Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

International Variation in Inequality of Opportunity: Insights from 72 Countries

Paolo Brunori

p.brunori@lse.ac.uk

International Inequalities Institute - London School of Economics & University of Florence
https://orcid.org/0000-0002-1624-905X

Francisco Ferreira
LSE

Vito Peragine
University of Bari

Pedro Salas-Rojo
CUNEF

Domenico Moramarco
University of Bari

Louis Sirugue
International Inequalities Institute - London School of Economics https://orcid.org/0009-0001-3687-2674

Luis Barajas
Inter-American Development Bank

Teresa Barbieri
Italian Ministry of Economy and Finance

Nancy Daza-Baez
University College of London

Gaurav Datt
Monash University

Vito de Sandi
LISER

Fabio Farella
Univeristy of Bari

Arturo Martinez Jr
Asian Development Bank

John Nguyen
Monash University

Albert Park

Asian Development Bank

Enza Simeone

University of Turin https://orcid.org/0009-0007-9226-9275

Pedro Torres Lopez

International Inequalities Institute - London School of Economics

Giorgia Zotti

University of Bari

Social Sciences - Article

Keywords: Inequality of opportunity, mobility, machine learning

Posted Date: August 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10409108/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

International Variation in Inequality of Opportunity: Insights from 72 countries*

Francisco H.G. Ferreira[†], Vito Peragine[‡], Paolo Brunori[†,§], Pedro Salas-Rojo[¶], Domenico Moramarco[‡], Louis Sirugue[†], Luis Barajas[#], Teresa Barbieri[||], Nancy Daza-Báez[**], Gaurav Datt[††], Vito de Sandi[‡‡], Fabio Farella[‡], Arturo Martinez Jr.[§§], John Nguyen[††], Albert Park[§§], Enza Simeone[¶¶], Pedro Torres López[†], Giorgia Zotti[‡]

July 15, 2026

Summary

Inequality is considered particularly unfair when it stems from factors beyond individual control, such as gender, ethnicity, or family background, rather than from personal effort or choices. This is what inequality of opportunity captures. There is growing evidence that it matters both intrinsically and instrumentally: people object most strongly to this kind of inequality, and it can prevent talented individuals from disadvantaged backgrounds from reaching their full potential, holding back innovation and growth. Measuring its extent poses several challenges, and until now no systematic cross-country evidence has been put forward. We provide comparable estimates of inequality of opportunity for 72 countries, representing two-thirds of the world's population, drawn from 196 household surveys. We show that, on average, factors beyond individual control account for 40.9% of total income inequality, ranging from 18.9% in Denmark to 76.7% in South Africa, with the United States among the most opportunity-unequal countries in the global North. These results offer the first global picture of how much of the gap between rich and poor is driven by inherited or predetermined factors, providing a shared basis for judging the fairness of economic outcomes worldwide. All estimates are available at www.geom.ecineq.org, along with data-visualization tools.

Keywords: Inequality of opportunity; mobility; machine learning.

*The authors are grateful to the VélezReyes+ Plataforma Filantrópica for generous financial support, as well as for technical support with website design. We are grateful to Elisa Mansur and her team at VélezReyes+ for many useful discussions. All errors are our own.

[†]International Inequalities Institute, London School of Economics and Political Science, [‡]University of Bari, [§]University of Florence, [¶]CUNEF University, Madrid, [#]National Planning Department of Colombia, [||]Italian Ministry of Economy and Finance, [**]University College London, [††]Monash University, [‡‡]Luxembourg Institute of Socio-Economic Research, [§§]Asian Development Bank, [¶¶]University of Turin.

1 Introduction

The concept of inequality of opportunity (IOp) recognizes that not all inequalities are alike, and regards as particularly unfair those arising from factors beyond individual control such as gender, race, or family background. Concerns about this unfair kind of inequality calls for systematic cross-country evidence on its magnitude. Drawing on data from 196 household surveys, this paper provides comparable estimates of inequality of opportunity for 72 countries, representing just over two-thirds of the world's population, over a period of up to 44 years. On average, factors beyond individual control predict 40.9% of overall inequality, with considerable variation across countries and, in some regions, over time.

This kind of inequality matters both intrinsically and instrumentally. Evidence from opinion surveys and behavioral experiments suggests that people object most strongly to inequality they see as arising from factors independent of individual choices and efforts (Konow, 2000; Almås et al., 2010; Cappelen et al., 2010, 2013; Pew Research Center, 2012). Instrumentally, unequal opportunities may prevent disadvantaged children from reaching their full potential and lead to a misallocation of talents, with negative repercussions on innovation, economic efficiency, and growth (Hsieh et al., 2019; Marrero and Rodríguez, 2013).

Inequality of opportunity isolates the unfair component of inequality from that attributable to differences in effort, much as one would target "bad" but not "good" cholesterol (Ferreira, 2022). The concept of IOp is related to but distinct from that of intergenerational mobility (IGM), which captures persistence of a single dimension across generations (Björklund and Jäntti, 2020; Corak, 2013). While several studies document cross-country variation in IGM (e.g., Neidhöfer et al. (2018); van der Weide et al. (2024); Muñoz and van der Weide (2025); McNamara et al. (2026)), no comparable studies exist on IOp. We fill this gap by providing the first large dataset of comparable IOp estimates across 72 countries, drawing on the unit-record level data from 196 representative household surveys.

Our results document striking variation across countries in the share of total income inequality that can be predicted by inherited factors such as place of birth, gender, ethnicity, and family background. Our preferred estimates range from 18.9% in Denmark (2011), to 76.7% in South Africa (2017). Latin American countries are substantially overrepresented at the top of the range, while European countries mostly lie at the bottom. The latest estimate for the United States is 41.6%, slightly above the mean value of 40.9% and higher than the median (38.4%). This challenges the narrative of a "land of opportunity", in line with the results of Chetty et al. (2014) on IGM.

We find considerable differences in trends over time with, for example, a substantial increase in IOp in the United States from 1978 to 2002 and a general decline in Latin American countries. We also confirm the existence of an opportunity "Great Gatsby curve", i.e., a positive association between IOp and cross-sectional income inequality, much as found for intergenerational mobility by Corak (2013). Similarly, we find evidence for an "opportunity

Kuznets curve", i.e., an inverted-U relationship between IOp and GDP per capita.

In addition, we estimate the relative contribution of individual background variables to inequality of opportunity in each country. This analysis reveals that fathers' characteristics matter more than mothers' in Western Europe, whereas the opposite is observed in Eastern and Northern Europe. While this paper provides a concise summary of key takeaways from our estimations, all results are available as an online database at www.geom.ecineq.org.

The remainder of the paper is organized as follows. Section 2 briefly reviews the theoretical framework underpinning the IOp literature, and Section 3 describes our empirical approach. Section 4 describes the data sources and variable definitions. Section 5 presents the results, and Section 6 concludes.

2 Theoretical framework

The canonical model used in the literature to measure IOp can be described as follows.[1] Suppose that all determinants of an outcome x can be classified into either circumstances C, factors beyond individual control that shape and constrain choice sets, or effort e, factors for which individuals can be held responsible. The outcome x is generated by a function g, such that:

$$x = g(C,\, e) \tag{1}$$

This is a reduced-form model in which neither opportunities themselves, nor the structural process by which outcomes are determined, are explicitly modeled. Under equality of opportunity, circumstances ought not to influence outcomes, either directly or through their influence on efforts. Any differences across outcome distributions conditional on circumstances are therefore prima-facie evidence of inequality of opportunity (Roemer, 1998; Ferreira and Peragine, 2016).

The normative foundations of this opportunity egalitarian theory rest on two independent principles. First, the compensation principle holds that all outcome inequalities due to circumstances are unfair and should be compensated by society. Second, the reward principle holds that outcome inequalities due to effort should be preserved. Two versions of the compensation principle yield distinct approaches: ex-ante and ex-post inequality of opportunity.

According to the ex-ante approach, opportunity sets available to individuals should have the same expected value regardless of circumstances (Van de Gaer, 1993). Consider a partition of the population into groups of individuals that share the same circumstances, called types. Ex-ante inequality of opportunity is measured as inequality between type means, assigning to each individual the average outcome of their type.

[1] Different variants of this model were proposed in theoretical contributions by Fleurbaey (1994); Roemer (1993); Van de Gaer (1993); Peragine (2002) and used in the first empirical analyses of inequality of opportunity: see Bourguignon et al. (2007); Checchi and Peragine (2010); Ferreira and Gignoux (2011). See Ferreira and Peragine (2016); Roemer and Trannoy (2016) for reviews of the literature.

According to the ex-post approach, all individuals exerting the same degree of effort should achieve the same outcome (Roemer, 1998). Consider a partition of the population into groups of individuals that belong to the same quantile in their respective type distributions. These groups are called tranches, and capture the relative degree of effort. Inequality of opportunity within this approach is measured as inequality within tranches.

Different versions of the reward principle have been proposed in the literature, expressing different attitudes to the outcome inequality observed within types. The most common is the utilitarian reward, which focuses on the sum (or the average) of outcomes in each group of individuals sharing the same circumstances, and remains neutral with respect to the way differences in effort are remunerated within these groups.[2]

3 Empirical framework

The derivation of a scalar measure of inequality of opportunity follows a two-step procedure. First, the actual distribution of the outcome of interest (e.g., income), call it $\boldsymbol{x}$, is transformed into a counterfactual distribution $\tilde{\boldsymbol{x}}$, which reflects only and fully the unfair inequality in $\boldsymbol{x}$, while all the fair inequality is removed. Second, a measure of inequality is applied to $\tilde{\boldsymbol{x}}$.

3.1 Concepts of inequality of opportunity

Different versions of the counterfactual distribution can be considered, either consistent with ex-ante or ex-post compensation and with different versions of reward, as described in Section 2. They express different and sometimes conflicting views on equality of opportunity, and the distributional rankings they generate may be different (see Ferreira and Peragine (2016) for a discussion). In this paper, we report measures for two concepts of inequality of opportunity.

Concept A, our headline estimate, refers to the extensively used *between-types* inequality, which arises from the combination of ex-ante compensation with utilitarian reward. It consists in computing inequality on a counterfactual distribution $\tilde{\boldsymbol{x}}_A$ that is obtained by replacing each individual outcome by the average outcome of the type to which the individual belongs (Bourguignon et al., 2007; Checchi and Peragine, 2010; Ferreira and Gignoux, 2011). This smoothing transformation is intended to remove all inequality within types.

Concept B refers to an alternative, ex-post measure, inspired by Roemer (1993) and implemented by Checchi and Peragine (2010) and Aaberge et al. (2011). It is based on the *within-tranches* counterfactual distribution ($\tilde{\boldsymbol{x}}_B$). This distribution is obtained by replacing individual outcomes in each tranche with the ratio between that outcome and the average outcome of the tranche. This normalization procedure removes inequalities between tranches and leave unchanged the inequality within tranches.

[2] Alternative formulations have been proposed, from inequality-averse reward (Ramos and Van de Gaer, 2016; Fleurbaey et al., 2024) which incorporates some aversion to inequality even within types, to intermediate and agnostic positions (Peragine, 2002; Peragine and Serlenga, 2008; Fleurbaey and Peragine, 2013).

3.2 Counterfactual distribution estimation

The fact that estimation typically occurs in samples rather than entire populations gives rise to a tradeoff between two different kinds of bias in selecting a prediction model to construct the counterfactual distribution. With too few variables and interactions, most inequality attributable to circumstances will go undetected and IOp estimates will suffer from a downward bias. With too many, the model will overfit the data generating an upward bias (Brunori et al., 2019).

Different prediction methods have been used in the past but, until recently, these generally relied on arbitrary or *ad hoc* specifications. In addition, constraining all countries to a given pre-specified model would not be appropriate as the set of relevant circumstances is surely context-dependent. We address these challenges using a suite of data-driven tree-based supervised machine learning techniques to select optimal prediction models.

Tree-based data-driven approach

Data-driven machine learning approaches, which select optimal partitions without imposing prior specifications, are naturally suited to the estimation of IOp from survey data. Tree-based methods, in particular, offer three main advantages.

First, they partition the regressor space into non-overlapping regions, assigning individuals to mutually exclusive groups defined by the interaction of their circumstances. This is directly consistent with the notion of types having access to different sets of opportunities, as discussed in Section 2.

Second, the possibility to tune the algorithm prevents the model from becoming too complex and overfitting the data (Chakravarty and Eichhorn, 1994; Brunori et al., 2019), while maximizing *out of sample* predictive power to minimize the downward bias frequently highlighted by researchers (Ferreira and Gignoux, 2011).

Third, unlike standard recursive partitioning algorithms, trees do not tend to overuse variables with many distinct values as splitting variables, ensuring that the selection of circumstances is driven by statistical significance rather than the number of categories (Hothorn et al., 2006).

Application to both concepts of IOp

For Concept A, we partition the population into types using a Conditional Inference Tree (CIT) algorithm, which searches for the partition that maximizes the statistical significance of differences in average outcome between subsamples (Hothorn et al., 2006). It is therefore especially well-suited for Concept A, which, as discussed above, uses type means to construct the counterfactual distribution $\tilde{\boldsymbol{x}}_A$. For Concept B, we follow Brunori et al. (2026) and use Transformation Trees (Trafotrees), which are analogous to CITs, but partition the population

using the full conditional distribution of the outcome rather than only its mean, which better suited to this concept (Hothorn and Zeileis, 2021).

Although single trees are low-bias estimators, they are known to have high variance, meaning that small changes in the data could lead to different partitions and unstable estimates. By aggregating predictions across many trees grown on different subsamples of the data, forests substantially reduce variance while preserving the low bias of individual trees. Hence, we use conditional inference random forests instead of conditional inference trees to compute our headline estimates for Concept A (Hothorn et al., 2006; Brunori et al., 2023). We are not aware of suitable methods to produce forests of Trafotrees that could be used for Concept B (see Brunori et al. (2026) for a discussion).

We briefly describe the conditional inference tree, random forest, and transformation tree algorithms in Appendix A.1 and A.2 (see also Brunori et al. (2023) and Brunori et al. (2026)).

3.3 Choice of summary measure

Once the counterfactual distribution has been obtained, $\tilde{\boldsymbol{x}}_A$ for Concept A or $\tilde{\boldsymbol{x}}_B$ for Concept B, an inequality index $I(.)$ is used to obtain a scalar estimate of inequality of opportunity. The Gini coefficient and the Mean Logarithmic Deviation (MLD) are most common in the literature and we align with this practice. We report the Gini as our headline measure because it is more sensitive to the central part of the distribution, where most of the between-type inequality occurs, whereas the MLD is more sensitive to the tails.

Two closely related versions of the IOp index can then be considered. The first one is the absolute measure of inequality of opportunity $\text{IO}_\text{A} = I(\tilde{\boldsymbol{x}})$. The second one is its relative counterpart, the ratio of IO_A to overall inequality:

$$\text{IO}_\text{R} = \frac{I(\tilde{\boldsymbol{x}})}{I(\boldsymbol{x})} \tag{2}$$

IO_R can be interpreted as the share of total inequality that can be predicted by circumstances, i.e., by inherited characteristics which people cannot be held responsible for.[3]

4 Data sources

4.1 Sample definitions

We gather data from 72 countries, drawn from 196 household surveys selected balancing two key desiderata: broad country coverage and methodological comparability. We restrict our attention to household surveys because administrative data on social security records across

[3]When the Gini (G) coefficient is used, IO_R is a lower bound measure of between-types inequality, since $\text{G} = \text{G}_\text{W} + \text{G}_\text{B} + \text{G}_\text{R}$, with $\text{G}_\text{R} \geq 0$ containing elements of both types of inequality (Bhattacharya and Mahalanobis, 1967; Lambert and Aronson, 1993; Fleurbaey et al., 2026).

generations are not available in most countries. Surveys were selected according to four criteria. First, the sample has to be representative of the entire country.[4] Second, it must contain information on a money-metric wellbeing indicator, either income or consumption expenditure, at the individual or household level. Third, it has to contain individual-level information on at least five of the following seven circumstance variables: area or region of birth, sex, ethnicity,[5] occupation of the father and/or mother, and education of the father and/or mother. The fourth criterion is a minimum of 1,500 individual observations with complete information and strictly positive outcomes.[6] All samples are restricted to individuals aged 18 or older.[7]

Details of all countries, time periods, and surveys are provided in Appendix Table B.1. Time coverage varies significantly across countries, from a single survey-year for some, such as Armenia (2016), Colombia (2010), or Mali (2019), to eight or more years for others, such as the United States, Peru, or Australia. Coverage is broad for Europe, North America, South Asia, Latin America and Caribbean, and East Asia and Pacific, each representing more than 75% of their regional population, but it remains limited for the Middle East and North Africa (0.1%). In total, we cover 72 out of 217 countries, representing 66.9% of the world's population in 2022.[8] Appendix Table B.2 describes the coverage of our data in more detail.

4.2 Variable definitions

The outcome variable is equivalized household disposable income, i.e., total household income divided by the square root of household size to account for economies of scale, measured at the individual level. Where income data is unavailable, we use equivalized household consumption expenditure instead.[9] This measure ignores intra-household inequality by construction, which mechanically reduces the importance of the sex circumstance variable.

Before estimation we adjust equivalized household income (or consumption) by subtracting the contribution of age and age squared in a regression approach to control for systematic

[4]The only exception to this is Argentina, where the 2014 ENES is representative only of the country's urban areas. In 2022, the urban share of Argentina's population was 92% (United Nations, 2025).

[5]In a few cases the "ethnicity" circumstance is proxied by religion, the language spoken at home, or similar definitions. These proxies are used when explicit ethnicity data is not available but other variables are sufficiently correlated with ethnic or cultural identity.

[6]The threshold of 1,500 observations is based on the share of observations used in the resampling process for conditional inference random forests, which is set at 0.632 (see Appendix A.1). To ensure that we retain at least 1,000 observations after each resampling, we calculate the required number of observations as 1,000/0.632 $\approx$ 1,500.

[7]This decision is based on the age of consent or responsibility, which is typically 18 in many countries and is associated with the legal right to vote and be tried as an adult. The dominant view in the literature is that any inequality observed among children younger than the age of consent is inequality of opportunity (Hufe et al., 2017).

[8]The classification of countries was retrieved from the World Bank Country and Lending Groups on the 1st of September 2023: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

[9]We use equivalized household consumption expenditure for Benin, Burkina Faso, Ivory Coast, Ghana, Guinea-Bissau, Indonesia, Mali, Malawi, Niger, Nigeria, Senegal, Sierra Leone, Togo, Timor-Leste, Tanzania, and Uganda.

correlations between life cycle and the outcome distribution (Solon, 1992). For comparability purposes, all monetary values are expressed in 2017 US dollars after adjusting by Purchasing Power Parity (PPP) and the Consumer Price Index (CPI). Both adjustments are described in Appendix A.3, along with additional details.[10]

As regards the circumstance variables, all are categorical in nature and the definition of categories is inevitably specific to each country. To reduce computational cost, categorical variables are limited to a maximum of 25 values. If a circumstance variable has 26 or more categories, they are merged until 25 different values are reached. Any threat to comparability posed by these country idiosyncrasies is substantially mitigated by the data-driven nature of the algorithms we use to partition the sample.

5 Results

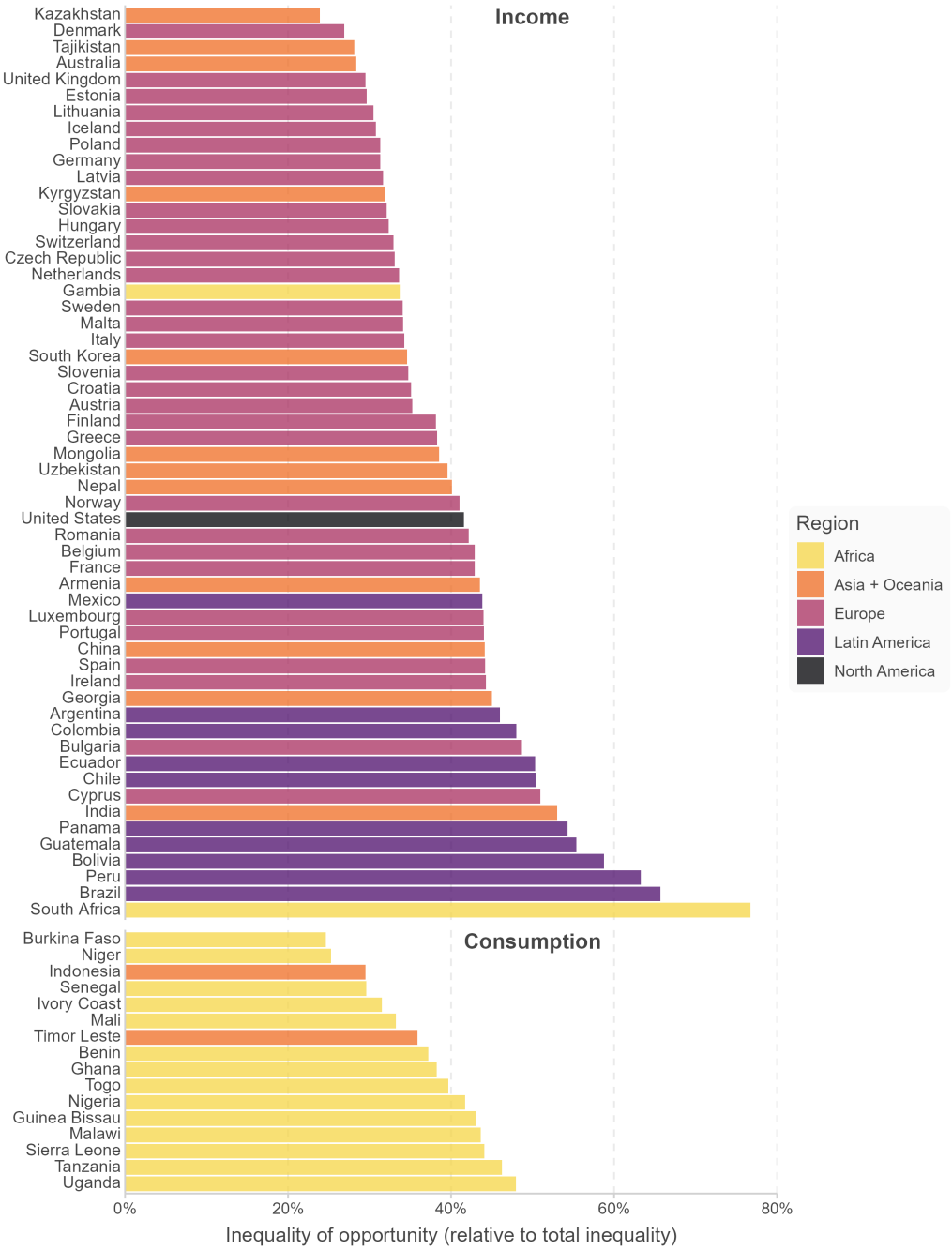
We discuss results using our preferred estimates, namely Concept A, the ex-ante approach based on the Gini index, estimated using conditional inference random forests. This choice reflects both normative and statistical considerations, as discussed in Sections 2 and 3. We provide estimates of Concepts A and B based on different inequality indices in relative and absolute terms in Appendix Tables B.3 and B.4. Estimates obtained from these different approaches are highly correlated, as illustrated in Appendix Figures B.2 and B.3. Although levels may differ, patterns and conclusions are consistent throughout, supporting the robustness of our headline findings.

5.1 Levels and trends

Figure 1 reports our preferred measure of inequality of opportunity for the latest survey available in each country. Underlying values can be found in Appendix Table B.3, Column 6, along with other measures of inequality of opportunity. Countries for which inequality of opportunity is measured in terms of income are shown on the top panel, and estimates obtained using consumption expenditure are shown on the bottom panel. These two groups of estimates are not directly comparable since consumption is typically more equally distributed than income (Alvaredo and Gasparini, 2015). This concern is attenuated for relative IOp, where the difference affects both the numerator and the denominator of the ratio. However, since these two components may not be affected to the same extent, we report income- and consumption-based estimates separately throughout.

[10]The use of an equivalence scale, the age adjustment, and our sample restrictions, may generate slightly different inequality indices from those available in other compilations. Appendix Figure B.1 plots our estimates of the Gini index of income inequality against those published by the World Bank (World Development Indicators). The two measures are highly correlated ($\rho = 0.91$), and results show relatively little deviation from the 45-degree line.

Figure 1: Inequality of opportunity – Relative to total inequality



Note: This figure plots the most recent estimates of inequality of opportunity across the 72 countries covered in this study. Estimates were computed according to Concept A, which computes inequality between groups of individuals sharing the same set of inherited characteristics (gender, race, place of birth, etc.). We determine the partition of groups non-parametrically with conditional inference random forests. A Gini inequality index is then estimated on a smoothed counterfactual distribution in which every individual is assigned the mean outcome of their type. To be expressed relative to total inequality, this measure is then divided by the Gini index estimated on the actual underlying distribution. The upper and lower panels display countries for which inequality of opportunity is estimated on income and consumption respectively.

Reading: In Brazil, the most recent estimate indicates that 65.7% of total income inequality can be predicted by inherited characteristics.

Source: Own calculations on multiple surveys listed in Appendix Table B.1.

Estimates for the most recent surveys indicate that the share of total inequality that can be predicted by inherited characteristics ranges from 23.9%, in Kazakhstan, to 76.7%, in South Africa. Across all survey years, estimates range from 18.9% in Denmark (2011), to 76.7% in South Africa (2017). There is considerable variation across regions, with Latin America substantially overrepresented at the top of the range, while Europe is overrepresented at the bottom. There is much more dispersion across Asia, with some countries like Kazakhstan and Indonesia at the bottom, and others like India or China reaching much higher values. It is difficult to compare Africa to the rest of the world since, as noted earlier, most African countries report consumption rather than income-based IOp.

The absolute counterpart of these relative IOp measures is shown in Appendix Figure B.4. The variability is, of course, higher when we look at absolute IOp, which is not normalized by the total inequality in the country. Absolute and relative IOp are, unsurprisingly, highly correlated ($\rho = 0.9$), but comparing them provides some interesting insights. For example, by looking at absolute IOp in Denmark one may conclude that, with a coefficient of 0.07, IOp is hardly a problem for this country. While this might be true when we compare it with other absolute values around the world, Figure 1 shows that IOp in Denmark still accounts for more than a quarter (26.9%) of its (rather low level of) total inequality. On the other hand, some countries, like Australia, move from the middle to the bottom part of the IOp distribution when passing from absolute to relative measures.

We investigate the evolution of relative and absolute IOp estimates over time in Figure 2 and Appendix Figure B.5 respectively. Continent-level trends are fitted through all country-year estimates available since 1998. Over the last two decades, IOp has remained relatively stable in North America, Europe, Asia, and Oceania. However, we observe a decline in IOp over the last decade in African and Latin American countries, suggesting a convergence between the global North and the global South.

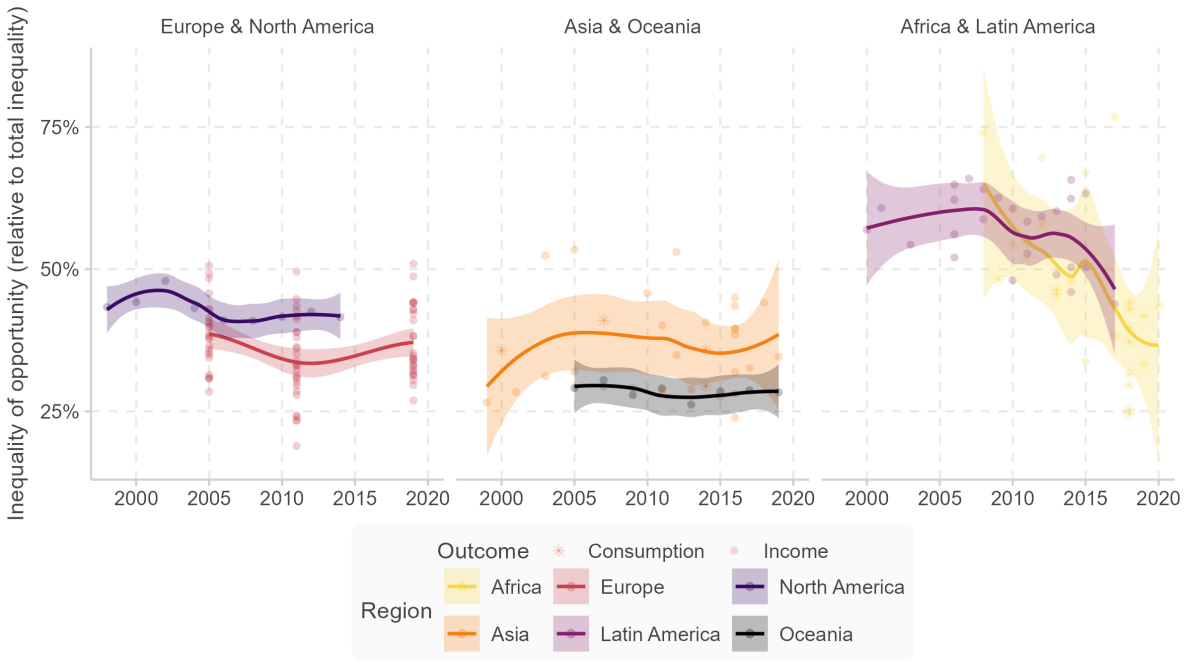
The Panel Study of Income Dynamics allows us to estimate inequality of opportunity in the United States for more than four decades. Results, presented in Appendix Figure B.6, show a steady increase in absolute IOp from 1978 to 2002, followed by a slight decline. Relative IOp, however, has remained relatively stable throughout this period, accounting for between 40 and 45% of total inequality, among the highest values for high-income economies. This challenges the image of the United States as a "land of opportunity", a conclusion also reached by Chetty et al. (2014), from the mobility side.

5.2 Empirical regularities

The opportunity Great Gatsby curve

Given the close conceptual and empirical associations between IOp and intergenerational persistence, one might expect to find a positive association between IOp and total inequality,

Figure 2: Trends in inequality of opportunity – Relative to total inequality



Note: This figure plots the trends in inequality of opportunity across 6 regions of the world. See notes to Figure 1 for a description of the estimation methods. Fitted lines and confidence bands were produced with locally estimated scatterplot smoothing (LOESS). The shape of the data points indicates whether the corresponding IOp coefficient was computed based on income (dots) or consumption expenditure (asterisks).

Reading: In the only North American country represented on this graph, the United States, 44% of total income inequality could be predicted by inherited characteristics in 2000.

Source: Own calculations on multiple surveys listed in Appendix Table B.1.

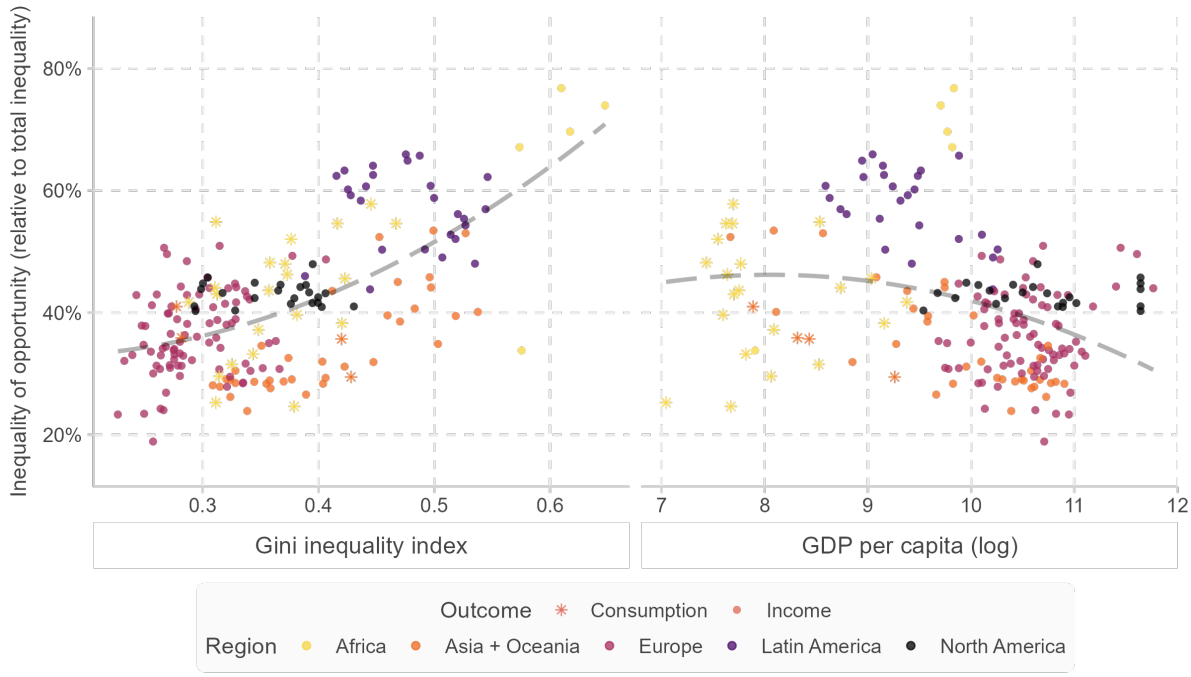
analogous to the Great Gatsby curve first plotted by Corak (2013).[11]

The left panel of Figure 3 plots inequality of opportunity against total income inequality measured two decades earlier, with each point corresponding to a country-year estimate. The results confirm a clear positive association: more unequal societies in one generation tend to exhibit higher inequality of opportunity in the next. It is also interesting to notice that a given level of income inequality is associated with more unequal opportunities in Latin American countries, which tend to cluster above the fitted line, than in Asian countries, which mostly lie below the line.

Appendix Figure B.7 shows a similar positive relationship with estimates of absolute IOp. Contrary to its relative counterpart, that positive association is partly mechanical, since absolute IOp corresponds to the between-type component of total income inequality.

[11] Some measures of intergenerational income persistence can even be interpreted as IOp under the simplifying assumption that parental income is the only circumstance beyond individual control (see Brunori et al. (2026)).

Figure 3: Opportunity Great Gatsby & Kuznets curves



Note: This figure plots inequality of opportunity across 72 countries and multiple time periods against total inequality measured two decades earlier ("Great Gatsby curve" – left panel) and log GDP per capita ("Kuznets curve" – right panel). The shape of the data points indicates whether the corresponding IOp coefficient was computed based on income (dots) or consumption expenditure (asterisks). Dashed lines represent the respective quadratic fits. See notes to Figure 1 for a description of inequality of opportunity estimation methods. All values are expressed in 2017 USD and adjusted by PPP and CPI as described in Appendix Section A.3.

Reading: Latin American countries have high shares of total income inequality that can be predicted by inherited characteristics (y-axis), high levels of total inequality (x-axis – left panel), and intermediate levels of GDP per capita (x-axis – right panel).

Source: Own calculations on multiple surveys listed in Appendix Table B.1 for inequality of opportunity and Gini index, and IMF (2024) for GDP per capita.

The opportunity Kuznets curve

We next examine how inequality of opportunity varies with economic development, proxied by GDP per capita. This question echoes the Kuznets hypothesis, which posits that inequality first tends to increase and then decrease as a country develops, leading to an inverted-U relationship between inequality and development. Given the positive association between IOp and total inequality documented above, a similar inverted-U pattern might be expected to carry over to inequality of opportunity. In addition, as argued by Ferreira et al. (2025), an inverse U-shape can emerge from economic development reducing the cost of accessing more productive technologies, and intergenerational transmission of income making it less costly for the children of wealthy parents to invest in new productive sectors.

The right panel of Figure 3 plots inequality of opportunity against the logarithm of GDP per capita, drawn from the World Economic Outlook database (IMF, 2024) and expressed in 2017 PPP US dollars to align with our IOp estimates. The pattern is suggestive of an inverted

U-shape relationship, in line with the mechanism proposed by Ferreira et al. (2025). Indeed, the coefficient on the squared term is negative and marginally significant (-1.13, p-value = 0.06). The corresponding pattern for absolute IOp, shown in Appendix Figure B.7, is more pronounced, which possibly reflects the mechanical relationship between absolute IOp and total inequality.

5.3 Most important determinants

In addition to documenting differences in the magnitude of IOp across countries and over time, we investigate variations in the nature of IOp by assessing the relative importance of individual circumstances. To do so, we rely on Shapley-Shorrocks decompositions. This method, which is described in detail in Appendix A.4, consists in calculating each variable's contribution to predicting variation in the outcome by assessing the average decline in explained outcome variability when the variable is excluded. A weighted average of these drops provides the Shapley value (Shapley, 1952; Shorrocks, 2013). To account for sample dependency, this process is repeated 100 times, and the results are averaged across iterations. Of course, resulting estimates are of descriptive interest but cannot be interpreted as causal.

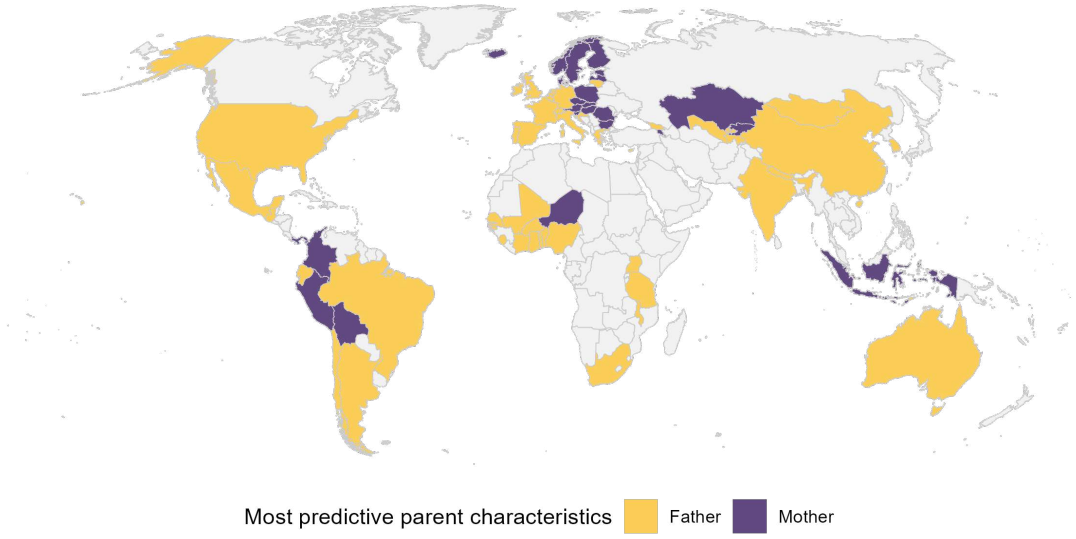
In Figure 4 we show for each country whether the education and occupation of the mother contribute more to inequality of opportunity than those of the father, or less. Interestingly, Europe seems rather neatly divided between the West, where father's characteristics dominate, and the East and North, where mother's characteristics dominate. Investigating this pattern is beyond the scope of this paper, but it may reflect the relatively high labor-force participation of women in former planned economies and Nordic countries during the 1970s and 1980s (Jaumotte, 2004; Campa and Serafinelli, 2019). In other continents, father characteristics dominate in most cases.

6 Conclusion

Not all forms of inequality are the same, and there is growing evidence that inequality of opportunity, which is mostly inherited and reflects factors beyond individual control, is particularly harmful. This paper provides comparable estimates of IOp across 72 countries, which together account for just over two-thirds of the world's population. We compute estimates on household-level microdata from 196 different household surveys, following definitions and protocols designed for maximum comparability. We use data-driven machine learning tools to compute these indices, given their superior properties in trading off upward (overfitting) and downward (omitted variable) estimation biases in the measurement of IOp. We discuss the results corresponding to our preferred measure of IOp, namely Concept A, which refers to the ex-ante approach based on the Gini index, using conditional inference random forests.

There is considerable variation in this measure across countries, ranging from 18.9% of to-

Figure 4: Main contributor to inequality of opportunity across countries



Note: This figure plots which of mothers' or fathers' characteristics (education and occupation) contribute the most to inequality of opportunity in each of the 72 countries covered in this study. See notes to Figure 1 for a description of inequality of opportunity estimation methods. Relative contribution is assessed with Shapley-Shorrocks decompositions, by comparing the average decline in explained outcome variability when each variable is excluded.

Reading: In France, the education and occupation of the mother contribute less to inequality of opportunity than those of the father.

Source: Own calculations on multiple surveys listed in Appendix Table B.1.

tal inequality (Denmark, 2011), to 76.7% (South Africa, 2017). Across the entire sample, IOp is positively correlated with overall inequality, confirming an "opportunity Great Gatsby curve". With respect to per capita GDP, IOp first rises and then falls, describing an inverted-U curve rather like that postulated by the Kuznets hypothesis for income inequality. There are sharp regional differences in IOp, with European (and some Asian) countries displaying lower levels, while Latin American (and some African) countries are at the upper range. Among high-income countries, the United States stands out for its high levels of inequality of opportunity. Our approach also allows us to quantify the contribution of individual background variables in each country. While mothers' characteristics are more important than fathers' in Eastern and Northern European countries, the opposite is observed in Western Europe.

This paper focuses on our preferred set of IOp estimates, but we compute several IOp measures that reflect different normative approaches, use different methods to partition the population into types, and different inequality indices. All estimates are available for download as an online database at www.geom.ecineq.org, where data visualization tools can help interested readers interact with the data.

References

Aaberge, Rolf, Magne Mogstad, and Vito Peragine, "Measuring long-term inequality of opportunity," *Journal of Public Economics*, 2011, *95* (3-4), 193–204.

Almås, Ingvild, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden, "Fairness and the development of inequality acceptance," *Science*, 2010, *328* (5982), 1176–1178.

Alvaredo, Facundo and Leonardo Gasparini, "Recent trends in inequality and poverty in developing countries," *Handbook of income distribution*, 2015, *2*, 697–805.

Bhattacharya, Nath and Bimalendu Mahalanobis, "Regional disparities in household consumption in India," *Journal of the American Statistical Association*, 1967, *62* (317), 143–161.

Björklund, Anders and Markus Jäntti, "Intergenerational mobility, intergenerational effects, sibling correlations, and equality of opportunity: A comparison of four approaches," *Research in Social Stratification and Mobility*, 2020, *70*, 100455.

Bourguignon, François, Francisco HG Ferreira, and Marta Menéndez, "Inequality of opportunity in Brazil," *Review of Income and Wealth*, 2007, *53* (4), 585–618.

Breiman, Leo, "Random forests," *Machine learning*, 2001, *45* (1), 5–32.

Brunori, Paolo and Guido Neidhöfer, "The evolution of inequality of opportunity in Germany: A machine learning approach," *Review of Income and Wealth*, 2021, *67* (4), 900–927.

_ , Francisco HG Ferreira, and Pedro Salas-Rojo, "Inherited inequality and the distribution of opportunities in the USA, China, India and South Africa," *Economica*, 2026.

_ , Paul Hufe, and Daniel Mahler, "The roots of inequality: Estimating inequality of opportunity from regression trees and forests," *The Scandinavian Journal of Economics*, 2023, *125* (4), 900–932.

_ , Vito Peragine, and Laura Serlenga, "Upward and downward bias when measuring inequality of opportunity," *Social Choice and Welfare*, 2019, *52* (4), 635–661.

Campa, Pamela and Michel Serafinelli, "Politico-Economic Regimes and Attitudes: Female Workers under State Socialism," *The Review of Economics and Statistics*, 2019, *101* (2), 233–248.

Cappelen, Alexander W, Erik Ø Sørensen, and Bertil Tungodden, "Responsibility for what? Fairness and individual responsibility," *European Economic Review*, 2010, *54* (3), 429–441.

_ , James Konow, Erik Ø Sørensen, and Bertil Tungodden, "Just luck: An experimental study of risk-taking and fairness," *American Economic Review*, 2013, *103* (4), 1398–1413.

Chakravarty, Satya R and Wolfgang Eichhorn, "Measurement of income inequality: Observed versus true data." In *Models and measurement of welfare and inequality*. pp. 28–32, Berlin, Heidelberg: Springer Berlin Heidelberg, 1994.

Checchi, Daniele and Vito Peragine, "Inequality of opportunity in Italy," *The Journal of Economic Inequality*, 2010, *8* (4), 429–450.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner, "Is the United States still a land of opportunity? Recent trends in intergenerational mobility," *American Economic Review*, 2014, *104* (5), 141–147.

Chow, Gregory C, "Tests of equality between sets of coefficients in two linear regressions," *Econometrica: Journal of the Econometric Society*, 1960, pp. 591–605.

Corak, Miles, "Income inequality, equality of opportunity, and intergenerational mobility," *Journal of Economic Perspectives*, 2013, *27* (3), 79–102.

Datt, Gaurav, John Nguyen, Pedro Salas Rojo, Francisco HG Ferreira, Paolo Brunori, Vito Peragine, Albert Park, Arturo Martinez Jr, and Joseph Bulan, "Measuring inequality of opportunity in Asia and the Pacific," International Inequalities Institute Working Paper, LSE, 2026.

Ferreira, Francisco H. G., "Not all inequalities are alike," *Nature*, 2022, *606*, 646–649.

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

_ **and Vito Peragine**, "Individual responsibility and equality of opportunity." In *The Oxford Handbook of Well-Being and Public Policy*. pp. 622–657, Oxford University Press, 2016.

_ **, Domenico Moramarco, and Vito Peragine**, "Economic development and inequality of opportunity: Kuznets meets the Great Gatsby?," *Structural Change and Economic Dynamics*, 2025, *76*, 94–114.

Fleurbaey, Marc, "On fair compensation," *Theory and Decision*, 1994, *36* (3), 277–307.

_ **and Vito Peragine**, "*Ex ante* versus *ex post* equality of opportunity," *Economica*, 2013, *80* (317), 118–130.

_ **, Domenico Moramarco, and Vito Peragine**, "Measuring inequality and welfare when some inequalities matter more than others," ECARES Working Papers, 2024.

_ **, Peter Lambert, Domenico Moramarco, and Vito Peragine**, "Inequality decomposition analysis, the Lorenz curve and the Gini coefficient: M. Fleurbaey et al.," *Social Choice and Welfare*, 2026, *66* (2), 489–511.

Hothorn, Torsten, "trtf: Transformation trees and forests," *CRAN*, 2023.

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

_ **and** _ , "Predictive distribution modeling using transformation forests," *Journal of Computational and Graphical Statistics*, 2021, *30* (4), 1181–1196.

_ **, Heidi Seibold, and Achim Zeileis**, "partykit: A toolkit for recursive partitioning," *CRAN*, 2023.

_ , **Kurt Hornik, and Achim Zeileis**, "Unbiased recursive partitioning: A conditional inference framework," *Journal of Computational and Graphical Statistics*, 2006, *15* (3), 651–674.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

Hufe, Paul, Andreas Peichl, John Roemer, and Martin Ungerer, "Inequality of income acquisition: The role of childhood circumstances," *Social Choice and Welfare*, 2017, *49* (3), 499–544.

IMF, "World Economic Outlook, October 2024," Washington, DC: International Monetary Fund, 2024.

Jaumotte, Florence, "Labour force participation of women: Empirical evidence on the role of policy and other determinants in OECD countries," *OECD Economic studies*, 2004, *2003* (2), 51–108.

Konow, James, "Fair shares: Accountability and cognitive dissonance in allocation decisions," *American Economic Review*, 2000, *90* (4), 1072–1092.

Lambert, Peter J and J Richard Aronson, "Inequality decomposition analysis and the Gini coefficient revisited," *The Economic Journal*, 1993, *103* (420), 1221–1227.

Marrero, Gustavo A and Juan G Rodríguez, "Inequality of opportunity and growth," *Journal of Development Economics*, 2013, *104*, 107–122.

McNamara, Sarah, Guido Neidhöfer, and Patrick Lehnert, "Intergenerational mobility fosters innovation in Europe," *Nature*, 2026, pp. 1–7.

Muñoz, Ercio and Roy van der Weide, "Intergenerational income mobility around the world," World Bank Policy Research Working Paper, 2025, *11166*.

Neidhöfer, Guido, Joaquín Serrano, and Leonardo Gasparini, "Educational inequality and intergenerational mobility in Latin America: A new database," *Journal of Development Economics*, 2018, *134*, 329–349.

Peragine, Vito, "Opportunity egalitarianism and income inequality," *Mathematical Social Sciences*, 2002, *44* (1), 45–64.

_ **and Laura Serlenga**, "Higher education and equality of opportunity in Italy," *Inequality And Opportunity: Papers From The Second Ecineq Society Meeting*, 2008, pp. 67–97.

Pew Research Center, "Trends in American values: 1987–2012. Partisan polarization surges in Bush, Obama years," https://www.pewresearch.org/wp-content/uploads/sites/4/legacy-pdf/06-04-12-Values-Release.pdf, 2012, June 4.

Ramos, Xavier and Dirk Van de Gaer, "Approaches to inequality of opportunity: Principles, measures and evidence," *Journal of Economic Surveys*, 2016, *30* (5), 855–883.

Roemer, John E, "A pragmatic theory of responsibility for the egalitarian planner," *Philosophy & Public Affairs*, 1993, pp. 146–166.

_ , *Theories of distributive justice*, Harvard University Press, 1998.

— **and Alain Trannoy**, "Equality of opportunity: Theory and measurement," *Journal of Economic Literature*, 2016, *54* (4), 1288–1332.

Shapley, Lloyd S, "A value for n-person games," *Proceedings of the National Academy of Sciences*, 1952.

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

Solon, Gary, "Intergenerational income mobility in the United States," *The American Economic Review*, 1992, pp. 393–408.

United Nations, "World urbanization prospects," *Population Division of the Department of Economic and Social Affairs*, 2025.

Van de Gaer, Dirk FG, "Equality of opportunity and investment in human capital," PhD Dissertation, Katholieke Universiteit Leuven, 1993.

van der Weide, Roy, Christoph Lakner, Daniel Gerszon Mahler, Ambar Narayan, and Rakesh Gupta, "Intergenerational mobility around the world: A new database," *Journal of Development Economics*, 2024, *166*, 103167.

Online Appendix

International Variation in Inequality of Opportunity: Insights from 72 countries

A Estimation methods **1**

A.1 Concept A: Ex-ante (between-type) inequality of opportunity 1

A.2 Concept B: Ex-post (within tranches) inequality of opportunity 3

A.3 Variable adjustments 6

A.4 Shapley-Shorrocks decompositions 7

B Additional tables and figures **9**

A Estimation methods

This section provides additional details on the empirical approach followed in this paper. Sections A.1 and A.2 describe the tree-based machine learning algorithms used to partition populations into types. Most recent resulting country estimates are presented in Appendix Tables B.3 and B.4. Section A.3 documents the adjustment of the outcome variables to account for age and purchasing power parity, along with additional details. Section A.4 describes the algorithm used to perform the Shapley-Shorrocks decompositions.

A.1 Concept A: Ex-ante (between-type) inequality of opportunity

The ex-ante IOp statistical approach we adopt was proposed by Brunori et al. (2023). It employs the Conditional Inference regression Trees (CIT) and Conditional Inference Random Forests (CForest) developed by Hothorn et al. (2006). A CIT is a supervised machine learning algorithm aimed at partitioning a regressor space to predict the variation of a dependent variable. The regression space is partitioned delivering a set of terminal nodes or leaves obtained by recursive binary splitting of the sample. The algorithm can be summarized as follows:

1. Set a confidence level $(1-\alpha)$.[12]

2. Test the correlation between the dependent variable (outcome) and each regressor (circumstance). If for all observed regressors, the Bonferroni-adjusted *p-value* of the correlation test is higher than the chosen critical value α, exit. Otherwise, go to step 3.

3. Among all regressors in which the null hypothesis of independence is rejected, select the variable whose correlation with the outcome has the smallest *p-value* as splitting variable $[c]$.

4. Consider how circumstance $[c]$ can be used to partition the sample into two subsamples $[s_i,\ s_{-i}]$. Let S_c denote the set of all possible binary splits of the sample based on $[c]$. For each possible binary partition, compute the *p-value* for the null hypothesis that the mean is the same in the two sub-samples $\left(\mathrm{p}^{[s_i,\, s_{-i}]}\right)$.

5. Choose $[s_i,\ s_{-i}]^* = argmin_{S_c}\ \mathrm{p}^{[s_j,\, s_{-j}]},\ \forall j$ as the most appropriate partition.

6. Repeat steps 2 – 5 for each resulting node (sub-sample) until the null hypothesis of step 2 cannot be rejected in any resulting sub-sample.

[12] We set $\alpha = 0.01$. We impose an additional stopping requirement, namely that each terminal node must have a minimum of 1% of the observations in the sample (or 50 if the sample size is smaller than 5,000). This country-specific minimum is set to minimize the effect of different sample sizes on the depth of the tree. See Brunori et al. (2023) for a discussion of the effect of sample size on IOp estimation. All remaining parameters are the default values in the "ctree" R function in the package "partykit" (Hothorn and Zeileis, 2015; Hothorn et al., 2023).

The output of this algorithm consists of an exhaustive partition of the sample into mutually exclusive groups. We treat these terminal nodes as empirical types $t = \{1, ..., T\}$, for which we compute the respective population weighted share $\hat{w}_t$ and the weighted mean $\hat{\mu}_t$.[13] Ex-ante IOp is estimated as $\widehat{IO}_\alpha = I(\tilde{x})$, where $\tilde{x}$ is a counterfactual distribution obtained by replacing the outcome of individual i belonging to type t with $\tilde{x}_{i,t} = \hat{\mu}_t$. We select the Gini coefficient as the reference inequality measure, but we also report mean logarithmic deviation (MLD) estimates.

In addition to producing an estimate of ex-ante IOp, the CIT approach has the significant added benefit that the partitioning process itself contains interesting information on the structure of inequality of opportunity within a particular observed population. We show a CIT visualization in Figure A.1, which is an example from Sweden, 2019. For simplicity, we normalize the expected outcome in each node dividing it by the sample mean, so a value higher than 1 can be interpreted as an expected outcome higher than the expected outcome in the entire population. For instance, type 4 in Figure A.1 has an expected outcome approximately twice as large as the sample average. The GEOM database available at https://geom.ecineq.org/ contains similar trees for all countries and time periods.

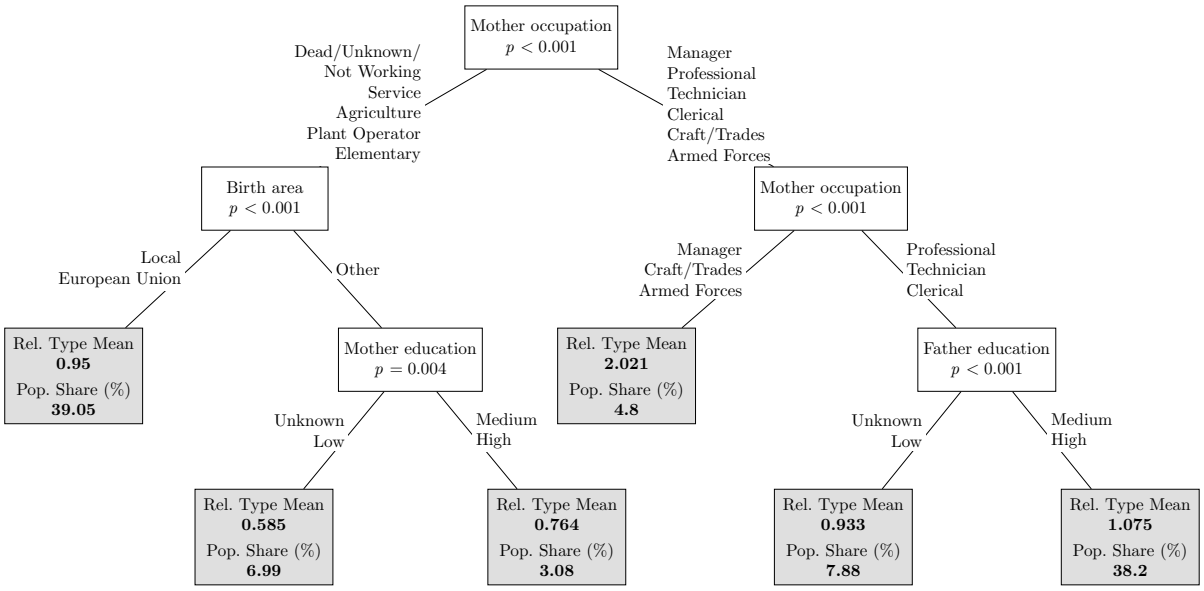
Ex-ante IOp using random forests

Like any other tree-based method, CITs are low-bias but high-variance learners. Therefore, an aggregation procedure can improve the reliability of estimates. In this situation, it is standard practice in the machine learning literature to use resampling methods such as random forests. A random forest draws different subsamples of the original data and computes a tree on each one. Under the appropriate aggregation procedures, this process smooths sample dependency and generates robust IOp estimates.

We implement this method by aggregating 200 CITs into a conditional inference random forest. Following Breiman (2001) and Hothorn et al. (2006), we set α to 1 (mincriterion, $1-\alpha = 0$), such that each tree is free to grow as much as it can. We use the default 0.632 share of each subsample drawn in every iteration. The minimum number of observations that we allow in each terminal node is 0.1% of the sample size, with the aim of maximizing comparability across surveys with different sample size (or 10, if the sample size was smaller than 1,000). All remaining tuning parameters are set to the default values in the "cforest" R function in "partykit" (Hothorn et al., 2023).

[13] We do not use weights to determine splits. Including sampling weights expands the sample size, such that individual observations turn into hundreds or thousands of identical values. As a result, the tree becomes very deep, as null hypotheses are easily rejected. Weights are used to calculate the values of the counterfactual distribution and to estimate IOp.

Figure A.1: Conditional Inference regression Tree example (Sweden, 2019)



Note: This figure illustrates how a conditional inference regression tree partitions the sample into types. When regressing income on every circumstance, the occupation of the mother is the one whose coefficient has the smallest p-value, reported in the corresponding node at the base of the tree. Among all possible binary splits of the sample according to mother education, the one yielding the most statistically significant difference splits categories listed on the left of the node from those listed on the right of the node. This process is iterated over every pair of subsamples until regressing the outcome on every circumstance generates no statistically significant coefficients. Terminal nodes, shaded with a gray color, report the average income of the corresponding type relative to the average income in the population, along with the population share of the type.

Source: Own elaboration on data from EUSILC (2019).

A.2 Concept B: Ex-post (within tranches) inequality of opportunity

To define the most appropriate partition in Roemerian types according to Concept B, we follow the method proposed by Brunori et al. (2026) based on the Transformation Trees (Trafotrees) introduced by Hothorn and Zeileis (2021). As indicated in Section 3, Trafotrees are analogous to CITs, but the sequence of tests used to grow the tree concerns the full conditional distribution of the outcome, instead of only the conditional expectation. Trafotrees estimate those distribution functions for each possible partition. The algorithm can be summarized as follows:

1. Set a confidence level $(1 - \alpha)$.[14]

2. Set a Bernstein polynomial order m.[15]

[14] As for the CIT, we follow the convention and set $\alpha = 0.01$. We impose the additional requirement that each terminal node must have a minimum of 1% of the observations in the sample (or 50 if the sample size is smaller than 5,000). This country-specific minimum is set to minimize the effect of different sample sizes on the depth of the tree. For robustness, we run a second tree relaxing the previous requirement, such that minimum observations imposed in each terminal node is 0.1%. If the types obtained are different from those with the node-size restriction, we store the plot for further inspection. All remaining parameters are the default set in the "*ctree*" R function in the package "partykit" (Hothorn et al., 2023).

[15] Unlike CITs, transformation trees require the practitioner to choose the order of the Bernstein polynomial

3. Approximate the shape of the unconditional distribution of the outcome using a Bernstein polynomial of order m. Store the $m+1$ parameter values.

4. Run an M-fluctuation test to assess the instability of the parameters conditional on the value of each circumstance.[16] If the Bonferroni-adjusted *p-value* of the association test is higher than the chosen critical value α, exit. Otherwise, continue to step 5.

5. Among all regressors in which the null hypothesis of independence is rejected, select the variable producing the smallest adjusted *p-value* as splitting variable $[c]$.

6. Consider how circumstance $[c]$ can be used to partition the sample into two subsamples $[C]$. Among all possible binary partitions, compute the *p-value* for the null hypothesis that the distribution is the same in the two sub-samples $(p^{[C]})$.

7. Choose $[C]^* = \left\{[C] : argmin\ p^{[C]}\right\}$ as the most appropriate partition.

8. Repeat steps 3 – 7 for each resulting node (sub-sample) until the null hypothesis of step 4 cannot be rejected in any resulting sub-sample.

We use the "trafotree" R function developed by Hothorn and Zeileis (2021). The output of the estimation consists of a partition in types, that allows us to obtain a parametric estimate of each type's cumulative income distribution function.[17] These parametric conditional distributions can be inverted to yield the predicted type quantile functions $\hat{x}_{qc} = F^{-1}\left(q, \hat{\theta}(c)\right)$, from which a measure of ex-post inequality of opportunity can be computed as $\widehat{IOp} = I_q\left(\frac{\hat{\mu}}{\hat{\mu}_q}\hat{x}_{qc}\right)$.

We do not use weights to determine splits. Including sampling weights expands the sample size, such that individual observations are turned into hundreds or thousands of identical values. As a result, the tree becomes very deep, as null hypotheses are easily rejected. Weights are used to calculate the values of the counterfactual distribution and to estimate IOp.

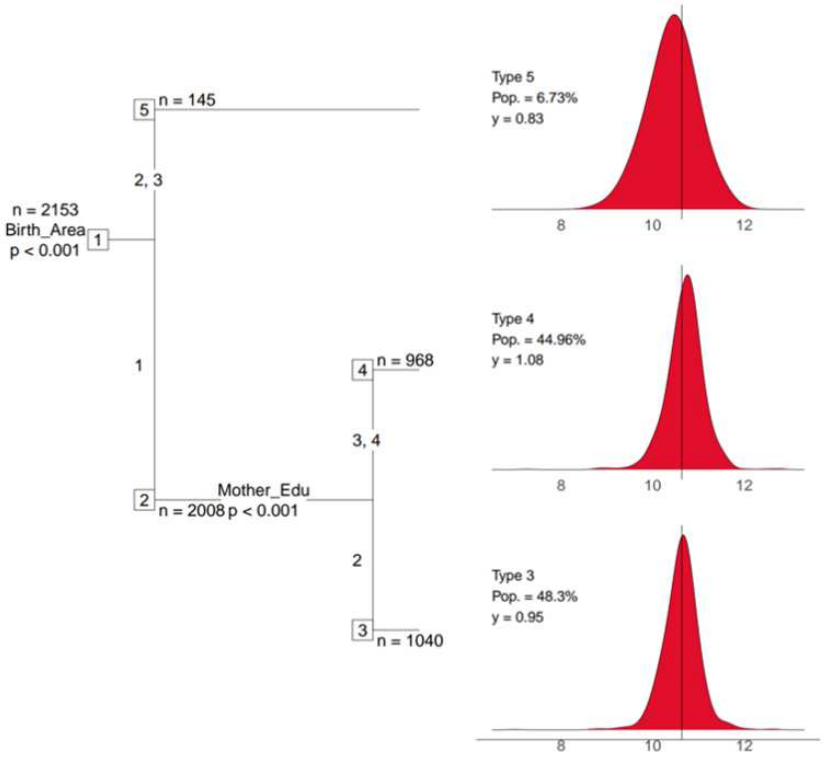
Just like CITs, Trafotrees can also be graphically represented, providing information on the structure of inequality of opportunity within a particular observed population, now based on differences on the outcome distributions. We show a Trafotree visualization in Figure A.2. The type-specific parametric CDFs obtained in Trafotree can be represented as in Figure A.3. They can also be aggregated into the overall density function as a mixture of type distributions.

used to approximate the type-specific conditional distribution functions. We choose that order by setting a minimum improvement in the aggregate out-of-sample log-likelihood of 0.1%. All other parameters are the default parameters in the "*Trtf*" R function in the package "Trtf" (Hothorn, 2023).

[16] This test can be understood as a test to detect a structural break in a time series (Chow, 1960). The cumulative sum of residuals is used to detect whether the parameters are unstable over time, while in this case the instability is investigated conditioning on the values or categories of each circumstance (see Hothorn and Zeileis (2021) for details).

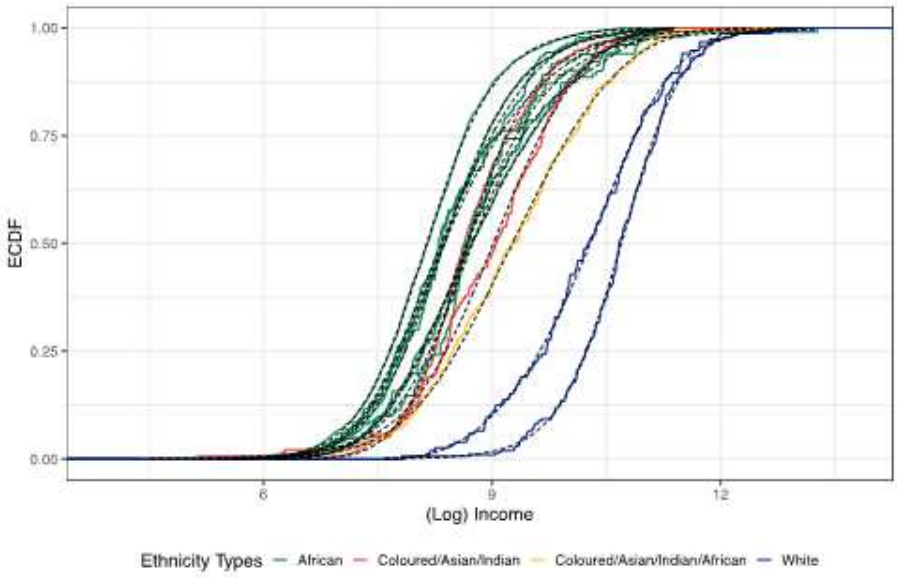
[17] The same interpolation was implemented in Brunori and Neidhöfer (2021) on a partition of types obtained with CIT.

Figure A.2: Transformation tree example (Denmark, 2011)



Source: Own elaboration on data from EUSILC (2011).

Figure A.3: Type CDFs example (South Africa, 2017)



Source: Brunori et al. (2026).

A.3 Variable adjustments

Age adjustment

Before estimation we adjust equivalized household income (or consumption) to control for systematic correlations between life cycle and the outcome distribution (Solon, 1992). We use a regression approach where, for each observation i, we subtract from the dependent variable x_i the predictions obtained from the regression of (log) dependent variable on age and age squared (Datt et al., 2026), as follows.[18]

$$Ln(x_j) = \alpha + \beta age_j + \gamma age_j^2 + \varepsilon_j \tag{3}$$

$$x_i^{adj} = exp(Ln(x_i) - \hat{\beta} age_i - \hat{\gamma} age_i^2) \tag{4}$$

PPP adjustment

All inequality measures we report are scale-invariant, but the database does contain scale-sensitive information, such as the mean incomes of different types. To enable cross-country and temporal comparisons in these variables, all monetary values are expressed in 2017 US dollars after adjusting by Purchasing Power Parity (PPP) and the Consumer Price Index (CPI). We use Stata to download the PIP (Poverty and Inequality Platform at the World Bank) series for CPI and PPP.[19] If the CPI value from PIP is missing, we use the CPI series provided by the World Bank (*Consumer Price Index (2010 = 100)*, downloaded on September 5th, 2023), after modifying the base year to 2017. The PPP values come directly from PIP with no further modifications. The national currencies are adjusted in this manner:

$$\frac{NationalCurrency}{PPP_{2017} \times \left(\frac{CPI_{year}}{100}\right)} = 2017\,USD \tag{5}$$

Additional details

Finally, for certain countries, our different time periods come from panel data where the same individuals or households are interviewed in multiple waves (e.g., Australia, South Africa and South Korea). In these cases, inconsistencies in responses regarding retrospective circumstances (such as a parent's educational attainment) can occasionally arise. To resolve these inconsistencies, we set the value to that reported in the first available wave. For example, if an

[18] To avoid including young adults who may earn only a small share of the household income and thereby introduce biases in the adjustment, the age-adjustment regression is run on household heads, indexed by j in Equation 3. After running the regression and adjusting incomes for all individuals i, we rescale the adjusted incomes to match the sample mean. We use the household head as reported in the survey; if this information is unavailable, we treat the respondent as the household head.

[19] We use the PIP Stata command ("ssc install pip") and execute "pip tables, table(ppp) clear" and "pip tables, table(cpi) clear" to obtain the PIP series for PPP and CPI (base 2017=100). The version we use was downloaded on September 5th, 2023. We thank Daniel Gerszon Mahler for his help.

individual reports in 2012 that her mother was illiterate, but in 2018 reports that her mother attended primary school, we assume the information from 2012 is correct, as it is (i) less prone to recall bias, and (ii) more likely to have held when the respondent was a young child. Additionally, when missing observations are encountered for time-invariant circumstances where information is available for the same individual from other waves of the panel, we use this available information, always prioritizing the oldest information (i.e., from earlier waves).

A.4 Shapley-Shorrocks decompositions

Since there is no guarantee that the contributions of all circumstance variables are additively separable, a plausible approach to identifying individual contributions is through a Shapley value decomposition (see Shapley (1952); Shorrocks (2013)). Intuitively, a Shapley value decomposition calculates the overall contribution of each variable c, included among explanatory variables, to the variability of some outcome x. The Shapley value of variable c is calculated as the average decline in the explained variability of x resulting from all possible combinations of ways in which it can be explained without including c among explanatory variables.[20]

We follow Brunori et al. (2026) and obtain the Shapley value decompositions as follows:

1. Draw a sub-sample of the full sample. To favor computational speed, the sub-sample should consist of 5,000 observations, or 90% of the original sample size if it was smaller than 5,000.

2. Estimate IOp in the sub-sample by estimating a CIT, allowing the tree to overfit and get deep ($\alpha = 0.9$), and minimum sample size in terminal nodes defined as in the conditional inference random forest (0.1% of the sample size).

3. Re-estimate IOp in the sub-sample for all possible elimination sequences of each circumstance. Elimination of one or more circumstances is obtained by replacing their values with a vector of **1**.

4. Estimate the difference between the overall IOp and the new IOp values obtained after different elimination sequences of each circumstance. Estimate the weighted average of these differences as the contribution of c.

5. Accounting for possible data or sample dependencies, repeat steps 1 – 4 one hundred times.

6. The final estimate of the contribution of c to IOp is the average contribution across 100 iterations.

[20] Note that according to Shapley (1952) each elimination sequence has a different probability that is used as weight to obtain such average decline.

Note that contributions of each circumstance are reported in relative terms. Absolute values are not directly comparable with IOp estimated with a single tree, because the sample sizes are smaller and confidence level is lower. We perform the decomposition estimating 100 trees on different sub-samples and we calculate average values across iterations. This procedure makes our estimates robust to the high variance typical of a single tree.

B Additional tables and figures

Table B.1: Primary data sources

Country	Year(s) covered	Survey and acronym	Research team
Argentina	2014	Encuesta Nacional sobre la Estructura Social (ENES)	CEDLAS
Armenia	2016	Life in Transition Survey (LITS)	ADB & Monash
Australia	2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019	Household, Income and Labour Dynamics in Australia (HILDA)	ADB & Monash
Austria	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Belgium	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Benin	2018	Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM)	Bari
Bolivia	2008	Encuesta de Hogares (EH)	CEDLAS
Brazil	2014	Pesquisa Nacional por Amostra de Domicilios Contínua Anual (PNAD)	CEDLAS
Bulgaria	2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Burkina Faso	2018	Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM)	Bari
Chile	2006, 2011, 2013, 2015	Encuesta de Caracterización Socioeconómica Nacional (CASEN)	CEDLAS
China	2010, 2012, 2014, 2016, 2018	China Family Panel Studies (CFPS)	LSE
Colombia	2010	Encuesta Nacional de Calidad de Vida (ENCV)	CEDLAS
Croatia	2011	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Cyprus	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Czech Republic	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Denmark	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Ecuador	2006, 2014	Encuesta Condiciones de Vida (ECV)	CEDLAS
Estonia	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Finland	2005, 2011, 2019	European Union Statistics on Income	LSE

		and Living Conditions (EU-SILC)	
France	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Gambia	2015	Integrated Household Survey (IHS)	Bari
Georgia	2016	Life in Transition Survey (LITS)	ADB & Monash
Germany	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Ghana	2013, 2017	Ghana Living Standard Survey (GLSS)	Bari
Greece	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Guatemala	2000, 2006, 2011	Encuesta Nacional sobre Condiciones de Vida (ENCOVI)	CEDLAS
Guinea Bissau	2018	Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM)	Bari
Hungary	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Iceland	2005	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
India	2005, 2012	India Human Development Survey (IHDS)	ADB & Monash
Indonesia	2000, 2014	Indonesian Family Life Survey (IFLS)	ADB & Monash
Ireland	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Italy	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Ivory Coast	2018	Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM)	Bari
Kazakhstan	2016	Life in Transition Survey (LITS)	ADB & Monash
Kyrgyzstan	2016	Life in Transition Survey (LITS)	ADB & Monash
Latvia	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Lithuania	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Luxembourg	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Malawi	2020	Malawi Fifth Integrated Household Survey (MFIHS)	Bari
Mali	2019	Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM)	Bari
Malta	2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE

Mexico	2017	Encuesta de Movilidad Social (EMOVI)	Provided by CEEY
Mongolia	2016	Life in Transition Survey (LITS)	ADB & Monash
Nepal	2003, 2011	Nepal Living Standards Survey (NLSS)	ADB & Monash
Netherlands	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Niger	2018	Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM)	Bari
Nigeria	2019	Nigeria Living Standards Survey (NLSS)	Bari
Norway	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Panama	2003	Encuesta de Niveles de Vida (ENV)	CEDLAS
Peru	2001, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	Encuesta Nacional de Hogares sobre Condiciones de Vida y Pobreza (ENAHO)	CEDLAS
Poland	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Portugal	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Romania	2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Senegal	2018	Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM)	Bari
Sierra Leone	2011, 2018	Sierra Leone Integrated Household Survey (SLIHS)	Bari
Slovakia	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Slovenia	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
South Africa	2008, 2012, 2015, 2017	National Income Dynamics Study (NIDS)	LSE
South Korea	1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019	Korean Labour and Income Panel Study (KLIPS)	ADB & Monash
Spain	2005, 2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Sweden	2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
Switzerland	2011, 2019	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE

Tajikistan	2016	Life in Transition Survey (LITS)	ADB & Monash
Tanzania	2009, 2011, 2013	National Panel Survey (NPS)	Bari
Timor-Leste	2007, 2014	Timor-Leste Survey of Living Standards (TSLS)	ADB & Monash
Togo	2018	Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM)	Bari
Uganda	2010, 2011, 2012, 2014	National Panel Survey (NPS)	Bari
UK	2005, 2011	European Union Statistics on Income and Living Conditions (EU-SILC)	LSE
USA	1970, 1972, 1974, 1976, 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014	Panel Study of Income Dynamics (PSID)	LSE
Uzbekistan	2016	Life in Transition Survey (LITS)	ADB & Monash

Note: This table provides information about the data sources we use. It includes the country name, the year, and the official name of each survey. The last column refers to the research team responsible for obtaining, cleaning, and harmonizing the data.
ADB & Monash University: Gaurav Datt, Arturo Martinez Jr., John Nguyen, Albert Park.
CEDLAS: Matías Ciaschi.
CEEY: Pedro Torres López.
LSE: Luis Barajas, Paolo Brunori, Nancy Daza-Báez, Francisco H.G. Ferreira, Pedro Salas-Rojo, Louis Sirugue, Pedro Torres López.
University of Bari: Teresa Barbieri, Vito de Sandi, Fabio Farella, Domenico Moramarco, Vito Peragine, Enza Simeone, Giorgia Zotti.
Source: Own elaboration.

Table B.2: Coverage

	Number of countries *(World Bank)*	Number of countries *(This study)*	Share of countries (%)	Share of population (%)
	Panel A: Geographical Macro Regions			
East Asia & Pacific	37	6	16.22	75.26
Europe & Central Asia	58	36	62.07	65.96
Latin America & Caribbean	42	10	23.81	82.9
Middle East & North Africa	21	1	4.76	0.11
North America[a]	3	1	33.33	89.53
South Asia	8	2	25	75.43
Sub-Saharan Africa	48	16	33.33	49.34
Total	*217*	*72*	*33.18*	*66.89*
	Panel B: Economic Macro Regions			
High Income	83	35	42.17	77.4
Upper-middle Income	54	14	25.93	82.26
Lower-middle Income	54	14	25.93	59.36
Low Income	26	9	34.62	22.93
Total	*217*	*72*	*33.18*	*66.89*

Note: The classification of countries was retrieved from the World Bank Country and Lending Groups. The population data refers to 2022 and comes from the World Bank Open Data Repository (variable SP.POP.TOTL). [a]Mexico is included in Latin America and Caribbean. North America includes Canada, the United States and Bermuda.

Source: Own elaboration.

Table B.3: Most recent estimates of inequality of opportunity – Based on Gini index

Country	Year	Variable	Sample Gini	Concept A: Ex-ante IOp		Concept B: Ex-post IOp	
				Absolute	Relative (%)	Absolute	Relative (%)
Argentina	2014	Income	0.388	0.179	46.0	0.177	45.6
Armenia	2016	Income	0.412	0.179	43.6	0.184	44.7
Australia	2019	Income	0.355	0.101	28.4	0.081	22.8
Austria	2019	Income	0.280	0.099	35.2	0.091	32.4
Belgium	2019	Income	0.243	0.104	42.9	0.107	44.1
Benin	2018	Consumption	0.348	0.130	37.2	0.133	38.0
Bolivia	2008	Income	0.500	0.294	58.8	0.341	68.1
Brazil	2014	Income	0.487	0.320	65.7	0.339	69.6
Bulgaria	2019	Income	0.406	0.198	48.7	0.190	46.7
Burkina Faso	2018	Consumption	0.379	0.093	24.6	0.099	26.1
Chile	2015	Income	0.492	0.248	50.3	0.309	62.8
China	2018	Income	0.497	0.219	44.1	0.292	58.8
Colombia	2010	Income	0.535	0.257	48.0	0.333	62.2
Croatia	2011	Income	0.306	0.107	35.1	0.118	38.6
Cyprus	2019	Income	0.315	0.160	50.9	0.171	54.5
Czech Rep.	2019	Income	0.239	0.079	33.1	0.073	30.5
Denmark	2019	Income	0.268	0.072	26.9	0.057	21.2
Ecuador	2014	Income	0.455	0.229	50.3	0.273	60.0
Estonia	2019	Income	0.280	0.083	29.6	0.081	29.0
Finland	2019	Income	0.287	0.109	38.2	0.093	32.6
France	2019	Income	0.286	0.123	42.9	0.118	41.4
Gambia	2015	Income	0.576	0.195	33.8	0.234	40.6
Georgia	2016	Income	0.469	0.211	45.0	0.178	38.0
Germany	2019	Income	0.279	0.087	31.3	0.074	26.5
Ghana	2017	Consumption	0.420	0.161	38.2	0.172	40.8
Greece	2019	Income	0.306	0.117	38.3	0.126	41.2
Guatemala	2011	Income	0.526	0.291	55.4	0.330	62.7
Guinea Bissau	2018	Consumption	0.312	0.134	43.0	0.137	43.9

Hungary	2019	Income	0.275	0.089	32.4	0.080	29.1
Iceland	2005	Income	0.263	0.081	30.8	0.075	28.3
India	2012	Income	0.527	0.279	53.0	0.327	62.1
Indonesia	2014	Consumption	0.428	0.126	29.5	0.101	23.7
Ireland	2019	Income	0.281	0.124	44.3	0.117	41.7
Italy	2019	Income	0.315	0.108	34.2	0.125	39.7
Ivory Coast	2018	Consumption	0.325	0.102	31.5	0.101	31.1
Kazakhstan	2016	Income	0.339	0.081	23.9	0.063	18.5
Kyrgyzstan	2016	Income	0.448	0.143	31.9	0.176	39.2
Latvia	2019	Income	0.337	0.107	31.7	0.103	30.7
Lithuania	2019	Income	0.341	0.104	30.5	0.103	30.2
Luxembourg	2019	Income	0.322	0.142	44.0	0.134	41.7
Malawi	2020	Consumption	0.357	0.156	43.6	0.170	47.6
Mali	2019	Consumption	0.344	0.114	33.2	0.122	35.4
Malta	2019	Income	0.266	0.091	34.1	0.091	34.4
Mexico	2017	Income	0.445	0.195	43.8	0.195	43.9
Mongolia	2016	Income	0.470	0.181	38.5	0.176	37.5
Nepal	2011	Income	0.538	0.216	40.1	0.263	48.9
Netherlands	2019	Income	0.255	0.086	33.6	0.106	41.6
Niger	2018	Consumption	0.311	0.079	25.3	0.082	26.4
Nigeria	2019	Consumption	0.288	0.120	41.7	0.132	45.7
Norway	2019	Income	0.273	0.112	41.0	0.107	39.3
Panama	2003	Income	0.527	0.286	54.3	0.335	63.5
Peru	2015	Income	0.422	0.267	63.3	0.286	67.8
Poland	2019	Income	0.282	0.088	31.3	0.092	32.8
Portugal	2019	Income	0.306	0.135	44.0	0.135	44.3
Romania	2019	Income	0.341	0.144	42.2	0.198	58.1
Senegal	2018	Consumption	0.314	0.093	29.6	0.102	32.5
Sierra Leone	2018	Consumption	0.311	0.137	44.1	0.144	46.2
Slovakia	2019	Income	0.232	0.075	32.1	0.086	37.1
Slovenia	2019	Income	0.249	0.086	34.7	0.091	36.7
South Africa	2017	Income	0.610	0.468	76.7	0.495	81.3
South Korea	2019	Income	0.351	0.121	34.6	0.145	41.4

Spain	2019	Income	0.329	0.145	44.2	0.159	48.3
Sweden	2019	Income	0.276	0.094	34.0	0.099	36.1
Switzerland	2019	Income	0.283	0.093	32.9	0.073	26.0
Tajikistan	2016	Income	0.309	0.087	28.1	0.051	16.6
Tanzania	2013	Consumption	0.373	0.173	46.3	0.171	45.8
Timor Leste	2014	Consumption	0.282	0.101	35.9	0.113	40.1
Togo	2018	Consumption	0.382	0.151	39.7	0.152	39.8
Uganda	2014	Consumption	0.371	0.178	48.0	0.177	47.7
United Kingdom	2011	Income	0.324	0.096	29.5	0.087	26.8
United States of America	2014	Income	0.395	0.164	41.6	0.165	41.7
Uzbekistan	2016	Income	0.460	0.182	39.5	0.092	19.9

Note: This table presents the most recent estimates of inequality of opportunity across the 72 countries covered in this study according to different methodological approaches. All estimates in this table use the Gini index to measure inequality on both the smoothed counterfactual distribution and the actual outcome distribution. Columns (1), (2), and (3) indicate the country, year, and outcome used to compute the estimate. Column (4) indicates the overall Gini inequality index. Column (5) indicates the inequality of opportunity index in absolute terms obtained following Concept A, i.e., the ex-ante approach and using conditional inference random forests to partition the population into types. Column (6) shows the same results relative to total inequality, i.e., divided by Column (4). Column (7) indicates the inequality of opportunity index in absolute terms obtained following Concept B, i.e., the ex-post approach and using transformation trees to partition the population into types. Column (8) shows the same results relative to total inequality, i.e., divided by Column (4).

Reading: Inequality of opportunity in Argentina in 2014, as measured according to Concept A based on the Gini coefficient, amounts to 0.179.

Source: Own calculations on multiple surveys listed in Appendix Table B.1.

Table B.4: Most recent estimates of inequality of opportunity – Based on Mean Logarithmic Deviation

Country	Year	Variable	Sample MLD	Concept A: Ex-ante IOp		Concept B: Ex-post IOp	
				Absolute	Relative (%)	Absolute	Relative (%)
Argentina	2014	Income	0.276	0.050	18.1	0.051	18.3
Armenia	2016	Income	0.309	0.056	18.1	0.091	29.5
Australia	2019	Income	0.222	0.015	7.0	0.011	4.9
Austria	2019	Income	0.155	0.016	10.3	0.016	10.1
Belgium	2019	Income	0.106	0.019	18.0	0.021	20.1
Benin	2018	Consumption	0.199	0.032	16.1	0.033	16.6
Bolivia	2008	Income	0.505	0.144	28.6	0.237	47.0
Brazil	2014	Income	0.427	0.163	38.1	0.191	44.8
Bulgaria	2019	Income	0.296	0.062	20.9	0.058	19.7
Burkina Faso	2018	Consumption	0.235	0.025	10.5	0.024	10.0
Chile	2015	Income	0.458	0.095	20.8	0.232	50.7
China	2018	Income	0.459	0.076	16.6	0.172	37.5
Colombia	2010	Income	0.547	0.106	19.4	0.252	46.1
Croatia	2011	Income	0.177	0.018	10.4	0.022	12.4
Cyprus	2019	Income	0.169	0.041	24.4	0.049	28.9
Czech Rep.	2019	Income	0.099	0.010	9.9	0.008	8.4
Denmark	2019	Income	0.142	0.008	5.6	0.005	3.6
Ecuador	2014	Income	0.377	0.082	21.8	0.142	37.7
Estonia	2019	Income	0.150	0.011	7.1	0.011	7.4
Finland	2019	Income	0.143	0.019	13.4	0.014	9.7
France	2019	Income	0.145	0.028	19.3	0.025	17.0
Gambia	2015	Income	0.670	0.058	8.7	0.112	16.7
Georgia	2016	Income	0.381	0.069	18.1	0.053	13.9
Germany	2019	Income	0.138	0.012	8.9	0.009	6.7
Ghana	2017	Consumption	0.314	0.042	13.2	0.051	16.4
Greece	2019	Income	0.166	0.022	13.6	0.032	19.2
Guatemala	2011	Income	0.519	0.136	26.3	0.251	48.4
Guinea Bissau	2018	Consumption	0.158	0.029	18.3	0.029	18.5

Hungary	2019	Income	0.146	0.012	8.5	0.010	7.0
Iceland	2005	Income	0.129	0.011	8.4	0.011	8.2
India	2012	Income	0.518	0.123	23.7	0.207	40.0
Indonesia	2014	Consumption	0.309	0.026	8.6	0.020	6.5
Ireland	2019	Income	0.136	0.024	17.6	0.022	15.9
Italy	2019	Income	0.207	0.019	9.2	0.026	12.6
Ivory Coast	2018	Consumption	0.173	0.022	12.6	0.019	11.3
Kazakhstan	2016	Income	0.203	0.010	5.2	0.008	4.0
Kyrgyzstan	2016	Income	0.362	0.032	8.9	0.168	46.3
Latvia	2019	Income	0.214	0.018	8.2	0.019	8.9
Lithuania	2019	Income	0.220	0.017	7.6	0.018	8.4
Luxembourg	2019	Income	0.194	0.032	16.5	0.031	16.0
Malawi	2020	Consumption	0.208	0.051	24.6	0.056	27.2
Mali	2019	Consumption	0.190	0.027	14.1	0.030	15.9
Malta	2019	Income	0.126	0.013	10.3	0.014	11.3
Mexico	2017	Income	0.349	0.060	17.1	0.067	19.2
Mongolia	2016	Income	0.412	0.053	12.8	0.057	13.8
Nepal	2011	Income	0.513	0.075	14.7	0.122	23.8
Netherlands	2019	Income	0.115	0.012	10.2	0.031	27.2
Niger	2018	Consumption	0.160	0.027	16.6	0.027	16.6
Nigeria	2019	Consumption	0.138	0.023	16.4	0.027	19.4
Norway	2019	Income	0.149	0.020	13.6	0.020	13.2
Panama	2003	Income	0.636	0.132	20.7	0.209	32.9
Peru	2015	Income	0.327	0.113	34.6	0.136	41.6
Poland	2019	Income	0.146	0.012	8.3	0.013	9.1
Portugal	2019	Income	0.172	0.030	17.4	0.030	17.5
Romania	2019	Income	0.233	0.033	14.0	0.074	31.8
Senegal	2018	Consumption	0.160	0.022	13.7	0.024	15.3
Sierra Leone	2018	Consumption	0.158	0.036	22.6	0.036	22.6
Slovakia	2019	Income	0.107	0.009	8.6	0.017	15.8
Slovenia	2019	Income	0.110	0.012	10.8	0.013	12.1
South Africa	2017	Income	0.690	0.360	52.1	0.413	59.8
South Korea	2019	Income	0.224	0.025	11.1	0.062	27.6

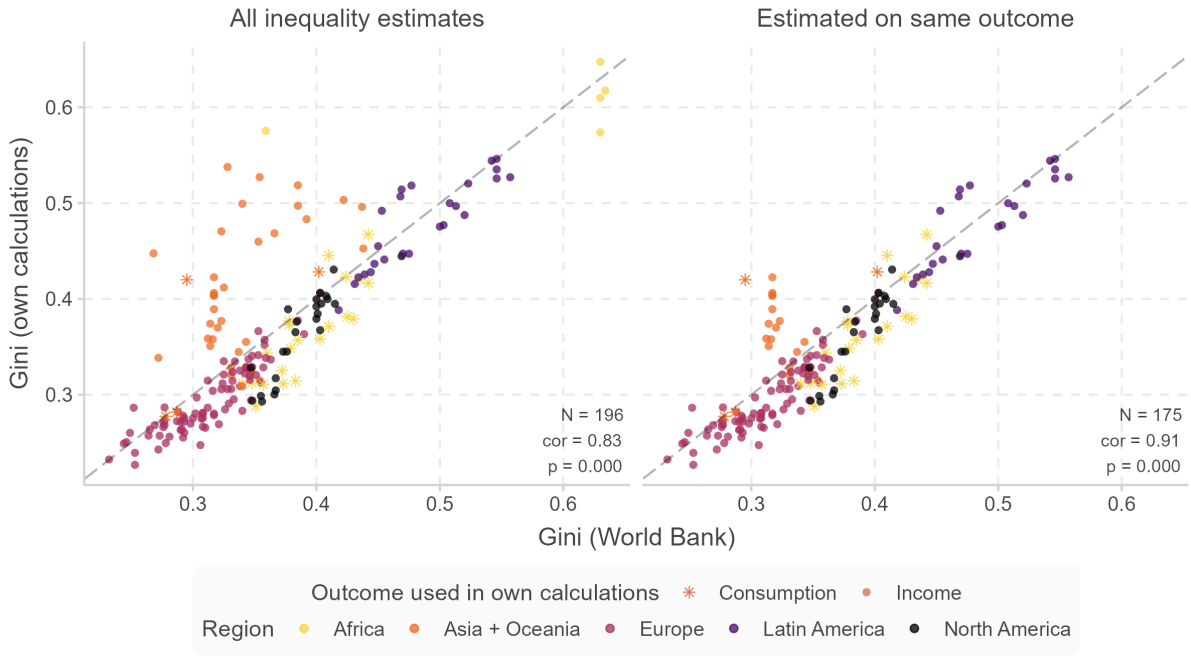
Spain	2019	Income	0.211	0.035	16.8	0.045	21.2
Sweden	2019	Income	0.166	0.017	10.0	0.021	12.5
Switzerland	2019	Income	0.145	0.015	10.1	0.009	6.5
Tajikistan	2016	Income	0.160	0.012	7.4	0.008	5.2
Tanzania	2013	Consumption	0.235	0.047	19.9	0.046	19.7
Timor Leste	2014	Consumption	0.129	0.017	13.4	0.021	16.1
Togo	2018	Consumption	0.242	0.037	15.4	0.037	15.2
Uganda	2014	Consumption	0.229	0.053	23.2	0.056	24.5
United Kingdom	2011	Income	0.187	0.015	7.9	0.016	8.5
United States of America	2014	Income	0.290	0.044	15.2	0.054	18.5
Uzbekistan	2016	Income	0.363	0.051	14.2	0.013	3.7

Note: This table presents the most recent estimates of inequality of opportunity across the 72 countries covered in this study according to different methodological approaches. All estimates in this table use the mean logarithmic deviation to measure inequality on both the smoothed counterfactual distribution and the actual outcome distribution. Columns (1), (2), and (3) indicate the country, year, and outcome used to compute the estimate. Column (4) indicates the overall mean logarithmic deviation. Column (5) indicates the inequality of opportunity index in absolute terms obtained following Concept A, i.e., the ex-ante approach and using conditional inference random forests to partition the population into types. Columns (6) shows the same results expressed relative to total inequality, i.e., divided by Column (4). Column (7) indicates the inequality of opportunity index in absolute terms obtained following Concept B, i.e., the ex-post approach and using transformation trees to partition the population into types. Column (8) shows the same results expressed relative to total inequality, i.e., divided by Column (4).

Reading: The inequality of opportunity index in absolute terms in Argentina in 2014, as measured according to Concept A based on mean logarithmic deviation, amounts to 0.050.

Source: Own calculations on multiple surveys listed in Appendix Table B.1.

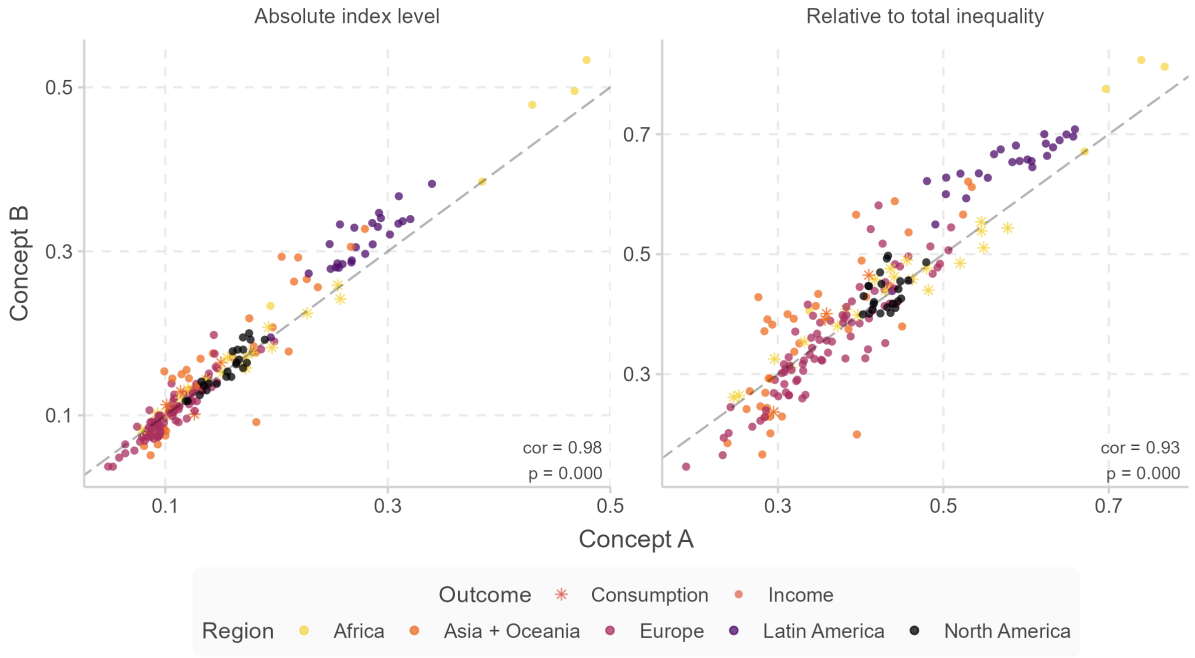
Figure B.1: Total inequality – Own versus World Bank estimates



Note: This figure compares the Gini inequality index we calculated from survey data on 72 different countries and multiple time periods to the corresponding estimates provided by the World Bank. The left panel displays results for all countries, while the right panel only shows those for which the same underlying concept of money-metric wellbeing is used, either income or consumption. The distinction between income (dots) and consumption (asterisks) as outcomes to compute IOp on the left panel refers to our own calculations.

Source: Own calculations on multiple surveys listed in Appendix Table B.1 and World Bank (World Development Indicators).

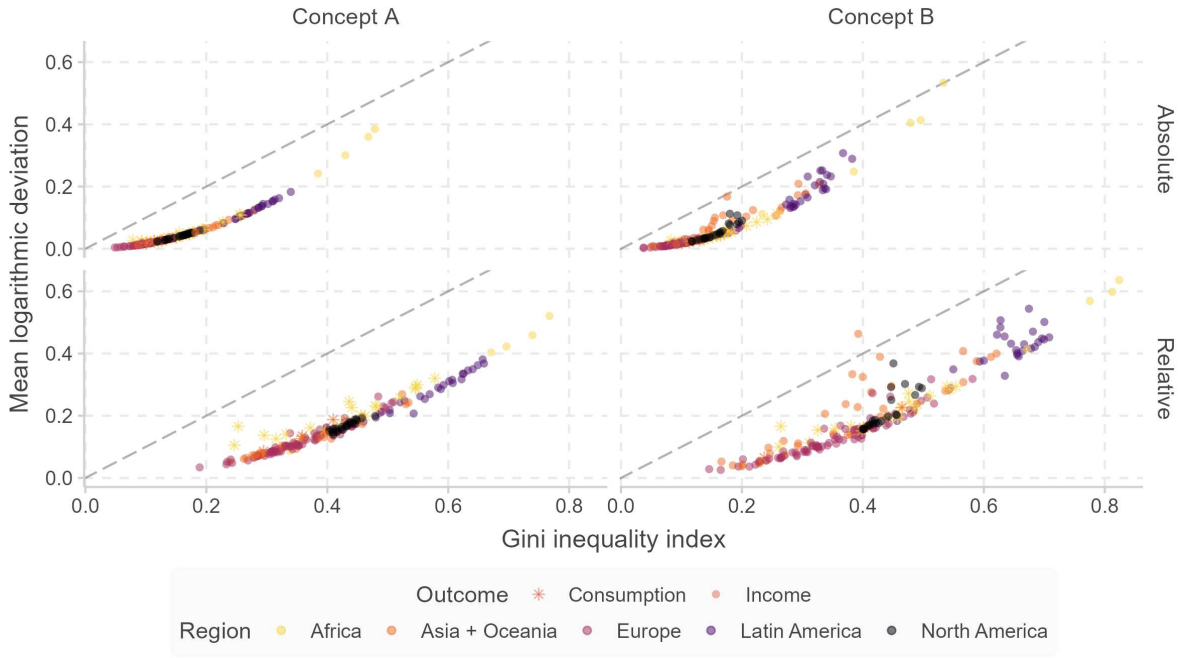
Figure B.2: Correlations between Concepts A and B of inequality of opportunity



Note: This figure plots estimates of the two concepts of inequality of opportunity against one another. The left panel displays the absolute levels of inequality of opportunity indices, while the right panel represents the corresponding measures relative to total inequality. Concept A refers to ex-ante conditional inference random forest IOp. Concept B refers to ex-post tree IOp. The shape of the data points indicates whether the corresponding IOp coefficient was computed based on income (dots) or consumption expenditure (asterisks). More details are provided in Section 3 and Appendix A. Inequality of opportunity estimates reported on this graph are based on the Gini inequality index. The correlation coefficient (cor) and the associated p-value (p) of each relationship are shown at the bottom-right.

Source: Own calculations on multiple surveys listed in Appendix Table B.1.

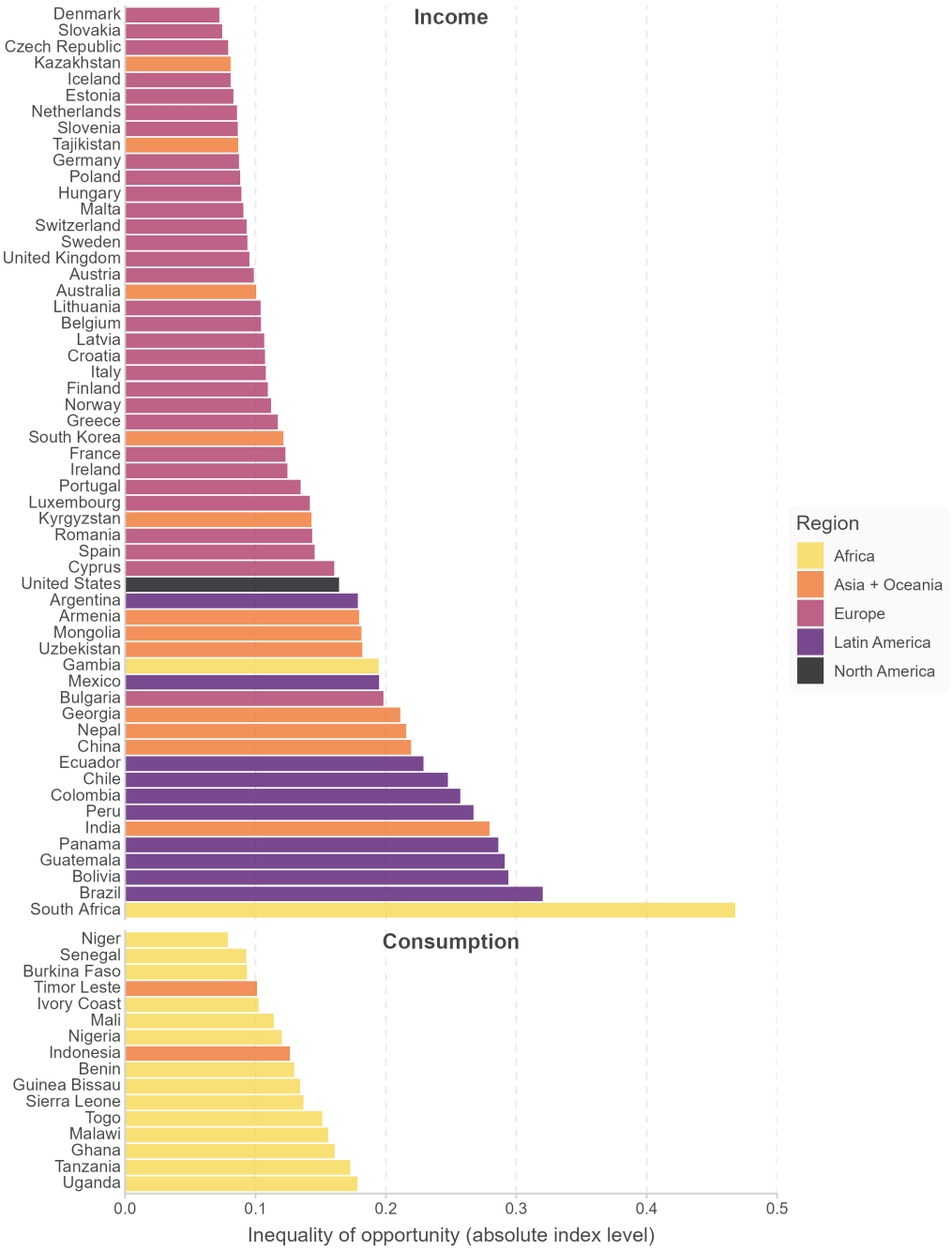
Figure B.3: Correlations between IOp measures based on Gini vs. MLD inequality indices



Note: This figure plots different measures of inequality of opportunity against one another depending on whether they were computed with the Gini inequality index or with mean logarithmic deviations. The upper panels display the absolute levels of inequality of opportunity indices, while the lower panels represent the corresponding measures relative to total inequality. Concept A refers to ex-ante conditional inference random forest IOp. Concept B refers to ex-post transformation tree IOp. The shape of the data points indicates whether the corresponding IOp coefficient was computed based on income (dots) or consumption expenditure (asterisks). More details are provided in Section 3 and Appendix A.

Source: Own calculations on multiple surveys listed in Appendix Table B.1.

Figure B.4: Inequality of opportunity – Absolute index level

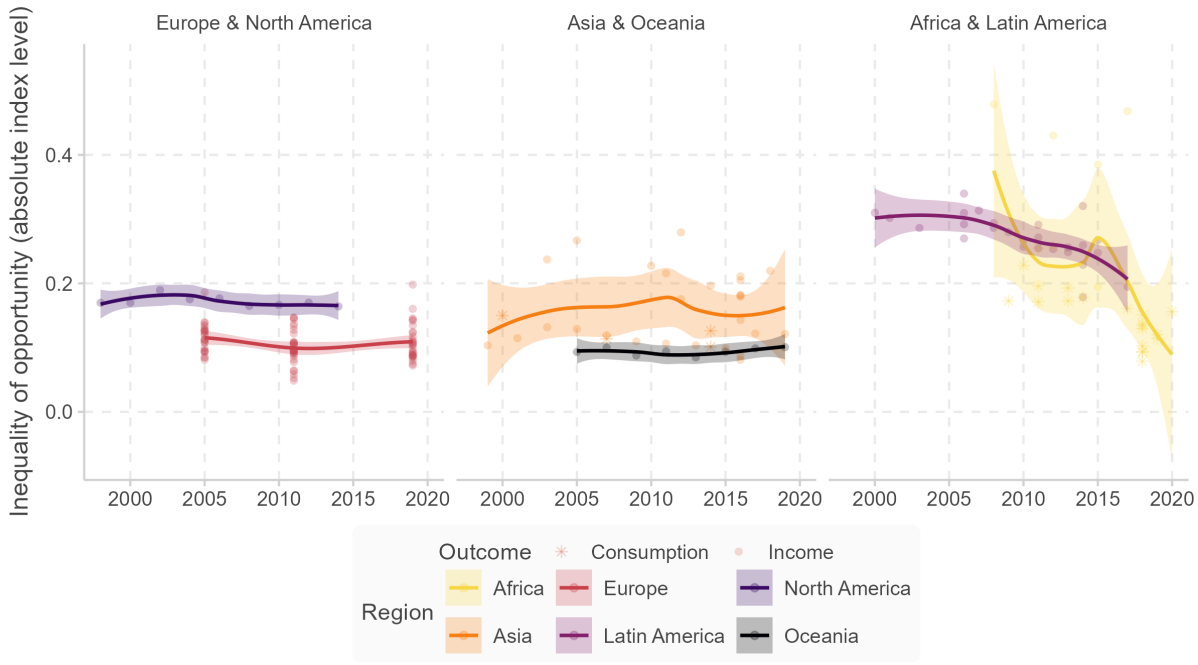


Note: This figure plots the most recent estimates of inequality of opportunity across the 72 countries covered in this study. Estimates were computed according to Concept A, which considers inequality between groups of individuals sharing the same set of inherited characteristics (gender, race, place of birth, etc.). We determine the partition of groups non-parametrically with conditional inference random forests. A Gini inequality index is then estimated on a smoothed counterfactual distribution in which every individual is assigned the mean outcome of their type. The upper and lower panels display countries for which inequality of opportunity is estimated on income and consumption respectively.

Reading: In Brazil, the most recent index of inequality of opportunity amounts to 0.32.

Source: Own calculations on multiple surveys listed in Appendix Table B.1.

Figure B.5: Trends in inequality of opportunity – Absolute index level



Note: This figure plots the trends in inequality of opportunity across 6 regions of the world. See notes to Figure B.4 for a description of the estimation methods. Fitted lines and confidence bands were produced with locally estimated scatterplot smoothing (LOESS). The shape of the data points indicates whether the corresponding IOp coefficient was computed based on income (dots) or consumption expenditure (asterisks).

Reading: In the only North American country represented on this graph, the United States, the inequality of opportunity index has remained relatively stable throughout the period, around 0.19.

Source: Own calculations on multiple surveys listed in Appendix Table B.1.

Figure B.6: Trends in inequality of opportunity in the United States

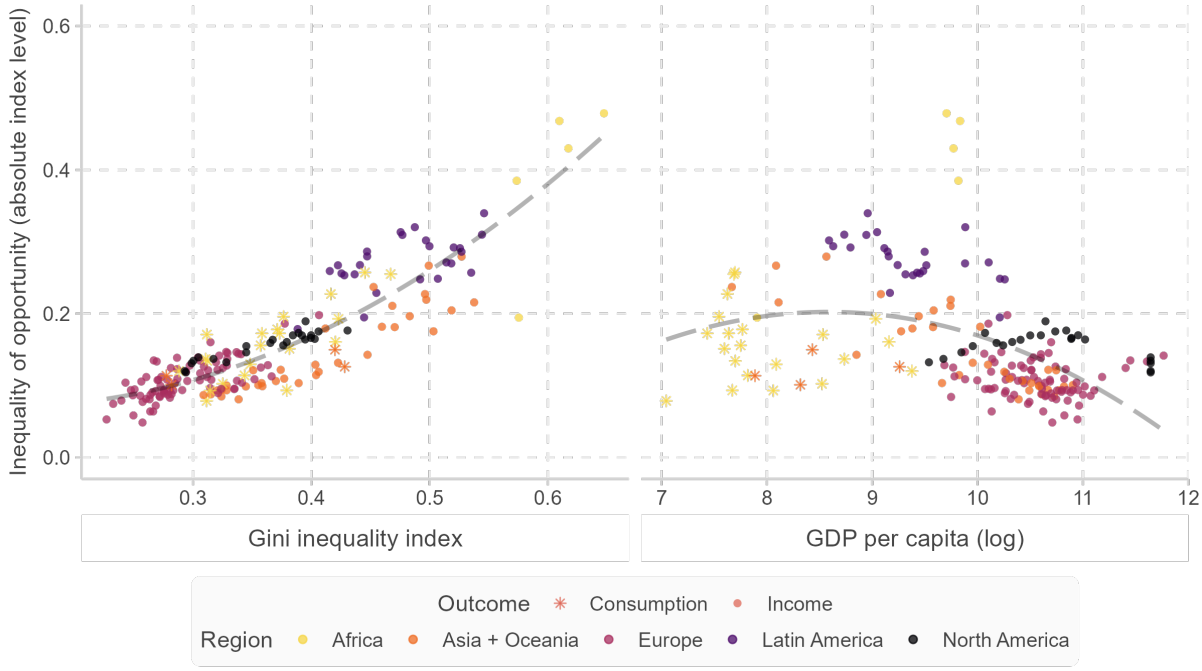


Note: This figure plots the trend in inequality of opportunity in the United States as a cubic fit through the time series. See notes to Figures 1 and B.4 and for a description of the estimation methods of absolute and relative inequality of opportunity.

Reading: In the United States, inequality of opportunity has increased from 0.13 in the 1970s to close to 0.18 in the early 2000s in absolute terms, while its ratio to total inequality has remained relatively constant, mostly between 40 and 45%.

Source: Own calculations on the Panel Study of Income Dynamics.

Figure B.7: Opportunity Great Gatsby and Kuznets curves – Absolute index level



Note: This figure plots inequality of opportunity across 72 countries and multiple time periods against total inequality measured two decades earlier ("Great Gatsby curve" – left panel) and log GDP per capita ("Kuznets curve" – right panel). The shape of the data points indicates whether the corresponding IOp coefficient was computed based on income (dots) or consumption expenditure (asterisks). Dashed lines represent the respective quadratic fits. See notes to Figure B.4 for a description of inequality of opportunity estimation methods. All values are expressed in 2017 USD and adjusted by PPP and CPI as described in Appendix Section A.3.

Reading: Latin American countries have high shares of total income inequality that can be predicted by inherited characteristics (y-axis), high levels of total inequality (x-axis – left panel), and intermediate levels of GDP per capita (x-axis – right panel).

Source: Own calculations on multiple surveys listed in Appendix Table B.1 for inequality of opportunity and Gini index, and IMF (2024) for GDP per capita.