

Contextual Factors Behind Income Inequality: A Predictive Analysis

Arda Aktas

Vrije Universiteit Amsterdam

Miguel Poblete Cazenave

`m.a.poblete.cazenave@vu.nl`

Vrije Universiteit Amsterdam

Narasimha Rao

Yale University

Research Article

Keywords: income inequality, XGBoost, Machine learning, political institutions, regional classification, cross-country analysis

Posted Date: September 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10832070/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Contextual Factors Behind Income Inequality: A Predictive Analysis

Arda Aktas^{1†}, Miguel Poblete-Cazenave^{1*}, Narasimha Rao^{2,3†}

^{1*}Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, De Boelelaan 1111, Amsterdam, 1081 HV, The Netherlands.

²Yale School of the Environment, Yale University, 195 Prospect Street, New Haven, CT 06511, Connecticut, USA.

³Energy, Climate, and Environment Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, Laxenburg, A-2361, Lower Austria, Austria.

*Corresponding author(s). E-mail(s): m.a.poblete.cazenave@vu.nl;

Contributing authors: a.aktas@vu.nl; narasimha.rao@yale.edu;

[†]These authors contributed equally to this work.

Abstract

Income inequality remains a central concern in economics and public policy, yet conclusions about its predictors often depend on both the factors considered and the measure of inequality used. This paper examines the predictive importance of a broad set of indicators covering structural factors, capital structure, financial-sector development, political institutions, regional classification, human capital, globalization, and technology. Using data for 88 countries from 1970 to 2017 and six inequality measures from the World Income Inequality Database (WIID), we apply a machine learning approach to assess the relative predictive importance of 35 candidate indicators across alternative income inequality measures.

The analysis proceeds in three stages: a full specification including all candidate indicators, a category-representative specification with one common indicator from each conceptual category, and an inequality-specific specification allowing the most predictive indicator within each category to vary across inequality measures.

Regional classification is the dominant predictor across all specifications and inequality measures, although its broad nature prevents identification of the specific regional characteristics underlying this result. Political and technology indicators provide the most consistent additional predictive information, whereas the importance of human capital, financial development, and globalization varies across inequality measures. Technology is particularly informative for top-income

concentration, while returns to capital are more relevant for measures capturing gaps between the upper and lower parts of the income distribution. Despite these differences, the category-representative specification retains most of the full model’s predictive information, showing that cross-country inequality can be analysed using a more parsimonious set of indicators without losing much predictive power.

Keywords: income inequality, XGBoost, Machine learning, political institutions, regional classification, cross-country analysis

1 Introduction

Income inequality has been a persistent economic and policy concern because income disparities affect people’s access to goods and services, including those necessary to maintain an adequate standard of living. For a few decades in the mid-20th century, income inequality appeared to decline in many countries while average incomes rose, creating the impression that sustained economic growth alone might gradually reduce the problem. Since the 1980s, however, this pattern has reversed, with income inequality increasing across many parts of the world, including high-income countries with sufficient resources to support broadly shared prosperity.

These developments have brought the longstanding question of what drives inequality back into focus. A large body of literature has examined this issue, with notable contributions dating back to the seminal work of Kuznets ([Kuznets 1955](#)), which relates economic growth to income inequality by arguing that inequality increases during the early stages of economic development and begins to decline after a certain level of income is reached, producing the well-known inverted U-shaped Kuznets curve. However, this proposition has become increasingly difficult to reconcile with observed trends, particularly in recent decades, as income inequality has risen rapidly in some of the world’s largest economies. During this period, the structure of production has changed considerably, along with major shifts in labor markets, financial systems, political institutions, technology, and globalization. Therefore, income growth alone is no longer sufficient to explain whether inequality rises or falls, leaving room for a range of other factors that may help explain different parts of the puzzle.

Among the various studies proposing different drivers of inequality in the recent literature, the work of Piketty and co-authors ([Piketty 2001](#); [Piketty and Saez 2003](#); [Piketty 2014](#)) has attracted particular attention. Their central argument is that if the rate of return on capital exceeds the rate of economic growth, incentives for capital accumulation will tend to drive inequality upward. While this view has been subject to some critique ([Krusell and Smith 2015](#); [Homburg 2015](#)), there appears to be sufficient theoretical and empirical evidence supporting some of its key elements. However, while the argument largely explains rising wealth inequality, it falls short of explaining the observed increase in income inequality and therefore remains far from providing a complete picture of inequality ([Acemoglu and Robinson 2015](#); [Góes 2025](#)).

A broader view considers a range of economic, institutional, and structural factors that affect income generation and distribution. Starting with the labor force, one of the most important factors contributing to inequality is variation in human capital. Workers differ in their productivity, and these differences are reflected in their wages, an idea that dates back to Adam Smith (Spengler 1977). Because individual productivity cannot be directly observed, other measures are often used to assess potential productivity differences (Folloni and Vittadini 2010; Abraham and Mallatt 2022). Nevertheless, the widening gap between wage growth and growth in GDP per capita points to factors beyond human capital as potential drivers of inequality (Brynjolfsson and McAfee 2014; Berlingieri et al. 2024).

Technology appears to be one such factor, either by increasing productivity or by reducing the role of labor in production (Caselli 1999; Acemoglu and Autor 2011). Globalization may also contribute to inequality by expanding the cross-border movement of goods, capital, and economic activity and increasing firms' ability to organize production across countries. Because capital is more mobile internationally than labor, firms may relocate production to countries with lower labor costs when workers there can perform same tasks at comparable levels of productivity (Helpman et al. 2010; Autor et al. 2013). Government and political institutions represent another important factor. Institutional frameworks can influence both development and inequality by affecting access to economic opportunities and determining how the gains from economic activity are distributed, as explored extensively by Acemoglu, Robinson, and co-authors (Acemoglu and Robinson 2005, 2012). These institutional arrangements, however, are also embedded in broader historical and regional contexts. Countries within the same region often share political, legal, and institutional legacies that may contribute to similar patterns of inequality (Andreescu et al. 2025). Finally, the financial sector may also play an important role. One possible channel links back to Piketty's argument: if capital owners receive higher returns in financial markets than in productive activities, capital will increasingly flow toward the financial sector, contributing to greater inequality (Greenwood and Jovanovic 1990; de Haan and Sturm 2017).

In sum, many of the factors associated with inequality are broad conceptual categories rather than single, clearly defined mechanisms. Human capital, for example, may be represented by average educational attainment, the distribution of education, health, or a composite human capital index, while political institutions may differ in their electoral, liberal, participatory, deliberative, or egalitarian dimensions. Similarly, technology, globalization, and financial development each encompass several distinct features that need not be equally informative about inequality. The conclusions reached may therefore depend on both the indicator used to represent each broader category and the measure of inequality being examined (Acemoglu and Robinson 2015). Consequently, identifying which indicators are most informative under alternative inequality measures is important for obtaining a clearer picture of inequality dynamics.

In this study, we use modern predictive techniques to identify which indicators within these broader conceptual categories are most informative for predicting income inequality in the following year across alternative inequality measures. Using a broad cross-country panel dataset and a diverse set of indicators capturing the main factors

discussed in the literature, we employ a machine-learning approach to evaluate their relative predictive importance. This approach is particularly suited to our setting because it can accommodate a large set of potentially correlated predictors, as well as nonlinear relationships and interactions, without imposing a predetermined parametric structure. Rather than focusing on causal identification or the estimation of structural coefficients, our objective is to assess which indicators consistently contain information about future inequality outcomes and whether their relevance varies across different parts of the income distribution.

Our contribution is threefold. First, we simultaneously evaluate a broad range of factors emphasized in the literature on income inequality, including political institutions, human capital, globalization, financial development, and technology. Second, we examine six commonly used measures of income inequality, allowing us to assess whether the factors associated with overall inequality differ from those associated with top-income concentration and relative income gaps. By considering several measures, the study provides a broader view of the factors associated with different parts of the income distribution than analyses based on a single measure. Third, by adopting a predictive rather than causal perspective, we provide complementary evidence on which contextual factors consistently contain information about inequality across countries and which are more relevant to particular measures of inequality.

2 Data and Variables

We combine information from multiple international datasets and compile a broad set of indicators capturing factors that have been associated with income inequality in the literature.

The indicators are described below and summarized in Table 1, where they are organized into thematic categories reflecting the main conceptual channels discussed in the literature. From the set of potential indicators, we include only those with observations covering at least 60% of the country–year sample. This threshold was chosen to retain sufficient data coverage while preserving representation across the different conceptual categories.

2.1 Income Inequality Measures

Income inequality can be measured in several ways, and because different measures capture different parts of the income distribution, empirical conclusions may vary depending on the measure used. This is reflected, for example, in the contrasting interpretations of inequality associated with Piketty (Piketty 2014) and the critique by Acemoglu and Robinson (Acemoglu and Robinson 2015)

To account for these differences, we employ six alternative measures of income inequality, all obtained from WIID and widely used in the empirical literature. They are defined below, together with references to seminal or widely cited studies employing each measure:

- The Gini coefficient (Gini 1936; Sen 1973)

- The income share held by the top 10% of the population ([Atkinson 2007](#); [Leigh 2007](#))
- The income share held by the top 1% of the population ([Piketty and Saez 2003](#); [Acemoglu and Robinson 2015](#))
- The Palma Ratio: the share held by the top 10% divided by the share held by the bottom 40% ([Palma 2011](#); [Cobham and Sumner 2014](#))
- The S80/S20 ratio (also called the 20:20 ratio): the share held by the top 20% divided by the share held by the bottom 20% ([United Nations Development Programme 1999](#); [Forbes 2000](#))
- The P90/P10 ratio: income share at the 90th percentile divided by income share at the 10th percentile of the income distribution ([Piketty 2001](#); [Autor et al. 2008](#))

The measures are drawn from the World Income Inequality Database (WIID) ([UNU-WIDER 2025](#)), compiled by the United Nations University World Institute for Development Economics Research (UNU-WIDER). WIID is one of the most comprehensive global inequality databases, providing harmonized Gini coefficients and income share statistics derived from household surveys, tax records, and national statistical sources.

2.2 Capital Structure

We include two variables calculated using data from the Penn World Tables (PWT) ([Feenstra et al. 2015](#)) to represent Piketty’s view of inequality as being driven by the structure of capital: the capital-to-labor ratio (KY_L), and the ratio of returns to capital to GDP growth (r_g), which shows whether returns to capital exceed the rate of economic growth.

2.3 Financial Sector

To measure financial-sector development, we use a set of variables from the World Bank’s Global Financial Development Database (GFDD) ([Beck et al. 2000](#); [World Bank 2019](#)). These include private credit to GDP ($di01$ and $di12$), deposit money banks’ assets to GDP ($di02$), liquid liabilities relative to GDP ($di05$), financial system deposits to GDP ($di08$ and $oi02$), the ratio of bank credit to bank deposits ($si04$), and total liquid liabilities ($oi07$).

2.4 Political Institutions

To assess which aspects of political institutions are most informative for predicting inequality, we use a set of indicators from the V-Dem database, including the electoral democracy index ($v2x_polyarchy$), the liberal component index ($v2x_liberal$), the participatory component index ($v2x_partip$), the deliberative component index ($v2xdl_delib$), and the egalitarian component index ($v2x_egal$).

These indicators provide alternative characterizations of political institutions, focusing on the quality of elections and political competition ($v2x_polyarchy$), the protection of minority rights, civil liberties, and constraints on government power ($v2x_liberal$), citizen participation in political life within and beyond elections

(*v2x_partip*), the role that public reasoning toward the common good plays in political decision-making (*v2xdl_delib*), and equality in political rights, resource distribution, and access to power across social groups (*v2x_egal*).

2.5 Regional Aggregation

We use several regional classifications to group countries that share broad historical, political, and geographic characteristics that may affect development and income inequality in similar ways within each region. These include the 23 World Bank subregions (*region_wb*) and the 18 United Nations subregions (*un_subregion*), both obtained from the *R countrycode* package (Arel-Bundock et al. 2018).

From the V-Dem dataset, we also include a politico-geographic region classification (*e_regionpol*), originally obtained from the Quality of Government Standard Dataset (QoG) (Teorell et al. 2013). This classification groups countries into 10 regions based on a combination of geographical proximity and regional boundaries defined by area specialists’ understanding of democratization.

We include two additional aggregation variables as controls, to test whether the predictive importance of regional aggregation reflects more specific historical or economic similarities. From the QoG dataset, we use *ht_colonial*, which indicates whether a country was formerly colonized and by which major colonial power. We also add the World Bank income-group classification (*incomegroup*) to test whether the predictive importance of the regional classifications for income inequality simply comes from similarities in countries’ income levels.

2.6 Human Capital

We use the human capital index from the Penn World Table (*hc*), which combines years of schooling with estimated returns to education, together with three indicators from the V-Dem database: average years of education among individuals aged 15 and older (*e_peaveduc*), an index of educational inequality among the adult population (*e_peedgini*), and life expectancy (*e_pelifeeex*). For some countries, the latter two indicators were complemented with data from the Wittgenstein Centre Human Capital Data Explorer (WCDE) (K.C. et al. 2024). Although the WCDE provides more comprehensive country coverage than the V-Dem database, its data are available only at five-year intervals and therefore require interpolation.

2.7 Globalization

We obtain globalization indicators from the KOF Globalization Index (Dreher 2006; Gygli et al. 2019), which covers economic, trade, and financial globalization and distinguishes between actual cross-border flows (de facto globalization) and the institutional framework governing international integration (de jure globalization).

2.8 Technology

We consider several indicators related to technological sophistication, productivity, and the contribution of technology to economic activity.

Economic Complexity Index (ECI), obtained from The Observatory of Economic Complexity (OEC) (Simoes and Hidalgo 2011; Observatory of Economic Complexity (OEC) 2026), shows a country’s stock of productive knowledge based on composition and sophistication of its export basket. It increases when a country exports many different products, particularly those that few other countries can produce. Therefore, higher values indicate a more advanced and knowledge-intensive productive system (Hidalgo and Hausmann 2009).

We also use two productivity-based measures obtained from the Penn World Table (PWT), version 11.0: standard Total Factor Productivity *ctfp* and welfare-adjusted TFP *cwtfp*. The first one measures how efficiently capital and labor are used to produce output and is therefore commonly interpreted as an indicator of technological efficiency. The second also incorporates terms-of-trade effects and better reflects differences in welfare and purchasing power across countries.

Because there is no consensus on how to calculate TFP, we also include an alternative measure, *tfp_alt*, calculated using the growth of total factor productivity obtained from the Total Economy Database (The Conference Board 2025). Unlike the TFP measure from the PWT that is calculated using the growth of the capital stock, this measure is obtained from the growth of capital services. From the same dataset, we also use two additional measures related to digitalization of the economy: *kgdpgric*, which shows the contribution of capital services provided by information and communication technology (ICT) assets to real GDP growth, and *kgdpgrnic*, which shows the contribution of non-ICT capital services to real GDP growth.

2.9 Other Structural Factors

We include a final set of indicators related to the broader economic, demographic, and state of unrest conditions of each country. Among these, gross domestic product (GDP) and total population, which represent overall economic size, country scale, and market size, are obtained from the World Income Inequality Database (WIID).

Urbanization rates (*e_miurbani*), which measure the share of the population living in urban areas and are considered an indicator of the spatial distribution of population and economic activity, are obtained from the Varieties of Democracy (V-Dem) dataset (Coppedge et al. 2025) and the World Development Indicators (World Bank 2026).

Finally, we add indicators of international armed conflict (*e_miinteco*) and internal armed conflict (*e_miinterc*), originally from the Uppsala Conflict Data Program (UCDP) (Sundberg and Melander 2013; Davies et al. 2025) and available in the V-Dem dataset only up to 2000, considering they may affect income inequality by disrupting economic activity and labor markets, destroying assets and infrastructure, displacing populations, and changing the allocation of public resources.

3 Methodology

To assess the predictive importance of the different indicators presented in Table 1 within each conceptual category across alternative measures of inequality, we employ a machine learning approach based on the *XGBoost* model (Chen and Guestrin 2016). *XGBoost* is a gradient boosting framework that is well suited to settings with multiple,

potentially correlated predictors and can accommodate nonlinear relationships and interaction effects without requiring these relationships to be specified in advance. This flexibility allows us to assess the predictive relevance of the indicators without imposing a predetermined functional form.

The analysis is conducted using a country–year sample restricted to observations with complete information on the selected indicators, covering 88 countries from 1970 to 2017. The dependent variable is defined as next-year inequality, measured using six alternative indicators (the Gini coefficient, the top 10% income share, the top 1% income share, the Palma ratio, the 20:20 ratio, and the P90P10 ratio), allowing us to evaluate one-step-ahead predictive performance. All models are estimated using a random 70% training sample over 50 bootstrap iterations to ensure the robustness of the results. The same training—hold-out split is used across specifications within each bootstrap iteration to ensure comparability.

We implement three complementary model specifications, which progressively refine the selection of predictors while preserving the conceptual category structure.

First, we estimate a baseline model that includes the full set of 35 candidate indicators simultaneously. This unrestricted specification allows us to evaluate and rank the relative predictive importance of each indicator for each inequality measure separately.

Second, to identify the indicator that has consistently more predictive power within each theoretical category across the alternative inequality measures, we construct a reduced specification based on cross-measure importance aggregation. More specifically, feature importance is scaled within each model and then averaged across the six inequality measures. Based on these average importance scores, we select one indicator from each of the following six conceptual categories: financial sector, human capital, globalization, political institutions, regional aggregation, and technology. In addition, we retain the most predictive baseline control variables. This specification preserves representation across the main theoretical categories while selecting a common set of predictors for all inequality measures.

Finally, we allow the optimal predictor within each of the six conceptual categories to differ across inequality measures. For each inequality measure, we perform a constrained model search in which we evaluate all possible combinations containing exactly one candidate indicator from each conceptual category, together with the retained baseline controls. For each combination, we estimate an *XGBoost* model on the training sample and select the specification with the lowest combined mean squared prediction error across the training and hold-out samples in each bootstrap iteration. The procedure is repeated over 50 bootstrap iterations with different samples to assess the stability of the selected specifications and their predictive performance. This approach identifies the best-performing inequality-specific predictor configurations while maintaining the theoretical category structure.

4 Results

4.1 Full Specification

Figure 1 presents the results from the full model, showing the indicators with the highest feature importance for each inequality measure and providing an unrestricted

assessment of their relative predictive importance based on *XGBoost* feature importance scores. The corresponding results for all 35 candidate indicators are reported in Figure A1. Importance values are normalized within each inequality measure by rescaling them to a range between 0 and 1, allowing comparison of indicators within each panel while preserving their relative ranking in predicting that particular inequality measure.

The most striking result is the dominant predictive importance of a country’s regional classification across all inequality measures. In every case, either the politico-geographic classification (*e_regionpol*) or the World Bank regional classification (*region_wb*) is the most important predictor. This suggests that broad regional characteristics associated with geographic and historical context are strongly linked to cross-country differences in inequality, whereas other forms of aggregation considered in the analysis, such as income level or colonial history, display considerably lower predictive importance.

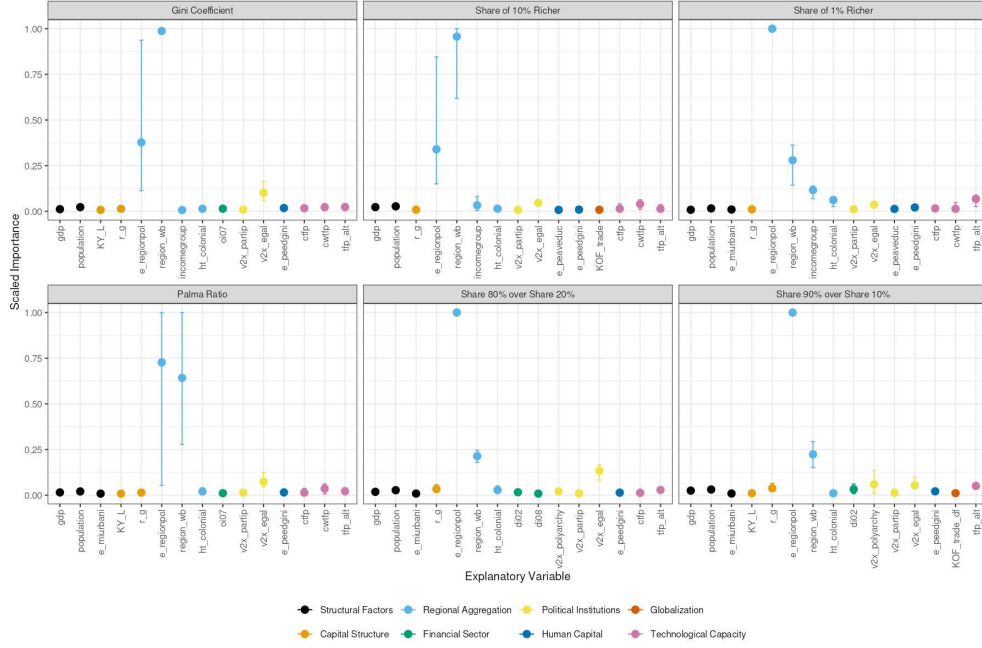
Beyond regional classification, indicators of political institutions also display relatively high predictive importance across several inequality measures, with the egalitarian component index (*v2x_egal*) standing out most clearly for the Gini coefficient, the Palma ratio, and the S80/S20 ratio. Technology indicators also appear among more informative predictors, particularly for the top 1% income share and the P90/P10 ratio. The other indicators are generally less important in the full specification, although the ratio of returns to capital to economic growth (*r_g*) is relatively more predictive for the S80/S20 and P90/P10 ratios.

4.2 Category-Representative Specification

To assess whether the predictive information contained in the full set of indicators can be retained in a more parsimonious specification, we construct a reduced model containing one representative indicator from each conceptual category. This also allows us to compare the predictive relevance of the different categories using a common specification in which each category is represented by a single indicator. We select the representative indicator from each category based on its average predictive importance across the six inequality measures, as estimated in the full model within each bootstrap iteration. We then select the indicators that are chosen most frequently across bootstrap iterations.

Figure 2 presents the results from the reduced specification, specifically, *di02* as the representative measure of financial-sector development, *v2x_egal* for political institutions, *e_peedgini* for human capital, *KOF_trade_dj* for globalization, and *tfp_alt* for technology. The model also includes the selected regional classification, indicators related to the structure of capital, and other retained indicators outside the six conceptual categories. Results showing the indicators selected across all bootstrap iterations are presented in Figure A2.

Consistent with the full specification, the politico-geographic regional classification (*e_regionpol*) remains the dominant predictor across all inequality measures. It is followed by the egalitarian component index (*v2x_egal*), which also consistently ranks among the most informative predictors. This suggests that political arrangements

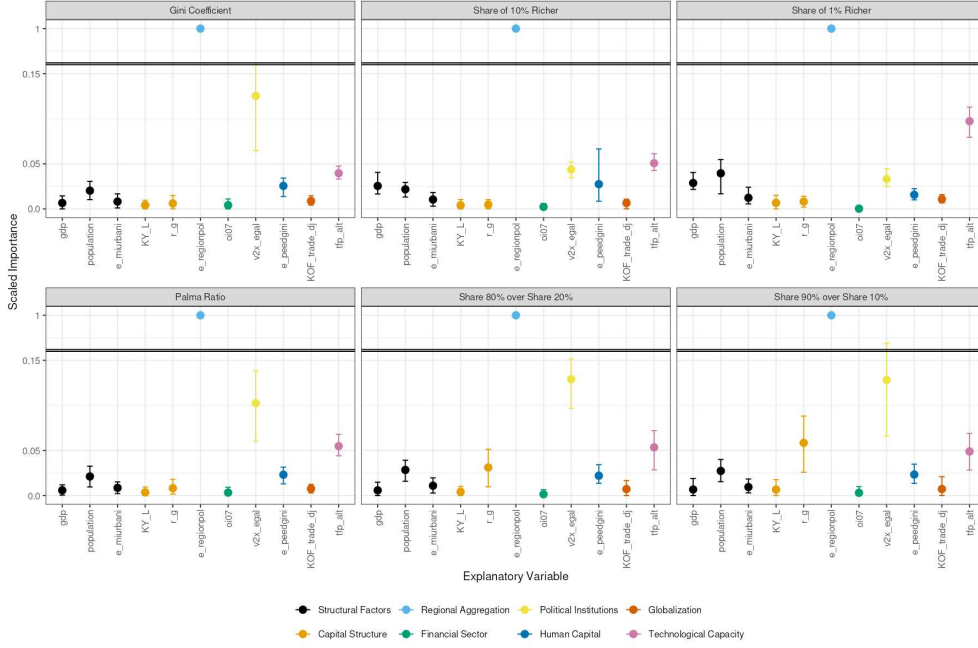


Notes: The figure reports scaled feature importance from XGBoost models estimated using the full specification with all 35 candidate indicators. Only 10 indicators with the highest importance for each inequality measure are shown. Importance is normalized within each inequality measure to range between 0 and 1, preserving the relative ranking of indicators within each panel. Lines represent 95% bootstrap confidence intervals. Each panel corresponds to a different inequality measure, and colors denote 8 conceptual categories.

Fig. 1: Scaled Feature Importance of the 10 Most Important Indicators Across Inequality Measures Under the Full Specification

related to equality in political rights, resource distribution, and access to power are informative predictors of income inequality across the alternative measures considered.

The alternative TFP measure (*tfp_alt*) is also among the more informative indicators, particularly for the top 1% income share. This points to an association between differences in productivity and technology and upper-tail income concentration. The educational inequality indicator (*e_peedgini*) has more modest predictive importance and shows relatively less variation across the inequality measures. The ratio of returns to capital to economic growth (*r_g*) is relatively more predictive for the S80/S20 and P90/P10 ratios. Finally, the globalization indicator (*KOF_trade_dj*) shows moderate predictive importance across several inequality measures, whereas the selected financial-sector indicator remains comparatively less prominent in the reduced specification.



Notes: The figure reports scaled feature importance from *XGBoost* models estimated using the reduced specification. The model includes one representative indicator from each conceptual category, selected based on average predictive importance across the six inequality measures, together with the retained baseline controls. Importance is normalized within each inequality measure to range between 0 and 1, preserving the relative ranking of indicators within each panel. Each panel corresponds to a different inequality measure, and colors denote the conceptual categories. Lines represent 95% bootstrap confidence intervals.

Fig. 2: Scaled Feature Importance using Category-Representative Predictors (Reduced Specification)

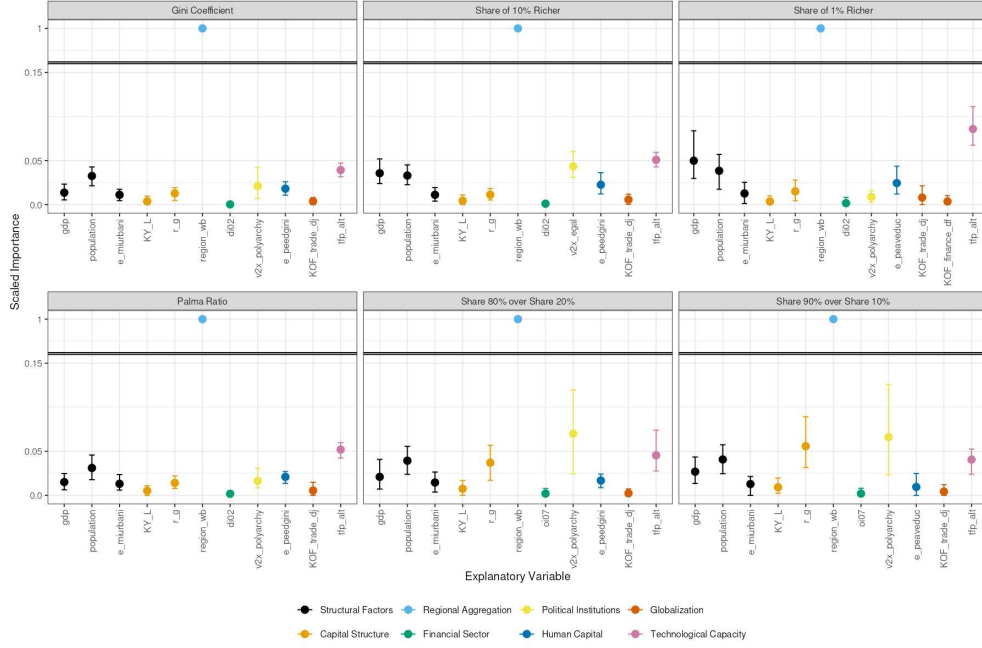
4.3 Inequality Measure-Specific Specification

The category-representative specification selects a common indicator from each conceptual category across all six inequality measures. The inequality measure-specific specification relaxes this restriction by allowing the selected indicator within each category to differ across inequality measures. This allows us to assess whether the patterns observed in the reduced model reflect broad average effects or measure-specific dynamics. For each inequality measure, we evaluate all possible combinations containing exactly one candidate indicator from each conceptual category, together with the retained indicators outside these categories. Figure 3 presents the indicators selected most frequently across bootstrap iterations, while Figure A3 reports the results across all iterations.

Regional classification remains the dominant predictor under the inequality measure-specific specification, with the World Bank regional classification (*region_wb*) selected for all six inequality measures. Political-institution and technology indicators also continue to rank among the most informative predictors, indicating that their

importance is not driven solely by the use of a common set of indicators in the reduced specification. The selected political indicator varies across outcomes: *v2x_polyarchy* is selected for most measures, while *v2x_egal* is selected for the top 10% income share. The alternative TFP measure (*tfp_alt*) is particularly important for the top 1% income share.

Human-capital indicators have more modest predictive importance and differ across measures, with educational inequality (*e_peedgini*) selected more frequently than average years of education (*e_peaveduc*). Financial-sector and globalization indicators remain comparatively less important. Overall, the specification reveals some measure-specific differences, while regional classification, political institutions, and technology remain the most consistently informative categories.



Notes: The figure reports scaled feature importance from *XGBoost* models estimated using the inequality measure-specific specification. For each inequality measure, the selected model is chosen from all possible combinations containing exactly one candidate indicator from each conceptual category, together with the retained baseline controls. Importance is normalized within each inequality measure to range between 0 and 1, preserving the relative ranking of indicators within each panel. Lines represent 95% bootstrap confidence intervals. Each panel corresponds to a different inequality measure, and colors denote the conceptual categories.

Fig. 3: Scaled Feature Importance using Inequality-Specific Predictors (Inequality Measure-Specific Specification)

4.4 Model Fit Across Specifications

Figure A4 compares observed inequality values with the predictions from the three model specifications. All specifications display a very high degree of fit, with R^2 values above 0.98 for each inequality measure. The inequality measure-specific specification (“Best Within Measures”) has the highest R^2 in all six cases, with predictions that lie more closely along the 45-degree line. The improvement is most visible for the $P90/P10$ ratio and, to a lesser extent, the top 1% income share and the $S80/S20$ ratio.

The category-representative specification (“Best Across Measures”) performs almost as well as the full specification despite including substantially fewer indicators. The differences in R^2 between these two specifications are very small across all inequality measures, indicating that the reduced model retains most of the predictive information contained in the full set of candidate indicators. Although allowing indicator selection to vary by inequality measure produces a further improvement in fit, the gains remain relatively modest. Overall, the comparison shows that a parsimonious set of category-representative indicators can provide predictions that are close to those obtained from both the full and inequality-measure-specific specifications.

5 Discussion

The results show both consistency and important differences in the predictors of alternative measures of income inequality. Across all model specifications and inequality measures, regional aggregation remains the single most important predictor of future inequality outcomes, with the politico-geographic classification playing a particularly prominent role. This finding is consistent with comparative evidence showing that countries form distinct and persistent regional patterns of inequality (Mann and Riley 2007; Andreescu et al. 2025). Similar differences are also visible descriptively in our data, as Figure A5 shows clear variation in income inequality across politico-geographic regions. However, because regional aggregation combines a wide range of political, institutional, economic, historical, and geographical characteristics, the analysis provides only limited insight into which region-specific factors account for its predictive importance. To test the importance of these other unobserved regional factors, we re-estimated otherwise identical *XGBoost* models after removing the regional classification indicators. Excluding region increases hold-out MSE and lowers hold-out R^2 for all six inequality measures, indicating that regional classification contributes important predictive information beyond the other indicators included in the analysis (Figure A6).

At the same time, among the political institution indicators, the egalitarian component index (*v2x_egal*) consistently ranks among the most informative predictors across inequality measures, particularly for the Gini coefficient, the Palma ratio, and the $S80/S20$ ratio. This suggests that equality in political rights, access to power, and resource distribution across social groups is closely associated with broader patterns of income inequality, consistent with research emphasizing the role of political institutions in shaping economic opportunities and the distribution of economic gains (Acemoglu et al. 2001; Acemoglu and Robinson 2005, 2012).

Technology also provides important predictive information, with the total factor productivity measure derived from the growth of capital services (*tfp_alt*) standing out particularly for the top 1% income share. This suggests that differences in productivity and technological capacity are especially closely associated with income concentration at the very top. This is consistent with findings from labour economics literature showing that technological change can alter the demand for skills and widen earnings differences across workers and occupations (Krueger 1993; Acemoglu 2002).

Human capital has modest and relatively stable predictive importance across inequality measures, although educational inequality (*e_peedgini*) appears somewhat less informative for the top 1% income share than for the other measures. Its selection as the representative human capital indicator among others suggests that how education is distributed may be more informative for predicting inequality than the national average alone. The literature shows that countries with higher average levels of educational attainment tend to display lower levels of inequality (Gregorio and Lee 2002), but at the same time, the effects of educational expansion depend on labour-market institutions, the composition of labour demand, and broader economic conditions (Katz and Murphy 1992; Autor 2014). Therefore, increases in education may not automatically reduce inequality when productivity gains and labour-market rewards remain concentrated in particular sectors or occupations (Makhlouf and Lalley 2023; Chang 2024).

Among the capital structure indicators, the ratio of returns to capital to economic growth (*r_g*) is relatively more informative for the S80/S20 and P90/P10 ratios, consistent with Piketty’s argument that high returns to capital can widen income gaps (Piketty 2014).

The results also provide a more nuanced picture of financial-sector development and globalization. Although both are widely discussed in the inequality literature, their predictive importance is less consistent than that of the other indicators. The selected financial-sector indicator (*di02*) remains comparatively weak across inequality measures, whereas trade globalization (*KOF_trade_dj*) shows moderate importance for several outcomes. Financial development may affect inequality through access to credit and financial assets (Greenwood and Jovanovic 1990; de Haan and Sturm 2017), while globalization may influence wages and employment through trade and import competition (Helpman et al. 2010; Autor et al. 2013).

Overall, the findings show that the predictors of income inequality are neither entirely common across measures nor completely measure-specific. Regional aggregation, political institutions, and technology provide the most consistent predictive information, while the relevance of human capital, returns to capital, financial-sector development, and globalization varies more across inequality measures. The results therefore highlight the importance of considering multiple measures of inequality, as relying on a single measure may fail to capture differences in the factors associated with different parts of the income distribution.

These findings should nevertheless be interpreted in light of several limitations. First, the analysis is predictive rather than causal, so the identified indicators should not be understood as direct causes of inequality. Second, although regional classification is highly predictive, it is broad and does not reveal which underlying

regional characteristics drive its predictive importance. Third, because the analysis uses country-level data, it cannot fully capture within-country mechanisms that generate inequality at the household or individual level. Finally, although we selected a broad range of indicators from the literature, they still do not capture all the factors that may affect inequality.

6 Conclusion

This study examines the predictive importance of 35 indicators across six measures of income inequality using *XGBoost* models for 88 countries from 1970 to 2017. We find that politico-geographic regional classification is the strongest predictor across all specifications. Political institutions and technology provide the most consistent additional information, while human capital plays a more modest but relatively stable role. Returns to capital are more informative for the S80/S20 and P90/P10 ratios, whereas financial-sector development and globalization are generally less prominent.

The findings show that the importance of different indicators depends partly on how inequality is measured. They also show that a relatively small set of carefully selected indicators retains most of the predictive information contained in the full specification, allowing for a more focused and parsimonious analysis of cross-country inequality.

From a practical perspective, the results suggest that informative predictions of income inequality can be obtained without relying on a large and complex set of indicators. The approach can also support inequality monitoring by identifying which indicators are most informative for different dimensions of the income distribution.

Supplementary information.

Acknowledgements.

Declarations

- **Funding:** This research received no external funding.
- **Conflict of interest/Competing interests:** The authors declare that they have no financial or non-financial competing interests that are directly or indirectly related to this work.
- **Ethics approval and consent to participate:** Not applicable.
- **Consent for publication:** Not applicable.
- **Data availability:**
- **Materials availability:** Not applicable.
- **Code availability:**
- **Author contribution:**

References

Abraham KG, Mallatt J (2022) Measuring human capital. *Journal of Economic Perspectives* 36(3):103–30. <https://doi.org/10.1257/jep.36.3.103>, URL <https://www.>

aeaweb.org/articles?id=10.1257/jep.36.3.103

- Acemoglu D (2002) Technical change, inequality, and the labor market. *Journal of Economic Literature* 40(1):7–72. URL <http://www.jstor.org/stable/2698593>
- Acemoglu D, Autor D (2011) Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, vol 4. Elsevier, p 1043–1171, [https://doi.org/https://doi.org/10.1016/S0169-7218\(11\)02410-5](https://doi.org/https://doi.org/10.1016/S0169-7218(11)02410-5), URL <https://www.sciencedirect.com/science/article/pii/S0169721811024105>
- Acemoglu D, Robinson JA (2005) *Economic origins of dictatorship and democracy*. Cambridge university press, Cambridge
- Acemoglu D, Robinson JA (2012) *Why nations fail: The origins of power, prosperity and poverty*. Profile London, London
- Acemoglu D, Robinson JA (2015) The rise and decline of general laws of capitalism. *Journal of Economic Perspectives* 29(1):3–28. <https://doi.org/10.1257/jep.29.1.3>, URL <https://www.aeaweb.org/articles?id=10.1257/jep.29.1.3>
- Acemoglu D, Johnson S, Robinson JA (2001) The colonial origins of comparative development: An empirical investigation. *American economic review* 91(5):1369–1401
- Andreescu M, Arias-Osorio M, Bauluz L, et al (2025) Equality and development: A comparative & historical perspective 1800-2025. *World Inequality Lab Working Papers* halshs-05514773, HAL, <https://doi.org/None>, URL <https://ideas.repec.org/p/hal/wilwps/halshs-05514773.html>
- Arel-Bundock V, Enevoldsen N, Yetman C (2018) countrycode: An r package to convert country names and country codes. *Journal of Open Source Software* 3(28):848. URL <https://doi.org/10.21105/joss.00848>
- Atkinson AB (2007) Measuring top incomes: Methodological issues. In: Atkinson AB, Piketty T (eds) *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries*. Oxford University Press, Oxford, p 18–42
- Autor DH (2014) Skills, education, and the rise of earnings inequality among the “other 99 percent”. *Science* 344(6186):843–851. <https://doi.org/10.1126/science.1251868>, URL <https://www.science.org/doi/abs/10.1126/science.1251868>, <https://www.science.org/doi/pdf/10.1126/science.1251868>
- Autor DH, Katz LF, Kearney MS (2008) Trends in u.s. wage inequality: Revising the revisionists. *The Review of Economics and Statistics* 90(2):300–323. <https://doi.org/https://doi.org/10.1162/rest.90.2.300>, URL <http://www.jstor.org/stable/40043148>

- Autor DH, Dorn D, Hanson GH (2013) The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review* 103(6):2121–68. <https://doi.org/10.1257/aer.103.6.2121>, URL <https://www.aeaweb.org/articles?id=10.1257/aer.103.6.2121>
- Beck T, Demirgüç-Kunt A, Levine R (2000) A new database on financial development and structure. *World Bank Economic Review* 14(3):397–405
- Berlingieri G, Blanchenay P, Criscuolo C (2024) The great divergence(s). *Research Policy* 53(3):104955. <https://doi.org/https://doi.org/10.1016/j.respol.2024.104955>, URL <https://www.sciencedirect.com/science/article/pii/S0048733324000040>
- Brynjolfsson E, McAfee A (2014) *The second machine age: Work progress and prosperity in a time of brilliant technologies*. WW Norton & company, New York
- Caselli F (1999) Technological revolutions. *American Economic Review* 89(1):78–102. <https://doi.org/10.1257/aer.89.1.78>, URL <https://www.aeaweb.org/articles?id=10.1257/aer.89.1.78>
- Chang FH (2024) The impact of education costs on income inequality. *International Review of Economics* 71(3):553–574. <https://doi.org/10.1007/s12232-024-00452-z>, URL <https://doi.org/10.1007/s12232-024-00452-z>
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp 785–794, <https://doi.org/10.1145/2939672.2939785>
- Cobham A, Sumner A (2014) Is inequality all about the tails?: The palma measure of income inequality. *Significance* 11(1):10–13. <https://doi.org/10.1111/j.1740-9713.2014.00718.x>, URL <https://doi.org/10.1111/j.1740-9713.2014.00718.x>
- Coppedge M, Gerring J, Knutsen CH, et al (2025) V-dem [country-year/country-date] dataset v15. *Varieties of Democracy (V-Dem) Project* <https://doi.org/10.23696/vdemds25>
- Davies S, Pettersson T, Sollenberg M, et al (2025) Organized violence 1989–2024, and the challenges of identifying civilian victims. *Journal of Peace Research* 62(4):1223–1240. <https://doi.org/10.1177/00223433251345636>, URL <https://doi.org/10.1177/00223433251345636>, <https://academic.oup.com/jpr/article-pdf/62/4/1223/65348942/00223433251345636.pdf>
- de Haan J, Sturm JE (2017) Finance and income inequality: A review and new evidence. *European Journal of Political Economy* 50:171–195. <https://doi.org/https://doi.org/10.1016/j.ejpoleco.2017.04.007>, URL <https://www.sciencedirect.com/science/article/pii/S0176268017301490>

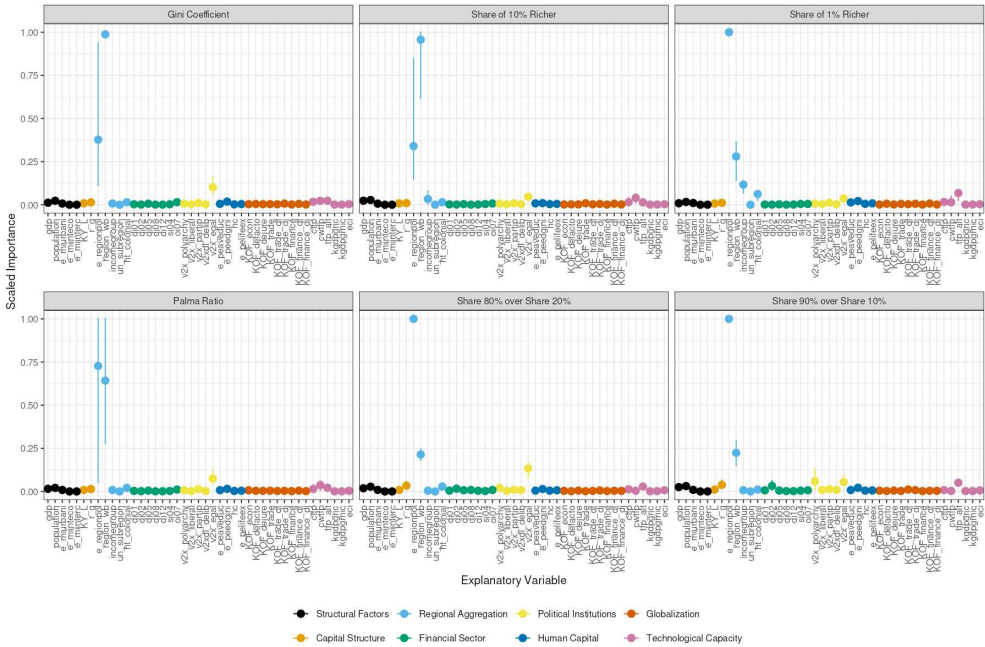
- Dreher A (2006) Does globalization affect growth? evidence from a new index of globalization. *Applied Economics* 38(10):1091–1110. <https://doi.org/10.1080/00036840500392078>, doi: 10.1080/00036840500392078
- Feenstra RC, Inklaar R, Timmer MP (2015) The next generation of the penn world table. *American Economic Review* 105(10):3150–82. <https://doi.org/10.1257/aer.20130954>, URL <https://www.aeaweb.org/articles?id=10.1257/aer.20130954>
- Folloni G, Vittadini G (2010) Human capital measurement: A survey. *Journal of Economic Surveys* 24(2):248–279. <https://doi.org/https://doi.org/10.1111/j.1467-6419.2009.00614.x>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6419.2009.00614.x>, <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-6419.2009.00614.x>
- Forbes KJ (2000) A reassessment of the relationship between inequality and growth. *The American Economic Review* 90(4):869–887. <https://doi.org/10.1257/aer.90.4.869>, URL <http://www.jstor.org/stable/117312>
- Gini C (1936) On the measure of concentration with special reference to income and statistics. In: Cowles Commission Research Conference on Economics and Statistics. No. 208 in General Series, Colorado College Publication, Colorado Springs, CO, p 73–80
- Góes C (2025) Testing piketty’s hypothesis on the drivers of income inequality: evidence from panel vars with heterogeneous dynamics. *Public Choice* 205(1):49–77. <https://doi.org/10.1007/s11127-025-01283-4>, URL <https://doi.org/10.1007/s11127-025-01283-4>
- Greenwood J, Jovanovic B (1990) Financial development, growth, and the distribution of income. *Journal of Political Economy* 98(5):1076–1107. URL <http://www.jstor.org/stable/2937625>
- Gregorio JD, Lee J (2002) Education and income inequality: New evidence from cross-country data. *Review of Income and Wealth* 48(3):395–416. <https://doi.org/https://doi-org.vu-nl.idm.oclc.org/10.1111/1475-4991.00060>, URL <https://onlinelibrary-wiley-com.vu-nl.idm.oclc.org/doi/abs/10.1111/1475-4991.00060>, <https://onlinelibrary-wiley-com.vu-nl.idm.oclc.org/doi/pdf/10.1111/1475-4991.00060>
- Gygli S, Haelg F, Potrafke N, et al (2019) The kof globalisation index – revisited. *The Review of International Organizations* 14(3):543–574. <https://doi.org/10.1007/s11558-019-09344-2>
- Helpman E, Itskhoki O, Redding S (2010) Inequality and unemployment in a global economy. *Econometrica* 78(4):1239–1283. <https://doi.org/https://doi.org/10.3982/ECTA8640>, URL <https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA8640>, <https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA8640>

- Hidalgo CA, Hausmann R (2009) The building blocks of economic complexity. *Proceedings of the National Academy of Sciences* 106(26):10570–10575. <https://doi.org/10.1073/pnas.0900943106>, URL <https://www.pnas.org/doi/abs/10.1073/pnas.0900943106>, <https://www.pnas.org/doi/pdf/10.1073/pnas.0900943106>
- Homburg S (2015) Critical remarks on piketty’s capital in the twenty-first century. *Applied Economics* 47(14):1401–1406. <https://doi.org/10.1080/00036846.2014.997927>, URL <https://doi.org/10.1080/00036846.2014.997927>, <https://doi.org/10.1080/00036846.2014.997927>
- Katz LF, Murphy KM (1992) Changes in relative wages, 1963–1987: Supply and demand factors. *The Quarterly Journal of Economics* 107(1):35–78. <https://doi.org/10.2307/2118323>, URL <https://doi.org/10.2307/2118323>, <https://academic.oup.com/qje/article-pdf/107/1/35/5429219/107-1-35.pdf>
- K.C. S, Dhakad M, Potančoková M, et al (2024) Updating the shared socioeconomic pathways (ssps) global population and human capital projections. Iiasa working paper, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, URL <https://pure.iiasa.ac.at/id/eprint/19487/>
- Krueger AB (1993) How computers have changed the wage structure: Evidence from microdata, 1984–1989. *The Quarterly Journal of Economics* 108(1):33–60. <https://doi.org/10.2307/2118494>, URL <https://doi.org/10.2307/2118494>, <https://academic.oup.com/qje/article-pdf/108/1/33/5361737/108-1-33.pdf>
- Krusell P, Smith AA (2015) Is piketty’s “second law of capitalism” fundamental? *Journal of Political Economy* 123(4):725–748. <https://doi.org/10.1086/682574>, URL <https://doi.org/10.1086/682574>, <https://doi.org/10.1086/682574>
- Kuznets S (1955) Economic growth and income inequality. *The American Economic Review* 45(1):1–28. URL <http://www.jstor.org/stable/1811581>
- Leigh A (2007) How closely do top income shares track other measures of inequality? *The Economic Journal* 117(524):F619–F633. <https://doi.org/10.1111/j.1468-0297.2007.02099.x>, URL <https://doi.org/10.1111/j.1468-0297.2007.02099.x>, <https://academic.oup.com/ej/article-pdf/117/524/F619/26505545/ej619.pdf>
- Makhlouf Y, Lalley C (2023) Education expansion, income inequality and structural transformation: Evidence from OECD countries. *Social Indicators Research* 169(1):255–281. <https://doi.org/10.1007/s11205-023-03161-2>, URL <https://doi.org/10.1007/s11205-023-03161-2>
- Mann M, Riley D (2007) Explaining macro-regional trends in global income inequalities, 1950–2000. *Socio-Economic Review* 5(1):81–115. <https://doi.org/10.1093/ser/mwl017>

- Observatory of Economic Complexity (OEC) (2026) Economic complexity index long time series (1964-2017). Accessed January 2026, <https://oec.world/en/rankings/legacy/eci>
- Palma JG (2011) Homogeneous middles vs. heterogeneous tails, and the end of the ‘inverted-u’: It’s all about the share of the rich. *Development and Change* 42(1):87–153. <https://doi.org/10.1111/j.1467-7660.2011.01694.x>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-7660.2011.01694.x>, <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-7660.2011.01694.x>
- Piketty T (2001) Income inequality in france, 1901–1998. *Journal of Political Economy* 111(5):1004–1042. <https://doi.org/10.1086/376955>, doi: 10.1086/376955
- Piketty T (2014) *Capital in the twenty-first century*. Harvard University Press
- Piketty T, Saez E (2003) Income inequality in the united states, 1913–1998*. *The Quarterly Journal of Economics* 118(1):1–41. <https://doi.org/10.1162/00335530360535135>, URL <https://doi.org/10.1162/00335530360535135>
- Sen A (1973) *On Economic Inequality*. Clarendon Press, Oxford
- Simoes AJG, Hidalgo CA (2011) The economic complexity observatory: an analytical tool for understanding the dynamics of economic development. In: *Proceedings of the 17th AAAI Conference on Scalable Integration of Analytics and Visualization*. AAAI Press, AAAIWS’11-17, p 39–42
- Spengler JJ (1977) Adam smith on human capital. *The American Economic Review* 67(1):32–36. URL <http://www.jstor.org/stable/1815877>
- Sundberg R, Melander E (2013) Introducing the ucdp georeferenced event dataset. *Journal of Peace Research* 50(4):523–532. <https://doi.org/10.1177/0022343313484347>, URL <https://doi.org/10.1177/0022343313484347>, <https://academic.oup.com/jpr/article-pdf/50/4/523/65743342/0022343313484347.pdf>
- Teorell J, Sundström A, Holmberg S, et al (2013) QoG Standard Dataset. <https://doi.org/10.18157/qogstdjan24>, URL <https://doi.org/10.18157/qogstdjan24>
- The Conference Board (2025) The conference board total economy database™. Data Central, The Conference Board, all series must be cited as: ”The Conference Board Total Economy Database™, September 2025”
- United Nations Development Programme (1999) *Human Development Report 1999*. Oxford University Press, New York
- UNU-WIDER (2025) World income inequality database (wiid). Version 29 April 2025 <https://doi.org/10.35188/UNU-WIDER/WIID-290425>

World Bank (2019) Global financial development report 2019/2020: Bank regulation and supervision a decade after the global financial crisis. World Bank Group, Washington DC, <https://doi.org/10.1596/978-1-4648-1447-1>

World Bank (2026) World development indicators. World Bank Open Data, accessed: April 22, 2026



Notes: The figure reports scaled feature importance from XGBoost models estimated with the full specification, including all 35 candidate indicators. Importance is normalized within each inequality measure to range between 0 and 1, preserving the relative ranking of indicators within each panel. Each panel corresponds to a different inequality measure, and colors denote the conceptual categories.

Fig. A1: Scaled Feature Importance Across Inequality Measures Under the Full Specification - All Variables Across All Bootstrap Iterations

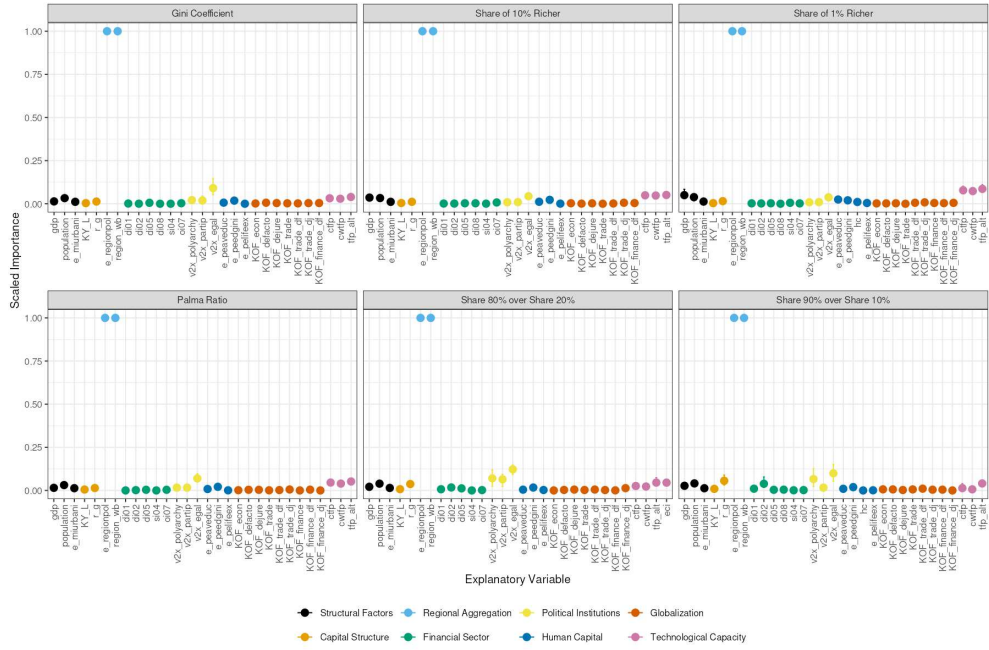
Table 1: Set of Explanatory Variables*

Variable	Description	Source
Structural Factors		
<i>gdp</i>	Gross Domestic Product (GDP)	WIID
<i>population</i>	Total Population	WIID
<i>e_miurbani</i>	Urbanization Rate: share of the population living in urban areas	V-Dem & WDI
<i>e_miinteco</i>	Country's participation in an international armed conflict	V-Dem & UCDP
<i>e_miinterc</i>	Country's Participation in an internal armed conflict	V-Dem & UCDP
Capital Structure		
<i>KY_L</i>	Capital to Labor ratio [‡]	PWT
<i>r-g</i>	Returns to capital over GDP growth [‡]	PWT
Financial Sector		
<i>di01</i>	Private credit by deposit money banks to GDP	GFDD
<i>di02</i>	Deposit money banks' assets to GDP	GFDD
<i>di05</i>	Liquid liabilities to GDP	GFDD
<i>di08</i>	Financial system deposits to GDP	GFDD
<i>di12</i>	Private credit by deposit money banks and other financial institutions to GDP	GFDD
<i>si04</i>	Bank credit to bank deposits	GFDD
<i>oi02</i>	Bank deposits to GDP	GFDD
<i>oi07</i>	Liquid liabilities in millions USD (2010 constant)	GFDD
Political Institutions		
<i>v2x_polyarchy</i>	To what extent is the ideal of electoral democracy in its fullest sense achieved	V-Dem
<i>v2x_liberal</i>	To what extent is the liberal principle of democracy achieved	V-Dem
<i>v2x_partip</i>	To what extent is the participatory principle achieved	V-Dem
<i>v2xld_delib</i>	To what extent is the deliberative principle of democracy achieved	V-Dem
<i>v2x_egal</i>	To what extent is the egalitarian principle achieved	V-Dem
Regional Aggregation		
<i>region_wb</i>	Regions as used to be in the World Bank Development Indicators	countrycode
<i>un_subregion</i>	United Nations sub-region	countrycode
<i>e_regionpol</i>	Politico-geographic classification of world regions	V-Dem
<i>ht_colonial</i>	Tenfold classification of former colonial ruler of the country, focused on 'Western overseas' colonialism	QoG
<i>incomegroup</i>	World Bank classification of countries by income level	WIID
Human Capital		
<i>e_peaveduc</i>	Average years of education among individuals aged 15 and older	V-Dem & WCDE [‡]
<i>e_peedgini</i>	How unequal is the level of education of the population older than 15	V-Dem & WCDE [‡]
<i>e_pelifeex</i>	Life Expectancy	V-Dem & WCDE [‡]
<i>hc</i>	Human capital index, based on years of schooling and returns to education	PWT
Globalization		
<i>KOF_econ</i>	KOF Economic Globalization Index	KOF
<i>KOF_defacto</i>	De facto component of the KOF Economic Globalization Index	KOF
<i>KOF_dejure</i>	De jure component of the KOF Economic Globalization Index	KOF
<i>KOF_trade</i>	KOF Trade Globalization Index	KOF
<i>KOF_trade_df</i>	De facto component of the KOF Trade Globalization Index	KOF
<i>KOF_trade_dj</i>	De jure component of the KOF Trade Globalization Index	KOF
<i>KOF_finance</i>	KOF Financial Globalization Index	KOF
<i>KOF_finance_df</i>	De facto component of the KOF Financial Globalization Index	KOF
<i>KOF_finance_dj</i>	De jure component of the KOF Financial Globalization Index	KOF
Technology		
<i>eci</i>	Economic Complexity Index, based on the diversity and sophistication of a country's export basket	ECI
<i>ctfp</i>	Total Factor Productivity (TFP) level at current purchasing power parity (PPT) (USA=1)	PWT
<i>cwtfp</i>	Welfare-relevant TFP levels at current PPPs (USA=1)	PWT
<i>tfp_alt</i>	Alternative TFP measure calculated using total factor productivity growth from TED [‡]	TED
<i>kgdpgric</i>	Contribution of capital services provided by ICT Assets to real GDP growth	TED
<i>kgdpgrnic</i>	Contribution of capital services provided by non-ICT Assets to real GDP growth	TED

Sources: World Income Inequality Database (WIID), Penn World Tables (PWT) v11.0, KOF Globalization Index (KOF), World Bank Global Financial Development Database (GFDD), World Development Indicators (WDI), Wittgenstein Centre Human Capital Data Explorer (WCDE), Total Economy Database™ (TED), Uppsala Conflict Data Program (UCDP), Varieties of Democracy (V-Dem) dataset, Quality of Government (QoG) Standard Dataset, and the *R* countrycode package. Variables are retained if coverage exceeds 60% of the country-year sample.

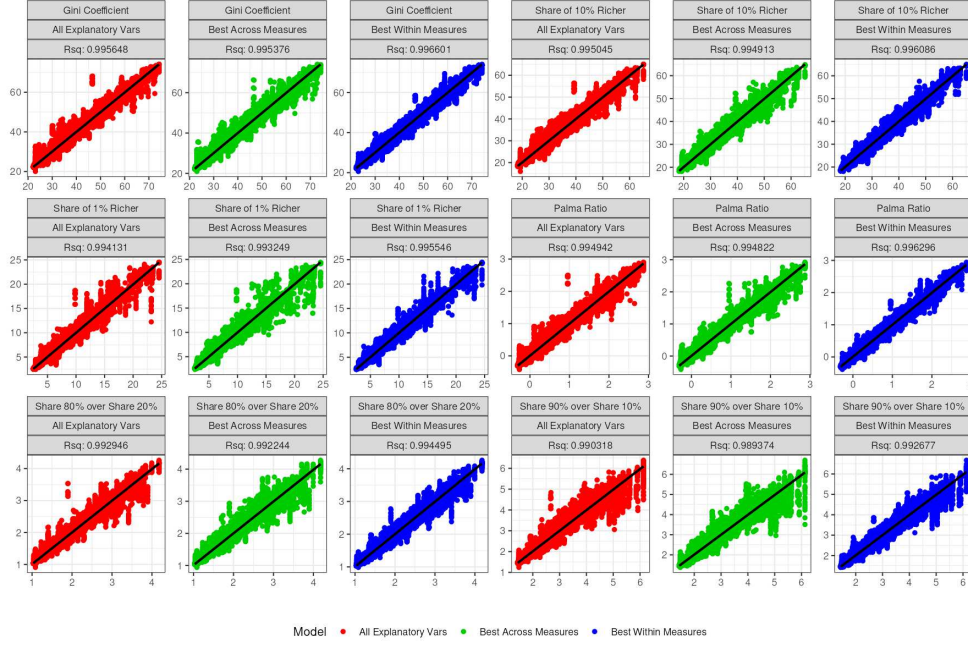
*The variables are grouped into thematic categories reflecting key theoretical channels through which inequality may be affected.

[‡]Variables calculated by the authors using the corresponding data.



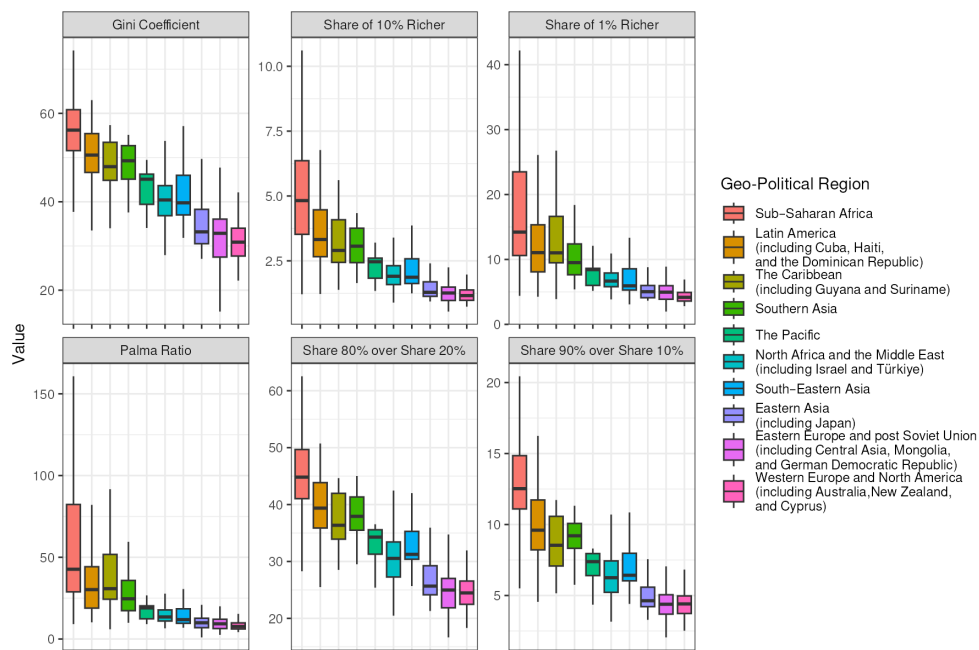
Notes: The figure reports scaled feature importance from *XGBoost* models selected under the inequality-specific specification. For each inequality measure, the model is chosen from all combinations containing exactly one candidate variable from each conceptual category, together with the retained general controls. The selected model minimizes the combined mean squared prediction error across the training and hold-out samples. Importance is normalized within each inequality measure (0–1). Colors denote thematic variable categories.

Fig. A3: Scaled Feature Importance using Inequality-Specific Predictors (Inequality Measure-Specific Specification - All Selected Variables Across All Bootstrap Iterations)



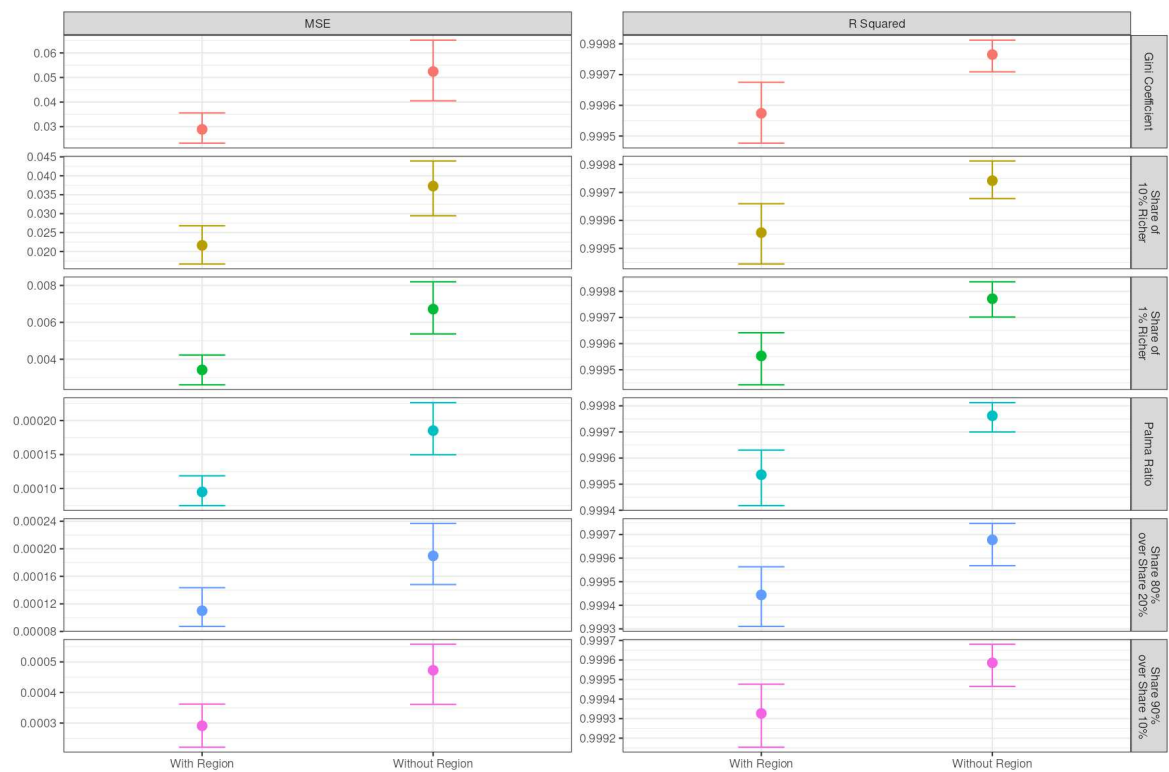
Notes: The figure plots predicted values against observed inequality values for each of three model specifications and six inequality measures. Predictions are obtained from the fitted *XGBoost* models under three specifications. “All Explanatory Variables” refers to the full specification, “Best Across Measures” to the reduced specification using category-representative predictors, and “Best Within Measures” to the inequality-specific specification. The solid line represents the 45-degree line corresponding to perfect prediction.

Fig. A4: Predicted versus Observed Inequality across Model Specifications



Notes: The figure shows the distribution of six income inequality measures across the ten politico-geographic regions defined by the V-Dem regional classification (*e_regionpol*). Boxes indicate the interquartile range, horizontal lines denote the median, and whiskers show the remaining distribution excluding outliers. Each panel corresponds to a different inequality measure. The figure is descriptive and does not imply a causal relationship between regional classification and income inequality.

Fig. A5: Distribution of Income Inequality Across Politico-Geographic Regions for Different Inequality Measures



Notes: The figure compares hold-out MSE and R^2 from otherwise identical *XGBoost* models estimated with and without the regional classification indicators. Hold-out MSE is the mean squared prediction error calculated using the 30% of observations not used to train the model; lower values indicate better predictive performance. Higher hold-out R^2 values indicate a closer correspondence between observed and predicted inequality. Points report average performance across bootstrap iterations, and lines represent 95% confidence intervals. The same training–hold-out split is used for both models within each iteration.

Fig. A6: Predictive Performance of XGBoost Models With and Without Regional Classification Across Inequality Measures