

# CROSS-NATIONAL VARIATION IN INCOME INEQUALITY AND ITS DETERMINANTS: AN APPLICATION OF BAYESIAN MODEL AVERAGING ON A NEW STANDARDIZED INEQUALITY DATA SET

---

Jiří Hasman, Josef Novotný\*

---

**Abstract:**

A new good quality standardized data set for cross-national comparisons of income inequality covering 147 countries has been constructed. The Bayesian model averaging and multiple imputation approach have been used to identify robust correlates from a larger pool of potential predictors of cross-national variation in inequality determined from previous literature. The results document significant macro-regional specificity both regarding levels and predictors of inequality. While globalization associates with lower inequality in Western countries, it has opposite effects in Latin America, Sub-Saharan Africa or post-communist countries. Age structure, the extent of social spending, or colonial history are another important factors with regionally specific impacts on inequality. By contrast, explanative power of some other traditional determinants of inequality has not been corroborated.

**Keywords:** Bayesian model averaging, cross-country analysis, income inequality, multiple imputation.

**JEL classification:** O15, D31, C11

## 1. Introduction

This paper focuses on the cross-national variation in income inequality and examines its determinants. We review available sources of data and compile a new standardized data set covering the most recent figures on income inequality for 147 countries. We overview existing literature offering a large number of sometimes competing theoretical explanations for cross-national variation in inequality and specify the measurable variables for our analysis. Based on these departures, the main goal of the empirical analysis is to compare the aggregate relevance of these alternative explanations one against the other.

Although the cross-national analysis of inequality is a traditional topic, this paper contributes to the existing evidence in the four respects. Firstly, as already mentioned, we use a new standardized inequality data set. Secondly, we apply Bayesian model averaging (BMA) to identify most relevant predictors of inequality when reducing the arbitrariness of model specification omnipresent in traditional regression approaches. Thirdly, we apply a multiple imputation approach to deal with the missing data uncertainty associated with incomplete data on independent variables. Fourth, having confirmed a significant

---

\* Jiří Hasman, Department of Social Geography and Regional Development, Faculty of Science, Charles University in Prague ([jirka.hasman@seznam.cz](mailto:jirka.hasman@seznam.cz)); Josef Novotný, Department of Social Geography and Regional Development, Faculty of Science, Charles University in Prague ([pepino@natur.cuni.cz](mailto:pepino@natur.cuni.cz)).

macro-regional resemblance in the levels of income inequality, we assume that different drivers of inequality may operate in different parts of the world. In addition to the global level analysis, we therefore separately examine the determinants of inequality in four broader regional categories of countries.

## 2. Standardized Inequality Data Set

More frequently than not incomparable inequality data (most typically expressed by the Gini coefficient) are used for cross-national comparisons. The main methodological problems stem from differing income definitions, issues of survey representativeness, or questions about the reference unit used, among other less influential issues (Atkinson and Brandolini, 2001; Deininger and Squire, 1996; Gottschalk and Smeeding, 1997). There is a trade-off between comparability and coverage of inequality data (Solt, 2009). To reach an acceptable coverage, inequality figures with different data conceptions metrics are typically adjusted by adding or subtracting a constant calculated from the information contained in the existing data (Deininger and Squire, 1996; Babones and Alvarez-Rivadulla, 2007). The usage of a universal constant has nevertheless been criticized because differences between various data conceptions may vary considerably across countries (Atkinson and Brandolini, 2001; Mukhopadhyaya, 2004; Solt, 2009).

When constructing our standardized inequality data set we started from the World Income Inequality Database (WIID), which is considered as the most extensive inequality database. However, as the most recent version of WIID does not include the latest data available, we used this database as a primary source only for a third of the countries (we also excluded low quality WIID data and excessively volatile observations). The remaining two thirds of inequality figures considered here are based either on the other smaller regional inequality collections or on information obtained directly from national statistical offices. The data set considered in this paper can be downloaded on: [http://web.natur.cuni.cz/~pepino/SN\\_Hasman\\_Novotny\\_PEPs.xlsx](http://web.natur.cuni.cz/~pepino/SN_Hasman_Novotny_PEPs.xlsx) with all relevant specifications can be obtained from the authors upon a request.

The inequality definition applied here is the net household equivalent income. We collected data on income inequality for 81 countries, while for the remaining countries only data on expenditure inequality was available. For half of these 81 countries, the net income inequality was directly reported, while for the rest it was gross income or, worse, the definition of income was not specified, and so it had to be estimated according to older WIID records. The Gini indices for countries with data based on expenditures or gross income were adjusted as follows: Firstly, all pairs of records for the same country and year but different definition were identified (68 pairs with both income and expenditure inequality observations and 57 pairs with both net and gross income inequality observations). The average differences within individual pairs of these observations were then calculated. Unlike earlier papers, we followed the assumption of a significant macro-regional similarity of income distributions (also corroborated below) and used macro-regional averages instead of the global level figures. Although the availability and geographical distribution of the above mentioned pairs of records allowed us to consider merely five broader macro-regions, the differences in constants for these regions shown in Table 1 are significant. The net income Gini indices for missing observations were adjusted accordingly.

**Table 1 | Macro-Regional Constants Used for Gini Index Adjustments**

Macro-regional groups	Expenditure → Income	Gross income → Net income
<b>North-West Europe, South Europe, Anglo-Saxon countries</b>	+0.45	-3.88
<b>Post-Soviet countries (excl. Baltic countries)</b>	+5.82	-
<b>East Asia, South-East Asia, South Asia, Middle East and North Africa</b>	+4.45	-1.61
<b>Latin America, Sub-Saharan Africa</b>	+7.19	-3.02
<b>Central-East Europe</b>	+2.62	-6.85

With respect to the reference unit, there were eight countries where observations for households were reported in a non-equalized form and one case where family was used. Analogously as above, the 85 pairs of records were identified for the same countries and years but differing in reference unit. Because of missing data for some macro-regions, here only a rough division into two groups of developing and developed countries was possible. Equalized household income inequality was found to be on average 3.88 Gini points and 1.32 Gini points lower than non-equalized household inequality in developed and developing countries, respectively. The Gini indices for the respective countries were adjusted accordingly.

Two recent attempts to construct standardized inequality databases are the Standardized World Income Inequality Database (SWIID) constructed by Solt (2009) and the Standardized Income Distribution Database (SIDDD) provided by Babones and Alvarez-Rivadulla (2007). The high correlation coefficient of 0.92 calculated between our data and the SWIID figures suggests good correspondence between these two data sets. The correspondence with SIDDD is weaker with correlations of 0.64 (to our data) and 0.67 (between SIDDD and SWIID).

We noted that SIDDD inequality data are considerably higher with respect to Anglo-Saxon, South-European, and North-West European countries and significantly lower for South Asia regions. These differences are not a matter of a few influential observations but reflect wider inconsistencies in the Gini indices reported in SIDDD for countries from these regions. The application of numerous interpolations and extrapolations, as well as the fact that the SIDDD figures are not newer than 2005, may provide two possible explanations.

Although we mentioned the tight correlation between our inequality data and those in SWIID, we have also found that figures in the latter data set are generally lower and that this is especially true for countries with more unequal income distributions. This may reflect the fact that data from the Luxemburg Income Study, which covers very few developing countries, served as the standard when constructing SWIID.

We also attempted to quantify the extent of geographical similarity in income distributions by decomposing the overall cross-national variation in Gini indices into its between-region and within-region components using the Theil index. These results suggest that 59% of the overall cross-national variation in Gini indices can be attributed

to differences between averages of five macro-regional groups (as in Table 1). Such macro-regional similarity of income inequality figures suggests that the role of particular determinants of inequality may also differ significantly between these regions.

### 3. Potential Determinants of Cross-National Differences in Income Inequality

The existing literature offers a large number of (sometimes) competing assumptions concerning possible explanations for cross-national differences in inequality. Here, we provide a brief overview of this evidence and determine the proxy measures for our analysis.

It is argued that many post-colonial societies still suffer from an institutional framework established directly or indirectly by their colonial powers. Differential socioeconomic trajectories, including differences in income distributions, have been explained by the presence of colonial settlers in combination with various ultimate determinants including environmental factors such as favourability of agricultural conditions for the production of cash crops, abundance of natural resources, or variation in settlers' mortality rates (Sokoloff and Engerman, 2000; Acemoglu *et al.*, 2002). In particular, Angeles (2005) has shown that the highest inequality can be found in societies with a 10–30% share of settlers who were able to control resources and who shaped institutions in their and their descendants' interest. By contrast, inequality tends to be lower both in societies where a greater proportion of colonists settled (as they gradually adopted European-like institutions) and in societies with only a minor share of colonists (as in many post-colonial Asian societies where settlers dealt mainly with administration and tax collection, while resources remained to a larger extent in the hands of local inhabitants).

Unequal capabilities established during the era of colonialism related to the appropriation of land, control of government, monopoly of armed forces, or involuntary movement of people have been considered an important explanation of today's inequality (Putterman and Weil, 2010). The initial concentration of land ownership in the hands of the elite is regarded as a major obstacle for the emergence of human-capital promoting institutions along a broader process of social modernization (Galor *et al.*, 2009). Among its other consequences, land inequality determined educational and skill heterogeneity and, through the latter, it affected modern income inequality (Glaeser, 2006; Gradstein, 2007).

Here, we will employ four proxy measures that can, at least partially, account for the cross-national variation in past colonial experience and other historical determinants. The first one is a simple dummy variable regarding whether or not a country was under colonial rule. The second measure of this type captures the proportion of European settlers (Japanese in the cases of South Korea and Taiwan) in the local population as reported in Angeles (2005). The final two historical variables are GDP *per capita* in 1820 and 1913 as estimated by Angus Maddison. These latter two variables are considered with regard to the institutional reversal attributed to European colonial intervention (Acemoglu *et al.*, 2002).

Regarding “modern” determinants of inequality, the level of economic development has been examined intensively as a determinant of income inequality, especially following the Kuznets's hypothesis about an inverted U-shaped trajectory of income inequality along the process of economic development. While the original argument focused primarily on a longitudinal development, the relationship has been analogically studied

in a cross-sectional manner. Here we include the present GDP *per capita* among the set of independent variables.

The long-term development of inequality can be seen as part of a broader modernization process with a variety of possible underlying mechanisms, including but not limited to economic ones (Glaeser, 2006; Tam, 2008). Economic modernization is traditionally viewed in close association with political modernization. Therefore, democratization and the level of democracy have also been examined frequently as predictors of inequality, while predicting moderating effects of the variables such as extent of participation in political and public life, protection of civil rights or good governance indicators (Li, Squire and Zou, 1998; Gradstein and Milanovic, 2004; Gradstein, 2007). However, it is also suggested that the relationship between democracy and inequality is not necessarily linear, and so the political Kuznets curve has been proposed as an analogy to the economic Kuznets curve (Acemoglu and Robinson, 2002; Tam, 2008). Importantly, the empirical evidence has indicated that the historical and geographical context can matter a lot. This has been illustrated for example by the dramatic increases in inequality in many post-communist societies along with their recent democratization, or by the experience of some of the Asian countries with undemocratic regimes but relatively equal distribution of income (Gradstein and Milanovic, 2004). In our analysis we employ two proxy indicators related to democracy and quality of governance – the Freedom in the World Index and the Corruption Perception Index.

Economic and political modernization is usually related to growing social expenditures and increasing social redistribution. Thus, the size of the social system can be presumed as another potential correlate of inequality (Beblo and Knaus, 2001; Goñi *et al.*, 2011). In addition to the first-round moderating effects of social redistribution, the literature suggests that there may also be some additional adverse effects on inequality. These are based on assumptions about voting preferences or negative behavioural effects of generous social spending on decisions regarding the labour market or savings and investment (Niehues, 2010). Overall, most of the existing empirical work in this area is limited to developed countries and the evidence seems to be inconclusive. Because of limited data availability, we eventually decided to include only one variable of this type – the social security expenditures in proportion to GDP.

A very traditional group of factors that may explain much of the cross-national variation in inequality concerns the extent of dualism and related structural characteristics of the economy (Bourguignon and Morrisson, 1998). This literature has highlighted the effects of wage and productivity differentials between traditional and modern sectors of the economy on income distribution. The process of urbanization in an inter-temporal view and the level of urbanization in a cross-sectional view have thus been regarded as important determinants of inequality. In many countries, however, the effects of the urban-rural divide have been in decline at the expense of other factors such as skill biased technological change (Cornia and Kiiski, 2001; Glaeser, 2006). The idea of a gradual structural transformation incorporates the shift of a major part of the labour force away from agriculture, hand in hand with increases in agricultural productivity. The variables that can account for the structural economic differences considered here are labour productivity in agriculture relative to the rest of economy, proportion of employment in the industrial sector, ratio of arable land to total area, and the rate of urbanization.

Another group of potentially important variables concerns various forms of socio-cultural heterogeneity. It can be assumed that, controlling for other factors, countries that are more homogeneous in terms of ethnicity, language, or religion among other socio-cultural characteristics will display lower economic inequality (Glaeser, 2006; Putterman and Weil, 2010). Some reasonable explanations refer to the association of socio-cultural differentiation with skill and educational stratification, and to various forms of minority discrimination. The inter-group animosity in more ethnically fractionalized societies may limit support for welfare and reduce the tendency to redistribute (Glaeser, 2006). We use two different variables of socio-cultural heterogeneity including the measure of ethnic fractionalization and the wider fractionalization coefficient computed as the geometric average of the measures of ethnic and religious fractionalization (Alesina *et al.*, 2003). Ethnic fractionalization may become additionally important as a driver of inequality if the country is rich in commercially attractive resources. Competition for these resources can, especially in institutionally weak and resource-dependent societies, catalyse rent seeking and fuel economic inequality (Bourguignon and Morrisson, 1990; Fum and Hodler, 2010). We include a simple variable indicating the extent of dependency on natural resources – the share of mining in GDP.

Another major factor emphasized in literature is education, though the direction of its effects on inequality is not clear. Not only can its impact depend on the level of education in question but also both the equalizing and un-equalizing effects of education on income distribution tend to operate in parallel (Ram, 1984; Gottschalk and Smeeding, 1997; Cornia and Kiiski, 2001). It is also argued that the explanative power attributable to the effects of educational variables has decreased significantly in recent decades (Bourguignon and Morrisson, 1998). Here, we employ two basic educational variables – average years of schooling and the literacy rate.

Andreski (1968) argued that countries with higher military participation tend to have more equal distribution of income. Kick *et al.* (2006) tested this hypothesis in a cross-sectional setting having found that the share of soldiers in the population is indeed negatively associated with inequality. Therefore, military service can be viewed as providing a redistributive and socially integrative function in terms of the opportunity of upward mobility for a relatively unskilled segment of population. Here we will consider this argument by examining two military variables – the share of soldiers in the population and military expenditures as a proportion of GDP.

Several authors believe that income inequality can be significantly influenced by the age structure and population dynamics when proposing several different underlying mechanisms. In many developing countries a large army of youth entering the labour market can be considered an important reason for higher income inequality (Galbraith and Kum, 2005). Similarly, the ratio of younger to older age groups of the economically active population can be positively associated with inequality through age dependent wage differentials (Higgins and Williamson, 1999). Alternatively, a larger retired population may contribute to lower inequality (Gottschalk and Smeeding, 1997). We include three variables to capture the age structure – shares of the population in three age categories: 15–24, 40–59, and 65 and above.

**Table 2 | Descriptive Statistics for Independent Variables**

Variable	N	Mean	Std. dev.	Source
<b>Former colony (dummy)</b>	147	0.63	0.48	Economy-point.org (2011)
<b>Percentage of European settlers*</b>	121	51.75	46.65	Angeles (2005)
<b>GDP per capita 1820 PPP</b>	45	776	338	Maddison (2008)
<b>GDP per capita 1913 PPP</b>	55	2,069	1,441	Maddison (2008)
<b>GDP per capita 2007 PPP</b>	147	12,496	14,221	CIC (2009)
<b>Freedom in the World</b>	147	3.177	1.835	Freedom House (2008)
<b>Corruption Perception Index</b>	146	4.003	2.130	Transparency International (2008)
<b>Social security expenditure (% of GDP)</b>	86	12.49	8.87	International Labor Organization (2008)
<b>Relative labour productivity in agriculture</b>	106	0.455	0.260	World Bank (2011); computed as in Bourguignon and Morrisson (1998)
<b>Urbanization rate (%)</b>	146	57	22	UNSD (2009b)
<b>Industry employment (%)</b>	122	20.0	8.1	World Bank (2011)
<b>Arable land share (%)</b>	146	15.90	13.31	FAO (2009)
<b>Ethnic fractionalization</b>	140	0.453	0.263	Alesina et al. (2003)
<b>Fractionalization coefficient</b>	140	0.400	0.217	Alesina et al. (2003)
<b>Mining share (% in GDP)</b>	142	9.51	13.29	UNSD (2010)
<b>Average years in schools</b>	133	12	3	UNSD (2009b)
<b>Literacy rate (%)</b>	131	83	20	UNSD (2009b)
<b>Soldiers (‰ population)</b>	142	4.0	3.8	U.S. Department of State (2005)
<b>Military expenditures (% GDP)</b>	138	7.5	5.5	U.S. Department of State (2005)
<b>KOF Index of Globalization</b>	126	64.25	17.35	Dreher (2006)
<b>Economic Freedom</b>	125	6.65	0.77	Fraser Institute (2011)
<b>15–24 age group share (%)</b>	126	17.52	3.53	UNSD (2009a)
<b>40–59 age group share (%)</b>	126	20.80	6.46	UNSD (2009a)
<b>65+ age group share (%)</b>	126	8.74	5.37	UNSD (2009a)
<b>Country area (km<sup>2</sup>)</b>	147	816,503	2,163,667	FAO (2009)
<b>Population density per km<sup>2</sup></b>	147	128	193	FAO (2009), CIC (2009)

CIC = Center for International Comparisons of Production, Income and Prices; FAO = Food and Agriculture Organization of the United Nations; UNSD = United Nations Statistics Division; \*Figures for South Korea and Taiwan refer to the share of Japanese settlers; Detailed description of the sources of data can be found in supplementary material accessible on: [http://web.natur.cuni.cz/~pepino/SM\\_Hasman\\_Novotny\\_PEPs.xlsx](http://web.natur.cuni.cz/~pepino/SM_Hasman_Novotny_PEPs.xlsx).

Much of the recent debate about income inequality has focused on the effects of globalization and economic liberalization (Cornia and Kiiski, 2001; Dreher and Gaston, 2008; Bergh and Nilsson, 2010). Although globalization is conceptualized in a variety of ways, it is mostly viewed as contributing to more unequal distribution of incomes, and

this notion is supported by some empirical evidence (Cornia and Kiiski, 2001; Bosch and Manacorda, 2010). The measure included in our analysis is the KOF Index of Globalization. Globalization is implicitly linked to economic openness and liberalization and, generally, to the degree of economic freedom. Therefore, in our study we also consider the Economic Freedom of the World 2011 Index, which attempts to measure the extent to which the policies and institutions of countries are supportive of economic freedom.

Finally, we account for differences in population density and country size, another two factors studied in the literature on the determinants of inequality (Long *et al.*, 1977; Beenstock, 2005). Detailed info about all of the proxy variables including the description of sources of the data appear in Table 2.

#### 4. Methods

From different strategies applied to the problem of variable selection under model uncertainty, a flexible option of choice that we used in this paper is Bayesian model averaging (BMA). In a regression framework, BMA does not determine any best combination of predictors but calculates the weighted posterior probability of estimates over many “best models” with a very good fit. To apply BMA, certain prior probability distribution for unknown parameters has to be chosen. Here we opted for uninformative prior distribution, assuming the equal position of all of the variables being compared. Because of prior uncertainty about model size, we chose random prior, which assigns the same prior probability to all model sizes. Since a relatively high number of variables considered in our analysis, we used Markov Chain Monte Carlo (MCMC) sampling. We preferred analytical posterior likelihoods computed from best models to MCMC frequencies, so we saved the 5,000 best models in each calculation. We also ran some control calculations of BMA with different settings.

Instead of omitting cases with missing observations (in the case of independent variables) or attempting to estimate their single values, we opted for the multiple imputation approach. This technique replaces each missing observation with a set of plausible values trying to account for the uncertainty about the right value to impute. Several full (imputed) data sets are obtained and then analysed separately (here using BMA), while the partial results are eventually combined into the final estimates of the required parameters. Here we worked with five imputed datasets, while the final results were obtained as described by Rubin (1987).

One important assumption followed in this paper is that the effects of individual predictors on inequality may differ from one part of the world to another. Surprisingly, this kind of spatial non-stationary has been largely neglected in the income inequality literature. At most, regional dummy variables for specific regions are included in regression models, while these dummies usually turn out to be significant (Bourguignon and Morrisson, 1998; Angeles, 2005). These findings, however, shed almost no light on explanations for such regional specificity. Here, we try to take a step forward by undertaking the BMA analysis both at the global level (with regional dummies included) and separately for four broader macro-regional country groups.



## 5. Results

In this section, we report main findings obtained for the global level analysis and for individual world regions. Space restrictions do not allow presenting all of the extensive results in tabulated form but they can be accessed on [http://web.natur.cuni.cz/~pepino/SN\\_Hasman\\_Novotny\\_PEPs.xlsx](http://web.natur.cuni.cz/~pepino/SN_Hasman_Novotny_PEPs.xlsx) among other supplementary materials to this paper.

At the global level, only three variables appeared in 100% of the best models (in the language of BMA it corresponds to the Posterior Inclusion Probability (PIP) equal to 1). These are the regional dummies for Latin America and Sub-Saharan Africa and the share of population aged 15–24. From the other variables, social expenditures and relative labour productivity in agriculture seem to be quite influential factors as well. The threshold of PIP above 0.1 has also been attained by the KOF Index of Globalization and literacy rate.

In order to identify factors that can make countries in Latin America and Sub-Saharan Africa so unique with respect to their income inequality levels, we firstly tried to repeat the global level BMA without regional dummies. A simple expectation was that the change in the PIP coefficients can indicate which variables are responsible for the specificity of Latin America and Sub-Saharan Africa. The highest change in PIP has been observed for the share of European settlers. This finding is well in accordance with the arguments stressing the colonial roots of income inequality in Sub-Saharan Africa and Latin American countries.

The findings obtained by comparison of the two variants of the global level analysis can be refined by the BMA procedure run only for the set of countries from Latin America and Sub-Saharan region. The results showed that the KOF index of globalization (indicating the positive relationship between globalization and inequality), the share of European settlers, and two age structure variables are the most robust predictors. In addition, relative labour productivity in agriculture has been found to be significant, which is not a complete surprise given the still high importance of agriculture in African and Latin American economies.

Turning to Asia and North Africa, the first notable finding is generally smaller PIP values also implying a small number of relevant variables identified. From individual predictors, the age structure variables have again emerged as important, though here it applies especially for the share of the elderly (65+) in population with a negative relationship with inequality. None of the historical variables exceeds 5% of the PIP coefficient, suggesting a lesser impact of colonialism on inequality in Asia and North Africa in comparison to Sub-Saharan Africa and Latin America. From the other variables, only the Freedom of the World Index exceeded the 10% PIP level with a positive sign regarding its relationship with inequality.

Regarding the group of post-communist countries, the BMA results understandably do not contain the historical variables of a colonial past. For this sub-set of countries the social expenditure variable has become the most robust concerning the PIP value, with a negative sign as expected. Again, the Freedom Index has fallen into the group of solid predictors suggesting that more free countries tend to have lower inequality, at least within the group of post-communist countries. For the first time, the economic level and the share of employment in industry are indicated as being significant candidates for determinants of income inequality in post-communist countries.

The results obtained for the group of Western countries show generally lower PIP values, with only social expenditures exceeding the PIP level of 0.3. In addition, the

Corruption Perception Index has been found important, revealing negative effects on inequality. A slightly lower PIP has been recorded for the KOF Index of Globalization that, however, attained a negative sign in the case of Western countries. Both military variables have become stronger in comparison to the previous regions. The signs of their effects are, nevertheless, in the opposite direction: while the military expenditure variable seems to have a positive impact on inequality, a negative relationship has been indicated with respect to the share of soldiers. Interestingly, and in contrast to previous findings, education and age structure seem to be inconsequential regarding the level of income inequality in Western countries.

To summarize the results, Table 3 provides an aggregate confrontation of most robust predictors (those exceeding 10% PIP level) identified by BMA on the global and regional level analyses.

**Table 3 | Ranks and Signs of Variables with PIP Values above 10%**

	Global level with regional dummies*	Latin America and Sub-Saharan Africa	Asia and North Africa	Post-communist Countries	Western countries
European settlers		4.(+)			
Per capita GDP 1,820		7.(-)			
Social expenditure in GDP	2.(-)			1.(-)	1.(-)
Population aged 15–24	1.(+)	3.(+)	3.(+)	6.(+)	
Population aged 40–59		5.(-)			
Population aged 65+			1.(-)	11.(+)	
KOF Index of Globalization	4.(+)	1.(+)		7.(+)	3.(-)
Freedom in the World Index**			2.(+)	4.(+)	
Corruption perception					2.(-)
Literacy	5.(+)	6.(+)		5.(-)	
Share of mining		8.(+)		8.(+)	
Relative labour productivity in agriculture	3.(-)	2.(-)			
Industry employment				3.(-)	
Fractionalization coeff.					6.(+)
Military expenditure				9.(+)	4.(+)
Share of soldiers					7.(-)
Per capita GDP 2007				2.(-)	
Country area					5.(+)
Population density				10.(-)	

\*Regional dummies for Latin America and Sub-Saharan Africa appear as the two most robust predictors.

\*\*Higher values mean fewer freedom.

## 6. Concluding Discussion

Using a new standardized inequality data set, we showed that both the levels and predictors of national income inequality vary considerably between world regions. Regarding those variables focusing on history of a colonial past, our analysis has indeed corroborated arguments referring to the adverse effects of the presence of European settlers in Sub-Saharan Africa and Latin America (as the two regions whose societies were hit hardest by the most extractive forms of colonialism), but not in other parts of the world.

From the modern correlates of inequality, the extent of social security expenditure has proved to be an important predictor of income inequality at the global level and for Western and post-communist regions, unequivocally with moderating effects on the distribution of net income. The extent of social expenditure is, however, not relevant for inequality in Africa, Latin America, and Asia – that is for regions with comparatively lower level of social security expenditure. With the exception of Western countries, the age structure has been confirmed as an important predictor of inequality, with differential impacts of specific age groups on income distribution. The share of the young, economically active population tends to have a positive relationship with income inequality, while the proportion of older groups has a mostly negative association (with the exception of the post-communist region with a weaker positive association between the share of elderly people and inequality). At least in part, this is what can be expected on the basis of previous literature. Speculatively, the findings on the impacts of social security and age structure might be interpreted together. The comparatively older population and more generous social security systems in Western countries can attenuate the effects of the demographic structure on income distribution in this region. Conversely, a younger population and relatively weaker social security in many Asian, African, and Latin American countries may be a reason for more pronounced impacts of age structure on inequality in these regions.

Interesting results have been obtained regarding the KOF Index of Globalization. While the involvement in globalization seems to be negatively associated with inequality within the Western world, the relationship reversed in Sub-Saharan Africa and Latin America. The processes of globalization work at different speeds in these regions with the opposite aggregate impacts of global integration on national income distribution. These findings generally fall in line with often quoted arguments stressing regionally unequal distribution of impacts of globalization in diverse domains. The evidence provided here on the robustness of Freedom of the World Index and the Corruption Perception Index supports arguments suggesting that democracy and quality of governance determine more equal distribution of income. We have found that this holds for all regional groups with the important exceptions of Latin America and Sub-Saharan Africa – *i.e.* except regions where the shortage of good governance is typically large.

The importance of education with respect to income inequality seems to be negligible, at least as far as the basic proxy variables of literacy and years of schooling are considered. Although the PIP values have in some cases exceeded the 10% threshold, the estimated effects are closed to zero. From the proxy variables capturing different aspects of cross-national variation in economic structure, perhaps the most interesting are results on the share of mining that have been identified as a significant predictor of inequality in Latin America and Sub-Saharan Africa and in the post-communist region. Importantly, in

both cases the indicated relationship has a positive sign, suggesting that the hypothesis about the adverse effects of extractive industries on income distribution in resource-based societies is relevant.

In addition, surprisingly weak evidence has been obtained in support of some traditional variables that are frequently mentioned in literature. For example, this is the case of socio-cultural fragmentation, which has not been found significant with the exception of the Western region, or of the present level of GDP, which has only been identified as a robust predictor of lower income inequality within the group of post-communist countries.

Finally, a note on the limitations of this study is necessary. The analysis attempted to deal with the problem of selecting variables under model uncertainty. The results thus should be interpreted primarily as the confrontation of one analysed variable against another, when considering their power to predict cross-national variation in inequality. This study has thus identified, which driers of inequality – from a larger pool of potential candidates – are most important in particular parts of the world. However, it is necessarily superficial regarding the underlying mechanisms operating behind the indicated relationships.

## References

- Acemoglu, D., Johnson, S., Robinson, J. A.** (2002), "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics*, Vol. 117, No. 4, pp. 1231–1294.
- Acemoglu, D., Robinson, J. A.** (2002), "The Political Economy of the Kuznets Curve." *Review of Development Economics*, Vol. 6, No. 2, pp. 183–203.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R.** (2003), "Fractionalization." *Journal of Economic Growth*, Vol. 8, No. 2, pp. 155–194.
- Andreski, S.** (1968), *Military Organization and Society*, 2nd Ed. London: Routledge & Kegan Paul.
- Angeles, L.** (2005), "Income Inequality and Colonialism." Centre for Growth and Business Cycle Research Manchester Discussion Paper Series 066.
- Atkinson, A. B., Brandolini, A.** (2001), "Promise and Pitfalls in the Use of "Secondary" Data-Sets: Income Inequality in OECD Countries as a Case Study." *Journal of Economic Literature*, Vol. 39, No. 3, pp. 771–779.
- Babones, S. J., Alvarez-Rivadulla, M. J.** (2007), "Standardized Income Inequality Data for Use in Cross-National Research." *Sociological Inquiry*, Vol. 77, No. 1, pp. 3–22.
- Beblo, M., Knaus, T.** (2001), "Measuring Income Inequality in Euroland." *Review of Income and Wealth*, Vol. 47, No. 3, pp. 301–333.
- Beenstock, M.** (2005), "Country Size in Regional Economic," in Felsenstein, D., Portnov, B. A., eds., *Regional Disparities in Small Countries*. Berlin Heidelberg New York: Springer, pp. 25–45.
- Bergh, A., Nilsson, T.** (2010), "Do Liberalization and Globalization Increase Income Inequality?" *European Journal of Political Economy*, Vol. 26, No. 4, pp. 488–505.
- Bosch, M., Manacorda, M.** (2010), "Minimum Wages and Earnings Inequality in Urban Mexico." *American Economic Journal: Applied Economics*, Vol. 2, No. 4, pp. 128–149.

- Bourguignon, F., Morrisson, C.** (1990), "Income Distribution, Development and Foreign Trade: A Cross-Sectional Analysis." *European Economic Review*, Vol. 34, No. 6, pp. 1113–1132.
- Bourguignon, F., Morrisson, C.** (1998), "Inequality and Development: The Role of Dualism." *Journal of Development Economics*, Vol. 57, No. 2, pp. 233–257.
- Cornia, G. A., Kiiski, S.** (2001), "Trends in Income Distribution in the Post-World War II Period." UNU World Institute for Development Economics Research Helsinki Discussion Paper 2001/89.
- Deininger, K., Squire, L.** (1996), "Measuring Income Inequality: A New Database." Paper presented at the Conference on Economic Growth Cambridge.
- Dreher, A., Gaston, N.** (2008), "Has Globalization Increased Inequality?" *Review of International Economics*, Vol. 16, No. 3, pp. 516–536.
- Fum, R. M., Hodler, R.** (2010), "Natural Resources and Income Inequality: The Role of Ethnic Divisions." *Economics Letters*, Vol. 107, No. 3, pp. 360–363.
- Galbraith, J. K., Kum, H.** (2005), "Estimating the Inequality of Household Incomes: A Statistical Approach to the Creation of a Dense and Consistent Global Dataset." *Review of Income and Wealth*, Vol. 51, No. 1, pp. 115–143.
- Galor, O., Moav, O., Vollrath, D.** (2009), "Inequality in Land Ownership, the Emergence of Human Capital Promoting Institutions, and the Great Divergence." *Review of Economic Studies*. Vol. 76, No. 1, pp. 143–179.
- Glaeser, E. L.** (2006), "Inequality," in: Weingast, B.R., Wittman, D.A., eds., *The Oxford Handbook of Political Economy*. Oxford: Oxford University Press, pp. 624–641.
- Goñi, E., López, J. H., Servén, L.** (2011), "Fiscal Redistribution and Income Inequality in Latin America." *World Development*, Vol. 39, No. 9, pp. 1558–1569.
- Gottschalk, P., Smeeding, T. M.** (1997), "Cross-National Comparisons of Earnings and Income Inequality." *Journal of Economic Literature*, Vol. 35, No. 2, pp. 633–687.
- Gradstein, M.** (2007), "Inequality, Democracy and the Protection of Property Rights." *Economic Journal*, Vol. 117, No. 516, pp. 252–269.
- Gradstein, M., Milanovic, B.** (2004), "Does Libertè = Egalité? A Survey of the Empirical Links between Democracy and Inequality with Some Evidence on the Transition Economies." *Journal of Economic Surveys*, Vol. 18, No. 4, pp. 515–537.
- Higgins, M., Williamson, J. G.** (1999), "Explaining Inequality the World Round: Cohort Size, Kuznets Curves and Openness." National Bureau of Economic Research Cambridge (MA) NBER Working Paper No. 7224.
- Kick, E. L., Davis, B., Kentor, J.** (2006), "A Cross-National Analysis of Militarization and Inequality." *Journal of Political and Military Sociology*, Vol. 34, No. 2, pp. 319–337.
- Li, H., Squire, L., Zou, H.** (1998), "Explaining International and Intertemporal Variations in Income Inequality." *The Economic Journal*, Vol. 108, No. 446, pp. 26–43.
- Long, J. E., Rasmussen, D. W., Haworth, C. T.** (1977), "Income Inequality and City Size," *The Review of Economics and Statistics*, Vol. 59, No. 2, pp. 244–246.
- Mukhopadhyaya, P.** (2004). "World Income Inequality Data Base (WIID) Review." *Journal of Economic Inequality*, Vol. 2, No. 3, pp. 229–234.
- Niehues, J.** (2010), "Social Spending Generosity and Income Inequality: A Dynamic Panel Approach." German Socio-Economic Panel Study Berlin SOEP Papers 336.
- Putterman, L., Weil, D. N.** (2010), "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality." *The Quarterly Journal of Economics*, Vol. 125, No. 4, pp. 1627–1682.

- Ram, R.** (1984), "Population Increase, Economic Growth, Educational Inequality, and Income Distribution: Some Recent Evidence." *Journal of Development Economics*, Vol. 14, No. 3, pp. 419–428.
- Rubin, D. B.** (1987), *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- Sokoloff, K. L., Engerman, S. L.** (2000), "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World." *Journal of Economic Perspectives*, Vol. 14, No. 3, pp. 217–232.
- Solt, F.** (2009), "Standardizing the World Income Inequality Database." *Social Science Quarterly*, Vol. 90, No. 2, pp. 231–242.
- Tam, H.** (2002), "An Economic or Political Kuznets Curve?" *Public Choice*, Vol. 134, No. 3–4, pp. 367–389.