranking of those factors differs between countries; however, these top factors are at the highest importance in both 2011 and 2019 at the aggregate level. We show that countries can be grouped into regions by the main factors driving inequality of opportunity - Southern European countries (country of birth), Central European countries (gender and highest education level of a parent), others (highest education level of a parent and financial situation). We also demonstrate that the importance score of various childhood circumstances is associated to the extent to which policies actively tackle these issues at the time of the respondents' childhood. Furthermore, a correlation between reduction in inequality of opportunity between 2011 and 2019 and improvement in education quality and wider social support coupled with improvement in governance effectiveness between 1995-2000 and 2003-2008 is established.

Keywords: Inequality of opportunity, childhood circumstances, inter-generational transmission of disadvantages, boosted forest, EU-SILC database

JEL Codes: D31, D63, E24, C39, I14, I24

*Monetary Policy Department, Latvijas Banka, K. Valdemāra iela 2a, Rīga, LV-1050, Latvija; e-mail: ludmila.fadejeva@bank.lv

†Monetary Policy Department, Latvijas Banka, K. Valdemāra iela 2a, Rīga, LV-1050, Latvija; e-mail: krista.kalnberzina@bank.lv

1 Introduction

In this paper we study the roots of inequality of opportunity in European countries. We use the module on intergenerational transmission of disadvantages from the survey European Union Statistics on Income and Living Conditions (EU-SILC) to identify the main factors leading to unequal wage outcomes for individuals due to the different circumstances which they were subjected to when growing up.

Studying inequality of opportunity and understanding its sources is central to developing sustainable social and economic policies. One of the reasons concerning the inconclusiveness of research on overall income inequality and its impact on growth is the narrow definition attributed to inequality. Importantly, overall income inequality stems from two sources that impact economic growth from opposite directions. First, inequality based on circumstances, or factors that cannot be controlled by the individual, like one's socioeconomic background or gender. Second, inequality of effort or factors that are an individual's responsibility, like professional choices and effort put in education. Economic growth is negatively influenced by 'inequality traps'" or inequality of opportunity that permanently exclude groups of individuals from economic, social, and political life, leading to a loss in growth and development potential. As shown by Bourguignon et al. (2007), Ferreira and Walton (2006), Marrero and Rodríguez (2010) and Aiyar and Ebeke (2020) inequality of opportunity based on individuals circumstances rather than inequality of income is linked to economic growth. Furthermore, in research concerning strength of nature versus nurture factors in determining the income of individuals, Black et al. (2019) show that while human capital linkages between parents and children appear to have stronger biological roots than environmental, earnings and income are, if anything, more environmental. Thus, government policies aimed at supporting families and providing equal access to education and health services provide conditions for stronger economic growth in the future.

The necessity to 'level the playing field' or equalise opportunities instead of equalising individual outcomes, in order to strive towards a just society, has created a vast literature that formalises the methods of assessing equality of opportunity. The module on the intergenerational transmission of disadvantages in EU-SILC allows for the exploration of this topic (see for example Marrero and Rodríguez 2012, Checchi et al. 2010, Brunori et al. 2013, Sentís et al. 2023, Carranza 2022). Most previous studies focus on an estimation of the inequality of opportunity measures while distinguishing between the ex-ante/ex-post approaches and non-parametric and parametric methods (for example, Ramos and Van de Gaer 2017).Significantly less focus has been placed on the factors determining inequality of opportunity (for example, Han 2022 for

South Korea, Sentís et al. 2023 for Spain).

Contrary to others, who aim to measure the level of inequality of opportunity, we seek to identify the sources of inequality of opportunity by identifying the factors leading to inequality of opportunity in European countries. We estimate the importance of childhood circumstances in each country and group countries by the main circumstances driving inequality of opportunity. In this paper, we target the analysis to a specific age cohort (25-40) characterised by the start of respondents' career paths, thus capturing the impact of circumstances in their early adult years. Additionally, it was chosen to focus on a younger cohort due to the relatively short policy time series, and our wish to explore correlations between the policy measures during respondents' childhood and the importance of circumstances driving inequality of opportunity.

Traditionally, the group of methods which allow for the evaluation of the importance of circumstances in inequality of opportunity are parametric methods (for example, Ferreira and Gignoux 2008). OLS regression or instrumental regression methods suffer from estimation biases, which are challenging to resolve due to the limited number of explanatory variables available in the survey. Brunori et al. (2021) show that machine learning methods, particularly the regression tree approach, improves estimations by lowering upward and downward biases. Han (2022) and Carranza (2022) are among those who apply the tree approach to analyse inequality of opportunity in recent research applications. We follow this approach and apply tree methods with boosting, which has proven to provide sound explanatory power in the regression class of tree models.

So far, there have been three waves of the EU-SILC module on intergenerational transmission of disadvantages (2005, 2011, 2019). Due to improved question comparability and a larger sample of countries participating in the last two waves (27 European countries), we have chosen to focus on 2011 and 2019. The novelty of our paper stems from the focus on circumstances determining the inequality of opportunity rather than the level and application of a novel estimation method - boosted trees tuning models are estimated individually for each country.

We also show that the importance score of various factors is driven by the extent to which policies actively tackle these issues at the time of the respondents' childhood. By analysing the change in inequality of opportunity and change in policies, we show that while there is a positive association between the change in inequality of opportunity and change in quality of education, the effect of change in government expenses on education and family support is more pronounced if accompanied with improvement in governance effectiveness.

We also show that the importance score of various factors is driven by the extend to which policies are tackling the issues at the time of respondents childhood. Looking at change in inequality of opportunity and change in policies, we show that while there is a positive association between the change in inequality of opportunity and change in quality of education, the effect of change in government expenses on education and family support is pronounced if accompanied with improvement in governance effectiveness.

The paper is structured as follows. Section 2 provides a brief literature review, Section 3 discusses a selection of questions from the EU-SILC module on intergenerational transmission of disadvantages for analysis. Section 4 provides an overview of hourly wage income inequality by sets of circumstances. Section 5 outlines the methodology of estimating circumstance effect in income inequality. Section 6 presents the simulation results and discusses the correlations between the importance of various circumstances and change in inequality of opportunity with policy indicators. Section 7 concludes the study.

2 Literature Review

Empirical literature measuring inequality of opportunity has been vast since 1998 when John Roemer (1998) defined inequality of opportunity in economic terms. Roemer's idea is that individuals' outcomes are defined by two factors: effort, which is a factor over which individuals have control, and circumstances, a factor which is given to the individual. The idea of equal opportunities suggests that differences of outcome due to different efforts are acceptable and necessary; however, differences in outcomes due to differences in circumstances are not. Policy aimed at creating equal opportunities would need to ensure that an equal degree of effort leads to a possibility of equal outcomes, regardless of an individual's circumstances. A full review of the philosophical debate and economic models used to estimate inequality of opportunity is proposed in Roemer and Trannoy (2016). Various approaches have been used to determine the level of inequality of opportunity. See among others Checchi et al. (2016), Andreoli et al. (2021), Brunori et al. (2013), Brunori et al. (2021), Saavedra-Chanduví et al. (2011), Fleurbaey et al. (2015), Carranza (2022). At the same time, there are few papers exploring the importance of the factors determining inequality of opportunity (for example Han 2022 for South Korea, Sentís et al. 2023 for Spain).

Generally, the measurements of inequality of opportunity stem from two premises (Fleurbaey and Peragine 2013). First, from the ex – post perspective, there is equality of opportunity if all individuals exerting the same effort can achieve the same outcome. This approach is estimated by considering the individuals with the same effort and measuring inequality within these groups of individuals. Second, the ex-ante perspective is evaluated by considering individuals of the same circumstances. From this perspective, there is equality of opportunity if all individuals face the same outcome regardless of their circumstances. Ramos and Van de Gaer (2017) lay out the considerations of the choice between the ex – ante and ex – post approaches.

In European countries, the inequality of opportunity can be evaluated using the EU-SILC database. Previous studies show that there are persistent differences in levels of inequality of opportunity between European countries. Marrero and Rodríguez (2012) group the Nordic, continental, and select Eastern European countries as low inequality of opportunity countries, and Mediterranean, Atlantic, and select Eastern European countries as high inequality of opportunity countries. Checchi et al. (2010) define three groups of European countries based on inequality of opportunity i) formerly centrally planned economies with the highest levels of inequality of opportunity, ii) most of continental Europe and iii) egalitarian Northern countries. Carranza (2022) , using EU-SILC data for 2005, 2011, and 2019, while grouping countries into 5 regions, showed that in all regions there was growth in inequality of opportunity in 2011 as compared to 2005, which was followed by a slight decline in 2019. The only two groups of countries where inequality of opportunity increased in 2019 (yet not statistically significantly) were Nordic and Baltic countries.[1]

As regards factors determining inequality of opportunity, there is a vast body of literature connecting labour market outcomes later in life to childhood family background. In their comprehensive review of the literature on intergenerational mobility, Mogstad and Torsvik (2021) investigate the difficulties of measuring intergenerational persistence in socio-economic outcomes, emphasizing the influence of family environment and genetic factors on children's futures. They highlight the significant role family background plays in shaping an individual's economic prospects, noting a concerning trend towards increasing persistence of inequality across generations.

Andreoli et al. (2021) examine inequality of opportunity using EU-SILC data distinguishing between age cohorts and linking the results to educational policy. They find a negative relationship between inequality of opportunity among children and the duration of compulsory education of the parents. Checchi et al. (2010) also show that inequality of opportunity exhibits negative correlation with fiscal redistribution and public expenditure in education.

[1] 1According to our estimates, the increase in inequality of opportunity in Baltic countries is due to Lithuania. There were no changes in Latvia and a slight decline in Estonia.

3 Data

The EU-SILC survey contains cross-sectional and longitudinal information on income, social exclusion, and living conditions of different types of households and individuals. In 2005, 2011, and 2019, the questionnaire included a special module of questions on intergenerational transmission of disadvantages. Due to improved question comparability in the module, and a larger sample of countries participating in the last two waves (27 European countries[2]), we focus on EU-SILC 2011 and 2019.

The module provides information on numerous factors of a person at the age 14. First, demographic characteristics in the household: number of children and adults, number of working people, presence of both mother and father. Second, information about the mother and father of the individual: country of birth, citizenship, highest level of education, activity status, occupation. Lastly, information about the well-being of the household: self-reported assessment of the financial situation, tenancy status, degree of urbanisation, if basic school needs were met, if the household could afford meals with meat/chicken/fish (or vegetarian equivalent) daily, if the household could afford one-week annual vacations away from home.

The detailed description of questions and transformations used in the analysis is given in appendix Table A1. The decision on the inclusion of variables into our analysis was based on several factors. First, if the share of responses for some variables is low, by excluding these variables, we can work with a larger data set. The example of such variables are questions of managerial positions of parents - 85% response share, occupation of parents - 86%. Second, a high correlation between the financial situation, food consumption, and annual vacations away from home variables allows us to keep just one of them to describe the financial situation of a household at the age of 14. Similarly, the variables country of birth and citizenship are highly correlated, which allows us to keep only the country of birth variable in the data set. Third, in order to keep maximally different aspects of household circumstances affecting inequality of opportunity, we have kept variables such as the number of children in the family, presence of both parents, activity status of parents, education level of parents, tenancy status, degree of urbanization, coverage of school needs, and gender in the final version of the data set.

A significant share of households could provide information for only one parent. To keep the sample as full as possible, we chose to aggregate information about parents using minimum or maximum values for both parents. As a result, we created a variable showing the highest level

[2] Austria, Belgium, Bulgaria, Switzerland, Cyprus, Czech Republic, Estonia, Greece, Spain, Finland, France, Croatia, Hungary, Ireland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia

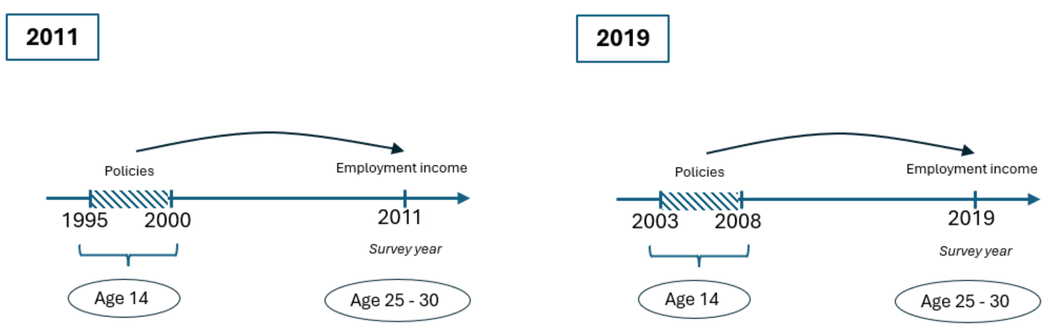
of education attained by a parent, a variable depicting the highest activity level of a parent (for example, if one parent was employed), and a variable showing if at least one of the parents was born in the country of the survey.

Table 1: Sample period

Year of the survey (age 25-40)	Year of registered circumstance (age 14)	Year of birth
2011	1985 - 2000	1971 - 1985
2019	1993 - 2008	1979 - 1993

The variable of interest in this analysis is gross employee cash (or near cash) income per hour. We estimate it using information on the number of months spent at full or part-time work and the number of hours actually worked per week in the main job. The sample is restricted to employees who worked for at least 3 months during the reference period. We also restrict the sample to relatively young individuals (age range from 26 to 40). Consequently, we analyse how the childhood circumstances in 1985 - 2000 (EU-SILC 2011) and in 1993-2008 (EU-SILC 2019) have influenced the income levels of individuals aged 25-40 at the moment of the survey (Table 1). The sample is restricted to limit income differences caused by severe changes in the socio-political situation during the end of the 20th century.

Figure 1: Timing scheme for policy analysis



For the change in policy analysis, we restrict the sample even further by selecting only individuals from 25 to 30 years old (Figure 1). This decision is driven by a short policy time series (most of the data starts in 1995 or 2000) and our wish to compare changes in the policies of non-overlapping periods.

4 Descriptive analysis

In this section, we provide a descriptive overview of hourly wage income inequality by sets of circumstances in Europe using the EU-SILC 2019 data (Figure 2).

We consider three circumstances as sources of inequality of opportunity: the level of income of the household (when the respondent's age was 14), highest education level reached by a parent, and gender. The respondents are separated in groups based on their circumstance: good, moderate, or bad financial situation; high, medium, or low parent education level; and gender - male or female. Inside each group people are ranged by their hourly adjusted wage income and divided into 10 groups. For each group, the median level is estimated and divided by country median. Thus, the obtained ratio shows by how much income in the specific decile of people differ for different childhood circumstances and gender. In a world without inequality of opportunity, ratios of groups defined by circumstances would be equal, i.e. the income median for people should be the same despite facing different circumstances in childhood and gender.

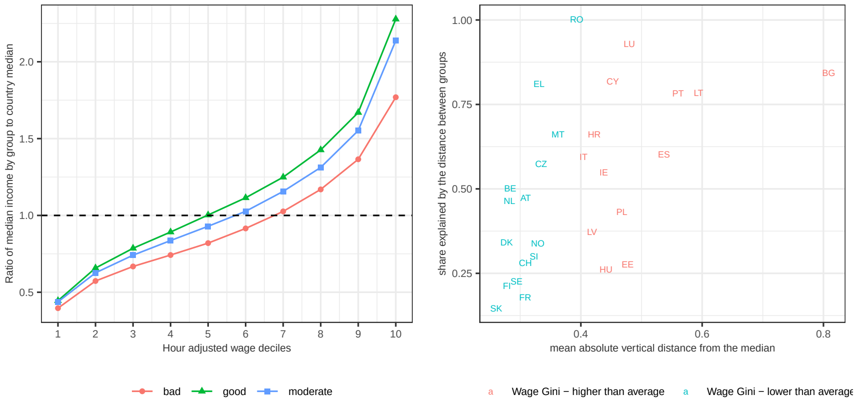
Figure 2 shows the inequality of hourly wage income in Europe. Figures to the left represent median wage income by income decile of each category and country relative to the country income median of the respective country. Each dot is estimated as a simple average across countries. The distance between the lines indicates the level of inequality of opportunity while the slopes of the lines indicate the inequality of hourly wage income relative to the country median hourly wage income.

Figures to the right summarizes country level information presented in the appendix (Figure A1, Figure A2, and Figure A3). On the horizontal axis, the mean absolute ratios of median income to country median for each country (mean vertical absolute distance from the line to the median) is shown. The vertical axis indicates the share of inequality of opportunity proxy (estimated as the average vertical distance between the lines) in the mean absolute ratio (mean vertical absolute distance from the line to the median). In other words, the vertical axis shows the share of wage inequality that is explained by the inequality of opportunity, while the horizontal axis proxy shows the overall inequality of hourly wage. Each dot is estimated as a simple average across income deciles for each country. To simplify the analysis, we separate the countries in low and high hourly wage income inequality countries by looking at the country specific Gini coefficient relative to the average Gini for all countries.[3]

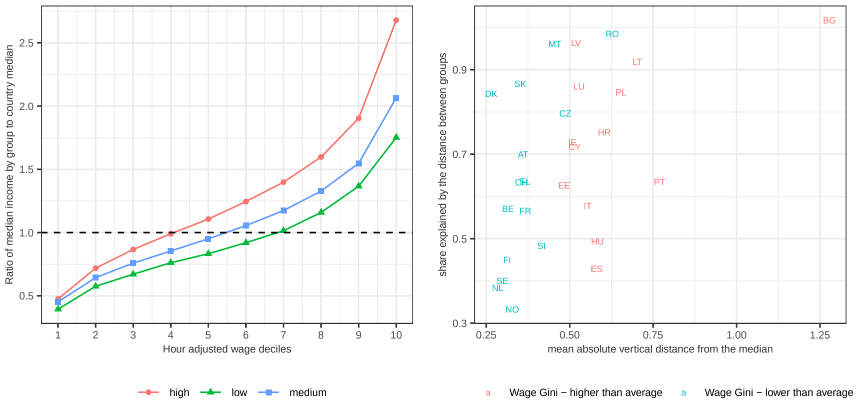
[3] Low wage income inequality countries are defined by a wage income Gini coefficient that is lower than the average. These are Slovakia, France, Finland, Sweden, Switzerland, Slovenia, Norway, Netherlands, Belgium, Austria, Czech Republic, Malta, Greece, Romania. A higher than average wage income Gini coefficient is in Hungary, Latvia, Estonia, Poland, Ireland, Italy Croatia, Spain, Portugal, Cyprus, Luxembourg, Lithuania, and Bulgaria.

Figure 2: Inequality of median wage income by income decile in 2019

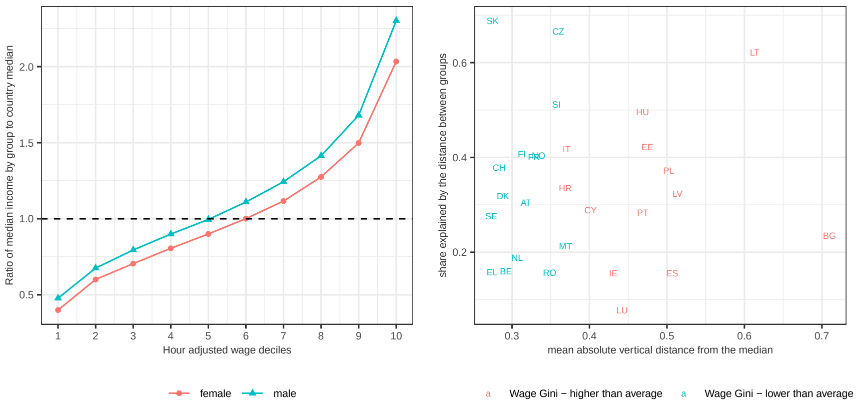
(a) by household financial situation at age 14



(b) by parents' education level



(c) by gender



Notes: the set of individuals includes people in the age category 26-40, who were employed for at least 12 months. Gross employee cash or near cash income (wage) is adjusted by hours worked per week.
(Graphs to the left): Income deciles are estimated separately for each country and category. Each point on the graphs represents a simple average over 29 countries. Ratios are obtained as median wage in the category of individuals to country median. (Graphs to the right): On the horizontal line - the mean absolute distance from the median: plot (a) - using category "Good", (b) - category "High", (c) - category "Male". On the vertical line - ratio between the average distance between groups and the mean absolute distance from the median. The distance between groups: (a) "Bad" - "Good", (b) "High" - "Low". Colours indicate the level of hourly wage Gini coefficient as compared to a simple average between the countries.
Source: authors' estimations using the EU-SILC 2019 database.

One of the main factors determining the inequality of opportunity later in life is the income level of the household. Figure 2a to the left shows that the level of household income while growing up has a significant impact on income levels in the future, which is depicted by the non-zero distance between the ratios of the median income across all income deciles. Furthermore, the distance between moderate to good is less than the distance from bad to moderate, suggesting that a very weak financial situation in childhood restricts the person's future income to a higher degree. On top of that, children that come from households with a moderate or good financial situation have less differences in the outcomes. The cross-country analysis shows significant heterogeneity among countries (Figure 2a, graph to the right). The household financial situation has low association with the hourly income later in life in Slovakia, France, Finland and Sweden. The association is significantly more pronounced in Bulgaria, Romania, Luxembourg, Cyprus, Bulgaria, Greece, Portugal, and Lithuania.

Following that, we consider the level of parental education (Figure 2b) and its correlation with the hourly wage of a person later in life. On average, the differences in income for different categories of parents' education level are more pronounced if compared to the financial situation measure. The data demonstrates that the higher the income decile, the larger the distance between different categories. Furthermore, the additional gain from the medium education level to high is greater than from low to medium. The cross-country analysis shows significant heterogeneity. Parental education has a high correlation with the overall wage income inequality in Bulgaria, Romania, Malta, Latvia, Lithuania, Luxembourg, and Poland. It is also high in Slovakia and Denmark, where the overall wage inequality is relatively low.

Gender is another important source of inequality of opportunity. Figure 2c shows that the distance between gender, male and female, groups is relatively stable along all income deciles. Comparing to the previous sets of circumstances, gender has a lower correlation with the overall inequality of wages. However, there are countries in Europe where gender is a significant source of inequality of opportunity, notably Slovakia, Slovenia, Czech Republic, Hungary, and Lithuania.

5 Methodology

There are two main approaches to estimating inequality of opportunity from an economic factor point of view - the ex-post and the ex-ante. The ex-post measure looks at the income in groups sharing the same effort (inequality of opportunity decreases if outcome inequality decreases among individuals at the same degree of effort). The ex-ante measure looks at the income in groups sharing the same circumstances. The EU-SILC provides a set of variables describing circumstances and allows for a convenient estimation using the ex-ante approach (see, for example, Marrero and Rodríguez 2012, Checchi et al. 2010, Brunori et al. 2013).

Bourguignon et al. (2007) proposed a regression-based method to estimate circumstances based on income level. Income y_i can be presented as

$$y_i = C_i\alpha + E_i\beta + u_i \tag{1}$$

$$E_i = HC_i + v_i \tag{2}$$

where α and β capture the effect of circumstance C_i and effort E_i. H is a matrix of coefficients that capture the effect of circumstances on effort. After substitution equation (2) into (1), we can obtain an expression only in terms of circumstances, which can be estimated by OLS regression.

$$y_i = C_i\gamma + z_i \tag{3}$$

where $\gamma = \alpha + H\beta$ and $z_i = u_i + v_i$. This way we can obtain a direct measure of income which can be explained by circumstances, and the error term would measure a combination of effort effect and pure residual.[4]

Empirically, methods can be divided into non-parametric and parametric methods (see, for example, Ferreira and Gignoux 2008). The traditional regression or instrumental variable methods suffer from estimation biases, which are hard to fix due to a limited number of explanatory variables available in the EU-SILC survey and a limited number of interaction terms explored. Brunori et al. (2021) and Han (2022) show that machine learning methods, particularly the regression tree approach, improves estimations by lowering those upward and downward biases. We follow this approach and apply the boosted tree model which is proven to have sound explanatory power in the class of regression tree models. To do so, we use *R* package *xgboost*, which is an efficient implementation of the gradient boosting framework from Chen and Guestrin (2016).

The majority of explanatory variables in our analysis are ordered values (level of education – low, medium, high) or discrete (levels of economic activity of households), thus its direct use in the tree model may be misleading (the distance between points might lack economic logic). To avoid this, we transform such variables into several binary dummies (see Table A1) and let

[4] Despite a rather large number of questions in EU-SILC describing the childhood experiences of a person at age 14, the set of circumstances is not exhaustive. Therefore, a part of inequality attributed to effort might be overestimated due to the omitted circumstances.

the model decide on the importance of each dummy separately. The importance of a variable is estimated by summing up the obtained importance values for the corresponding dummies. To restrict the artificial importance increase, which can occur if a variable is split into too many dummies (as compared to the other variables in the regression), we restrict the number of dummies created from one variable to five. The majority of variables have three dummies.

For each country/year we allow the model to tune three parameters: the size of sample (0.55-0.75), the number of trees (2-100), and the tree depth (3-10). We estimate regression trees using tree levels of learning rate 0.3, 0.1, and 0.05. Other parameters are set at the default values. Root mean squared error (RMSE) is used to pick the best specifications. Table A2 shows the final sets of parameters. The tuned sample size and number of trees vary by country. However, for the majority of countries, the tree depth is tuned to be shallow (3-4 layers), which is in line with the idea of the boosted forest framework. By construction, it builds an ensemble of shallow and weak successive trees with each tree learning and improving on the previous. When combined, these many weak successive trees produce a powerful *committee* [5].

Figures A5 and A6 graphically explain tuning for the number of trees. Specifically, the higher the learning rate, the smaller the number of trees chosen. With the learning rate 0.3 (default value), RMSE minimisation algorithm chooses 10 trees on average. The learning rate 0.05 requires around 75 trees. The goodness of fit, measured as RMSE for training and testing samples, on average, improves with a lower learning rate. Therefore, in the final version of the model, we use learning rate 0.05. Table A3 presents a comparison of RMSE and R square from training and testing samples in 2019. To provide a benchmark for the explanatory power of the obtained models, we compare R square and RMSE with the results of linear regression including the full set of binary variables for the testing and training samples. On average, a boosted tree specification with the learning rate 0.05 has a marginally higher predictive power than linear regression. However, the improvement in goodness of fit by using boosted trees is marginal. The benefit of the current exercise is in allowing all possible interaction terms in detecting the importance of circumstance factors and, therefore, restricting unobserved component biases.

The two main sets of results are obtained from the boosted tree analysis: first, the decomposition of the overall hourly wage inequality into inequality due to exogenous circumstances and inequality due to individual effort; and second, the importance level of factors in predicting the inequality of opportunity, the main focus of our study.

A decomposable measure of inequality (such that the within and between terms sum to

[5] http://uc-r.github.io/gbm_regression

total inequality) should be used to disentangle the effects. A generalized entropy index is an example of such a measure, and we use a *mean logarithmic deviation (MLD)* as a special case of the generalized entropy index with α=0. Checchi and Peragine (2010) show that the MLD allows for the presentation of total income inequality as a sum of income inequality due to effort (effort inequality) and income inequality due to circumstances (inequality of opportunity). The inequality of hourly wage due to opportunities is estimated using the predicted wages from the estimated tree regressions. Total inequality is estimated from the initial hourly wage variables. Therefore, effort inequality can be estimated as a difference between the two.

Circumstance importance is determined by calculating the relative influence of each circumstance: whether it was selected to split during the tree building process, and how much the squared error (over all trees) improved (decreased) as a result[6]. The importance of circumstances within a country are normalised to sum up to one. To create clusters of countries by similarity in importance of circumstances driving the inequality of opportunity, we use a k-means approach.

6 Results

Our estimations exploring hourly wages show that on average the inequality of opportunity in Europe did not change in 2019 as compared to 2011 (Figure A7 and Figure A8) and is in line with the estimates by Checchi and Peragine (2010), Carranza (2022), and Commission et al. (2023), who also use MLD decomposition of income inequality. The share of inequality of opportunity in inequality of income is heterogeneous between different European countries and ranges between 6%-18% in 2019 (on average 11% in both 2011 and 2019).

The average importance of factors explaining inequality of opportunity in Europe did not change between 2019 and 2011 (Figure 3). The five main childhood factors that influence current income are the highest education obtained by parents, financial situation of the household, parents' country of birth, gender, and degree of urbanisation.[7] The ranking of these factors differs between countries (Figure 4).

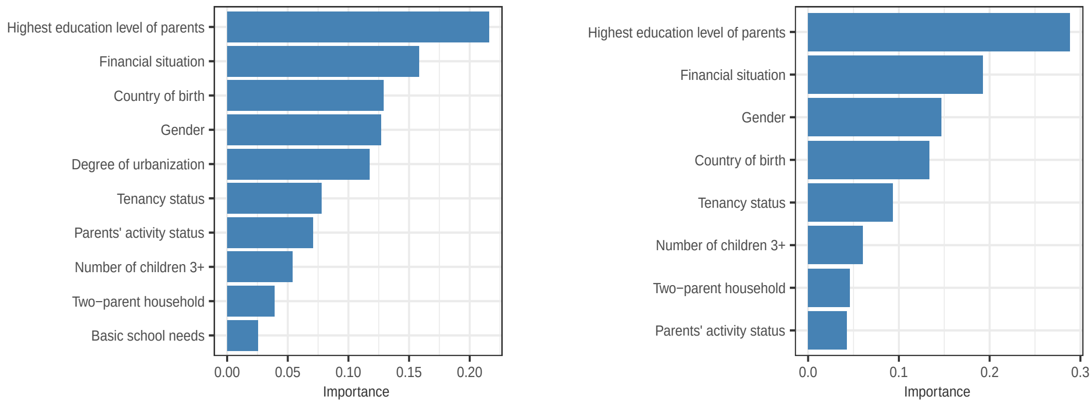
In the majority of European countries, the main circumstance associated with inequality of income later in life is the education level of parents (Figure 4). The financial situation of the household during the respondent's childhood has a lower importance than the education level of parents on average. At the same time, both factors are quite correlated, and there is a large

[6] https://h2o-release.s3.amazonaws.com/h2o/rel-yau/3/docs-website/h2o-docs/variable-importance.html

[7] The question on degree of urbanisation or basic school needs is available only in the EU-SILC 2019 module, thus these variables are not present in the regressions for 2011.

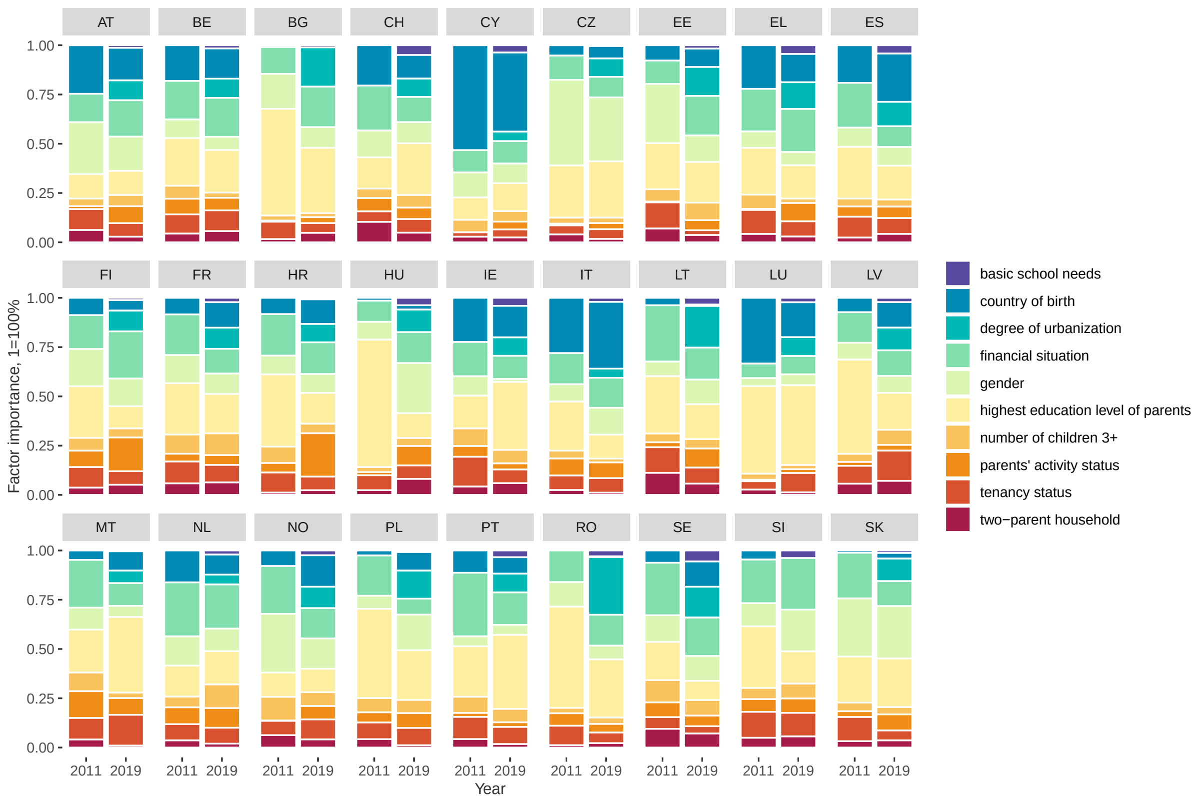
Figure 3: Importance of circumstances

(a) 2019 (b) 2011



Note: the figure shows the mean over all country models for the corresponding year.
Source: authors estimations using the EU-SILC 2019 and EU-SILC 2011 database.

Figure 4: Importance of circumstances by country

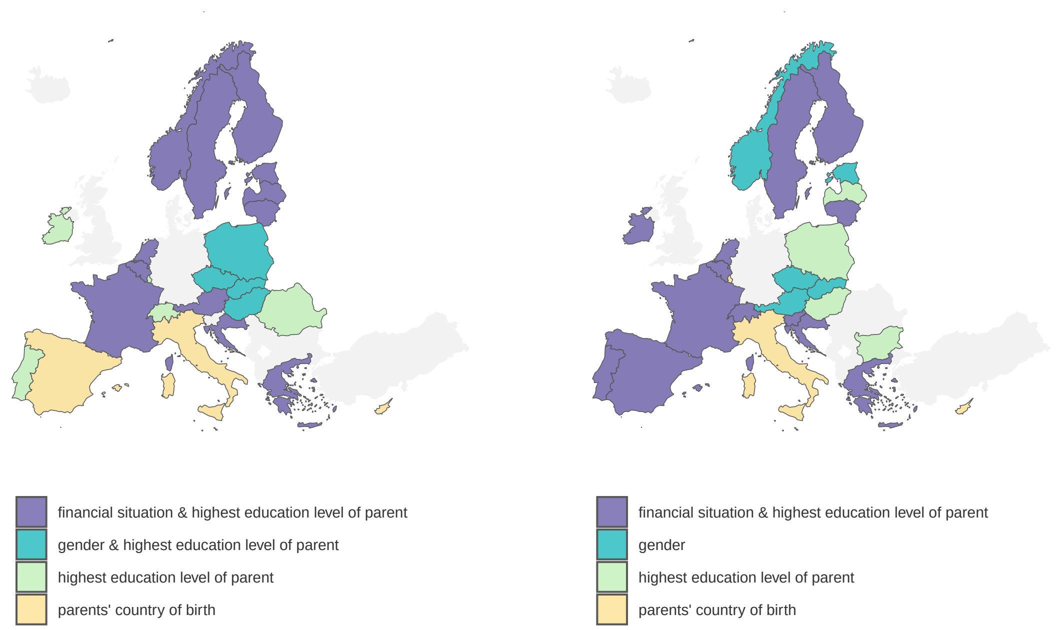


Note: Information on country of birth for Bulgaria and Slovenia is not available for 2019 in the database, and for Romania in 2011. Also, information regarding the degree of urbanisation in Slovenia is not available.
Source: authors' estimations using the EU-SILC 2019 and EU-SILC 2011 databases.

group of countries where both factors are of the highest importance in determining inequality of opportunity (Figure 5). However, there are countries where other circumstances hold equal or even greater importance. For example, in 2019, gender is one of the main factors determining the inequality of opportunities in Czech Republic, Hungary, Slovenia, and Slovakia. Country of birth has a very high impact in Cyprus, Italy, Luxembourg and Spain, which most probably can be explained by the high immigration flows to these countries.

Figure 5: Regions by importance of circumstance

(a) 2019 (b) 2011



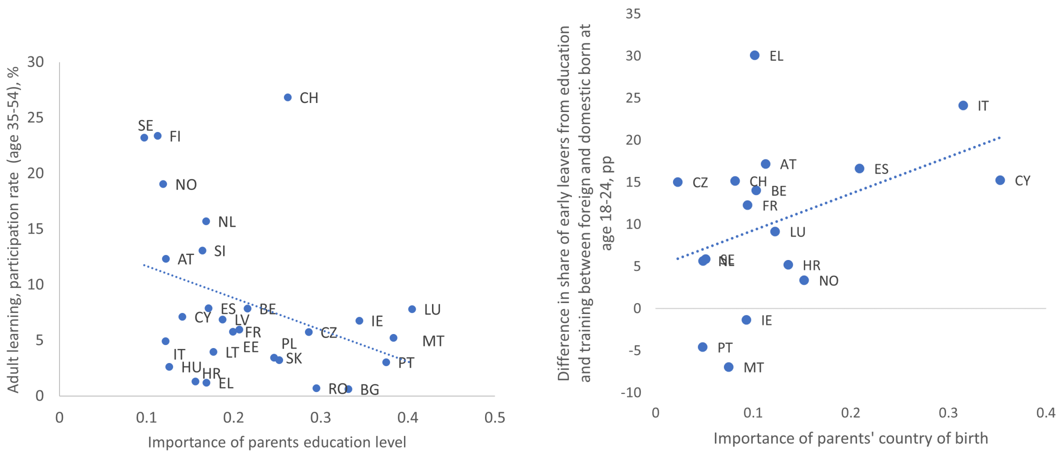
Note: clusters of regions are defined using k-mean analysis for 4 clusters. Cluster names indicate the circumstance with the largest mean importance level for the defined country group. The mean values of circumstance importance for 2011 and 2019 are presented in Table A4.
Source: authors' estimations using the EU-SILC 2019 and EU-SILC 2011 databases.

The introduction of the new explanatory variable *degree of urbanisation* in the 2019 database lowered the importance of other factors in explaining inequality of opportunity, thus, signalling important discrepancies in regional inequality (the effect was particularly strong in Bulgaria and Romania). One of the drivers of rural-urban discrepancies driving inequality of opportunities could be quality of school education. As shown by Echazarra and Radinger (2019) based on 2015 PISA results, students in city schools score higher in science than students in rural schools. This difference is roughly equivalent to one year of schooling. Furthermore, they show that the rural gap is even more visible in students' transitions to higher levels of education, while a much lower number of children from rural areas are expected to complete a university degree.

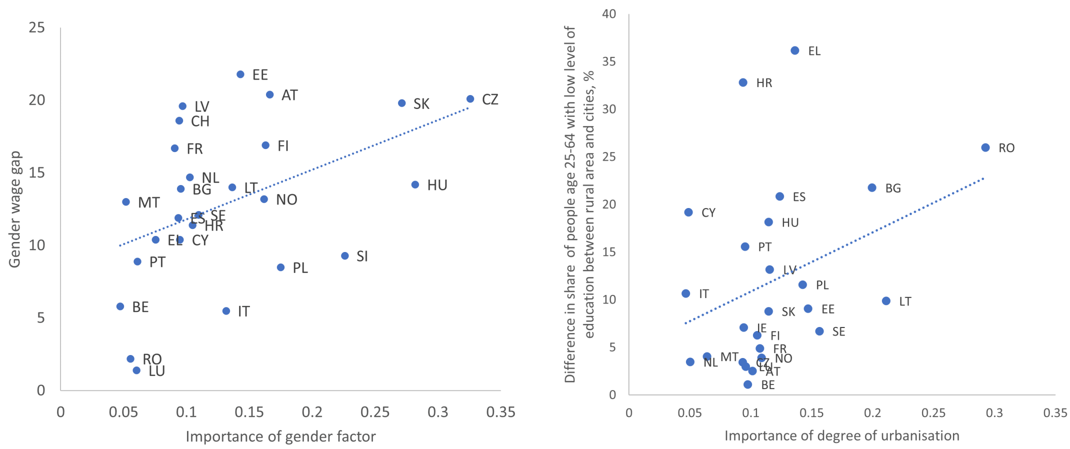
Another observation regarding the ranking of factors is that the importance of the activity status of parents became larger in 2019 and, on average, outranked the importance of two-parent households and many-sibling households (Figure 3). This may signal that it is generally becoming easier for a single parent household and household with many children in Europe to provide an adequate level of education and support for children, thus ensuring better opportunities in the future.

Figure 6: Importance of circumstance factor in 2019 and social and education indicators in 2004-2008

(a) Parents' education and country of birth



(b) Gender and urbanisation degree



Note: authors' calculations using the 2019 EU-SILC survey, Eurostat. Periods for the data sets: adult learning participation (ages 35-54) - 2004-2008; gender wage gap - 2018; difference in share of early leavers from education and training between foreign and domestic born from age 18-24 are average levels of 2004-2008; difference in share of people age 25-64 with less than primary, primary, and lower secondary education between rural areas and cities are the average levels of 2004-2008.
Source: authors estimations using the EU-SILC 2019 and EUROSTAT tables $TRNG\_LFSE_01$, $edat_lfs_9913$, sdg_05_20, $edat\_lfse\_02$.

Next, we look at the association between the importance score of the main factors defining the inequality of opportunity in 2019 and various policy indicators. As was shown in Table 1, in order to see the effect of policy on the respondent's income from the age of 25-40 in 2019. an analysis of policies in place during 1993-2008 is required. Limited availability of the statistical data describing social policies or government support measures at the time when participants of the survey were 14 years old prevents us from doing a similar analysis for the 2011 survey. Thus, we first restrict ourselves to correlations and use the obtained importance scores of factors by countries from 2019 and the average values for policies measures during the period of respondents' childhood.

Figure 6 plots the importance score in 2019 against the historical social and education indicators. We show that government focus on education policies might be of primary importance in determining if parents' education or country of birth drives inequality of opportunity in a selected country. In Figure 6a(figure on the left), we show that there is a negative association between participation rate in adult learning of parents during 2004-2008 and importance score of parents' education level in 2019. Also, if the share of early leavers from education and training (2004-2008) does not differ much between foreign and domestic born children from age 18-24, the importance score of parents' country of birth in 2019 tends to be lower (figure on the right). The importance of degree of urbanisation in 2019 positively correlates with a larger difference in the share of people aged 25-64 with a low education level between rural areas and cities during 2004-2008. These three examples show that policies supporting education and life-long learning for all might help reduce inequality of opportunity. Figure 6b (figure on the left) shows that the current wage gap is strongly associated with the importance score of the gender factor. This draws attention to the importance of the gender factor in driving inequality of opportunity in the countries where the gender wage gap is high.

Finally, to compare the effect of change in policy on change in level of inequality of opportunity between 2011 and 2019, we shrink the sample even further (to age 25-30, see Figure 1). We want to measure whether the change in government spending on educational and social policies when a person was 14 years old are correlated with a change in inequality of opportunity at the age 25-30. The inequality measure can be estimated at the country level, as we did at the first part of the result section, or it can be estimated for a group of respondents (measuring a within group inequality of opportunity). To create comparable groups between the countries, we use 3 criteria - gender, financial situation in the household at the age 14, highest education level of parents. To ensure that the number of observations in each group for each country is large enough, we combined the previously used groups describing childhood circumstances into larger ones - financial situation of a household (1. bad and moderate, 2. good), parents' education level (1. low and medium, 2. high). For each group in each country, we used the predicted hourly wage to estimate the weighted mean log deviation measure of inequality. This way, for each country and survey year, we estimate 8 within group inequality of opportunity measures. Next, assuming that groups are comparable between the survey waves, we estimate a change in the inequality of opportunity measures between 2019 and 2011 for each group by country.

Next, we estimated change in policy measures between the periods 2000-1995 and 2008-2003

Table 2: Change in within group inequality of opportunity (between 2019 and 2011) as a result of policies in respondents' childhood (between 2003-2008 and 1995-2000)

	(1)	(2)	(3)	(4)
Change in PISA reading score	-0.0217	-0.0281*	-0.0303**	-0.0303**
	(-1.56)	(-1.90)	(-2.08)	(-2.08)
Change in government expenses on family and children, % from total expenses	-0.207**	-0.212**	-0.134	-0.134
	(-2.42)	(-2.48)	(-1.07)	(-1.07)
Change in government expenses on education, % from total expenses	0.191**	0.162**	0.172**	0.172**
	(2.61)	(2.21)	(2.27)	(2.27)
Change in government effectiveness	0.324**	0.673***	0.894***	0.894***
	(2.29)	(2.77)	(3.12)	(3.12)
Change in expenses on education × Change in government effectiveness		-0.930*	-1.204**	-1.204**
		(-1.76)	(-2.26)	(-2.26)
Change in expenses on family and children × Change in government effectiveness			-0.465	-0.465
			(-1.18)	(-1.18)
Constant	-0.142	-0.128	-0.133	-0.133
	(-1.51)	(-1.46)	(-1.44)	(-1.44)
Group fixed effects	yes	yes	yes	yes
Number of observations	120	120	120	120
Adjusted R^2	0.151	0.166	0.166	0.166

Notes: t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01
Source: authors' estimations. See Table A5 for a detailed description of the variables used in regressions.

(see Table A5 for a detailed variable description). Since government support for educational policies and family support policies can directly limit the negative effect of bad childhood circumstances by providing equal education opportunities and extra material support for those in need, we decided to focus on these policies. Additionally, we look at the change in PISA reading scores by countries in 2000 and 2009, which allows for the control of quality of education and measures the ability of 15-year-olds to use their reading.[8] We also control for changes in the worldwide governance indicators (Kaufmann et al., 2010), in particular government effectiveness, which reflects perceptions of the quality of public services, the quality of the civil service, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.

The regression coefficients are presented in Table 2. We show that while there is a positive association between the change in inequality of opportunity and change in quality of education (as measured by PISA scores). The effect of change in government expenses on education and family support (as a share of overall government expenses) is not clear cut. The effect is strong and significant if accompanied with improvement in governance effectiveness. Thus, policy measures should strive not only towards improvement in monetary indicators, but also

[8]Unfortunately, PISA scores for math and science are only available from 2003. Thus, we decided to focus exclusively on reading scores.

towards a wider range of governance action improving the overall quality of institutions and policy making.

7 Conclusions

The equality of opportunity concept seeks to offset differences in outcomes attributable to luck, but not those differences in outcomes for which individuals are responsible (Roemer and Trannoy, 2016). Economic policy needs to act beyond the simple redistribution of income. Contributing to a fairer society, it should strive to provide the same development opportunities for all individuals. Understanding the share of inequality of opportunity in the overall inequality of income and identifying its sources is central to developing sustainable social and economic policies.

In this paper, we have explored the underlying factors of inequality of opportunities in Europe using the module on intergenerational transmission of disadvantages from the EU-SILC surveys in 2011 and 2019. Boosted tree regression models are applied to overcome the estimation biases

We show that a significant part of hourly wage income inequality (around 11%) stems from unequal circumstances in childhood. Factors such as parents' education, the financial situation of a household, country of birth, gender, and degree of urbanisation when a respondent was 14 years old are of the highest importance in determining hourly wage income in the future. The average importance and composition of circumstances vary significantly among countries, allowing for the identification of country specific policies needed to reduce inequality of opportunity.

In the majority of European countries, the main circumstance associated with inequality of income later in life is the education level of parents and the financial situation of a household. However, there are countries where other circumstances are of equal or even greater importance. For example, gender is one of the main factors in determining inequality of opportunities in Central Europe, and country of birth for parents in Southern Europe.

It is vital that government policy target vulnerable groups in order to close the opportunity gap for these groups in the future. We show that the importance score of various factors is driven by the extent to which policies tackle the issues at the time of the respondents' childhoods. We show that if the share of early leavers from youth education for people born outside the country is lower and the share of people in adult learning is higher - the importance attributed to the parents' education or country of birth factors is lower. Reducing the wage gap and providing access to high quality education in rural regions reduces the importance ranking of gender and urbanisation factors in determining the inequality of opportunity. We also show that a reduction in inequality of opportunity between 2011 and 2019 is associated with an improvement in education quality and wider social support coupled with improvement in governance effectiveness. Therefore, investment in education, adult learning programmes, support programmes, gender equality, and social integration policies together with a wider range of governance action improving the overall quality of institutions and policy making today will enhance equal opportunities for individual development in the future.

References

Aiyar, S. and C. Ebeke (2020). Inequality of opportunity, inequality of income and economic growth. *World Development 136*(C).

Andreoli, F., A. Fusco, I. Kyzyma, and P. Van Kerm (2021). New estimates of inequality of opportunity across European cohorts (and some insights on the long-term impact of educational policy). Working paper, International Association for Research in Income and Wealth.

Black, S. E., P. J. Devereux, P. Lundborg, and K. Majlesi (2019, February). Understanding Intergenerational Mobility: The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviors. Working Papers 201904, School of Economics, University College Dublin.

Bourguignon, F., F. Ferreira, and M. Walton (2007, August). Equity, efficiency and inequality traps: A research agenda. *The Journal of Economic Inequality 5*(2), 235–256.

Brunori, P., F. H. G. Ferreira, and V. Peragine (2013, January). Inequality of opportunity, income inequality and economic mobility: Some international comparisons. IZA Discussion Papers 7155, Institute of Labor Economics (IZA).

Brunori, P., P. Hufe, and D. G. Mahler (2021, August). The roots of inequality: Estimating inequality of opportunity from regression trees and forests. IZA Discussion Papers 14689, Institute of Labor Economics (IZA).

Carranza, R. (2022, October). Upper and lower bound estimates of inequality of opportunity: A cross-national comparison for Europe. *Review of Income and Wealth.*

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

Checchi, D., V. Peragine, and L. Serlenga (2010, Jun). Fair and unfair income inequalities in Europe. Working Papers 06_10, Centre for Household, Income, Labour and Demographic economics (CHILD).

Checchi, D., V. Peragine, and L. Serlenga (2016, November). Inequality of opportunity in Europe: Is there a role for institutions? In L. Cappellari, S. W. Polachek, and K. Tatsiramos (Eds.), *Inequality: Causes and Consequences*, Volume 43 of *Research in Labor Economics*, pp. 1–44. Emerald Publishing Ltd.

Chen, T. and C. Guestrin (2016, August). Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International conference on knowledge discovery and data mining, KDD'16.

Commission, E., S. A. Directorate-General for Employment, Inclusion, S. Filauro, F. Palmisano, and V. Peragine (2023). *The evolution of inequality of opportunity in the EU*. Publications Office of the European Union.

Echazarra, A. and T. Radinger (2019). Learning in rural schools. (196).

Ferreira, F. H. G. and J. Gignoux (2008, July). The measurement of inequality of opportunity: theory and an application to Latin America. Policy Research Working Paper Series 4659, The World Bank.

Ferreira, F. H. G. and M. Walton (2006, January). Inequality of opportunity and economic development. Policy Research Working Paper Series 3816, The World Bank.

Fleurbaey, M. and V. Peragine (2013, January). Ex-ante versus ex-post equality of opportunity. *Economica 80*(317), 118–130.

Fleurbaey, M., V. Peragine, and X. Ramos (2015, October). Ex post inequality of opportunity comparisons. IZA Discussion Papers 9407, Institute of Labor Economics (IZA).

Han, S. (2022, December). Identifying the roots of inequality of opportunity in South Korea by application of algorithmic approaches. *Palgrave Communications 9*(1), 1–10.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2010, September). The worldwide governance indicators: methodology and analytical issues. Policy Research Working Paper Series 5430, The World Bank.

Marrero, G. A. and J. G. Rodríguez (2010). Inequality of opportunity and growth. Working Papers 154, Society for the Study of Economic Inequality (ECINEQ).

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

Mogstad, M. and G. Torsvik (2021, May). Family Background, Neighborhoods and Intergenerational Mobility. NBER Working Papers 28874, National Bureau of Economic Research, Inc.

Ramos, X. and D. Van de Gaer (2017, November). Is inequality of opportunity robust to the measurement approach? Working Papers 450, ECINEQ, Society for the Study of Economic Inequality.

Roemer, J. E. (1998). *Equality of Opportunity.*

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

Saavedra-Chanduví, J., J. R. Molinas, R. P. De Barros, and F. H. G. Ferreira (2011). *Measuring Inequality of Opportunities in Latin America and the Caribbean.* Number 361 in IDB Publications (Books). Inter-American Development Bank.

Sentís, M. C., S. de la Rica, and L. Gorjón (2023). Opportunity bias in Spain: Empirical evidence, drivers and trends. Working Papers 2023/1, ISEAK.

Appendix

Figure A1: Inequality of median wage income by household income situation at age 14 and income decile

Notes: the set of individuals includes people in the age category 26-40, who were employed for at least 12 months. Gross employee cash or near cash income (wage) is adjusted by hours worked per week. Income deciles are estimated separately for each category of the financial situation of the household when the respondent was approximately 14 years old. Each point represents a ratio between the median wage in the particular group of individuals and country median.
Source: authors' estimations using the EU-SILC 2019 and 2011 databases.

Figure A2: Inequality of median wage income by highest education level of parents and income decile

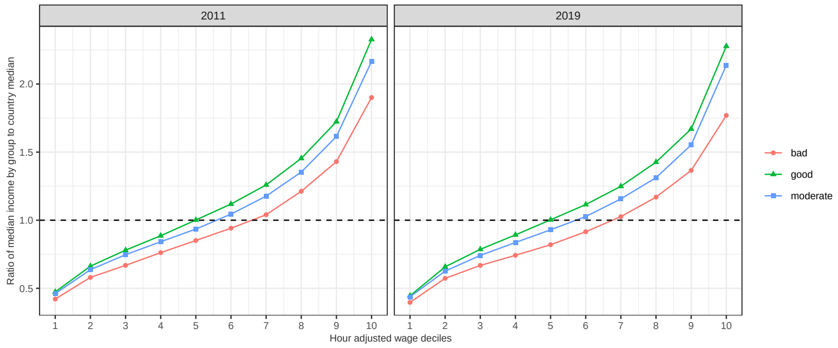
Notes: tthe set of individuals includes people in the age category 26-40, who were employed for at least 12 months. Gross employee cash or near cash income (wage) is adjusted by hours worked per week. Income deciles are estimated separately for each category of the parents' highest education level. Each point represents a ratio between the median wage in the particular category of individuals and country median.
Source: authors' estimations using the EU-SILC 2019 and 2011 databases.

Figure A3: Inequality of median wage income by gender and income decile

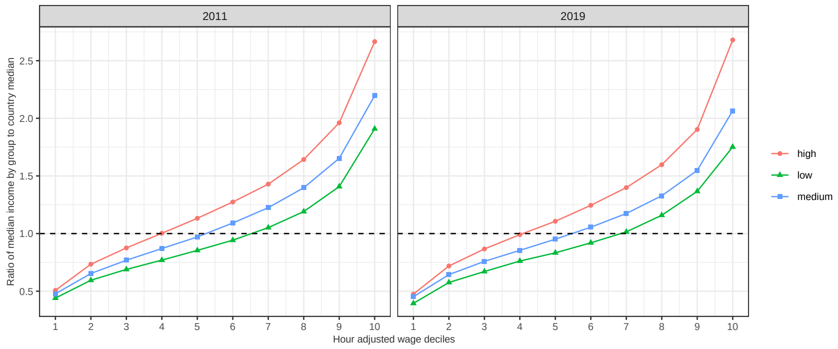
Notes: the set of individuals includes people in the age category 26-40, who were employed for at least 12 months. Gross employee cash or near cash income (wage) is adjusted by hours worked per week. Income deciles are estimated separately for each gender category. Each point represents a ratio between the median wage in the particular group of individuals and country median.
Source: authors' estimations using the EU-SILC 2019 and 2011 databases.

Figure A4: Inequality of median wage income by income decile in 2011 and 2019

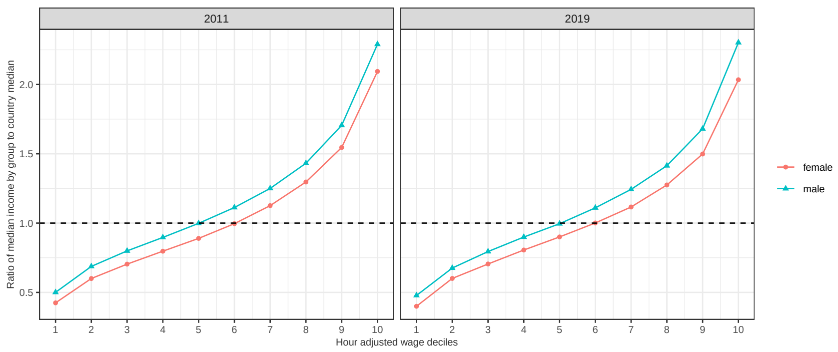
(a) by household income level at age 14



(b) by parents' education level



(c) by gender



Notes: the set of individuals includes people in the age category 26-40, who were employed for at least 12 months. Gross employee cash or near cash income (wage) is adjusted by hours worked per week. Income deciles are estimated separately for each country and category. Each point represents a simple average over 29 countries. Country ratios are obtained as median wage in the particular group of individuals (by household financial situation during childhood and income decile) to country median.
Source: authors' estimations using the EU-SILC 2019 and 2011 databases.

Table A1: Questions used from the module on intergenerational transmission of disadvantages

Variable	Questions
rb090	Sex
	(1) Male
	(2) Female
pt030_group	Number of children when respondent was roughly 14 years old
	(1) 1-2
	(2) 3 or more
pt190_group	Financial situation of the household when respondent was around 14 years old
	(1) Bad (very bad, bad)
	(2) Moderate (moderately bad, moderately good)
	(3) Good (good, very good)
pt210	When you were around 14 years old, the dwelling you lived in
	(1) Owned
	(2) Rented
	(3) Accommodation was provided for free
pt250	Degree of urbanisation when respondent was around 14 years old
	(1) City (more than 100 000 inhabitants)
	(2) Town or suburb (10 000 to 100 000 inhabitants)
	(3) Rural area, small town or village (less than 10 000 inhabitants)
pt260	Basic school needs met when respondent was around 14 years old
	(1) Yes
	(2) No - due to financial reasons
	(3) No - other reason
max_pt110_pt120	Highest education attained by the father/mother when respondent was around 14 years old
	(1) Low level
	(2) Medium level
	(3) High level
min_pt130_pt160	Activity status of the father/mother when respondent was around 14 years old
	(1) Employed (full or part-time)
	(2) Self-employed
	(3) Unemployed
	(4) Retirement
	(5) Other
min_pt060_pt090	Country of birth of the father/mother
	(1) Born in the respondent's present country of residence
	(2) Born in another EU-27 country
	(3) Born in another European country
	(4) Born outside Europe
max_pt024_pt023	Presence of both parents in the family
pt010 (for 2011)	(1) Yes
	(2) No

Notes: Variable max_pt110_pt120 shows the highest level of parents' education (for example, 3 - at least one parent has a high level of education). Variable min_pt130_pt160 shows the smallest value of parents' activity (for example, 1- at least one parent works full time). Variable min_pt060_pt090 shows the smallest value from the classification of country of birth used in the 2011 questionnaire (for example, 1 - at least one of the parents was born in the country of the survey).
Source: the EU-SILC questionnaire and authors' assumptions

Table A2: Model setting parameters

	2019									2011		
	trees			tree depth			sample size			trees	tree depth	sample size
learning rate	0.05	0.1	0.3	0.05	0.1	0.3	0.05	0.1	0.3	0.05	0.05	0.05
AT	79	32	12	3	3	3	0.56	0.58	0.60	79	3	0.68
BE	84	39	14	3	3	3	0.72	0.67	0.74	79	3	0.56
BG	69	29	9	3	3	3	0.56	0.57	0.72	63	3	0.59
CH	79	44	14	3	3	3	0.59	0.74	0.75	74	3	0.59
CY	74	32	12	3	3	3	0.73	0.60	0.67	74	3	0.74
CZ	79	42	12	3	3	3	0.67	0.75	0.62	79	3	0.71
EE	100	50	12	3	3	3	0.56	0.58	0.66	69	3	0.66
EL	89	50	12	3	3	3	0.73	0.74	0.73	69	3	0.64
ES	89	37	12	3	3	3	0.56	0.64	0.70	79	3	0.56
FI	79	39	9	3	3	3	0.60	0.67	0.68	74	3	0.70
FR	89	50	14	3	3	3	0.74	0.71	0.71	79	3	0.58
HR	74	39	12	3	3	3	0.74	0.69	0.70	89	3	0.63
HU	74	34	9	3	3	3	0.75	0.59	0.59	89	3	0.72
IE	63	32	9	3	3	3	0.71	0.74	0.69	69	3	0.67
IT	74	37	12	3	3	3	0.72	0.62	0.64	84	3	0.63
LT	69	39	9	3	3	3	0.61	0.62	0.67	84	3	0.73
LU	69	34	12	3	3	3	0.74	0.67	0.69	84	3	0.64
LV	63	37	9	3	3	3	0.57	0.58	0.68	63	3	0.63
MT	74	37	12	3	3	3	0.74	0.71	0.64	84	3	0.55
NL	79	37	14	3	3	3	0.75	0.59	0.58	79	3	0.57
NO	69	34	12	3	3	3	0.67	0.64	0.69	69	3	0.75
PL	84	42	12	3	3	3	0.63	0.66	0.62	74	3	0.57
PT	79	42	12	3	3	3	0.64	0.69	0.58	74	3	0.55
RO	84	37	12	3	4	3	0.56	0.67	0.62	74	3	0.66
SE	69	39	9	3	3	3	0.67	0.56	0.72	63	3	0.70
SI	79	34	17	3	3	3	0.61	0.67	0.67	74	3	0.72
SK	79	37	12	3	3	3	0.66	0.62	0.55	79	3	0.61

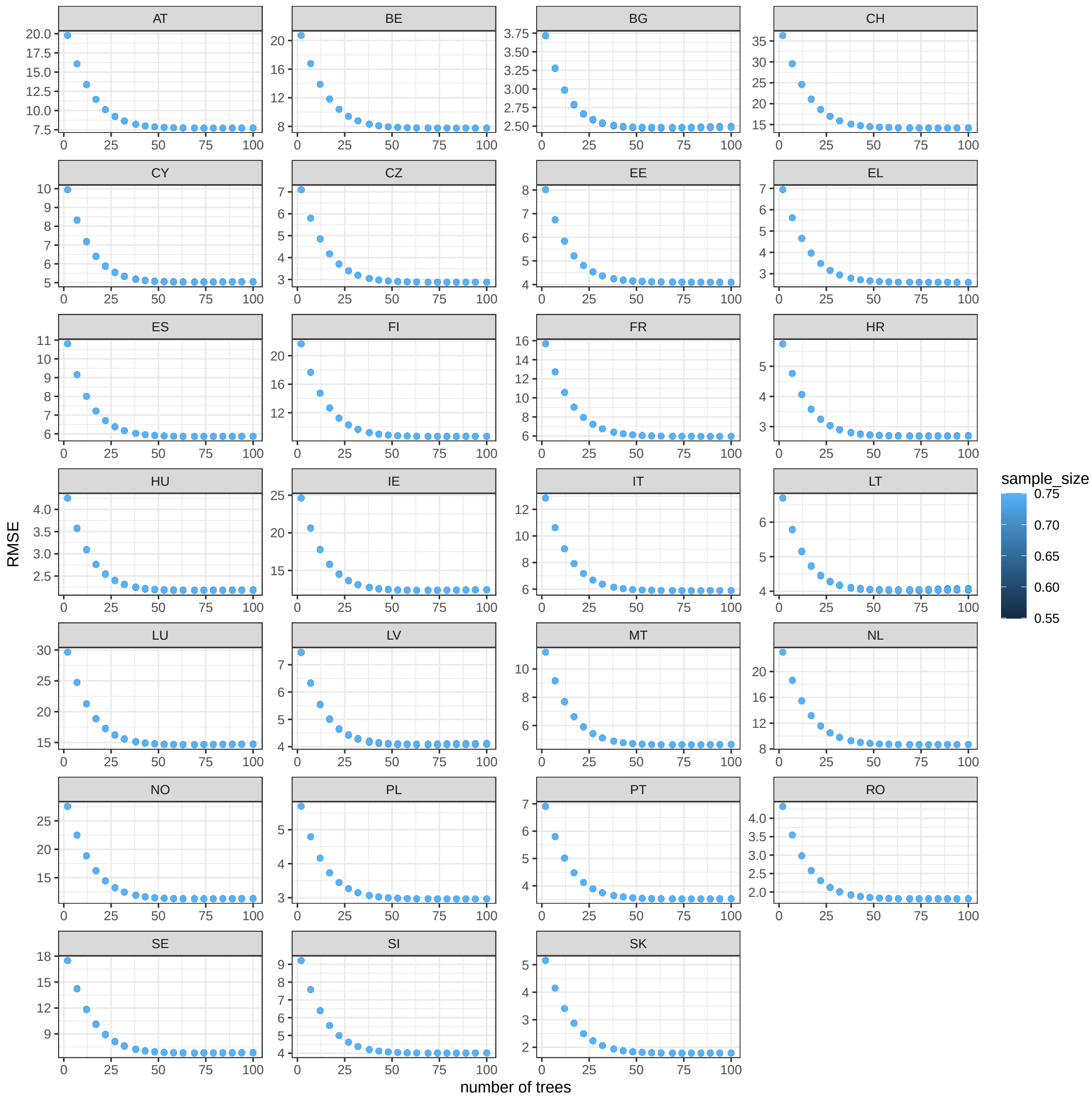
Notes: boosted tree model with binary variable specification.
Source: authors' estimations using the EU-SILC 2019 and 2011 databases.

Table A3: Goodness of model fit for training and testing samples, 2019

	Linear binary regression				Boosted regression (learning rate 0.05)			
	RMSE		R square		RMSE		R square	
	training	testing	training	testing	training	testing	training	testing
AT	7.64	7.48	0.086	0.054	**7.38**	**7.44**	**0.159**	**0.058**
BE	7.63	7.64	0.050	0.060	**7.36**	7.68	**0.135**	0.051
BG	2.46	2.38	0.135	0.131	**2.41**	2.40	**0.177**	0.117
CH	13.99	13.83	0.067	0.037	**13.50**	13.87	**0.145**	0.034
CY	4.94	4.52	0.111	0.099	**4.80**	4.57	**0.173**	0.079
CZ	2.81	2.78	0.147	0.117	**2.76**	2.80	**0.185**	0.106
EE	4.06	4.37	0.099	0.051	**3.91**	4.41	**0.172**	0.037
EL	2.58	2.66	0.059	0.074	**2.52**	2.66	**0.112**	**0.076**
ES	5.83	5.52	0.070	0.069	**5.70**	5.53	**0.115**	0.065
FI	8.56	8.67	0.053	0.025	**8.23**	**8.63**	**0.143**	**0.035**
FR	5.97	6.11	0.057	0.050	**5.70**	6.14	**0.153**	0.045
HR	2.63	2.43	0.058	0.087	**2.53**	**2.41**	**0.156**	**0.103**
HU	2.12	2.35	0.104	0.047	**2.01**	2.38	**0.220**	0.032
IE	12.06	11.81	0.112	0.023	**11.68**	**11.40**	**0.184**	**0.054**
IT	5.83	5.92	0.077	0.054	**5.76**	5.94	**0.106**	0.046
LT	3.97	3.21	0.135	0.095	**3.82**	**3.13**	**0.214**	**0.110**
LU	14.54	15.52	0.111	0.125	**14.00**	15.85	**0.187**	0.097
LV	4.01	4.02	0.098	0.122	**3.84**	**3.99**	**0.197**	**0.159**
MT	4.61	4.63	0.118	0.092	**4.40**	**4.58**	**0.207**	**0.118**
NL	8.55	7.67	0.049	0.008	**8.22**	**7.51**	**0.143**	**0.023**
NO	11.08	10.54	0.061	0.029	**10.63**	**10.48**	**0.163**	**0.035**
PL	2.96	2.97	0.066	0.074	**2.89**	**2.96**	**0.117**	**0.084**
PT	3.51	3.47	0.070	0.057	**3.42**	**3.45**	**0.122**	**0.068**
RO	1.79	1.84	0.170	0.118	**1.74**	1.84	**0.220**	0.115
SE	6.59	6.48	0.049	0.017	**6.27**	6.54	**0.177**	**0.031**
SI	3.95	4.08	0.049	0.060	**3.87**	4.09	**0.095**	0.059
SK	1.74	1.82	0.132	0.130	**1.69**	1.83	**0.195**	0.128
Average								
	5.79	5.73	0.089	0.070				
learning rate 0.05					**5.59**	**5.72**	**0.162**	**0.073**
learning rate 0.1					**5.61**	5.74		
learning rate 0.3					**5.63**	5.77		

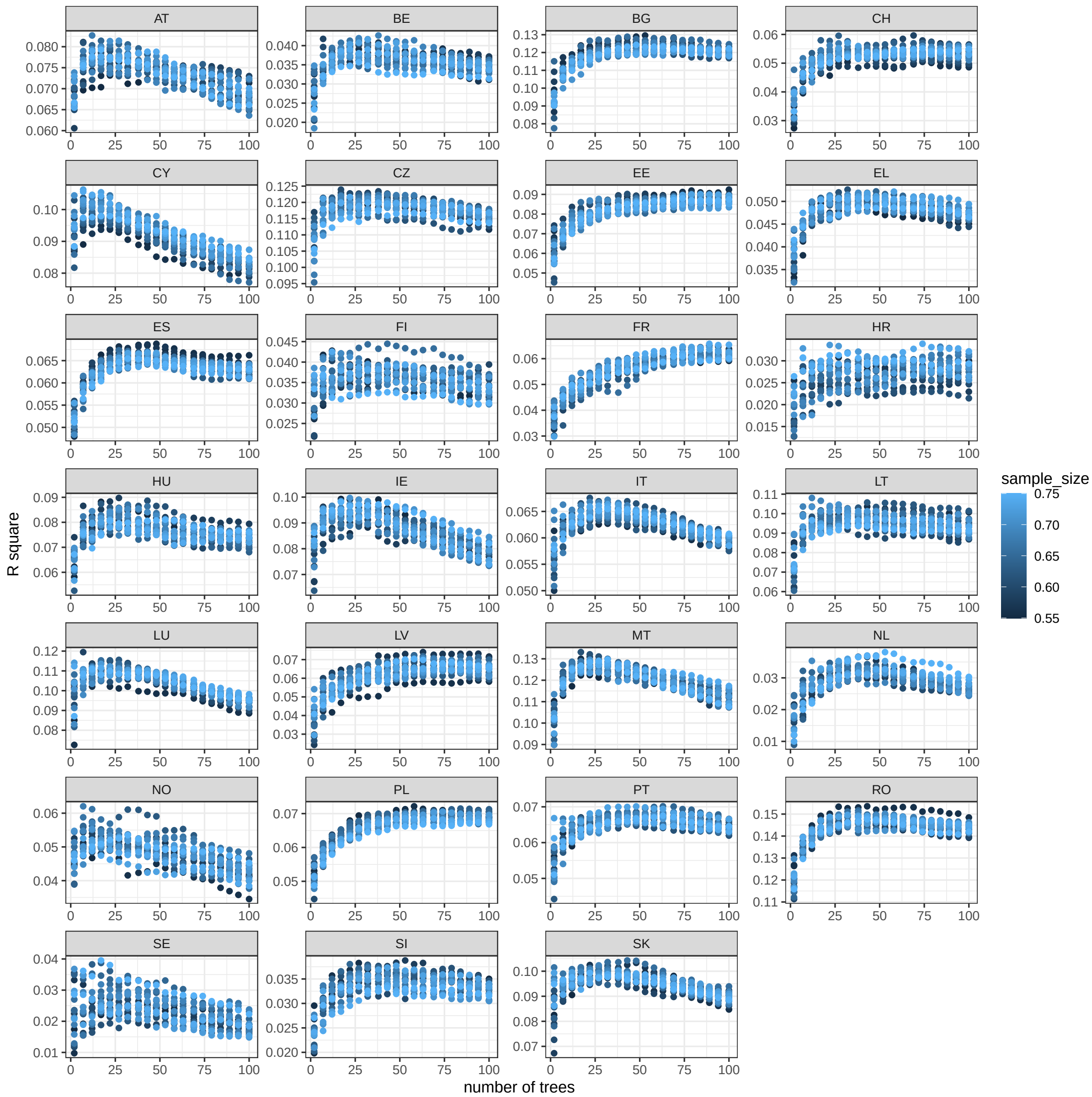
Notes: both the boosted tree model and linear model use binary variable specification. Linear model includes all possible binary variables without interaction terms. The values marked in bold are the ones showing better results of boosted regression models over the linear binary regression.
Source: authors' estimations using the EU-SILC 2019 database.

Figure A5: Tuning of tree size and sample size parameter (RMSE), 2019



Note: The selection of the model was based on the smallest RMSE. Graphs depict the grid over tree size and sample size for a tree depth of 3 and the learning rate 0.05.
Source: authors' estimations using the EU-SILC 2019 database.

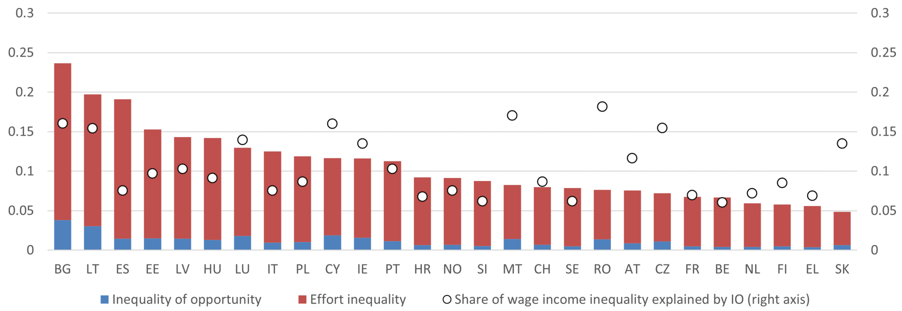
Figure A6: Tuning of tree size and sample size parameter (R square), 2019



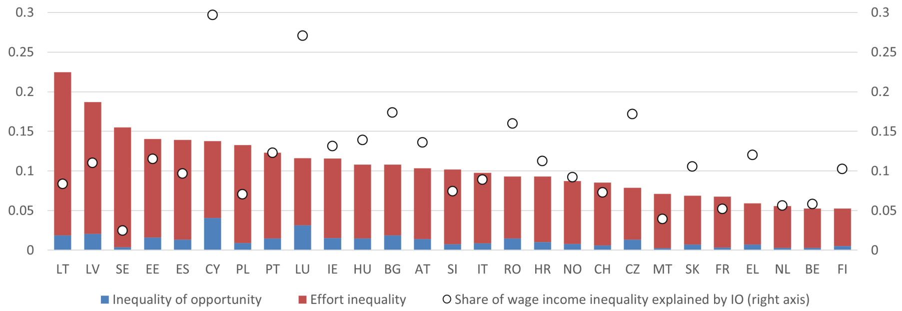
Notes: The selection of the model was based on the smallest RMSE. Graphs depict the grid over tree size and sample size for a tree depth of 3 and the learning rate 0.05.
Source: authors' estimations using the EU-SILC 2019 database.

Figure A7: Inequality decomposition of hourly wage income (mean log deviation)

(a) 2019



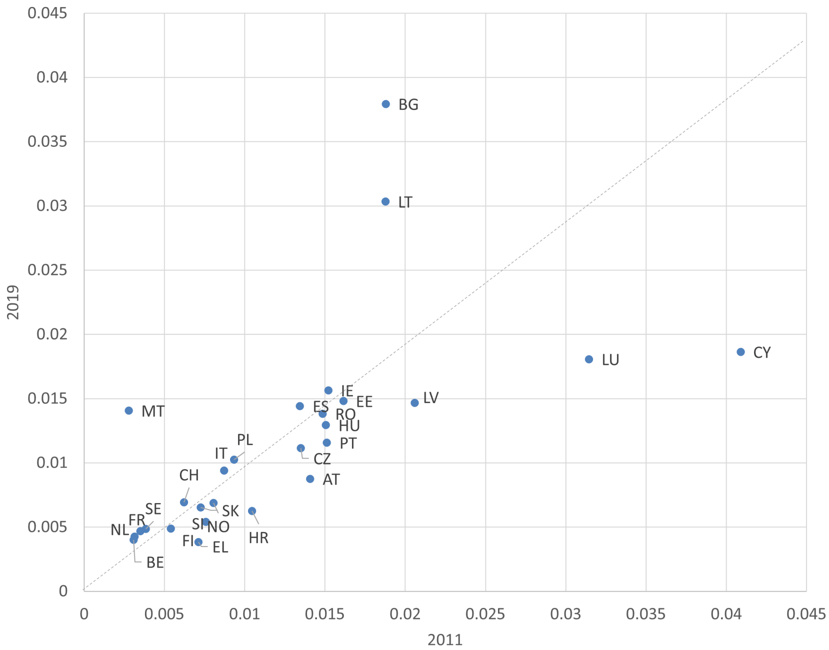
(b) 2011



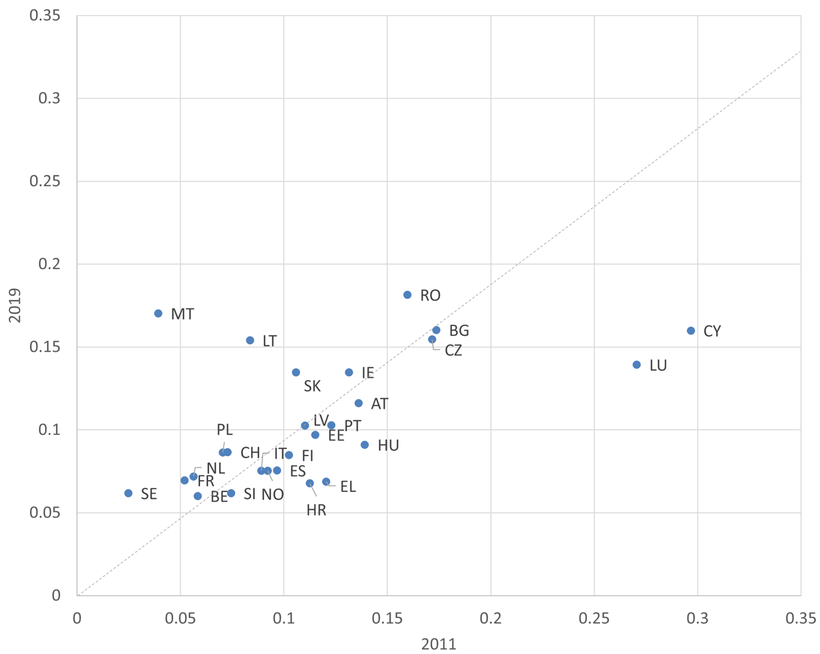
Source: authors' estimations using the EU-SILC 2019 and EU-SILC 2011 databases.

Figure A8: Comparison of inequality of opportunity results, 2011 and 2019

(a) Inequality of opportunity (IO) estimates



(b) Share of hourly wage income inequality explained by IO, 1=100%



Notes: the initial data for this plot is presented in Figure A7. Figure A8a plots IO estimates using EU-SILC 2019 and 2011 databases. Figure A8b plots the share of hourly wage income inequality explained by inequality of opportunity.
Source: authors' estimations using the EU-SILC 2019 and 2011 databases.

Table A4: Mean circumstance importance values of country clusters from k-means analysis

2019	Mean values of circumstance importance by cluster			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Activity status of parents	0.087	0.071	0.059	0.040
Basic school needs	0.025	0.021	0.033	0.025
Country of birth	0.125	0.037	**0.328**	0.129
Degree of urbanisation	0.109	0.161	0.073	0.087
Financial situation	**0.179**	0.131	0.123	0.122
Gender	0.115	**0.203**	0.111	0.045
Highest education level of parents	**0.168**	**0.230**	0.145	**0.377**
Number of children 3+	0.068	0.042	0.037	0.047
Tenancy status	0.079	0.065	0.065	0.102
Two-parent household	0.045	0.038	0.026	0.025
Total	1	1	1	1
Countries	AT, BE, CH, EE, EL, FI, FR, HR, LV, NL, NO, SE	CZ, HU, LT, PL, RO, SK	CY, ES, IT	IE, LU, MT, PT
Within cluster sum of squares by cluster	0.137	0.118	0.023	0.024
Between SS / Total SS	59%			

2011	Mean values of circumstance importance by cluster			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Activity status of parents	0.060	0.012	0.031	0.024
Country of birth	0.125	0.094	**0.382**	0.030
Financial situation	**0.231**	0.172	0.115	0.151
Gender	0.113	**0.318**	0.085	0.105
Highest education level of parents	**0.241**	0.196	**0.269**	**0.529**
Number of children 3+	0.072	0.059	0.045	0.041
Tenancy status	0.105	0.095	0.046	0.085
Two-parent household	0.053	0.054	0.027	0.035
Total	1	1	1	1
Countries	BE, CH, EL, ES, FI ,FR, HR, IE, LT, MT, NL, PT, SE, SI	AT, CZ, EE, NO, SK	CY, IT, LU	BG, HU, LV, PL
Within cluster sum of squares by cluster	0.190	0.095	0.104	0.041
Between SS / Total SS	65%			

Notes: K-means using non-standardised data and a pre-defined number of clusters.
Source: authors' estimations using the EU-SILC 2019 and 2011 databases.

Table A5: Variables included in the regression between change in inequality of opportunity and change in policy measures during a respondent's childhood.

Name	Details	Source
Change in within strata inequality of opportunity	Indicator shows the change in inequality of opportunity between 2019 and 2011 by strata estimated from hourly wage for employed people at age 25-30 (percentage points). Inequality of opportunity is estimated as mean log deviations for income determined by circumstances at childhood. A change in inequality of opportunity between 2011 and 2019 estimated as a simple difference between the mean log deviation measures multiplied by 100.	Authors' estimations using EU-SILC 2011 and EU-SILC 2019 microdata
Change in government effectiveness	Indicator shows the change in the government effectiveness indicator between the two periods (1996-2000 and 2003-2008). The value of the indicator for the corresponding periods are estimated as simple averages. The change between periods is an estimate as a difference between the period averages. Government effectiveness reflects perceptions of the quality of public services, the quality of the civil service, and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies. Estimates of governance performance from the WGI database range from approximately -2.5 (weak) to 2.5 (strong).	Worldwide Governance Indicators (WGI)
Change in government expenses on education	Indicator shows the change in share of government expenditure on education from the total expenditure between two periods (1995-2001 and 2003-2008). The value of the indicator for the corresponding periods are estimated as simple averages. The change between periods is an estimate of the difference between the period averages (percentage points).	EUROSTAT database, table GOV_10A_EXP
Change in government expenses on family and children	Indicator shows the change in share of government expenditure on family and children from the total expenditure between two periods (1995-2001 and 2003-2008). The value of the indicator for the corresponding periods are estimated as simple averages. The change between periods is an estimate, of the difference between the period averages (percentage points).	EUROSTAT database, table GOV_10A_EXP
Change in PISA reading score	PISA is the OECD's Programme for International Student Assessment. PISA measures 15-year-olds' ability to use their reading, mathematics, and science knowledge and skills to meet real-life challenges. The indicator shows a change in reading indicator for countries between 2009 and 2000.	OECD database, PISA 2000, PISA 2009