Society for the Study of Economic Inequality

Working Paper Series

Playing the birth lottery in Europe

Annaelena Valentini
Paolo Brunori
Francisco H. G. Ferreira
Pedro Salas-Rojo

ECINEQ 2026 695

ECINEQ 2026 695
2026 June
www.ecineq.org

Playing the birth lottery in Europe

Annaelena Valentini
University of Florence
Paolo Brunori
University of Florence, III-LSE
Francisco H. G. Ferreira
III-LSE
Pedro Salas-Rojo
CUNEF

Abstract

We analyse the extent to which a personâ€™s country of origin -alongside other factors beyond their control, such as their parentsâ€™ education and occupation- are predictive of adult incomes in Europe. Interpreting the joint predictive power of inherited circumstances as a measure of inequality of opportunity, we employ data-driven methods to estimate inequality of opportunity for household disposable incomes, treating Europe as a single entity. To ensure representativeness, we combine data from EUROSTAT and three different household surveys to construct a sample that represents the population of Europe, accounting for country-of-birth population shares within countries. We estimate overall inequality in Europe at 39 Gini points in 2019, with inequality predicted by ascriptive characteristics accounting for a full 23 Gini points. The country where a person was born accounts for 64% of the latter figure, emerging as the most significant predictor compared to other factors such as parental occupation (26%) and parental education (9%). The level of inequality of opportunity observed in Europe as a whole is comparable to that in China and India and significantly higher than estimates for the United States.

Keyword: inequality of opportunity, place of birth, migration, income distribution, Europe.

JEL Cassification: D31, J60, O52, O54

Playing the birth lottery in Europe[1]

Annaelena Valentini[1], Paolo Brunori[1,2], Francisco H. G. Ferreira[3,4], Pedro Salas-Rojo[5,2,6]

[1] Department of Economics and Management, University of Florence, Via delle Pandette 9, 50127 Florence, Italy

[2] International Inequalities Institute, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom

[3] Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom

[4] IZA-LISER, Esch-sur-Alzette, Luxembourg

[5] CUNEF Universidad, Leonardo Prieto Castro 2, 28040 Madrid, Spain

[6] EQUALITAS, Spain

Corresponding author: Paolo Brunori - p.brunori@lse.ac.uk

Date: 28 May 2026

[1] We are grateful to Torsten Hothorn and Achim Zeileis for their kind and useful support. We also thank useful comments from Paul Hufe, Flaviana Palmisano, participants at the 10th ECINEQ meeting (Aix-en-Provence) and the INEQUALITREES workshop (IFO-Munich).

Abstract: We analyse the extent to which a person's country of origin -alongside other factors beyond their control, such as their parents' education and occupation- are predictive of adult incomes in Europe. Interpreting the joint predictive power of inherited circumstances as a measure of inequality of opportunity, we employ data-driven methods to estimate inequality of opportunity for household disposable incomes, treating Europe as a single entity. To ensure representativeness, we combine data from EUROSTAT and three different household surveys to construct a sample that represents the population of Europe, accounting for country-of-birth population shares within countries. We estimate overall inequality in Europe at 39 Gini points in 2019, with inequality predicted by ascriptive characteristics accounting for a full 23 Gini points. The country where a person was born accounts for 64% of the latter figure, emerging as the most significant predictor compared to other factors such as parental occupation (26%) and parental education (9%). The level of inequality of opportunity observed in Europe as a whole is comparable to that in China and India and significantly higher than estimates for the United States.

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

JEL codes: D31, J60, O52, O54

1 Introduction

Equality of opportunity is a central principle in modern welfare states and a cornerstone objective of the European integration project. Above and beyond efficiency gains, European institutions explicitly aim at ensuring that people's life chances are not determined by arbitrary circumstances such as place of birth, nationality, or social origin.[2] Freedom of movement and non-discrimination are meant to promote both fairness and convergence across individuals living within a unified economic space. Whether Europe effectively operates as a single space of equal opportunities, however, remains an open empirical question.

This paper asks the following question: to what extent does country of birth shape individuals' economic opportunities in Europe when the continent is treated as a single unit of analysis?[3] If freedom of movement and institutional integration successfully equalize opportunities, the country of origin should play a limited role in predicting economic outcomes later in life. Conversely, if large opportunity gaps persist across countries of birth, inequality of opportunity may remain substantial at the continental level, even if individual European countries exhibit relatively low levels of inequality of opportunity.

Over the past two decades, a growing empirical literature has documented the extent of inequality of opportunity (IOp) in Europe, focusing primarily on national or sub-national contexts (Marrero and Rodríguez, 2012 and Checchi et al., 2016). These studies typically find that European countries perform relatively well in equalizing opportunities when compared to other regions of the world. More recent contributions have adopted data-driven approaches to estimate IOp within countries and regions, highlighting the importance of non-linearities and interactions among circumstances (Brunori et al., 2023a; Brunori et al., 2023b; Ferreira et al., 2026). Some recent pan-European evidence also

[2] These principles are formally enshrined in the European treaties, including the Treaty of Rome (1957), the Treaty on the Functioning of the European Union (Articles 3, 14, and 45), and the Charter of Fundamental Rights of the European Union (Article 21), which prohibit discrimination based on nationality or social origin and promote social and territorial cohesion.

[3] Throughout the paper we use country of birth and country of origin interchangeably.

abstracts from migration. For instance, Filauro et al. (2023) estimate inequality of opportunity across European countries focusing on native populations, thereby excluding the role of cross-country movements.

While this literature provides valuable insights, it largely abstracts from cross-country mobility and treats nation states as the natural unit of analysis. This assumption is particularly restrictive in the European context. According to recent estimates, more than 63 million individuals born in Europe (8.5% of the continent's population) were living in a country other than their birthplace in 2020, with the majority remaining within the continent (UN DESA, 2020). In the presence of such widespread international migration, focusing exclusively on within-country inequality of opportunity may underestimate the role of birthplace as a determinant of life chances. Europe is not a single country, but the European project does envisage convergence in living standards, and we can shed light on the extent to which that goal is being achieved.

We contribute to the literature by estimating inequality of opportunity at the European level, explicitly accounting for country of origin as a circumstance beyond individual control. Treating Europe as a single opportunity space, we estimate how much of total income inequality is predicted by circumstances determined at birth, including parental education, parental occupation, sex, and -crucially- the country of origin. To assess whether the importance of country of birth reflects selective migration or persistent place-based differences across European economies, we compare these results with estimates obtained using country of residence and all other circumstances. To the best of our knowledge, this is the first study to provide a comprehensive estimate of inequality of opportunity in Europe that incorporates country of origin and migration into the analysis.

Our approach builds on the global inequality of opportunity framework proposed by Milanovic (2015). Milanovic shows that a large share of global income inequality is associated with circumstances beyond individual control, using country of residence as a proxy for background and assuming a negligible role for migration. By focusing on Europe, on the one hand we restrict the focus from global to continental, on the other we are able to relax these assumptions. We explicitly account for migration by using country of origin rather than country of residence, and we expand the set of circumstances to include parental background and sex -dimensions that play a central role in the equality of opportunity framework (Ferreira and Peragine, 2013; Roemer and Trannoy, 2015), which are absent from Milanovic's analysis.

From a methodological perspective, we depart from standard parametric approaches to IOp estimation and adopt a data-driven framework based on Conditional Inference Trees and Random Forests (Hothorn et al., 2006; Brunori et al., 2023b). These methods allow for flexible, non-linear interactions among circumstances and provide a statistically grounded implementation of the concept of *types*, that is, *groups of origin,* central to the ex-ante equality of opportunity approach (Van de Gaer, 1993; Roemer, 1998). To overcome a computational limitation that prevents the estimation of conditional inference regression trees with categorical variables including more than 29 categories, we introduce a modification of the original algorithm that makes regression tree methods computationally feasible when dealing with variables characterized by a large number of unordered categories, such as country of origin (78 categories in our case).

We assemble a unique dataset by combining information from European Union Statistics on Income and Living Conditions (EU-SILC), the German Socio-Economic Panel (GSOEP), the UK Household Longitudinal Study (UKHLS), and population census data (EUROSTAT). Using a stratified sampling procedure, we construct a dataframe that is representative of the European population while preserving the distribution of countries of origin across countries of residence. This dataset includes approximately 150,000 individuals living in 29 European countries and originating from 78 different countries or regions.

Our findings reveal a striking pattern. Overall income inequality in Europe in 2019 amounts to a Gini coefficient of 0.39. Inequality of opportunity -measured as inequality in income predicted by circumstances at birth- reaches a Gini of 0.23 using the random forest algorithm, accounting for almost three-fifths of total inequality. This level is comparable to estimates for China and India and substantially higher than those reported for the United States (see Ferreira et al., 2026). Most notably, country of origin alone accounts for approximately 64% of total inequality of opportunity, making it by far the most important circumstance variable. While individual European countries generally display low levels of inequality of opportunity, aggregating across countries reveals deep disparities in opportunities associated with birthplace.

The remainder of the paper is organized as follows. Section 2 presents the theoretical framework and the methodology used to estimate inequality of opportunity. Section 3 describes the data and the construction of the representative European sample. Section 4 reports the main results, including the decomposition of inequality of opportunity and the role of migration and regional disparities. Section 5 concludes.

2 Theory and Methods

2.1 Inequality of Opportunity

The equality of opportunity model (EOp) proposed by Roemer (1993, 1998) considers a population of individuals indexed by $i \in (1, \dots, N)$, and a variable y measuring an economic outcome of interest, such as income. The outcome distribution is a function of the set of the individual's circumstances -factors beyond her control, c_i- and the amount of effort exerted (e_i) such that:[4]

$$y_i = f(c_i, e_i) \qquad (1)$$

Circumstances are defined as a finite vector of J elements and are, by definition, exogenous to individual choices. Following Roemer, any population can be divided into a set Π of exhaustive and mutually exclusive groups or types, such that individuals belonging to the same type T_k share the same circumstances.

We follow the ex-ante approach to Equality of Opportunity (EOp; Van de Gaer, 1993; Roemer 1998), which has dominated the empirical literature in the past two decades (Ramos and Van de Gaer, 2016). Ex-ante EOp is achieved when the expected outcome is the same across all types:

[4] We ignore the role of luck in this paper. For a discussion, see Lefranc et al. (2009).

$$E(y_i|i \in T_k) = E(y_j|j \in T_l) \quad \forall\, T_k, T_l \in \Pi \tag{2}$$

Consequently, disparities in mean outcomes across types indicate that exogenous circumstances *are* predictive of adult incomes, thus violating the equal opportunity principle. Inequality of opportunity (IOp) measures quantify the extent of this violation. In practice, IOp is estimated by constructing a counterfactual smoothed distribution $\hat{y}_i$, where each individual receives the average outcome of the type she belongs to:

$$\hat{y}_i = E(y_i|i \in T_k), \forall i, k$$

Applying a suitable dispersion measure *I(·)* -such as the Gini index or the Mean Log Deviation- to $\hat{y}_i$ delivers a measure of between-types inequality, that is, inequality attributed to the set of observed circumstances:

$$\mathrm{IOp} = I(\hat{y}_i) \tag{3}$$

Our preferred inequality measure *I(·)* is the Gini index, due to its many well-known properties and to the fact that its influence function makes it more sensitive to income gaps in the 'middle' of the distribution, rather than the tails. That enables it to better capture differences across type means, since the central limit theorem ensures these cluster 'towards the middle'.

However, the Gini is of course not perfectly additively decomposable. In the context of IOp estimation this implies that relative IOp, $\boldsymbol{I(\hat{y}_i)/I(y_i)}$, is not exactly equal to the "between-type share of total inequality". For this reason, we warn the reader to consider $\boldsymbol{I(\hat{y}_i)/I(y_i)}$ *as a relative* measure of inequality rather than *as a share*. Note, however, that when the absolute measure $\boldsymbol{I(\hat{y}_i)}$ is given by the Gini coefficient, it corresponds exactly to the between-groups Gini term $\boldsymbol{G_B(\hat{y}) = \frac{1}{2\mu}\sum_{k\neq l}\frac{n_k n_l}{n^2}|\mu_k - \mu_l|}$, where *k* and *l* denote types, as before, *n* refers to group sizes, and $\boldsymbol{\mu}$ to group means. That is the first term in the standard Gini decomposition of (Bhattacharya and Mahalanobis, 1967):

$$\boldsymbol{G(y) = \frac{1}{2\mu}\sum_{i\neq j}\frac{n_i n_j}{n^2}|\mu_i - \mu_i| + \sum_i \frac{n_i^2 \mu_i}{n^2 \mu} G(y_i) + R_{BM}} \tag{4}$$

Since the residual term R_{BM}, which contains both within-group and between-group inequality, is always non-negative, $G(\hat{y}_i)/G(y_i)$ is a lower-bound estimate of that 'between-share of inequality'.[5]

2.2 IOp Estimation

Given Equation 3, the problem of estimating ex-ante IOp reduces to obtaining the smooth vector $\hat{y}_i$. For years, most ex-ante IOp empirical analyses employed the *parametric approach* (Bourguignon et al., 2007; Ferreira and Gignoux, 2011), that consists of running an OLS regression to approximate the function $y = f(c)$ and then produce a smooth vector as $\hat{y}_i = \hat{f}(c_i)$. IOp estimates obtained with this approach were often considered lower bounds because, in practice, the whole set of circumstances the researcher ideally wants to include as circumstances outside individual control is unobservable and adding unobservable regressors could only (weakly) increase the explained variability of $\hat{y}_i$ (Ferreira and Gignoux, 2011). However, attempts to overcome this source of bias by including the most exhaustive list of control variables, their non-linear transformations and interactions can lead to upward biased estimates (Brunori et al., 2019).

Selecting an appropriate model specification must therefore appropriately trade off these two potential sources of bias: the downward omitted variable bias and the upward overfitting bias. This need has fostered a set of data-driven methodological contributions that use information on the data structure to generate statistically grounded and non-arbitrary type partitions (see Li Donni et al., 2015; Brunori et al., 2023a; Brunori et al., 2023b). The remainder of the section explains how one family of regression tree algorithms, the Conditional Inference Trees and Random Forests, are suited to estimate ex-ante IOp.

2.2.1 Conditional Inference Regression Trees

[5] As Lambert and Aronson (1993) note, "The overlaps term […] is at once a between groups and a within groups effect: it measures a between groups phenomenon, overlapping, that is generated by inequality within groups" (p.1224). They also show that the term can be represented as a sub-area of the area between the Lorenz Curve and the line of perfect equality and hence cannot be negative.

Regression tree algorithms are built to predict an output variable by splitting the sample data into non-overlapping subgroups, by partitioning the regressors' space. From the wide variety of available regression trees, Conditional Inference Trees (CITs, Hothorn et al., 2006), have recently been used by Brunori et al. (2023b) to estimate ex-ante IOp. The algorithm, in the IOp context, can be summarized as follows:

1. Select α, the preferred level of statistical significance of the tree.
2. Test the null hypothesis of independence between the income (y) and every circumstance (c). For each circumstance, the algorithm stores the Bonferroni-adjusted p-statistic associated to the independence test (p_{adj}).
3. Select the circumstance with the lowest Bonferroni-adjusted p-value, i.e., the one with the strongest association with the outcome. If $p_{adj} > \alpha$, the algorithm stops. Otherwise, continue by setting the selected circumstance as the splitting variable.
4. Choose the splitting value that divides the sample into two sub-samples. For each possible partition the algorithm performs a difference-in-means t-test and obtains a p-value. The partition associated with the smallest p-value is selected to split the sample in two nodes.
5. Steps 2 - 4 are repeated for each resulting sub-sample or node until the null hypothesis of independence cannot be rejected for the critical α selected in Step 1.

Once the algorithm stops, each observation receives its expected income, equivalent to the average income of the subgroup it belongs to. Note that in the IOp context these subgroups can be seen as *types,* since all individuals in them share an identical set of ascribed characteristics, or circumstances. This way, this prediction itself produces the counterfactual distribution $\hat{y}_i$ that allows the IOp estimation as in Equation 3. The statistical tests palliate the arbitrariness associated with traditional approaches to IOp, while the binary splitting makes trees particularly suitable to fit non-linear data-generating processes.

The CIT algorithm bears a technical shortcoming. In Step 4, when non-ordered categorical circumstances are used as regressors, the algorithm tests across all possible groupings of categories. Clearly, the number of combinations grows exponentially with the number of categories in the regressor. To make the problem computationally feasible, CITs limit the number of possible unordered categories to 30, which already delivers $2.65 * 10^{32}$ possible combinations. As discussed in the data section, the key circumstance in this paper is the individual's country of origin, which takes 78 possible values. Including this variable directly as an unordered categorical variable poses an unsolvable limit for the computation of CITs.

We modify the CIT algorithm to address this issue. Before testing for possible splits, we re-order the values of the categorical variable (country of origin) according to the category-specific mean income. This recursive reordering makes the original limitation computationally trivial, because it allows us to treat the unordered variable as an ordered categorical variable. Since binary splits in CITs are based on differences in means in the resulting subgroups, by construction, it is impossible for the resulting types to group categories with relatively lower or higher means together.[6]

2.2.2 Random Forest and Shapley value decomposition

Regression trees are characterized by low bias but high model variance, implying that the tree structure is potentially quite sensitive to the specific sample observed (Hastie et al., 2009). CITs are no exception, so a common practice to mitigate this aspect consists of obtaining predictions from *random forests*. First, several sub-samples are drawn from the original data, and trees are grown in each one of them.[7] The individual prediction $\hat{y}_i$ is obtained after averaging across all individual tree structures. We follow the random forest methodology proposed by Hothorn et al. (2006) and Hothorn and Zeileis (2008), including the modification described above to deal with a high number of unordered categories in

[6] This option is not available in the current partykit package available online. We discussed its consistency with Torsten Hothorn and Achim Zeileis, designers of the CITs algorithm, who have confirmed its suitability. A support function to replicate this procedure is provided in https://github.com/pedrosalasrojo/trick_tree.

[7] This process, similar to bootstrapping, is actually called *bagging*, as sub-samples are drawn without replacement.

the regressors.

The final step in our analysis is to decompose the overall estimates of inequality of opportunity into the contributions of each observed circumstance variable. Those contributions are in general not expected to be additively separable so, among the many possible approaches, we follow the literature and apply a Shapley value decomposition (Shapley, 1953; Shorrocks, 2013). These decompositions should clearly not be interpreted causally but may be of descriptive interest in understanding the nature of IOp in a population.

In our context, the Shapley value decomposition estimates the overall contribution of a circumstance c to IOp as the average decrease in IOp across all possible combinations that exclude c as a factor in estimating $\hat{y}$. Our Shapley value decomposition estimation follows Brunori et al. (2023a) and can be summarized as:

1. Draw a sub-sample sized *B*.[8]
2. Estimate a deep CITs on the sub-sample lowering the confidence level $(1-\alpha)$ close to 0, so the tree is allowed to explore many possible partitions.Measure IOp as between-terminal nodes inequality (see Equation 3).
3. Repeat by systematically generating all possible sequences in which each circumstance c_j can be eliminated from the model.
4. After each elimination sequence, estimate IOp. Store its difference with respect to overall IOp.
5. The stored values across all elimination sequences for each circumstance c_j are averaged and stored.
6. Repeat steps 1-6 a number *n* times.

To avoid the sensitivity of estimates based on a single tree, we follow a random forest approach and estimate the contribution of c_j to IOp as the average contribution from *n=100* bagged samples.

[8] We use the default selected by the cforest function in partykit, 62.3% of the full sample (Hothorn and Zeileis, 2015)

3. Data

While most inequality, mobility, and IOp analyses focus on within and between-countries comparisons, in this paper we attempt for the first time an estimation of continental IOp. Our estimates of European IOp are based on a pool of European countries treated as a whole region. Although the area covered by the analysis -which includes the United Kingdom- may seem large in scope, the European continent, with over 500 million inhabitants and 5 million square kilometres, is relatively small when compared to China and India, which respectively have twice the land area and roughly the same, but almost three times the population.

We combine data from three distinct national household survey sets: the European Union Statistics on Income and Living Conditions (EU-SILC, 2019), the German Socio-Economic Panel (GSOEP, 2019), and Understanding Society - the UK Household Longitudinal Study (UKHLS, 2019). These three surveys allow us to cover 29 European countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Serbia, Slovakia, Spain, Sweden, Switzerland, and the United Kingdom.[9] For simplicity, throughout the remainder of the paper, we will refer to the 29 countries covered by our data as "Europe".

The EU-SILC survey is conducted in member countries of the European Union plus some others and collects information on various sources of income and socioeconomic factors relevant to living conditions. In our sample of 29 countries, Norway, Serbia, Switzerland and the United Kingdom are not in the European Union.[10] While the survey is released annually, three modules provide information about circumstances variables, corresponding to 2005, 2011, and 2019. We use the 2019 wave because, beyond commonly used circumstances, it

[9] Slovenia and Romania are excluded because the data is not suitable for the analysis.
[10] Although the Brexit referendum was held in 2016, the United Kingdom formally belonged to the European Union until January 2020.

collects information that can be used to study the country of origin.

The EU-SILC 2019 data does not include the United Kingdom, so we rely on the Understanding Society survey (UKHLS, waves 10 and 11) as an alternative data source. We also employ an alternative data source for Germany, the German Socio-Economic Panel (GSOEP), because in the EU-SILC 2019 the information on some circumstances is not provided in detail.

The outcome of interest is the annual disposable household income, adjusted by Purchasing Power Parity conversion factors to 2019 USD.[11] We account for equivalences of scales by dividing incomes by the square root of the household size. Finally, accounting for life-cycle income dynamics, we adjust the dependent variable by removing age-related variations.[12] It should be noted that all three of these variable transformations are likely to lead to lower estimates of inequality, relative to a benchmark of household per capita market incomes unadjusted for life-cycle differences.

The EU-SILC does not provide detailed information on the respondent's country of origin, but it provides information that can be used as a proxy. First, it includes a variable stating whether the respondent was born in the country of residence, in another European country, or outside the EU. If the individual was born in the country of residence no additional adjustments are needed. For those born abroad, we rely on the information about the mother's country of birth, available in the 2019 wave. While imperfect, this proxy captures key cultural and socioeconomic background dimensions of the children. First because the probability that mother and children, both first-generation migrants are in fact born in the same country, second because the most relevant cultural heritages such as religion, language and social norms are mostly transmitted from mothers to children (Caneva and Pozzi, 2014).[13] As a result, our primary data includes 181 different places of origin.

[11] We use the World Bank PPP Conversion Factors for Private Consumption (LCU per International $).

[12] For the age adjustment, we run a regression of incomes against age and age squared. Our outcome of interest is the sum of the constant and the residuals of the regression.

[13] To what extent is using the country of birth of the mother as a proxy for the country of birth of respondents who report being born outside their country of residence accurate? To address this, we have analysed data

We complete the IOp analysis with other circumstances, namely the respondents' sex (2 categories), their parents' education level (3 categories, based on the International Standard Classification of Education ISCED 2011) and their parents' occupation when they were 14 years old (11 categories, based on the 1-digit International Standard Classification of Occupations ISCO-08). The complete list of circumstances and respective categories used in the analysis can be found in Table A1 in the Appendix. All remaining descriptive statistics are available upon request.

Pooling together different data sources presents a considerable challenge, because the data is collected to be nationally representative, rather than at the continental level. Sample sizes vary across countries, and the weighting scheme accounts for sex, education and other socioeconomic factors, but not for the country of origin.[14] To address these issues we restructure our data employing a stratified sampling procedure based on two key aspects: the population share of each country included in our dataset, and the distribution of first-generation immigrants by country of origin within each country. This procedure can be summarized as follows:

1. We use census data to obtain information on the population living in every European country in 2019, as well as the distribution of individuals by country of birth within each country.[15]

---

from the GSOEP and UKHLS, which include information on the country of birth for both respondents and their parents. In Germany and the UK, we can precisely identify respondents' countries of birth and assess the degree of inconsistency encountered when using the mother's country of birth as a proxy for respondents' countries of origin. The share of individuals born in a country different from that of their mother is limited to 14%. Thus, our variable for country of origin accurately captures the country of birth for Germany and the UK and is expected to correctly identify respondents' countries of birth in approximately 85% of cases in other countries. However, since the information is available for UK and Germany, which represent around 30% of the sample, we reduce the potential distortion using the reported individual's country of birth.

[14] Details about sampling strategy are available for EU-SILC https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:EU_statistics_on_income_and_living_conditions_(EU-SILC)_methodology_%E2%80%93_sampling&oldid=272118#Sampling_frame ,
SOEP https://www.diw.de/documents/publikationen/73/diw_01.c.841084.de/diw_ssp1106.pdf,
and UKHLS https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/clustering-and-stratification/ .

[15] For most countries we rely on 2011 census data reported in the Eurostat database about country population shares by country of birth. For Germany, Ireland, Greece, Croatia, Poland, Portugal, data are based on 2011 census data about resident by their citizenship rather than country of birth. For Serbia and Malta, the available data are, respectively, the 2022 and 2021 census.

2. We stratify the population by country of residence and, within each country, by country of birth.
3. To avoid issues related to very small population sizes, we set a minimum threshold. Only countries of birth accounting for at least 0.1% of the European population are considered separately. Countries below this threshold are aggregated into broader regions based on the United Nations geoscheme.[16] This results in 62 individual countries and 17 regional groupings, for a total of 79 possible countries or regions of origin.[17]
4. Based on population shares we determine the target number of observations for each country-of-residence (or country of origin) cell. The overall sample size is defined using Germany, the most populated country in the data, as a reference. We then construct a synthetic sample that matches these targets. Observations are randomly drawn without replacement when the number available in the original data exceeds the target and randomly replicated when the available number is insufficient.[18] When no observations are available for a given origin group within a country, individuals from the same origin group are drawn from countries with similar mean income levels. The original survey weights are preserved.[19] We are thus left with 78 possible values in the "country of birth"

[16] For instance, individuals born in Curaçao or Jamaica represent less than 0.1% of the population living in the original sample, so we include them in the macro region "Other Caribbean". See the table with the definition of the aggregates in Table A3 in the Appendix.

[17] The single countries used are: Afghanistan, Albania, Argentina, Austria, Bosnia and Herzegovina, Bangladesh, Belgium, Bulgaria, Brazil, Switzerland, China -including Hong Kong-, Colombia, Cyprus, Czechia, Germany, Denmark, Dominican Republic, Algeria, Ecuador, Estonia, Greece, Spain, Finland, France, Croatia, Hungary, Ireland, India, Iraq, Iran, Italy, Kosovo, Sri Lanka, Lithuania, Luxembourg, Latvia, Morocco, North Macedonia, Malta, Nigeria, Netherlands, Norway, Peru, Philippines, Pakistan, Poland, Portugal, Romania, Serbia, Russia, Sweden, Slovakia, Senegal, Syria, Tunisia, Turkey, Ukraine, United Kingdom, United States, Venezuela, Vietnam, South Africa. The macro regions we define are: Other Caribbean, Other Central Africa, Other Central American, Other Central Asia, Other East Africa, Other East Asia, Other Europe, Other Middle East, Other North Africa, Other North American, Other Oceania, Other South American, Other South Africa, Other South Asia, Other South-East Asia, Other West Africa, and Others.

[18] As a result, some countries of birth are relatively too small to be represented with the sample sizes for individual countries of residence, and the macro region "Other South Africa" is relatively too small across all countries of residence, resulting in no observations from this region. This leaves 16 regional groupings, leading to a final number of 78 categories for country of origin.

[19] Table A2 shows that this procedure does not affect representativeness along other dimensions, such as sex and age, for which the survey weights are calibrated.

variable.

After restricting the sample to keep only working-age individuals (between 25 and 65 years old), our final sample representing Europe is composed of 149,170 observations. Table 1 shows, for each country, the sample size, the population share, the share of foreign-born population, the average equivalized income, the standard deviation and the Gini index. As expected, Germany, France, Italy, the UK and Spain account for the biggest population shares. In terms of country of origin, we also find a wide heterogeneity, with countries like Poland or Croatia showing small foreign-born shares, and others like Sweden (19%) or Luxembourg (44%) showing a high proportion. Income-related statistics also align with our expectations.

Table 1: Composition and summary statistics of the representative European sample

Country (ISO-2)	Sample size	Sample share (%)	Share of Foreign Born (%)	Mean Income	Gini (Income)
Europe	149,170	100	10.50	21,242	0.39
Austria	2,610	1.75	19.85	29,684	0.27
Belgium	3,372	2.26	17.17	28,116	0.25
Bulgaria	2,262	1.52	10.96	5,677	0.39
Switzerland	2,480	1.66	28.75	51,316	0.28
Cyprus	249	0.17	22.09	21,478	0.34
Czech Republic	3,117	2.09	4.17	11,677	0.24
Germany	24,441	16.38	12.24	25,091	0.37
Denmark	1,679	1.13	10.90	37,590	0.22
Estonia	389	0.26	15.17	13,067	0.27
Greece	2,995	2.01	2.80	9,675	0.33
Spain	13,706	9.19	13.32	17,886	0.32
Finland	1,595	1.07	5.58	31,736	0.25
France	18,910	12.68	12.60	26,387	0.30
Croatia	1,189	0.80	1.09	8,301	0.29
Hungary	2,858	1.92	3.39	6,485	0.30

Ireland	1,458	0.98	13.10	32,427	0.32
Italy	17,868	11.98	11.53	21,170	0.32
Lithuania	819	0.55	4.76	10,140	0.35
Luxembourg	171	0.11	44.44	47,413	0.29
Latvia	564	0.38	12.06	9,689	0.35
Malta	139	0.09	14.39	18,832	0.28
Netherlands	5,069	3.40	13.02	28,796	0.27
Norway	1,566	1.05	15.71	46,691	0.25
Poland	11,102	7.44	0.12	7,913	0.30
Portugal	3,015	2.02	4.64	11,653	0.33
Serbia	1,699	1.14	5.83	3,877	0.35
Sweden	3,004	2.01	19.04	29,820	0.24
Slovakia	1,593	1.07	3.26	9,444	0.24
United Kingdom	19,251	12.91	7.59	19,616	0.30

Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares come from census data. Monetary values in 2019 USD. The shares refer to the population share with respect to the total. Sd stands for the Standard deviation of income.

4 Empirical analyses

4.1 European IOp estimates

We run the CIT and the random forest algorithms on the representative European sample of 149,170 individuals.[20] The CIT algorithm partitions the population into 149 types, with the smallest comprising 0.07% of the sample and the largest type representing 7% of the sample. Table 2 shows that, while income inequality across the entire sample is equal to 39 Gini points, absolute IOp estimated on the smoothed distribution $\hat{y}$ (see equation 3), reaches 24 points. This implies that the six circumstances considered predict a relative inequality of 62% of the overall Gini. The corresponding metrics obtained from the random

[20] The CITs parameters are the following: α is 0.01, and the minimum number of observations at each terminal node (minbucket bucket) to 100. The remaining parameters in the algorithm are set to the default values proposed in Hothorn and Zeileis (2015).

forest, which smooths out the variation in the prediction model, are an absolute IOp Gini of 0.23 and a relative IOp measure of 59% of total inequality. These estimates are broadly similar to the global inequality of opportunity figures reported by Milanovic (2015), suggesting that differences across European countries play a quantitatively important role when Europe is treated as a single opportunity space.[21] The comparison should be interpreted with caution given differences in data and methodology.

Table 2: European income inequality and inequality of opportunity

Overall Gini	Absolute IOp Gini (tree)	Rel. IOp Gini (tree)	Absolute IOp Gini (forest)	Rel. IOp Gini (forest)
0.39	0.24	0.62	0.23	0.59

Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The tree is estimated with α = 0.01, minbucket = 100. The random forest is estimated with α = 0.1, minbucket = 50, and 200 bagged subsamples sized 10% the original sample size. IOp stands for Inequality of Opportunity, Rel. stands for "Relative" and is obtained after dividing IOp over Overall Gini.

How do these Europe-wide levels of inequality of opportunity compare globally? In terms of absolute IOp levels, the results in Table 2 are similar to those of Latin American countries such as Brazil (2014), Panama (2008), Colombia (2010), and Peru (2014), which show IOp estimates around or slightly above 0.22 Gini points. These estimates are obtained using comparable data and a similar methodological approach (see Ferreira et al., 2026), but they refer to individual countries rather than to an entire continent. Interestingly, while overall disposable income inequality is like that of the United States (0.39 Gini points), IOp in that country is six Gini points lower than in Europe, lying between 0.16 and 0.17 Gini points in the 2000–2014 period. Absolute IOp in Europe is also very close to that of other large countries such as China (0.22 Gini points, 2018), although it remains below India (0.27 Gini

[21] Milanovic (2015) reports values between 60% and 65% across alternative specifications.

points, 2012). These results may appear surprising, especially considering that total inequality in China and India is 0.09 and 0.11 Gini points higher than in Europe, respectively.

To explore this apparent puzzle, Table A4 in the Appendix shows IOp estimates for every single European country in our sample. Countries like Denmark, Finland and Norway show low IOp values, lying below 0.04 Gini points, while Bulgaria, Luxembourg or Germany are characterized by more unequal opportunities, with IOp values lying above 0.12 Gini points.[22] Averaging across the 29 country-specific Absolute IOp estimates delivers a value of 0.07 Gini points, less than one third of our continental baseline estimate obtained with random forest. Even though single European countries usually fare well in terms of equalizing opportunities when compared with other areas of the globe, the convergence across countries is far from complete.

A natural question is whether the importance of country of birth reflects migration from poorer countries rather than persistent differences across European economies. We examine the role of migration in two ways that happen to produce very similar results.

First, we repeat the analysis using country of residence instead of country of birth. The resulting estimate of absolute inequality of opportunity is 0.228 Gini points, compared to 0.225 in the baseline specification. The similarity of these estimates suggests that the results are not driven by migration, but rather by persistent cross-country differences within Europe.

Second, we examine whether the IOp estimate changes under a counterfactual scenario without migration. To do so, we predict individual income and reassign incomes to Europeans living outside their country of birth. Specifically, individuals residing abroad are assigned the random forest-predicted income of individuals who live in their country of origin and share the same birth circumstances. The results indicate that, in this counterfactual scenario, inequality of opportunity across Europe remains very similar, albeit

[22] We have checked that inequality estimates by country closely align with those in the Global Estimates of Opportunity and Mobility Project (GEOM), with a correlation coefficient of 0.93. IOp estimates should not be directly compared because we use a different set of circumstances and data cleaning process.

slightly higher: 0.236 Gini points compared with 0.225 in the baseline.

We now turn our attention to the relative contribution of the country of origin, along with other birth circumstances, to the IOp value presented in Table 2. We present the results of the Shapley value decomposition in Table 3. The country of origin emerges as the most important factor, accounting for 63.6% of IOp. This value contrasts with the relevance of this circumstance in other prominent countries, such as China (48% of IOp, 2018), Brazil (13%, 2014), the USA (15%, 2014) or South Korea (27%, 2019).[23]

As noted in Section 3, our country-of-origin circumstance has 78 different categories, enabling the algorithm to find complex interactions and non-linearities across categories.[24] By comparison, the number of categories of the variable "region of birth" for the countries listed above ranges from 6 in the USA to 24 in China and 27 in Brazil. Additionally, in these cases, all migrants are grouped into a single "other" category. We expect that adopting a richer variable similar to ours in these studies would increase IOp estimates in these countries as well.

Table 3: Contribution of Circumstances

Circumstance	Contribution to IOp (%)
Country of origin	63.6
Father's occupation	17.3
Mother's occupation	8.7
Father's education	5.0
Mother's education	4.1
Sex	1.3

Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The Shapley value decomposition is estimated with a random forest procedure, with α = 0.1, minbucket = 50, and 100 bagged repetitions sized 10% the original sample size. IOp stands for Inequality of Opportunity.

[23] Results from GEOM database, see Ferreira et al., (2026).

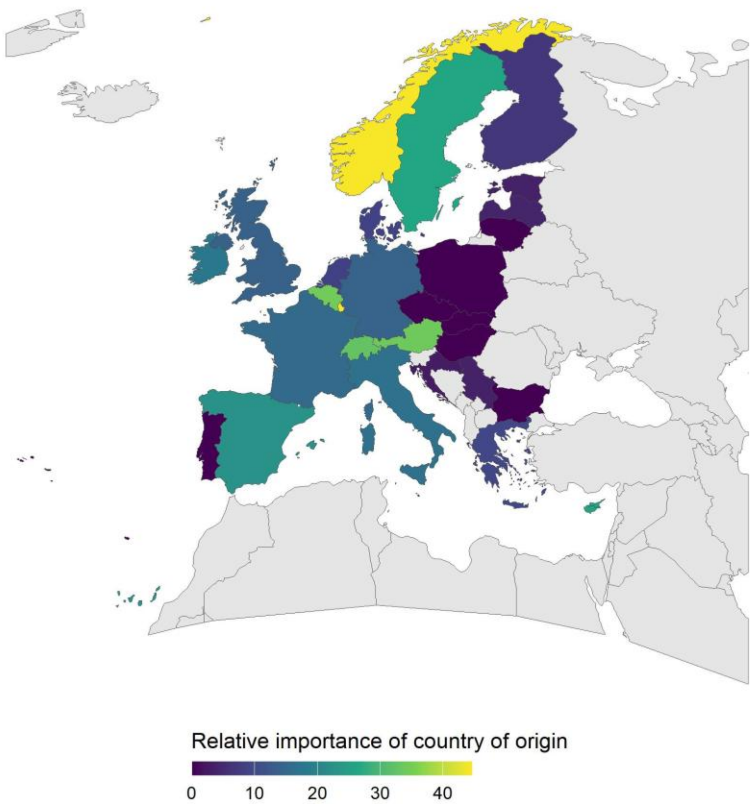
[24] CITs, as explained in Hothorn et al. (2006) are not biased towards regressors with many different categories.

As in Brunori et al. (2023b), the occupation of the father (17.3%) and the mother (8.7%) are found to be relevant contributors to IOp, while parental education accounts for 9.1% of overall inequality of opportunity, with the contribution of sex being rather small (1.3% of total IOp). This apparently surprisingly small value is generated by the outcome variable - equivalised household income-, which smooths out intra-household inequalities. The contribution of sex to inequality of opportunity (IOp) reflects the inequality associated with households where the share of females differs from 50%, such as single-parent households. Clearly, a completely different picture would emerge if we were to consider personal income (Leythienne and Pérez-Julián, 2022).

We repeat the Shapley value decomposition for each country in our sample, thus exploiting the heterogeneity of the diverse data-generating processes.[25] Figure 1 highlights that the relative contribution of country of origin to IOp varies across countries, being especially relevant in Austria, Belgium, Luxembourg, Norway, Sweden, Switzerland and Cyprus compared to other nations. The variability across countries mirrors the differences in the share of migrants living in each country, with few exceptions such as the UK, Germany, and Italy, where the relative share of migrants is lower (see Figure A1 in the Appendix). When measuring the absolute contribution of country of origin to IOp, the picture shifts slightly. Germany and Spain emerge among the countries where the influence of country of birth is high (Figure A2 in the Appendix).

[25] The Shapley value decomposition for each single country is reported in Table A5 in the Appendix.

Figure 1: Relative importance of country of birth



Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The Shapley value decomposition is estimated with a random forest procedure, with α = 0.1, minbucket = 50, and 100 bagged repetitions sized 0.632 the original sample size. The Shapley decomposition values reported in the map range between 0% and 45%.

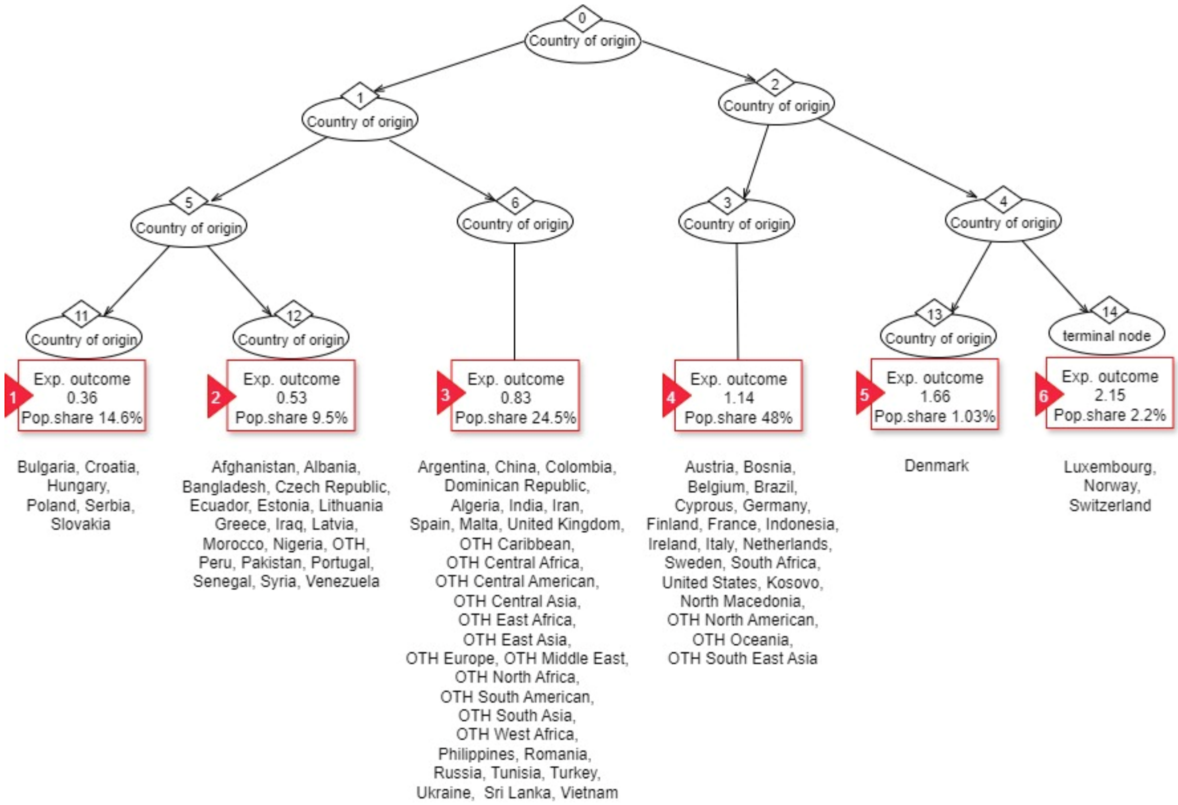
4.2 The income of European residents

The CIT algorithm identifies 149 types, and this complexity limits a complete graphical representation. Despite inherent limitations, including instability of the tree structure (Moramarco et al., 2024), a graphical representation provides valuable insights.[26]Figure 2 focuses on the upper part of the tree, presenting country of origin as the key determinant of income disparities. Six clusters emerge in the first three splits of the CITs, with average

[26] Beyond the model variability and dependence to the sample characteristics, the deterministic binary splitting process limits the creation of an increasing share of all the possible types and may generate terminal nodes with the same expected outcomes.

incomes indexed to the European mean (21,242 USD in 2019). Being born in Denmark, Luxembourg, Norway, or Switzerland yields incomes equivalent to 166%-215% of the mean. At the other end, individuals from Bulgaria, Croatia, Hungary, Poland, Serbia, and Slovakia earn only 36% of the sample average, thus highlighting stark disparities by country of origin. Inequality among these six nodes already reaches 0.20 Gini points.

Figure 2: European Opportunity Tree: upper structure

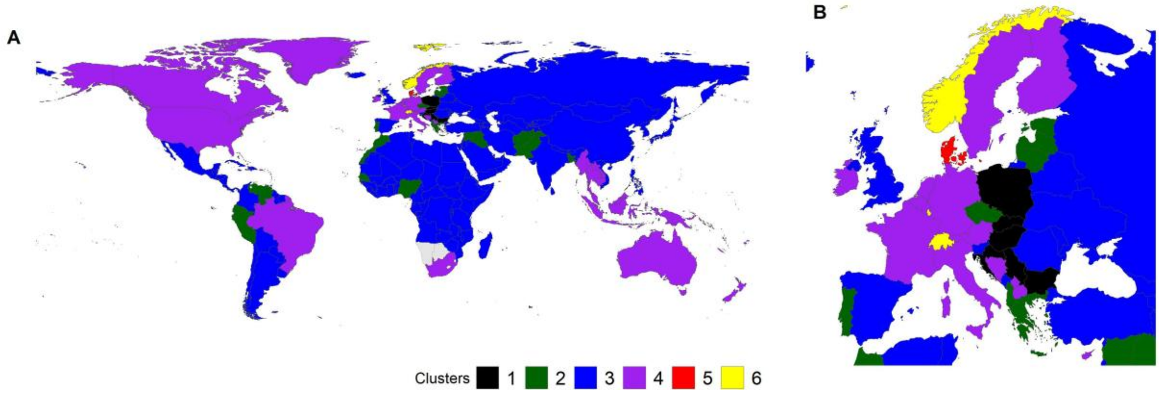


Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The tree is estimated with α = 0.01, minbucket = 100. The upper number in the terminal nodes indicate the predicted outcome of the group (Exp. Outcome, normalized to 1 = sample outcome mean, 21,242 in 2019 USD). The lower number in the terminal nodes (Pop. Share) shows the share of the population belonging to that sub-sample.

Figure 3 facilitates the readability of the tree in Figure 2 by mapping these six clusters. Panel A shows the global distribution of clusters while Panel B provides a zoomed-in view of Europe for greater readability. European individuals born in Eastern European countries

(Cluster 1 in Figure 2, representing 14.6% of the population), earn, on average, less than other citizens whose origin is located in developing countries. This highlights the fundamental interaction between country of origin and country of destination in the continent: a large share of respondents coming from the global South migrate towards relatively richer countries in Europe, obtaining access to more opportunities than Europeans from poorer countries.[27]

Figure 3: European Opportunity Tree: upper structure



Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The map illustrates the countries of birth as grouped in the six upper clusters obtained from the CITs algorithm. The groups are named using a numerical code that corresponds to those in the red triangles in Figure 2. The tree is estimated with α = 0.01, minbucket = 100.

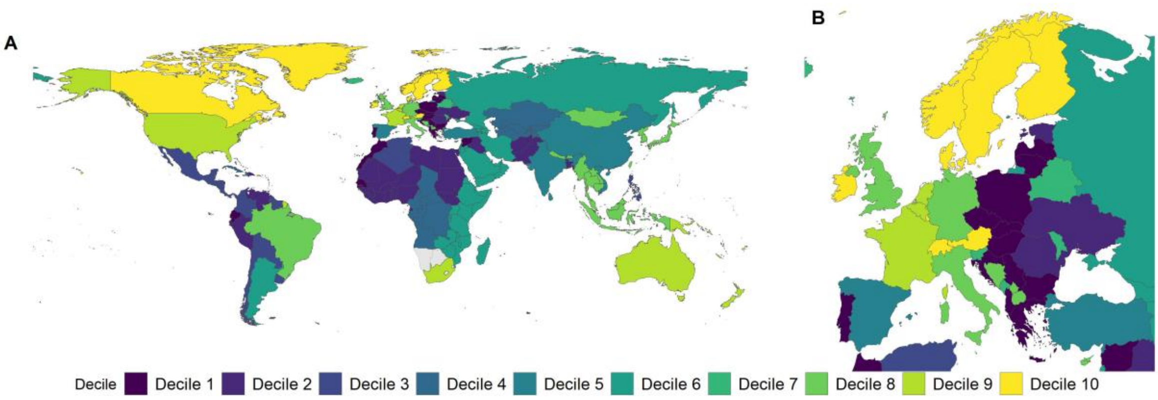
Figure 4 further illustrates the income variability and opportunities shaped by migration and regional disparities, showing the deciles of the income prediction obtained with the random forest depending on the respondent's country of origin. Panel A shows the world map, while Panel B zooms in on Europe. The active selection and relative homogeneity of migrants across regions is confirmed: migrants from Southeast Asian countries and Oceania are mostly located in the 8th and 9th deciles of the predicted income, while those born in North and West Africa, as well as in many Latin American countries, rank in the bottom three deciles. This figure also highlights the high variability across European countries of origin, with Spain and the Western Economies falling well behind those born in Scandinavia and

[27] We do not engage here in the complex discussion of how migrants are selected into destination countries or the extent to which they are representative of migrants from their countries of origin.

Central Europe.

Our data prevents us from further analysing aspects associated with positively selected migration in Europe. Such a comparison needs to account for migrants' characteristics and is beyond the scope of our analysis. Still, for descriptive purposes only, the interested reader may find a comparison between the predicted incomes obtained by individuals born outside their countries of residence and the GDP per capita of their countries of origin. Figure A3 in the Appendix focuses on individuals born outside Europe, while Figure A4 focuses on those born in Europe. Those born in East and Central Africa and Afghanistan experience, through migration, the greatest relative improvement in their opportunities. By contrast, those born in North America, Russia, Japan, Oceania and oil-exporting countries have an expected outcome, on average, well below the GDP per capita of their country of origin.

Figure 4: Average Predicted Income by Country of Origin



Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The random forest is estimated with α = 0.1, minbucket = 50, and 200 bagged subsamples sized 10% the original sample size. Table A6 in the Appendix reports the exact values of average predicted income by country of origin.

These aggregate values hide, in some cases, profound variations across recipient countries. Figure A5 in the Appendix shows that, for example, individuals from Morocco, Algeria, Syria or Romania earn similar incomes across the most frequent destination countries. In contrast, those with a Chinese background earn, on average, much more in Germany, the Netherlands and Sweden than those with the same background who settle in Spain, France or Italy. Something similar applies to Indians, who fare relatively worse in Spain and Italy than in Sweden or the United Kingdom. These results are purely descriptive and do not account for positive or negative selection by country of origin in the recipient countries but still reflect remarkable underlying disparities far from the normative principles driving the European Union described in the Introduction section.

Finally, we examine how migrants, both from within and outside Europe, compare to the native populations in terms of predicted income. This is, once again, a comparison to be interpreted with caution, since it cannot account for differences in migrants' characteristics relative to those of the destination-country population. On average, migrants from non-European countries, as well as those from Eastern Europe, Spain, and Portugal, tend to have predicted incomes below the national averages in their countries of residence (Figure A6 in the Figure Appendix). Although we do not account for factors such as education, skill levels, time spent in the host country, or language proficiency, all of which are critical for labour market integration, these findings suggest that migrants, especially those from outside Europe or less affluent European countries, face substantial challenges in achieving income parity with native populations. This is consistent with Dustmann and Frattini (2011), who find that immigrants from most regions are more likely than natives to fall into the lower deciles of the earnings distribution, except for those from North America and Oceania. While they identify Latin American migrants as particularly disadvantaged, our analysis reveals that migrants from most African countries (excluding South Africa) and Central Asia are among the most disadvantaged relative to natives.

5 Conclusions

Equality of opportunity and cross-country convergence are among the core objectives of the European integration project. European treaties and institutions emphasize non-discrimination, freedom of movement, and social and territorial cohesion as means to ensure that individuals' life chances are not determined by arbitrary circumstances such as place of birth or social origin. While substantial progress has been achieved within individual countries, this paper shows that, when Europe is treated as a single entity, true equality of opportunity remains elusive.

Using survey and census data for 2019, we construct a representative sample of the European population that explicitly accounts for individuals' countries of origin. Adopting the ex-ante equality of opportunity framework and employing data-driven methods based on Conditional Inference Trees and Random Forests, we estimate inequality of opportunity at the continental level. Our results indicate that inequality associated with circumstances beyond individual control amounts to approximately 0.23 Gini points -nearly three-fifths- of total income inequality in Europe. This level is comparable to that observed in continent-sized countries such as China and India, and substantially higher than estimates for the United States.

A central finding of the paper is the dominant role played by country of origin. Approximately 64% of measured inequality of opportunity is attributable to differences in countries of birth, far exceeding the contributions of parental occupation, parental education, or sex. While most European countries display relatively low levels of inequality of opportunity when analysed in isolation, aggregating across countries reveals deep disparities in opportunities associated with birthplace. From an equality-of-opportunity perspective, focusing exclusively on within-country analyses can therefore be misleading in highly integrated regions, as it masks substantial inequalities operating at broader spatial scales.

Beyond its substantive findings for Europe, this paper contributes to the literature on inequality of opportunity by highlighting the importance of the unit of analysis in empirical measurement. When circumstances vary systematically across regions and mobility is non-negligible, national estimates of inequality of opportunity may significantly understate the extent of opportunity inequality. Our results illustrate how treating Europe as a single opportunity space fundamentally alters the assessment of equality of opportunity relative to country-level analyses.

Several robustness checks -including alternative definitions of birthplace and a counterfactual scenario without internal migration- confirm that the main findings are not driven by specific modelling choices. These exercises suggest that persistent disparities across countries of origin remain a key source of inequality of opportunity in Europe, even in the presence of free mobility. In addition, estimates based on country of residence are very similar to those obtained using country of birth, suggesting that the results are not primarily driven by selective migration but rather by persistent differences across European countries.

Overall, the evidence presented in this paper indicates that, despite extensive institutional integration, Europe cannot yet be regarded as a fully integrated opportunity space. Instead, it resembles a collection of unequal opportunity regimes, where individuals' economic prospects continue to depend strongly on their country of birth. Whether future convergence, mobility patterns, or policy coordination will succeed in reducing these disparities remains an open question.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT to improve the structure and clarity of some paragraphs. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

Pedro Salas-Rojo acknowledges financial support from the Velez Reyes Foundation and grant César Nombela 2025-T1/PH-HUM-36094. The remaining authors declare that this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bhattacharya, N., & Mahalanobis, B. (1967). Regional disparities in household consumption in India. *Journal of the American Statistical Association*, *62*(317), 143-161.

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

Brunori, P., Ferreira, F., & Salas-Rojo, P. (2023a). *Decomposing inequality with machine learning: The case of opportunity in South Africa* (IZA Discussion Paper No. 17203). IZA.

Brunori, P., Hufe, P., & Mahler, D. (2023b). The roots of inequality: Estimating inequality of opportunity from regression trees and forests. *Scandinavian Journal of Economics, 125*(4), 900–930.

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

Caneva, E., & Pozzi, S. (2014). The transmission of language and religion in immigrant families: A comparison between mothers and children. *International Review of Sociology, 24*(3), 436–449.

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

Checchi, D., Peragine, V., & Serlenga, L. (2016). Inequality of opportunity in Europe: Is there a role for institutions? In *Inequality: Causes and consequences* (Vol. 43, pp. 1–44). Emerald Group Publishing.

Dustmann, C., & Frattini, T. (2011). Immigration: The European experience (Development Studies Working Paper No. 326). Centro Studi Luca d'Agliano.

European Commission. (2003). *Commission Regulation (EC) No 1982/2003 implementing Regulation (EC) No 1177/2003 of the European Parliament and of the Council concerning Community statistics on income and living conditions (EU-SILC)*. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32003R1982

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

Ferreira, F. H., & Peragine, V. (2013). *Equality of opportunity: Theory and evidence* (Policy Research Working Paper No. 7217). World Bank.

Ferreira, F. H. G., Peragine, V., Brunori, P., Salas-Rojo, P., Moramarco, D., Barajas Prieto, L., Barbieri, T., Daza Baez, N., Datt, G., de Sandi, V., et al. (2026). *Global estimates of opportunity and mobility: A database* (III Working Paper No. 158). International Inequalities Institute, London School of Economics and Political Science.

Filauro, S., Palmisano, F., & Peragine, V. (2023). *The evolution of inequality of* *opportunity in Europe*. European Commission.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

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

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

Lambert, P. J., & Aronson, J. R. (1993). Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal*, *103*(420), 1221-1227.

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

Leythienne, D., & Pérez-Julián, M. (2022). *Gender pay gaps in the European Union: A statistical analysis based on Structure of Earnings Survey 2022 data*. Eurostat.

Li Donni, P., Rodríguez, J. G., & Rosa Dias, P. (2015). Empirical definition of social types in the analysis of inequality of opportunity: A latent classes approach. *Social Choice and Welfare, 44*, 673–701.

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

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

Moramarco, D., Brunori, P., & Salas-Rojo, P. (2023). *Biases in inequality of opportunity estimates: Measures and solutions* (Working Paper No. 145). International Inequalities Institute.

Ramos, X., & Van de Gaer, D. (2016). Approaches to inequality of opportunity: Principles, measures and evidence. *Journal of Economic Surveys*, *30*(5), 855-883.

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

Roemer, J. E., & Trannoy, A. (2015). Equality of opportunity. In A. B. Atkinson & F. Bourguignon (Eds.), *Handbook of income distribution* (Vol. 2, pp. 217–300). Elsevier.

Salas-Rojo, P., & Rodríguez, J. G. (2022). Inheritances and wealth inequality: A machine learning approach. *The Journal of Economic Inequality, 20*(1), 27–51.

Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games II* (pp. 307–317). Princeton University Press.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. *Journal of Economic Inequality, 11*(1), 99–126.

United Nations Department of Economic and Social Affairs (UN DESA), Population

Division. (2020). *International migration 2020 highlights* (ST/ESA/SER.A/452). United Nations.

Van de Gaer, D. (1993). *Equality of opportunity and investment in human capital*. PhD. Dissertation, Katholieke Universiteit Leuven.

Appendix

Table Appendix

Table A1: Circumstances and Categories

Respondent's country of origin:
181 values (classification of Countries using SCL Geo codes), grouped into 62 countries of birth and 16 macro regions (Table A3 in the Appendix)

Respondent's sex:
- (1) Male
- (2) Female

Father's and mother's education, based on International Standard Classification of Education 2011 (ISCED-11):
- (1) Low level (less than primary, primary education, or lower secondary education)
- (2) Medium level (upper secondary education and post-secondary non-tertiary education)
- (3) High level (short-cycle tertiary education, bachelor's or equivalent level, master's or equivalent level, doctoral or equivalent level)

Father's and mother's main occupation, based on International Standard Classification of Occupations, published by the International Labour Office ISCO-08:
- (0) Armed Forces Occupations
- (1) Managers
- (2) Professionals
- (3) Technicians and Associate Professionals
- (4) Clerical Support Workers
- (5) Services and Sales Workers
- (6) Skilled Agricultural, Forestry and Fishery Workers
- (7) Craft and Related Trades Workers
- (8) Plant and Machine Operators and Assemblers
- (9) Elementary occupations
- (10) Unemployed, deceased or not living with the respondent

Table A2: Sample composition after resampling

Country	N Original	N Sample	Avg. Original Age	Avg. Age Sample	Female Original (Share)	Female Sample (Share)
Austria	6647	2610	47	46	0.53	0.52
Belgium	7846	3372	46	46	0.52	0.52
Bulgaria	8827	2262	48	48	0.50	0.48
Switzerland	8205	2480	47	46	0.52	0.52
Cyprus	5569	249	46	45	0.53	0.49
Czech Republic	9541	3117	47	46	0.52	0.52
Germany	24446	24441	45	47	0.51	0.52
Denmark	5621	1679	48	48	0.52	0.52
Estonia	7499	389	46	46	0.51	0.53
Greece	19172	2995	48	48	0.52	0.53
Spain	21063	13706	47	47	0.51	0.50
Finland	11775	1595	48	47	0.50	0.48
France	12866	18910	47	47	0.52	0.52
Croatia	9863	1189	48	48	0.50	0.48
Hungary	7621	2858	49	48	0.54	0.53
Ireland	4960	1458	46	47	0.53	0.53
Italy	21761	17868	47	47	0.51	0.52
Lithuania	6077	819	48	48	0.55	0.56
Luxembourg	5545	171	45	46	0.52	0.49
Latvia	5594	564	47	47	0.53	0.53
Malta	4813	139	45	43	0.50	0.49
Netherlands	14229	5069	48	48	0.52	0.53
Norway	7174	1566	46	46	0.49	0.47
Poland	22862	11102	47	48	0.55	0.54
Portugal	17460	3015	48	48	0.53	0.51
Serbia	8551	1699	46	46	0.50	0.49
Sweden	5653	3004	46	46	0.51	0.51
Slovakia	8164	1593	46	46	0.53	0.53
United Kingdom	20664	19251	47	47	0.56	0.55

Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. N stands for the number of observations. Avg. stands for "Average". We compare the sample composition in the original data (Original) and our analysis sample (Sample).

Table A3: Definition of countries of birth aggregates

Macro Area/Region of Birth	Countries
Other Caribbean (OTHCaribbean)	Anguilla, Antigua and Barbuda, Aruba, Bahamas, Barbados, Bonaire Saint Eustatius and Saba, British Virgin Islands, Caribbean, Cayman Islands, Cuba, Curaçao, Dominica, Former Netherlands Antilles, Grenada, Haiti, Jamaica, Montserrat, Sint Maarten, St. Kitts and Nevis, St. Vincent and the Grenadines, St-Barthélemy, St-Martin, Trinidad and Tobago, Turks and Caicos Islands, U.S. Virgin Islands
Other South American (OTHSouthAmerican)	Bolivia, Chile, Falkland Islands, French Guiana, Guiana, Paraguay, Suriname, Uruguay
Other Central American (OTHcentralAmerican)	Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Mexico, Panama
Other North American (OTHNorthAmerican)	Canada, Greenland, Saint Pierre and Miquelon, Puerto Rico, Bermuda
Other West Africa (OTHWestAfrica)	Burkina Faso, Benin, Togo, Ghana, Côte d'Ivoire, Liberia, Sierra Leone, Guinea-Bissau, Guinea, Niger, Cape Verde, Gambia, Mauritania, Saint Helena, Mali
Other North Africa (OTHNorthAfrica)	Egypt, Libya, Sudan, South Sudan, Western Sahara
Other Central Africa (OTHCentralAfrica)	Angola, Cameroon, Central African Republic, Chad, Congo, Gabon, Democratic Republic of the Congo, Equatorial Guinea, São Tomé and Principe
Other East Africa (OTHEastAfrica)	Burundi, Comoros, Djibouti, Eritrea, Ethiopia, Kenya, Madagascar, Malawi, Mayotte, Mauritius, Mozambique, Rwanda, Seychelles, South Sudan, Somalia, Somaliland, Tanzania, Uganda, Zambia, Zimbabwe

Continues in next page.

Table A3 (cont): Definition of countries of birth aggregates

Other Middle East (OTHMiddleEast)	Arab Emirates, Armenia, Azerbaijan, Bahrain, Georgia, Israel, Jordan, Kuwait, Lebanon, Oman, Palestine, Qatar, Saudi Arabia, Yemen
Other Central Asia (OTHCentralAsia)	Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan
Other East Asia (OTHEastAsia)	Japan, North Korea, South Korea, Mongolia, Taiwan
Other South East Asia (OTHSouthEastAsia)	Brunei, Cambodia, Indonesia, Lao PDR, Malaysia, Myanmar Burma, Myanmar, Singapore, Thailand, Timor-Leste
Other South Asia (OTHSouthAsia)	Bhutan, Maldives, Nepal
Other Oceania (OTHOceania)	American Samoa, Ashmore and Cartier Island, Australia, Australia and New Zealand, Cook Island, Federated States of Micronesia, Fiji, French Polynesia, Guam, Kiribati, Marshall Island, Melanesia, Micronesia, Nauru, New Caledonia, New Zealand, Niue, Norfolk Island, Palau, Papua New Guinea, Pitcairn Island, Polynesia, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu, Wallis and Futuna Island
Other Europe (OTHEurope)	Aland, Andorra, Belarus, Faroe Island, Gibraltar (UK), Guernsey, Iceland, Isle of Man, Jersey, Liechtenstein, Monaco, Montenegro, Moldova, San Marino, Slovenia, Vatican
Others (OTH)	Multiple categories or groups that represent undefined or aggregate countries

Source: Own elaboration. The table reports the definition of the macro regions used to aggregate the frequencies of countries of birth that represent less than 0.1% *of the population living in Europe. It is based on the United Nations geoscheme. United Nations. Standard country or area codes for statistical use (M49). United Nations Statistics Division.* https://unstats.un.org/unsd/methodology/m49/

Table A4: Inequality of Opportunity by Country

Country	Gini	IOp	Relative IOp (Share of Gini)
Austria	0.28	0.07	0.25
Belgium	0.25	0.08	0.32
Bulgaria	0.40	0.14	0.35
Switzerland	0.29	0.08	0.28
Cyprus	0.31	0.11	0.35
Czech Republic	0.24	0.05	0.21
Germany	0.32	0.12	0.38
Denmark	0.26	0.03	0.12
Estonia	0.29	0.06	0.21
Greece	0.31	0.08	0.26
Spain	0.33	0.10	0.30
Finland	0.26	0.02	0.08
France	0.29	0.07	0.24
Croatia	0.29	0.06	0.21
Hungary	0.29	0.05	0.17
Ireland	0.29	0.09	0.31
Italy	0.33	0.07	0.21
Lithuania	0.35	0.07	0.20
Luxembourg	0.32	0.13	0.41
Latvia	0.34	0.06	0.18
Malta	0.28	0.06	0.21
Netherlands	0.26	0.03	0.12
Norway	0.24	0.03	0.12
Poland	0.29	0.06	0.21
Portugal	0.31	0.09	0.29
Serbia	0.34	0.08	0.24
Sweden	0.26	0.04	0.15
Slovakia	0.23	0.04	0.17
United Kingdom	0.31	0.07	0.23

Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The table reports the overall income inequality, inequality of opportunity and relative inequality of opportunity in income estimated, through conditional random forest, within country. The random forest is estimated with α = 0.1, minbucket = 50, and 200 bagged subsamples. The rest of parameters are set as in Hothorn and Zeileis (2015). The estimates are based on the original data set relative to each country (not the one we get after the re-sampling to get the sample representative of European citizens). We have checked that inequality estimates by country closely align with those in the Global Estimates of Opportunity and Mobility Project (GEOM), with a correlation coefficient of 0.93. IOp estimates should not be directly compared because we use a different set of circumstances.

Table A5: Shapley Value Decomposition by Country

Country	Sex	Country of Origin	Father's Education	Mother's Education	Father's Occupation	Mother's Occupation
Austria	17.75	100.00	32.57	42.14	58.54	43.54
Belgium	13.07	100.00	46.25	43.65	53.84	34.21
Bulgaria	2.30	0.00	81.89	82.97	100.00	97.96
Switzerland	25.46	100.00	69.45	51.11	36.30	23.96
Cyprus	9.29	100.00	60.83	46.77	90.11	79.06
Czech Republic	24.08	0.00	67.79	60.05	65.40	100.00
Germany	16.86	59.86	75.88	93.21	100.00	69.80
Denmark	30.81	23.51	21.35	18.11	100.00	72.70
Estonia	9.08	13.26	73.90	74.75	80.92	100.00
Greece	8.22	29.01	74.08	47.28	100.00	50.05
Spain	3.60	72.05	61.71	37.29	100.00	43.07
Finland	8.01	21.27	47.95	50.65	100.00	83.42
France	12.59	42.02	63.29	27.82	100.00	23.47
Croatia	4.09	8.56	60.60	55.99	100.00	76.71
Hungary	6.32	0.00	65.64	63.40	68.97	100.00
Ireland	10.74	64.17	59.20	79.42	100.00	48.34
Italy	14.52	57.78	61.33	46.04	100.00	65.29
Lithuania	17.41	0.00	86.02	78.87	88.42	100.00
Luxembourg	6.87	100.00	30.79	26.35	37.45	22.98
Latvia	10.09	13.72	62.66	49.07	100.00	99.20
Malta	5.15	32.13	73.67	59.17	100.00	25.70
Netherlands	5.10	18.00	44.54	25.00	100.00	15.20
Norway	8.45	100.00	19.28	30.54	7.97	60.02
Poland	7.37	0.00	52.94	56.98	89.76	100.00
Portugal	2.74	0.00	25.11	21.76	100.00	55.21
Serbia	2.64	12.96	79.72	72.42	100.00	87.08
Sweden	8.82	54.43	8.47	18.61	15.84	100.00
Slovakia	18.04	0.00	49.94	61.19	85.52	100.00
United Kindom	26.91	52.64	51.44	61.72	100.00	84.40

Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The table reports the Shapley decomposition values of the income inequality of opportunity in income estimated within each country. The highest contributions are set to 100 to enhance interpretability. The rest of values are indexed accordingly.

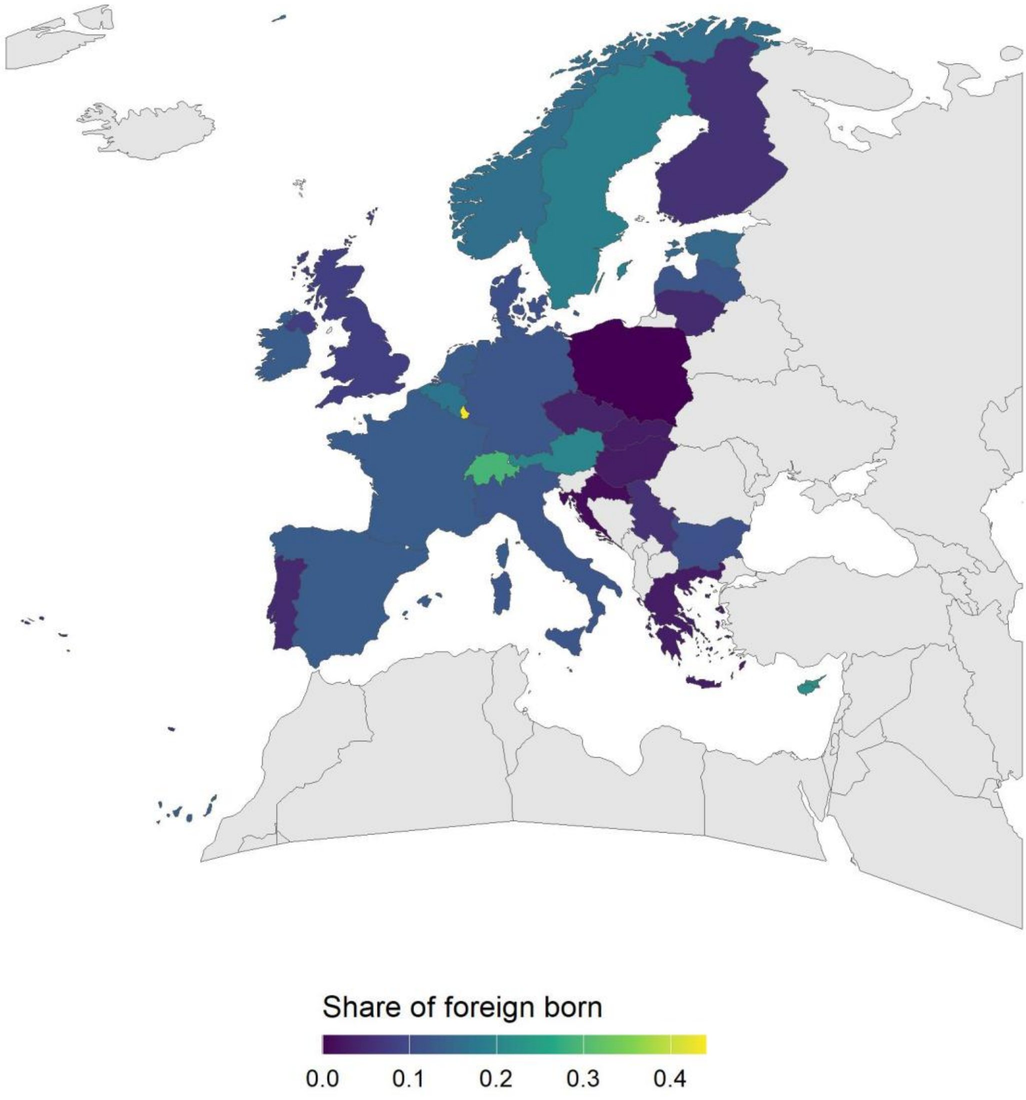
Table A6: Average Predicted Income by Country of Origin

Country of Origin	Average Predicted Income (% of the sample mean)	Country of Origin	Average Predicted Income (% of the sample mean)
Switzerland	215.66	OTHCaribbean	86.84
Norway	212.97	Turkey	85.68
Luxembourg	212.20	OTHCentralAfrica	84.49
Denmark	211.51	India	84.20
OTHNorthAmerican	146.17	Algeria	83.45
OTH	145.63	Estonia	82.35
Finland	143.56	OTHCentralAmerican	81.35
Austria	143.55	Ukraine	81.17
Sweden	143.52	Colombia	80.19
Ireland	140.73	Nigeria	78.70
OTHOceania	140.45	OTHSouthAmerican	78.58
Netherlands	136.39	Tunisia	77.78
Belgium	129.90	Philippines	72.88
France	128.65	OTHNorthAfrica	71.82
United States	123.46	Romania	70.64
OTHSouthEastAsia	123.22	OTHWestAfrica	68.18
Germany	122.87	Dominican Republic	67.94
South Africa	118.86	Venezuela	61.61
North Macedonia	115.05	Peru	61.60
OTHEastAsia	110.98	Pakistan	61.08
Kosovo	107.65	Albania	60.77
China	107.27	Senegal	60.76
Iran	106.05	Ecuador	60.59
Italy	105.74	Bangladesh	60.26
Bosnia and Herzegovina	93.73	Morocco	60.21
OTHSouthAsia	93.20	Afghanistan	59.12
Malta	91.86	Iraq	58.95
Brazil	90.74	Syria	58.57
Argentina	90.74	Portugal	57.90
Cyprus	90.50	Czechia	57.42
United Kingdom	89.64	Latvia	57.11
OTHEurope	89.48	Lithuania	52.73
OTHCentralAsia	89.38	Slovakia	49.12
OTHMiddleEast	88.94	Greece	42.60
Vietnam	88.58	Croatia	42.46
OTHEastAfrica	88.55	Poland	42.09
Russia	88.48	Serbia	32.41
Spain	88.22	Bulgaria	31.95
Sri Lanka	86.95	Hungary	31.87

Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The table reports the income predicted through conditional random forest averaged by country of origin and expressed as percentage of the European average income (21,242 2019 USD).

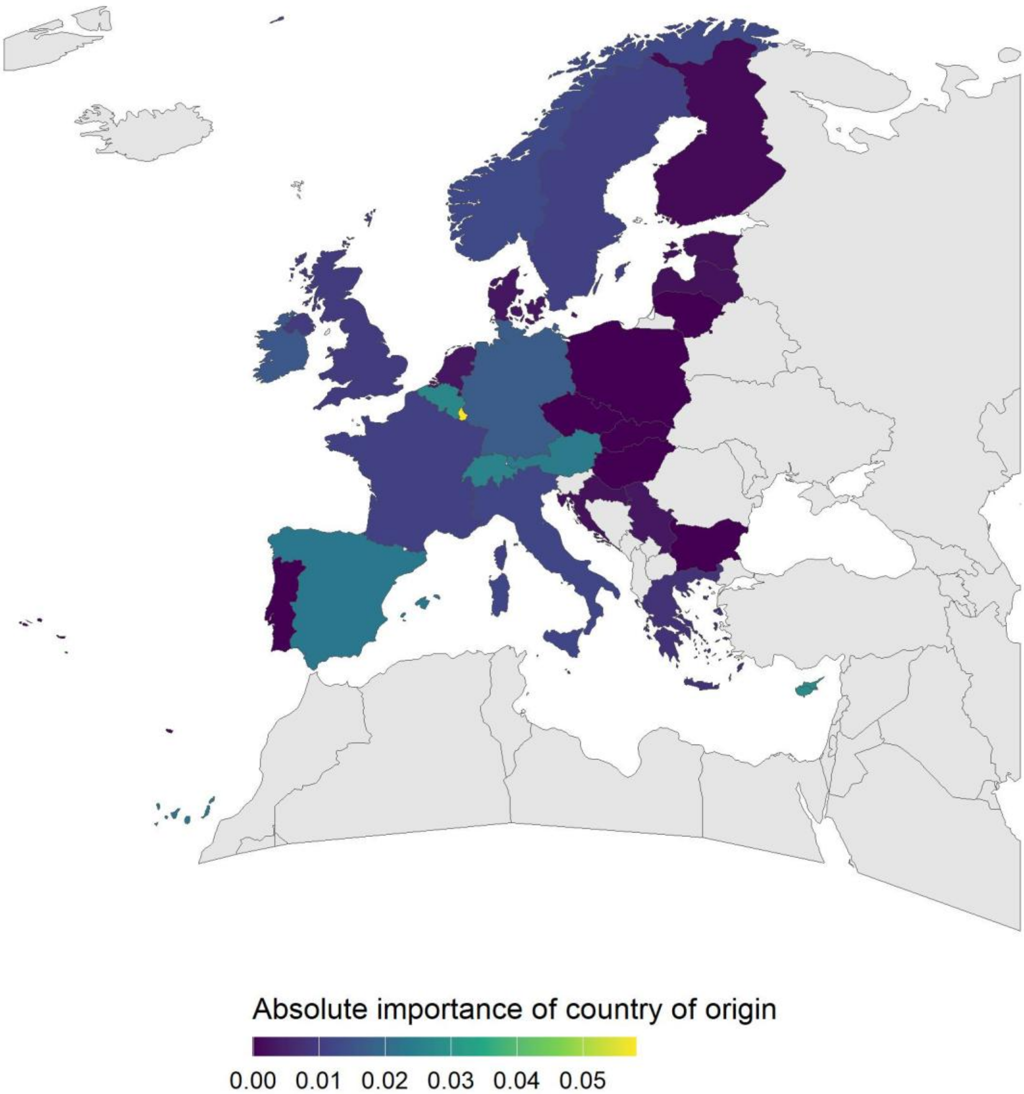
Figure Appendix

Figure A1: Share of migrants



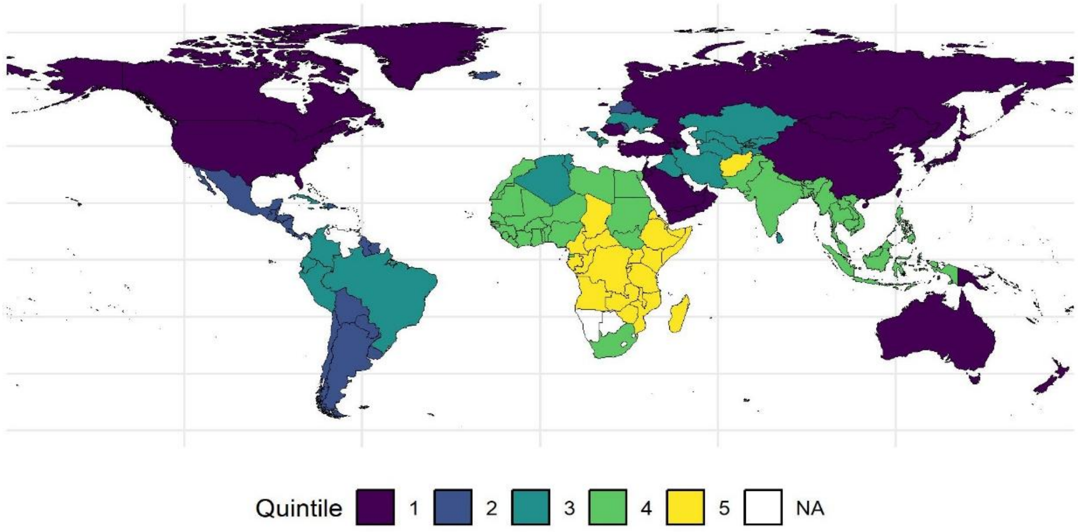
Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The share of migrants range between 0 and 0.44.

Figure A2: Absolute importance of Country of birth



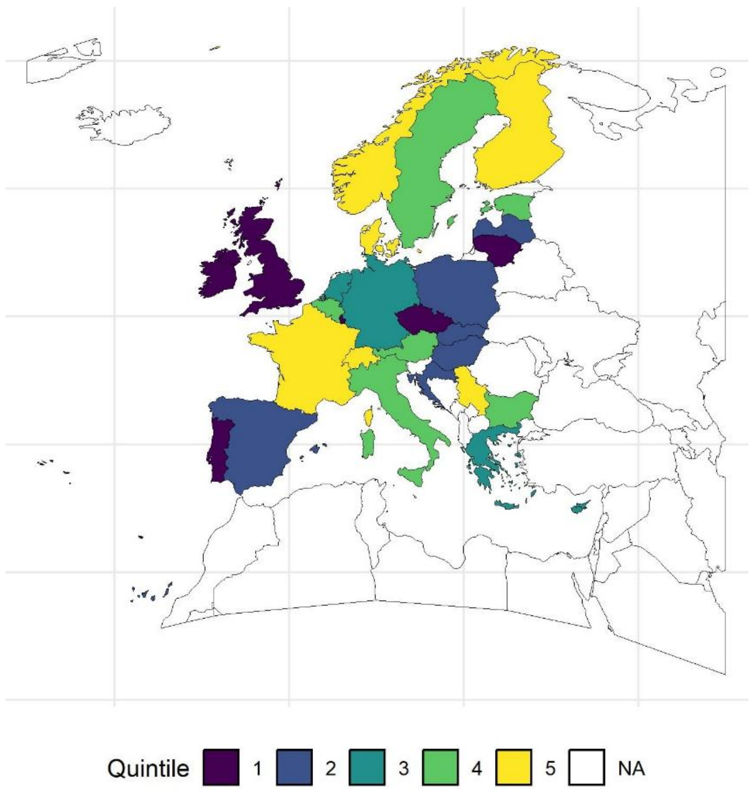
Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The Shapley value decomposition is estimated with a random forest procedure, with α = 0.1, minbucket = 50, and 100 bagged repetitions sized 0.632 the original sample size. The absolute importance is computed as the product between the Shapley decomposition value of the variable country of birth and the country's absolute level of inequality of opportunity. It ranges between 0 and 0.06 Gini points.

Figure A3: Expected income and GDP per capita by country of origin (non-European migrants)



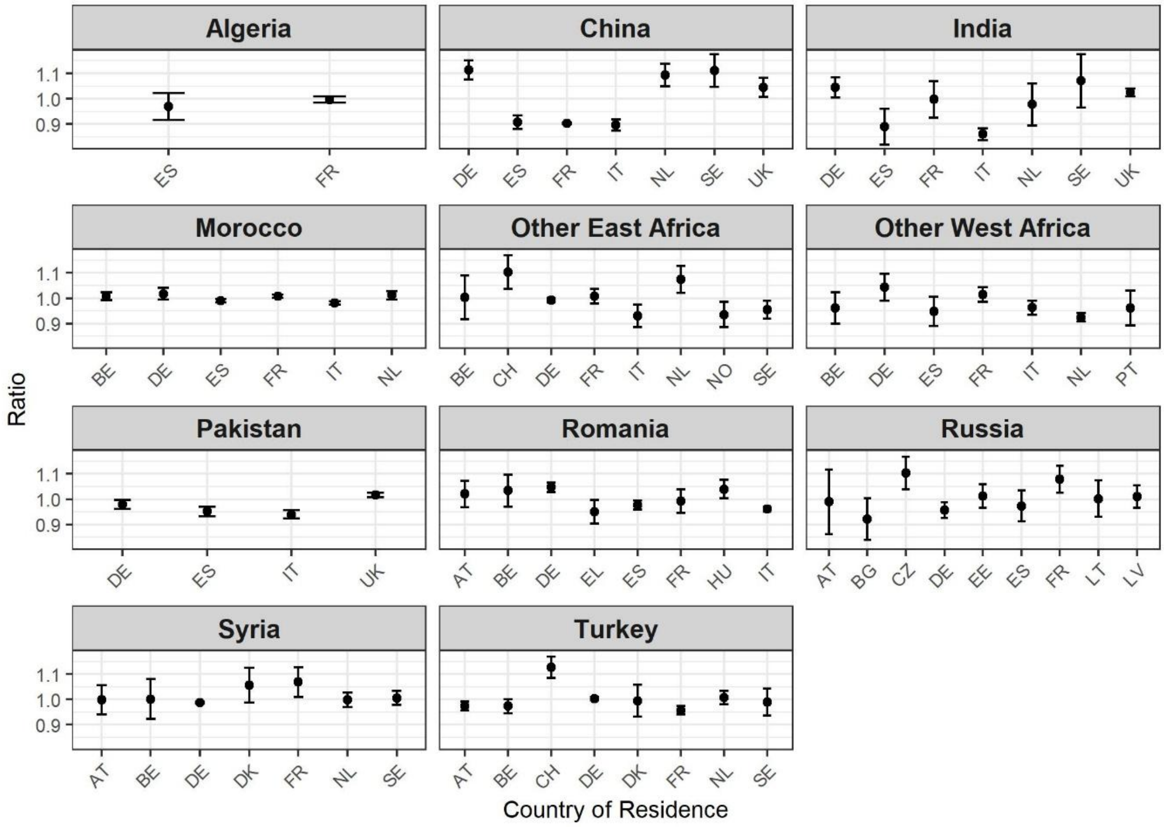
Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The map shows the ratio, by quintiles, between the average predicted income the migrants earn in Europe, and the GDP per capita of their country of origin.

Figure A4: Expected income and GDP per capita by country of origin (European migrants)



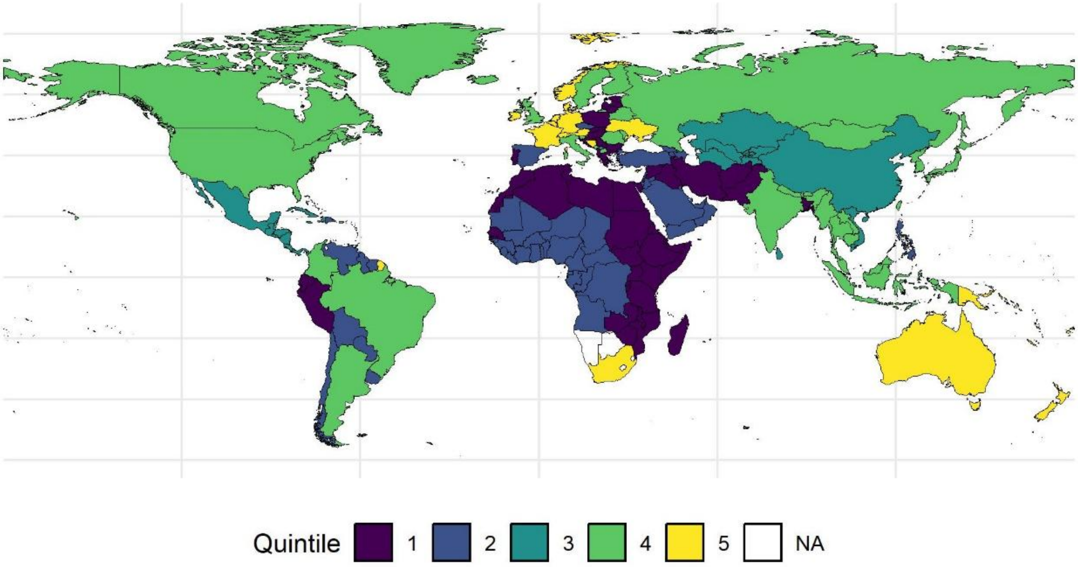
Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The map shows the ratio, by quintiles, between the average predicted income the migrants earn in Europe, and the GDP per capita of their country of origin.

Figure A5: Heterogeneity by country of destination



Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. For the most frequent countries of origin of non-European migrants, we show the ratio between the average predicted income in each country of residence and the average predicted income for the group in Europe as a whole. To simplify the exposition we show, in the vertical axis, the value of the ratio, and in the horizontal axis the countries of residence, grouping the plots by country of origin. Confidence intervals are estimated at a 95% level of significance.

Figure A6: Gap between migrants and natives predicted incomes by country of origin



Source: Own elaboration. The data comes from EU-SILC, UKHLS, GSOEP (2019), and population shares from census data. The map shows the average ratio for each country of origin of the migrants, by quintiles, between the individual predicted income of migrants and the average predicted income of the natives in the country of residence. The average values corresponding to the different quintiles of the ratio distribution: 0.57, 0.72, 0.82, 0.96, 1.24.