EGE AKADEMİK BAKIŞ / *EGE ACADEMIC REVIEW*

Cilt 22 • Sayı 4 • Ekim 2022
SS. 473/486
Doi: 10.21121/eab.1063549
Başvuru Tarihi: 26.01.2022 • Kabul Tarihi: 28.06.2022

The Nexus Between Income Inequality and Tax Composition: A Cross-Country Perspective

Oya EKİCİ[1]

ABSTRACT

We perform an empirical analysis to investigate the relationship between income inequality and the tax composition of countries. We first group the countries with respect to the composition of their taxes on income and taxes on goods and services as a share of gross domestic product. The clustering method is employed to identify country groups. Then, we examine the effect of being in a category on the countries' income inequality represented by the Gini index. To deal with the endogeneity issue, we use the instrumental variable method in the analysis. We find that the composition of tax revenues of countries is associated with the Gini index and the countries that impose a higher tax on income relative to tax on goods and services expose a lower level of income inequality; whereas the countries that place a higher tax on goods and services expose a higher level of income inequality. As a policy implication, structuring an effective tax composition will ultimately help the economies reduce inequality. Understanding the potential of the distributive effect of fiscal policy will contribute to managing taxation tools better and thus improving economic development.

Keywords: Gini index, Income Inequality, Taxes on Income, Taxes on Goods and Services, Clustering.

JEL Classification Codes: H20, H21, D63b.

INTRODUCTION

According to the World Inequality Report 2018, income inequality in many countries has risen in recent decades. However, it does not grow at the same rate across countries due to their different national policies and institutional structures (Alvaredo et al., 2018). Countries' fiscal policies are primary instruments that affect their national income distributions. Therefore, there is an increased interest in analyzing the tax systems and implication differences of countries.

Although a significant part of government revenues comprises tax revenues in many countries, their taxation structures expose important country variations (see Figure 1). Revenues from tax fall into two categories – indirect taxes, which comprise the taxes on goods and services, sales and trade, and direct taxes which involve the taxes on income, profits and property. For advanced economies, tax revenues rely considerably on income taxes. In the case of developing countries, these revenues heavily count on taxes collected from consumption and trade. But Turkey atypically has experienced a major shift from income to consumption taxes. As for low-income countries, direct taxes contribute a comparatively small portion to their revenues (ICTD, 2019; McNabb, 2017; Prichard et al., 2014). As countries become more developed, the share of direct taxes to gross domestic product (GDP) rises (Acosta-Ormaechea and Yoo, 2012; Bahl and Bird, 2008).

We focus in particular on the countries' internal income distributions and inequalities with an approach considering the overall development levels of countries. Income inequality is defined as a situation in which the distribution of income among individuals shows disparity in an economy. It is generally measured by Gini index. When seen from the countries' Gini indices data in the world development indicators table, no patterns are observed in the inequality level of the countries over time.

Many theoretical and empirical papers (Agnello et al., 2014; Ball et al., 2013; Ciminelli et al., 2019; Martinez-Vazquez et al., 2012; Muinelo-Gallo and Roca-Sagales, 2013) refer to the tax system as one of the most direct tools used to reduce income inequality. The studies of Adam et al. (2015) and Pickering and Rajput (2018) consider reverse causality and examine the effect of income inequality on tax policies. Similarly, Borge and Rattso (2004) explore the income distribution effect on tax structures and provide evidence that unequal income distribution shifts the existing tax burden to property tax. Drucker et al. (2017) refer to this reverse direct effect from net inequality on tax composition as political influence.

[1] Istanbul University, Faculty of Political Sciences, Department of Business Administration, Quantitative Methods, Istanbul University, SBF, Department of Business Administration, Beyazit Main Campus, Istanbul, oyaekici@istanbul.edu.tr

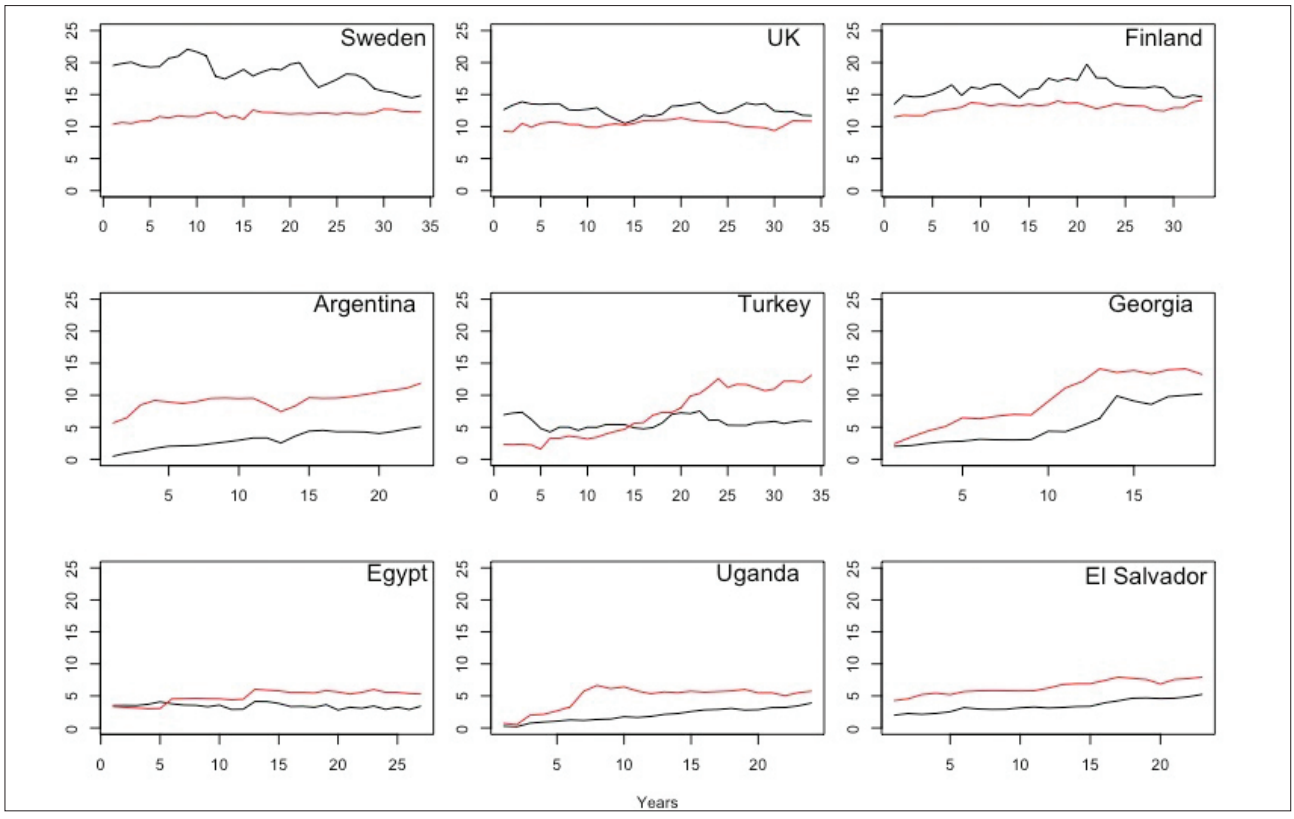

Figure 1. Selected countries' income (black) and consumption (red) taxes over time (% of GDP).

Fiscal policies of countries and their tax compositions, to some extent, are determined institutionally, and countries' tax preferences impact the level of income inequality. The associated empirical literature (Borge and Rattso, 2004) tests the hypothesis within countries. Drucker et al. (2017) and Akgun et al. (2017) use cross-country data to assess the tax structure effect on income inequality. Alternatively, considering the correspondence level, we explore preferred tax compositions and the resulting inequality from a cross-country perspective. Unlike other empirical studies, we perform a macro-level analysis that allows us to involve a wide range of country data and address the endogeneity problem with instrumental variable analysis.

In this study, we present evidence on the relationship between the tax composition of countries and their income inequality levels by analyzing cross-country data. For this, we cluster countries based on their tax composition: taxes on income and taxes on goods and services (as % of GDP). Next, we examine whether being in a cluster is associated with income inequality levels. Our empirical analysis reveals that income inequality is associated with the tax structure and compared to consumption tax: as the income tax increases, inequality also increases. With this study we articulate the theory of the relationship between tax policy and income inequality from a cross-county perspective. As a policy implication, structuring an effective tax composition ultimately helps the economies reduce inequality.

The paper is organized as follows: The next section provides a literature review, a problem-focused theoretical background, and the primary hypothesis of this study. Then, it describes the data and provides details about our method to investigate possible factors affecting income inequality. The paper continues with a section giving the estimation results and evaluation of the empirical findings. The last section discusses the overall findings of this study and offers some concluding remarks.

LITERATURE REVIEW

As is known, OECD has member countries non-homogeneous in terms of their development levels and account for almost 60% of world GDP. With this form, it provides a central knowledge source for reliable and representative data (OECD, 2022). The trends in income inequality of OECD countries highlight a long-term rise in income inequality. The growing disparities in income distributions of the countries lead the economies to be financially more fragile. It might also have the effect to undermining economic growth. There is a broad literature working on this effect. Some of them are the studies of Cingano (2014), Shin (2012), Voitchovsky (2005), and Mo (2000). Regarding economic and social welfare, the economies follow up on the behavior of income inequality changes. From this aspect, income inequality is one of the critical indicators that matter for an economy and requires getting to the bottom of it.

The factors affecting income inequality have been extensively investigated by different studies. The educational regime is a primary equalizer in the long run, as World Bank (2002) report and Glomm and Ravikumar (2003) show in detail. For the European Union region, Rodrigez-Pose and Tselios (2009) examine the causal factors of income inequality and indicate that higher inequality in educational attainment is associated with higher income inequality. The other factors that they determine are population aging, unemployment, female participation in the labor force, urbanization, agriculture, and industry. One can go over more detail the other determinants of income inequality and see the studies of, for example, Iacoviello (2008), Aghion et al. (2018), Hailemariam et al. (2021), and Jones and Kim (2018).

Additionally, for the countries at different stages of development, there is ample investigation on the changes in this indicator. But first and foremost, cross-country analyses depend on the availability of reliable and comparable databases. To estimate the existing income inequality database's accuracy and comparability, Solt (2020) revised the evaluations of standardized world income inequality database, as one of the most preferred sources. In reference to applications, Gottschalk and Smeeding (2000) make a cross-national comparison, and suggest that the level and trend in inequality in rich nations are different from those in the developing world. When Atkinson (2003) focuses nine OECD countries and considers the various hypotheses, he concludes a disparity on the income distribution changes of OECD countries. On the side of developing countries, Ravallion's (2014) research unveils that while inequality between developing countries is decreasing, average inequality within developing these countries is increasing. Roser and Cuaresma (2016) work with an average Gini index for 32 industrialized economies. Using that panel data set, they reveal that imports from low-income countries, democratization, and technological and educational interplays affect inequalities in industrialized countries.

Alternative to these determinants, taxation is one of the most widely accepted policy instruments for governments that need to control income distribution. The paper of Stiglitz (2010) emphasizes the importance of a distinctive tax design for countries that differ in structure and policy purposes. Similarly, Besley and Persson (2014) illustrate that low, middle, and high-income countries have various tax formations. Moreover, these heterogeneous structures drive the countries to change or broaden their tax base to increase their revenues. Gordon and Li (2009) mention the variability of tax structures across the countries to elucidate the differences in social preferences, such as policies on military spending, education, insurance, or investments. Thus, the researchers point out that governments of developed and developing countries use taxation instruments dissimilarly. These papers classify countries concerning their income level and then analyze their tax structure. However, considering the discrepancies between countries, we group them based on their tax compositions with a novel approach.

The heart of the matter discussed in this paper is the relationship between income inequality and tax composition, which is broadly analyzed by a substantial number of studies. While Poterba (2007), Martinez-Vazquez et al. (2012), Ball et al. (2013), Ciminelli et al. (2019), and Agnello et al. (2014) use the models searching the effect of the tax system on income inequality, conversely, Pickering and Rajput (2018), and Adam et al. (2015) investigate the effect of income inequality on the tax structure. Alternatively, Muinelo-Gallo and Roca-Sagales (2013) contribute by considering the mutual relationships between income inequality and economic growth in the determination of fiscal policy. All these researches by Adam et al. (2015), Poterba (2007), Pickering and Rajput (2018), Agnello et al. (2014), Ball et al. (2013), Ciminelli et al. (2019), Martinez-Vazquez et al. (2012), and Muinelo-Gallo and Roca-Sagales (2013) mention the endogeneity issue in common among tax composition and income inequality. To eliminate this, we propose a novel approach and add a new variable reflecting the tax structure. Then, in the estimation of the model, we use the instrumental variable method, which gives reliable results.

MORE FOCUSED THEORETICAL BACKGROUND AND HYPOTHESIS

As many previous studies remarks, the tax policy is one of the principal tools to prevent growing income inequality. (Agnello and Sousa, 2014; Ball et al. 2013; Cubero & Hollar, 2010; Woo et al., 2017). There is an extensive literature devoted to the examinations of this relationship. The prior economic research on income inequality has primarily focused on indicating the effect of the taxation levels of countries in micro level. The key argument that we develop in this paper is on the ability of reflecting the effect of governments' tax policy preferences. The determining the situation of countries' different tax compositions enable us to create an overall categorization of tax practices. So, we extend previous studies and consider the general tax composition with a single variable, Cluster, which is based on an objective clustering procedure and generated from a cross-country data. Then we examine the cluster effect with instrumental variables regression method. Hence, we state our main hypothesis.

***Hypothesis**: The tax compositions and the relevant practices of the countries have effect on income inequality in the countries. Higher dependence on income taxes results in lower levels of income inequality when compared to the indirect taxes*.

The other factors affecting the income inequality arise from a large body of literature analysis. Introducing the level of GDP, population, unemployment, investments, age dependency ratio (Kanbur, 2000; Tanzi, 1989; Devarajan et al., 1996; Jarven, 2013) allows for controlling the effect of macroeconomic factors. Besides, the effect of the various types of government expenditures

are discussed in the prior studies (Clements et al., 2015; Devries et al., 2011; Easterly, 1995; Fournier and Johansson, 2016; Martinez-Vazquez et al., 2012; Meltzer and Richard, 1983). We followed the literature in deciding the coverage of control variables.

DATA AND METHODS

For the empirical analysis, we use ordinary least squares (OLS) regression, generalized methods of moments (GMM) model and instrumental variables (IV) regression. We primarily run linear and ordered logistic regression models. By considering two-way causality, we account for the association between income inequality and tax compositions of countries. Before focusing on each country's income inequality change in response to their tax structure and vice versa, we group countries according to their two main tax components—income and consumption taxes—as a share of GDP and analyze these groups' association with income inequalities. Transforming the covariates into categories that reflect countries' tax structures did not help with the issue of endogeneity.

To deal with the endogeneity, we use instrumental variables regression. The method requires to introduce control variables so that the regression can find variables that is correlated with categories but uncorrelated with the error term of the model. Besides, in our analysis, including control variables enables the examination of other exogenous factors in income inequality.

Variable generated from clustering analysis

Our tax data (ICTD UNU-WIDER, 2019) consists of countries' total taxes on income (individual and corporate) and total taxes on goods and services (including VAT and sales taxes). This data is available for 189 countries. For each year, the number of available data changes: for example, for 2012, 64 countries are considered. From 2006 to 2012, for seven years, we clustered tax data (see Table 1 for the statistics of the clustering procedure). For overall descriptive statistics of Cluster variable (the variable generated from clustering analysis) see Table 2.

Based on the bivariate tax data, we group the countries with k-means clustering. The standard method using the Hartigan-Wong algorithm defines the clusters that have the minimum total intra-cluster variations. To exemplify our clustering outcome, we demonstrate the clusters defined for 2012 in Figure 2. Extreme unbalance data is not seen among the clusters in the covered years.

The three clusters shown in Figure 2 broadly correspond to the meaningful groups of countries with similar incomes: high-income, upper- or lower-middle income, and low-income economies. Also, these meaningful classifications are held over when different time periods are analyzed. For the periods under review, the optimal cluster number is three, which is confirmed by a Bayesian clustering analysis. Testing the main hypothesis of this study relies on the significance of the effect of the Cluster variable.

Dependent Variable

For our income inequality data (World Inequality Database, 2018), the Gini index is used as one of the most quoted measures. It quantifies the area between the Lorenz curve and 45-degree line that represents perfect equality. A Gini index of zero (0%) indicates perfect income equality, whereas an index of 1 (100%) corresponds to perfect income inequality. First row of Table 2 represents the descriptive statistics of the Gini index data (World Inequality Database, 2018).

Control variables

In order to clearly identify and control the cluster effect and to reflect the impact of other potential variables, we add a number of control variables. In one group of control variables, macroeconomic characteristic is represented by *GDP* per capita PPP in current international dollar units (*GDP*), unemployment rates as percent of total labor force (*Unemployment*), total population (*Population*), gross fixed capital formation as percent of GDP (*FixCapForm*), age dependency ratio as percent of working-age population (*AgeDep*) and household final consumption expenditure as percent of GDP (*HConsExp*). In the other group of control variables, government spending characteristic

Table 1. Clustering statistics for the years

Years	# of countries in Cluster 1	# of countries in Cluster 2	# of countries in Cluster 3	Total number of countries	Total number of countries for available data
2012	18	25	21	64	45
2011	13	31	22	66	46
2010	15	20	28	63	43
2009	18	27	25	70	49
2008	14	29	25	68	43
2007	18	22	22	62	46
2006	19	23	20	62	44
Total					**316**

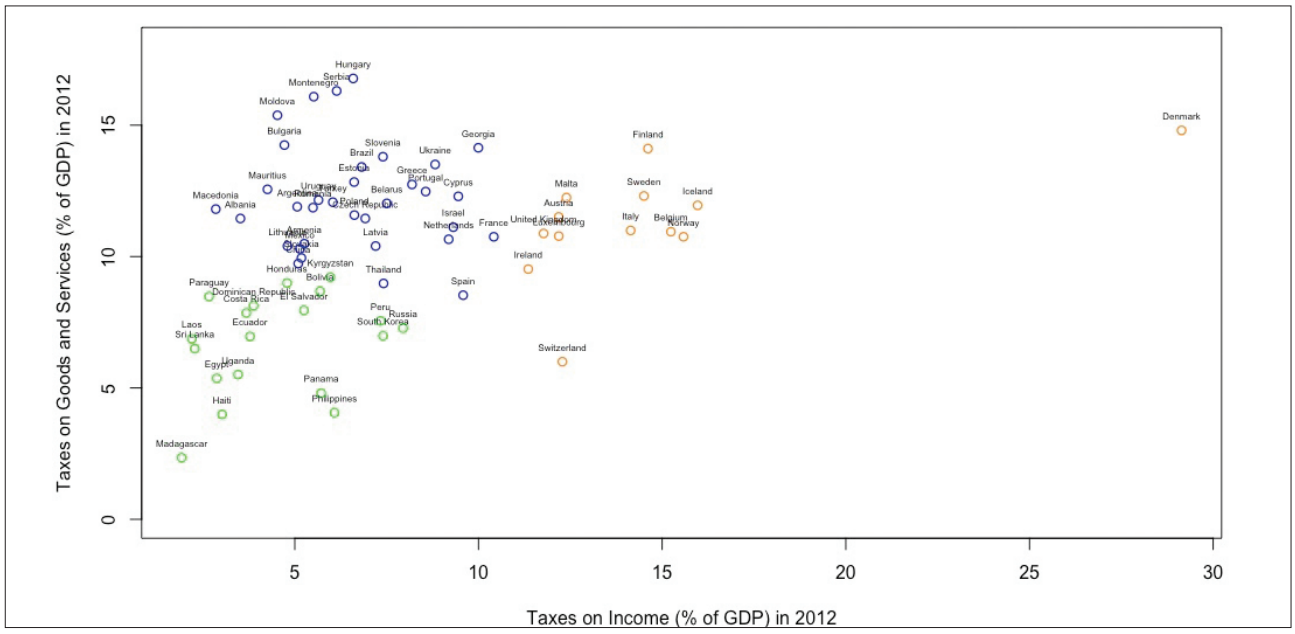

Figure 2. Country clusters for 2012 data.

Table 2. Descriptive Statistics

Variables	n	Mean	Sd	Median	Min	Max	Type of variable
Gini Index	316	33.11	5.94	32.60	23.70	63.40	numeric
Clusters	316	1.74	0.78	2.00	0.00	3.00	categorical
Population	316	16.00	1.64	15.86	11.39	21.02	numeric
GDP	316	9.99	0.72	10.15	7.20	11.43	numeric
Unemployment	315	8.24	4.51	7.27	0.58	26.67	numeric
FixCapForm	315	23.93	6.09	22.67	11.97	48.41	numeric
AgeDep	316	20.82	6.25	22.40	5.50	32.24	numeric
HConsExp	315	59.37	13.03	57.18	31.33	95.80	numeric
HealthExp	278	5.59	1.86	5.93	0.76	8.87	numeric
EducExp	278	5.21	1.18	5.36	2.58	8.03	numeric
SocProtExp	277	14.64	4.75	14.69	2.20	24.84	numeric
PubSerExp	279	6.06	2.37	5.57	0.23	17.11	numeric
PubOrdExp	253	1.89	0.65	1.84	0.00	5.15	numeric
EnvirExp	277	0.68	0.38	0.65	-0.26	2.11	numeric
RCultExp	278	1.19	0.56	1.15	0.14	3.57	numeric
EcoAffExp	278	5.14	2.28	4.64	1.64	25.04	numeric
AgricExp	241	0.73	0.71	0.58	-0.15	4.53	numeric
DefenExp	246	1.40	1.14	1.25	0.00	8.58	numeric
D2012	316	0.14	0.35	0.00	0.00	1.00	dummy
D2011	316	0.15	0.35	0.00	0.00	1.00	dummy
D2010	316	0.14	0.34	0.00	0.00	1.00	dummy
D2009	316	0.16	0.36	0.00	0.00	1.00	dummy
D2008	316	0.14	0.34	0.00	0.00	1.00	dummy
D2007	316	0.15	0.35	0.00	0.00	1.00	dummy

Note: Population and GDP data are in logarithmic scale.

is represented by government expenditures on health (*HealthExp*), education (*EducExp*), social protection (*SocProtExp*), public services (*PubSerExp*), public order (*PubOrdExp*), environment (*EnvirExp*), recreation, culture and religion (*RCultExp*), economic affairs (*EcoAffExp*), agriculture, fishing, forestry and hunting (*AgricExp*) and defense (*DefenExp*). All control variables under government spending characteristic are in percent of GDP. The data of control variables is retrieved from World Bank's World Development Indicators (World Bank, 2020).

We created a dummy variable for each year. The year dummy takes the value of one for the relevant year and zero otherwise. Down from the third row of Table 2 reports descriptive statistics for control variables and dummy variables.

Methods

Given the data, we preliminarily estimate a linear model and ordered logistic regression model to account for reverse causality. The models have the following forms:

$$Gini_i = \alpha_0 + \alpha_1 Cluster_i + \varepsilon_i$$
$$logit[Prob(Cluster_i \leq j \mid Gini_i)] = \beta_0 + \beta_1 Gini_i + \eta_i \quad (1.1)$$

where $Gini_i$ refers to ith countries' Gini coefficient and $Cluster_i$ refers to ith countries' tax structure cluster. α_0, α_1 denote, respectively, the intercept and tax structure effect. In the logistic regression model, j stands for the country category and β_0, β_1 represent, respectively, the intercept and inequality effect. Finally, ε_i, η_i are the error terms of the models. We initially use the models to test the causal link between tax structure and income inequality. The regular OLS regression does not consider heterogeneity across time and the omitted variables bring the bias problem of estimator. Therefore, we add control variables to the simple linear regression model and estimate OLS model with,

$$\begin{aligned} Gini_i &= \alpha_0 + \alpha_1 Cluster_i + \alpha_2 Population_i + \alpha_3 GDP_i \\ &+ \alpha_4 Unemployment_i + \alpha_5 FixCapForm_i + \alpha_6 AgeDep_i \\ &+ \alpha_7 HConsExp_i + \alpha_8 HealthExp_i + \alpha_9 EducExp_i + \alpha_{10} PubSerExp_i \\ &+ \alpha_{11} EnvirExp_i + \alpha_{12} RCultExp_i + \alpha_{13} EcoAffExp_i \\ &+ \alpha_{14} AgricExp_i + \alpha_{15} DefenExp_i + \sum_{j=1}^{k-1} \beta_j Dum_{ji} + \varepsilon_i \end{aligned} \quad (1.2)$$

where $Gini_i$ is the dependent variable, $\alpha_0, \ldots, \alpha_{h+1}$ are $1+h$ unknown parameters of h covariates. The β_j coefficient of time dummy Dum_{ji} is to measure potential change over time. By adding the dummy, we check for the unobserved time effects in the OLS model. Thus, the model provides a fixed effect OLS results. For fixed OLS method, since the data is not severely unbalanced, we do not need to remove the subjects or reduce the number of time periods or countries.

Due to the potential endogeneity between income inequality and the clusters we estimate both GMM and IV regression models. We fit GMM with the number of observations $i = 1, \ldots, n$

$$Gini_i = \gamma X_i + \varepsilon_i \quad (1.3)$$

where the dependent variable $Gini_i$ is q-*by*-*1* vector, X_i is a q-*by*-s matrix and ε_i is is q-*by*-*1* vector of errors. γ is unknown parameter vector with s-*by*-*1* dimension. GMM contains potential endogenous and exogeneous variables.

We adopt instrumental variable approach and identify the IV Regression model,

$$\begin{aligned} Gini_i &= \delta_0 + \delta_1 X_{i,Cluster} + \delta_2 X_{i,PubOrdExp} + \delta_3 W_{i,HConsExp} \\ &+ \delta_4 W_{i,HealthExp} + \delta_5 W_{i,EducExp} \\ &+ \delta_6 W_{i,PubSerExp} + \delta_7 W_{i,EcoAffExp} \\ &+ \delta_8 W_{i,AgricExp} + \sum_{j=1}^{k-1} \lambda_j W_{i,Dum_j} + \varepsilon_i \end{aligned}$$

and

$$\begin{aligned} Z = (&Z_{i,Population}, Z_{i,GDP}, Z_{i,Unemployment} \\ &, Z_{i,FixCapForm}, Z_{i,AgeDep}, Z_{i,SocProtExp} \\ &, Z_{i,EnvirExp}, Z_{i,RCultExp}, Z_{i,DefenExp}) \end{aligned} \quad (1.4)$$

with the number of observations $i = 1, \ldots, n$ where $Gini_i$ is the dependent variable, $\delta_0, \ldots, \delta_{m+1}$ are $1+m+r+k$ unknown regression coefficients. $W_{i,HConsExp}, \ldots, W_{i,AgricExp}, W_{i,Dum_j}$ are r+k exogenous variables which are uncorrelated with the error term ε_i. The λ_j coefficient of k dummy variable W_{i,Dum_j} is to account for the effect over years. In the model, $X_{i,Cluster}$ and $X_{i,PubOrdExp}$ are m endogenous variables that are instrumented by Z where consisting p variable $Z_{i,Population}, \ldots, Z_{i,DefenExp}$.

In all models above, we regress the lagged values of all explanatory variables.

FINDINGS

First, we interpret our clustering results. The three defined clusters are significative. Compared with the other two clusters, Cluster 1 consists of the countries with higher income tax relative to consumption tax and higher total tax revenue. Cluster 3 consists of the countries with lower income tax relative to consumption

tax and lower total tax revenues. Lastly, Cluster 2 consists of the countries with lower income tax relative to consumption tax but higher than the third cluster, and total tax revenue share is between the first and third clusters. Thus, $Cluster_i$ is our ordered categorical variable with the values 1, 2, and 3.

are unobservable effect which contributes to error term. Therefore, as a next step we fit an extended OLS model (1.2) which involves the control variables and the year-specific dummy variables. The estimation results of the model (1.2) are given in Table 3. In OLS the parameter a_1 let us to test our study's main hypothesis. Regarding

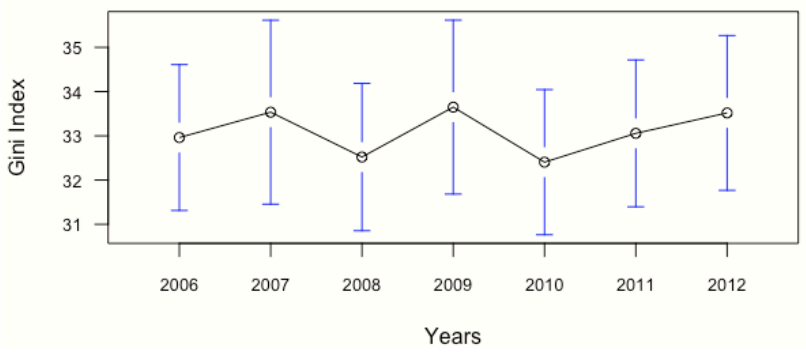

Figure 3. Heterogeneity across Years

It is plausible to assume the heterogeneity of the parameters across countries. Because, Gini indices varies for different level developed countries and based on this distinctness we achieve to classify the countries. On the other hand, we do not have reason to expect heterogeneity across the years. Our data exploration justifies this (see Figure 3).

If we start with preliminary results of simple linear regression and logistic regression analysis, we see that clusters associated with income inequalities and tax structures affect the income distribution for a given year (see Appendix A1 for Table A1). We regress the previous year's tax structure clustering to the existing year's Gini index. The results show that as the cluster number increases, which means that as the imposed income tax share decreases and tax revenues in total decreases, the Gini index—that is, income inequality—also increases. In this way, we treat examining the relationship for every year as a way of testing the causality. Replication of the analysis for different years confirms the findings. Consistently, the logistic regression analysis results indicate that as the inequality increases compared to consumption tax, the income tax share in the tax composition rises. We regress the previous year's Gini index to the existing year's tax structure clustering. We see that a one-unit increase in the Gini index is expected to result in a 0.22 increase in the odds of being in a higher number of clusters (see Appendix A2 for Table A2). For example, 6.16 is the expected odds of being defined as a Cluster 1 country. As the goodness-of-fit measure, the LR chi-square of 27.79 with a p-value of <0.0001 implies that the logistic model is statistically significant, as compared to the null model.

In the Methods section, we mentioned the issues of omitted variable bias. In the simple linear regression model, the other covariates are not included. Also, there

the Cluster effect on income inequality is significant at $p < 0.05$ level. The *Cluster* and *Gini* are positively related. As *Cluster* number increases *Gini* also increases. The control variables in macroeconomic characteristics, *Population*, *GDP*, *Unemployment* and *HConsExp* are positively related to *Gini* and significant. *AgeDep* has negative sign and significance. The other group of control variables in government spending characteristic, *HealthExp* and *EducExp* are negatively related to *Gini* and both are significant. The year specific fixed effects are jointly insignificant. We do not move forward in the interpretation on findings from OLS.

It is possible that *Gini* and *Cluster* variables to be affected by some of the control variables. In the case that these effects exist, OLS estimates become inconsistent. Hence, we employ instrumental variable analysis, and fit GMM (1.3) and IV regression model (1.4) to estimate *Cluster* effect. Since the standard errors of IV estimates are not consistent, we compute the sandwich estimator of covariance matrix for robust -heteroskedasticity consistent - standard errors.

Table 3. Results of Empirical Analysis (dependent is Gini Index)

	OLS		**GMM**		**IV Regression**	
Variables	*Estimate* (α)		*Estimate* (γ)		*Estimate*(δ, λ)	
Intercept	-20.40 (10.98)		42.66 *** (4.03)		41.21 *** (4.84)	
Cluster	0.85 * (0.42)		2.68 *** (0.90)		3.06 *** (0.95)	
Population	0.77 *** (0.18)					
GDP	3.87 *** (0.80)					
Unemployment	0.20 ** (0.07)					
FixCapForm	0.07 (0.06)					
AgeDep	-0.24 *** (0.06)					
HConsExp	0.21 *** (0.03)		0.17 *** (0.03)		0.14 *** (0.03)	
HealthExp	-0.78 *** (0.21)		-1.14 *** (0.29)		-0.99 *** (0.25)	
EducExp	-0.73 ** (0.27)		-0.87 *** (0.28)		-0.75 * (0.34)	
SocProtExp						
PubSerExp	-0.06 (0.11)		-0.55 *** (0.11)		-0.53 *** (0.13)	
PubOrdExp			-4.65 *** (0.65)		-4.03 *** (0.78)	
EnvirExp	-0.34 (0.69)					
RCultExp	-0.49 (0.56)					
EcoAffExp	-0.04 (0.12)		0.17 * (0.07)		0.14 * (0.12)	
AgricExp			-2.78 *** (0.61)		-2.71 *** (0.56)	
DefenExp	-0.004 (0.22)					
Dum2012	-0.42 (0.87)		-0.19 (0.85)		-0.37 (1.02)	
Dum2011	-0.25 (0.88)		-0.02 (1.07)		-0.60 (1.05)	
Dum2010	-0.07 (0.90)		0.30 (1.07)		0.09 (1.12)	
Dum2009	0.54 (0.86)		1.54 (1.12)		1.13 (1.05)	
Dum2008	-0.49 (0.88)		-0.08 (1.19)		-0.15 (1.03)	
Dum2007	-0.49 (0.86)		-0.37 (0.92)		-0.29 (0.97)	
					W.Inst. (Clust)	11.95 ***
					W.Inst. (PubOrdExp)	11.53 ***
The Tests for The Models	R^2	0.56			*Wu-Hausman*	32.62 ***
	F -Stat (p-val)	13.39 *** (<0.001)	*J-test (p-val)*	9.78 (0.20)	*Sargan (p-val)*	13.07 (0.07)
	Obs.	208	*Obs.*	204	*Obs.*	195

Notes: Robust standard errors in brackets. *** for p<0.001 ; ** for p<0.01; * for p<0.05.
Source: Author's calculations

In model specification, to decide on which variables are under the exogenous set and which are under the instrumental variables set, we apply the classification that relies on the distinction between productive and non-productive expenditures (discussed by Adam and Bevan, 2005 and Chu et al. 2020).

For the instrumental variables we determine two group of relevant variables sets: From the macroeconomic indicators; GDP, population, unemployment, gross fixed capital formation for investments, age dependency ratio and from the non-productive government spending; expenditures on social protection, environment recreation, culture & religion and defense. On the other hand, for the exogenous variables set we specify the productive type of government spending and cover the variables of the government expenditure on health, education, economic affairs, general public services and agriculture, fishing, forestry and hunting.

Based on the J-statistic of GMM model, we fail to reject the null hypothesis of instrumental variables are exogeneous. This suggest that our instrumental variables are valid. From F-test of IV regression, we reject the null hypotheses of the instruments are weak, so we inferred that the instrumental variables *Cluster* and *PubOrdExp* are sufficiently strong. Our decision remains unchanged when we consider the critical values for F test suggested in Stock et al. (2002). We apply Wu-Hausman test to check for the difference between OLS and IV regression results. We reject the null hypothesis implies that there is significant difference between two approaches and we need to employ IV regression method. For the validity check of instrumental variables, we use Sargan test. We get a similar result to J test of GMM, we fail to reject the null hypothesis; one or more instruments are invalid. The test results suggest that our instrumental variables are valid. Considering all these diagnostics results of IV analysis we move forward to interpret our estimated parameters of the IV model (1.4).

First, to explore the main hypothesis of this study we check the δ_1 estimation of *Cluster* variable in Column 3 of Table 3. Accordingly, *Cluster* variable significantly affects *Gini* and *Cluster* number has positive effect on the level of *Gini* index. Relying more heavily on consumption taxes is expected to increase the income inequality in the economy. As for the control variables under government spending characteristic, *HealthExp*, *EducExp*, *PubSerExp*, *AgricExp* have statistically significant negative effect on *Gini*. Government expenditures on essential considerations like health, education, agriculture and public services explicitly have a reducing impact on the level of income inequality. We check the cases where the sign of the effects flip. The column 3 in Table 3 suggests that among the government spending type of control variables, only *EcoAffExp* has significantly small effect but in opposite direction. Since this expenditure provides relatively an indirect type of support, it's not clear enough how it's allocated among the segments of the economy. Finally, the net effect of *HconsExp* on *Gini* index is positive and significant. The higher level of household final consumption expenditure is expected to be associated with a larger income inequality level. This is because the revenue from consumption taxes increases and it leads a less equal distribution in income. We might have considered the household consumption expenditure as endogenous, but it's out of the scope of this study. The estimation results of the dummies are insignificant across time specifications.

We have been limited by the weaknesses of cross-country income inequality, tax and government expenditures data. To work with standard-consistent data, we had to restrict the both time and country coverage.

The findings of this study put emphasis on the policies under which tax composition is beneficial to lowering income inequality of the countries. Developing countries rely more heavily on consumption taxes, whereas developed countries levy more tax on income. For that reason, determining the relevance of countries' income inequality levels and their tax compositions is essential in identifying policy interventions that countries' decision-makers should consider to reduce inequality. The required policy adjustment is feasible for developing economies. However, it is unlikely to be achieved for low-income countries without specifying other supportive policy tools. Since the policies that proposing higher proportion of income taxes requires a more powerful economic structure.

CONCLUSION

Tax composition of the countries play a pivotal role in controlling income inequality of countries by policies which change the distribution of tax among the segments of the society. To reduce income inequality, the commonly proposed adjustments are based on tax-related policies. Beyond this much discussed association in economic theory, we examine the relationship between tax structure and income inequality. We discuss the problem at the macro level and contribute to the literature by defining tax structures with a data mining method, k-means. The analysis generated meaningful clusters of countries regarding their tax composition. A preliminary logistic regression analysis indicates that clusters associated with income inequalities and tax structures affect the income distribution for a given year. We then explore the relationship between the clusters that represent different tax structures and income inequality levels and analyze them by using IV regression. At the macro level, our findings emphasize that the tax implications of countries are

associated with their income inequality levels. The outcomes of the models in detail reveal that leaning excessively on consumption taxes is expected to increase the income inequality in the economy. The control variables related to government expenditures on essential considerations like health, education, agriculture, and public services explicitly have a reducing impact on the level of income inequality. Among the government spending type of control variables, only economic affairs expenditures have a significantly small effect but the reverse. Lastly, the higher level of household final consumption expenditure is expected to be associated with a higher income inequality level.

To some extent, the tax compositions of countries follow the policy choices of governments. Therefore, to reduce income inequality, policymakers should find a way to increase the share of tax gained from income in emerging countries, which requires stronger corporate and individual income structures. This, consequently, addresses the need for structural reforms in the economies which impose higher level indirect taxes when compared to the direct taxes.

REFERENCES

Acosta-Ormaechea, S., Yoo, J. (2012). Tax Composition and Growth: A Broad Cross-Country Perspective. *IMF-Working-Papers*. 12/257. https://doi.org/10.5089/9781616355678.001.

Adam, A., Kammas, P., Lapatinas, A. (2015). Income inequality and the tax structure: Evidence from developed and developing countries. *Journal of Comparative Economics*, 43(1): 138-154. https://doi.org/10.1016/j.jce.2014.05.006.

Adam, C. S., & Bevan, D. L. (2005). Fiscal deficits and growth in developing countries. *Journal of Public Economics*. 89(4): 571–597. https://doi.org/10.1016/j.jpubeco.2004.02.006.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R., Hémous, D. (2018). Innovation and top income inequality. The *Review of Economic Studies*, 86(1):1–45. https://doi.org/10.1093/restud/rdy027.

Agnello, L., Sousa, R.M. (2014). How does fiscal consolidation impact on income inequality? *Review of Income and Wealth*, 60(4): 702–726. https://doi.org/10.1111/roiw.12004.

Akgun, O., Cournede, B., Fournier, J.M. (2017). The Effects of the Tax Mix on Inequality and Growth. *OECD Economics Dep.-Working-Papers*, No.1447, OECD Publishing, Paris. https://doi.org/10.1787/18151973.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., Zucman, G. (2018). World Inequality Report 2018, *World Inequality Lab*. Retrieved from: http://wir2018.wid.world.

Atkinson, A.B. (2003). Income Inequality in OECD Countries: Data and Explanations. *CESifo Economic Studies*, 49(4): 479–513, https://doi.org/10.1093/cesifo/49.4.479.

Bahl, R. W., Bird, R. M. (2008). Tax Policy in Developing Countries: Looking Back and Forward. *National Tax Journal*, 61(2): 279-301. https://doi.org/10.17310/ntj.2008.2.06.

Ball, L., Furceri, D., Leigh, D., Loungani, P. (2013). The distributional effects of fiscal consolidation. *IMF-Working-Papers*, 13/151. https://doi.org/10.5089/9781475551945.001.

Besley, T. & Persson, T. (2014). Why Do Developing Countries Tax So Little? *Journal of Economic Perspectives*, 28 (4): 99-120, https://doi.org/10.1257/jep.28.4.99.

Borge, L.E., Rattso, J.J. (2004). Income distribution and tax structure: Empirical test of the Meltzer-Richard hypothesis. *European Economic Review*, 48(4): 805-826.

Chu, T.T., Hölscher, J. & McCarthy, D. (2020). The impact of productive and non-productive government expenditure on economic growth: an empirical analysis in high-income versus low- to middle-income economies. *Empirical Economics*, 58: 2403–2430. doi.org/10.1007/s00181-018-1616-3.

Ciminelli, G., Ernst, E., Merola, R., Giuliodori, M. (2019). The composition effects of tax-based consolidation on income inequality. *European Journal of Political Economy*, 27: 107-124. https://doi.org/10.1016/j.ejpoleco.2018.08.009.

Cingano, F. (2014). Trends in Income Inequality and its Impact on Economic Growth. *OECD Social, Employment and Migration Working Papers*, No. 163, OECD Publishing. http://dx.doi.org/10.1787/5jxrjncwxv6j-en

Clements, B., Mooij, R., Gupta, S. and Keen, M. (2015), Inequality and Fiscal Policy (Washington: International Monetary Fund). https://doi.org/10.5089/9781513567754.071.

Cubero, R., and Hollar, I.V. (2010). Equity and Fiscal Policy: The Income Distribution Effects of Taxation and Social Spending in Central America. *IMF Working Paper*. 10(112). https://doi.org/10.5089/9781455200719.001.

Devarajan, S., Swaroop, V., and Zou, H. (1996). The Composition of Public Expenditure and Economic Growth. *Journal of Monetary Economics*. 37(2–3): 313–344. https://doi.org/10.1016/S0304-3932(96)90039-2.

Devries, P., Guajardo, J., Leigh, D., Pescatori, A. (2011). A New Action-based Dataset of Fiscal Consolidation. *IMF Working Papers* 11(128). International Monetary Fund. https://doi.org/10.5089/9781455264407.001.

Drucker, L., Krill, Z., Geva, A. (2017). The Impact of Tax Composition on Income Inequality and Economic Growth. *Ministry of Finance Publications and Reviews*. Retrieved from: https://mof.gov.il/en/PublicationsAndReviews/EconomicData/ DocLib/Article_082017_Eng.pdf

Easterly, W. (1995). What do cross-country studies teach about government involvement, prosperity, and economic growth? Comments on: Joel Slemrod. *Brookings Papers on Economic Activity* 2: 373–431. https://doi.org/10.2307/2534615.

Fournier, J.M. and Johansson, A. (2016). The Effect of the Size and the Mix of Public Spending on Growth and Inequality. *OECD Working Paper*. 1344. Organization for Economic Cooperation and Development. https://doi.org/10.1787/18151973.

Glomm, G. and Ravikumar, B. (2003). Public education and income inequality. *European Journal of Political Economy*, 19(2), 289-300. https://doi.org/10.1016/S0176-2680(02)00178-7.

Gordon, R., & Li, W. (2009). Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of public Economics*, 93(7-8): 855-866. https://doi.org/10.1016/j.jpubeco.2009.04.001.

Gottschalk, P., & Smeeding, T. M. (2000). Empirical evidence on income inequality in industrialized countries. In A. B. Atkinson & F. Bourguignon (Eds.), *Handbook of Income Distribution*, 1: 261–308. Amsterdam: North-Holland.

Hailemariam, A., Sakutukwa, T. & Dzhumashev, R. (2021). Long-term determinants of income inequality: evidence from panel data over 1870–2016. *Empirical Economics*, 61: 1935–1958. https://doi.org/10.1007/s00181-020-01956-7.

Iacoviello, M. (2008). Household debt and income inequality, 1963–2003. *Journal of Money, Credit and Bank*, 40(5): 929–965. https://doi.org/10.1111/j.1538-4616.2008.00142.x.

ICTD/UNU-WIDER. (2019). Government Revenue Dataset. Retrieved from: https://www.wider.unu.edu / project/government-revenue-dataset.

Jerven, M. (2013). Comparability of GDP estimates in sub-Saharan Africa: The effect of revisions in sources and methods since structural adjustment. *Review of Income and Wealth*. 59: 16–36. http://dx.doi.org/10.1111/roiw.12006.

Jones, C.I., & Kim, J. (2018). A Schumpeterian model of top income inequality. *Journal of Political Economy*, 126(5):1785–1826.

Kanbur, R. (2000). *Income Distribution and Development*. Handbook of Income Distribution, edited by A.B. Atkinson and F. Bourguignon, Elsevier. ISBN: 9780444816313.

Martinez-Vazquez, J., Vulovic, V., Dodson, B.M. (2012). *The impact of tax and expenditure policies on income distribution*: Evidence from a large panel of countries. Hacienda Publica Espanola, 200(1): 95–130. Retrieved from: https://www.ief.es/docs/destacados/publicaciones/revistas/hpe/200_Art4.pdf

Meltzer, A., Richard, S. (1983). Tests of a rational theory of the size of government. *Public Choice*. 41: 403–418. https://doi.org/10.1007/BF00141072.

McNabb, K. (2017). Toward Closer Cohesion of International Tax Statistics. *WIDER-Working-Paper* 184/2017. http://dx.doi.org/10.35188/UNU-WIDER/2017/410-0.

Mo, P. H. (2000). Income Inequality and Economic Growth. *Kyklos*, 53(3), 293-315. https://doi.org/10.1111/1467-6435.00122

Muinelo-Gallo, L., Roca-Sagales, O. (2013). Joint determinants of fiscal policy, income inequality and economic growth. *Economic Modeling*, 30:814–824. https://doi.org/10.1016/j.econmod.2012.11.009

OECD (2022). About OECD. Retrieved from: https://www.oecd.org/about/

Pickering, A., Rajput, S. (2018). Inequality and the composition of taxes. *International Tax Public Finance*, 25: 1001-1028. https://doi.org/10.1007/s10797-017-9476-x.

Poterba, J. (2007). Income inequality and income taxation, *Journal of Policy Modeling*, 29: 623-633. https://doi.org/10.1016/j.jpolmod.2007.05.010.

Prichard, W., Cobham, A., Goodall, A. (2014). The ICTD Government Revenue Dataset. *ICTD-Working-Paper* 19. http://dx.doi.org/10.2139/ssrn.2496442.

Ravallion, M. (2014). Income inequality in the developing world. *Science*, 344(6186): 851–855. doi:10.1126/science.1251875.

Rodríguez-Pose, A. and Tselios V. (2009). Education and Income Inequality in the regions of the European Union. *Journal of Regional Science*, 49(3), 411–437. https://doi.org/10.1111/j.1467-9787.2008.00602.x.

Roser M., Cuaresma J. C. (2016), Why is Income Inequality Increasing in the Developed World? *Review of Income and Wealth*, 62: 1–27, https://doi.org/10.1111/roiw.12153.

Shin, I. (2012). Income inequality and economic growth. *Economic Modelling*, 29(5), 2049-2057. https://doi.org/10.1016/j.econmod.2012.02.011

Solt, F. (2020). Measuring Income Inequality across Countries and over Time: The Standardized World Income Inequality Database. *Social Science Quarterly*, 101(3): 1183–99. https://doi.org/10.1111/ssqu.12795.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, 20: 518–529. https://doi.org/10.1198/073500102288618658.

Stiglitz, J.E. (2010). "CHAPTER 1. Development- Oriented Tax Policy". *Taxation in Developing Countries: Six Case Studies and Policy Implications*, edited by Roger Gordon, New York Chichester, West Sussex: Columbia University Press, 11-36. https://doi.org/10.7312/gord14862-004.

Tanzi, V. (1989). The impact of macroeconomic policies on the level of taxation and fiscal balance in developing countries. *IMF Staff Papers*. 36: 633–656. https://doi.org/10.2307/3867050.

Voitchovsky, S. (2005) Does the Profile of Income Inequality Matter for Economic Growth?. *Journal of Economic Growth*, 10, 273–296. https://doi.org/10.1007/s10887-005-3535-3

Woo J., Bova, E., Kinda, T. and Zhang, Y.S. (2017). Distributional Consequences of Fiscal Adjustments: What Do the Data Say?. *IMF Economic Review*. 65(2): 273–307. https://doi.org/10.1057/s41308-016-0021-1.

World Bank. (2002). Achieving Education for All by 2015: Simulation Results for 47 Low-Income Countries. Washington, DC: Human Development Network, Africa Region and Education Department, World Bank: 76.

World Bank, World Development Indicators. (2020). https://databank.worldbank.org/ source /world-development-indicators/preview/on.

World Inequality Database (2018). https://wid.world/world/

Appendix A1. Table A1

Table A1. Results of Linear Regression Analysis (dependent variable is Gini index)

Coefficients	Estimate	Std. Error	t-value	Pr(>|t|)		Year
	20.79***	2.39	8.68	<0.0001	0.45	2013
(cluster effect)	7.62***	1.16	6.59	<0.0001		

*** for p<0.001

Source: Author's calculations

Appendix A2. Table A2

Table A2. Results of Logistic Regression Analysis (dependent is Clusters)

	Estimate	Std. Error	Wald	p-value
(Gini effect)	0.221***	0.051	4.30	<0.0001
Intercepts:				
1/2	6.16**	1.65	3.74	0.002
2/3	9.91***	2.12	4.69	<0.0001
Goodness-of-fit statistic				
LR chi-square				27.79***
(p-value)				<0.0001

** for p<0.01 and *** for p<0.001

Source: Author's calculations