Multiple Linear Regression Analysis of Macroeconomic variables behind Income Inequality in Hungary

R. Boringa

Studentnumber: S3391906

Campus Fryslân, Rijksuniversiteit Groningen

BSc. Global Responsibility & Leadership

Capstone Thesis

Majors: Responsible Governance & Responsible Planet

Supervisor: MSc. V.L. Kedari

Date: 10/06/2022

ABSTRACT:

Income inequality has been associated with detrimental health, social and economic consequences. Currently income inequality is rising globally. Hungary has experienced a relatively mild increase in income inequality compared to other states. By understanding the impact of macroeconomic variables on income inequality across comparable states this paper aims to improve the understanding their dynamics. This study has therefore constructed several multiple linear regressions, aiming to understand the significant associations with different measures of income inequality and their directionality and strength. The macroeconomic variables used in this study include: Government Expenditure, Economic growth, Volume of Imports, Volume of Exports, Total investment and Capital Openness. The macroeconomic variables were identified through the construction of a Directed Acyclic Graph, based upon an analysis of academic literature. The only variable found to be significantly related to income inequality in Hungary is capital openness, which for every 1 increase in capital openness saw an increase of 0.027 (± 0.003) in the Gini coefficient. Across the other states assessed government expenditure, imports and investment were found to be significantly related to income inequality. As the results per state varied, this paper recommends local, tailored solutions for income inequality.

Keywords: *Hungary, Income Inequality, Economy, Macroeconomic, Statistics, Analysis*

Wordcount: 9604

Index

A Linear Regression Analysis of Macroeconomic variables behind Income Inequality in Hungary after the Fall of Communism

Income inequality in Hungary has had a turbulent recent history. As a communist state the transition from a socialist redistributive economy to a free market economy must have given rise to enormous inequality within its society. As income inequality is negatively associated with population health, wellbeing and economic growth, this could have crippled the Hungarian state (Pickett & Wilkinson, 2015) (Aiyar & Ebeke, 2020). This is however not the case. Hungary has seen a relatively mild transition from communism, while industrial economies have seen income inequality rise substantially since the mid-1970s (Aiyar & Ebeke, 2020). By inferring clues from the data of its economy, policy makers might be able to improve inequality conditions in comparable states and prevent the negative health, societal and economical effects associated with income inequality (Detollenaere et al., 2018) (Kim, 2015).

After the second World War, the Soviet Union established communist regimes in the areas it occupied in eastern Europe. Communism is an ideology that aims to establish a system of government that commands a state planned redistributive economy. Communist governments often had far reaching powers in the economy, setting prices, salaries and determining production quotas. Eastern European communist states, including Hungary, had relatively low rates of income inequality compared to capitalistic states of similar development before 1989 (Bandelj & Mahutga, 2010). Income inequality occurs naturally when people earn different incomes. In states that experience high income inequality the richest population group has a larger share of total income when compared to the poorer population groups. In states that have relatively low income inequalities, the income share a population has is more proportionate to the size of the population. A variable that is often used to measure the amount of income inequality a population experiences is the Gini coefficient. After the fall of communism in Hungary democratization began to occur, but the economic transition had already begun (Bandelj & Mahutga, 2010).

Income inequality in Hungary

The centrally planned economic system in Hungary, although similar to other communist states did have some differences. After 1968, with the introduction of the ‘New Economic Mechanism’, companies were given more power in economic decision making.

However, the government continued to regulate prices, wages, investment and interest rates (ŽÍdek, 2014). While these measures could not erase inequalities, they did have a significant impact, as can be seen in the graph below.

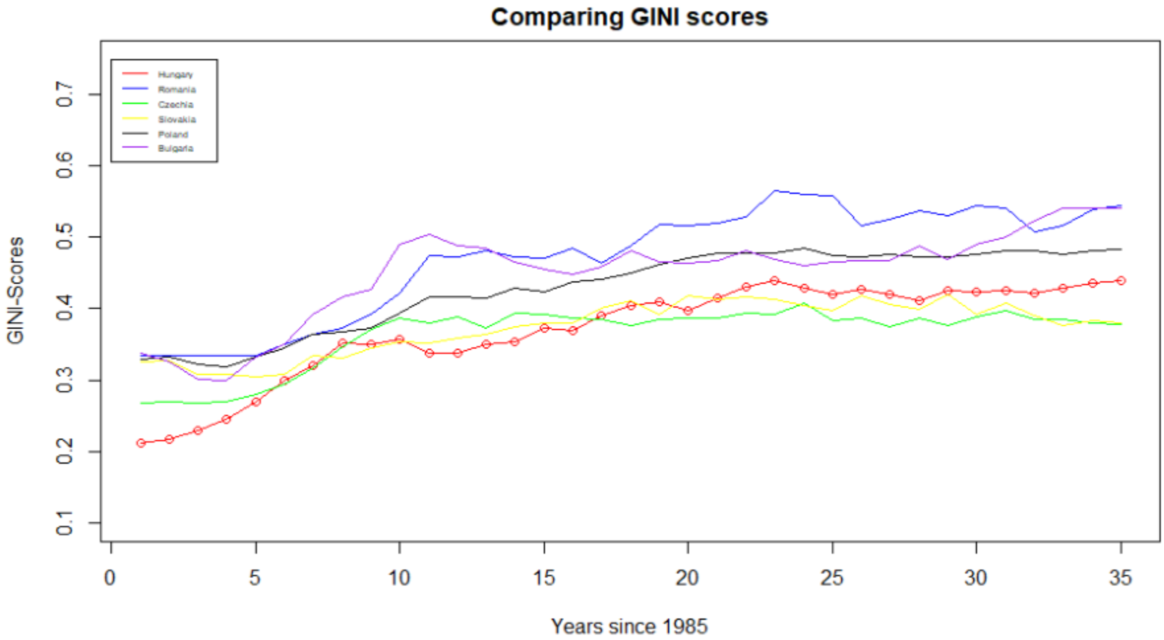

Figure 1: GINI Scores of various Eastern European States Post-Communism. Hungary in Red. Data retrieved from the IMF World Economic Outlook dataset.

The plot visualizes the trend in income inequality as measured by the Gini coefficient for several eastern European states. The increase in income inequality, that occurred after the fall of the communist governments, can be clearly seen in the first section of the graph. The rapid increase in income inequality can be attributed to the abandonment of government efforts to reduce inequality (Bandelj & Mahutga, 2010). The communist government of Hungary, under public pressure, abandoned some of the policies in place that kept income inequality low. Hungary became a member of the IMF in 1982 and created commercial banks in 1987. As of 1989, 63 percent of all prices were already governed by the market (ŽÍdek, 2014). In 1989 the communist regimes mentioned in this paper, transitioned to democratic free market economies. The measures taken by the communist government of Hungary resulted in a relatively high government expenditures as percentage of GDP, even when compared to other communist states, reaching 63.7 percent of GDP in 1989 (ŽÍdek, 2014).

During this period of transition foreign direct investment or FDI rose, as the economies of the region experienced globalization. Besides FDI, income inequality rose substantially (Mihaylova, 2015). As globalization allows for a more optimal use of resources it can accelerate economic growth, however not everyone benefits from this increase in economic growth equally, and as such income inequality tends to rise (Georgantopoulos & Tsamis, 2011). For the first seven years after the transition, income inequality in Hungary rose rapidly. Income inequality then stabilized until 2000, after which it continued to rise, however in contrast with the increase in income inequality before 2000, the rise was slower (Mihaylova, 2015). In 2019, the year at the end of the scope of this study, Hungary had a Gini coefficient of 0.44 (WID.WORLD, n.d.).

Consequences of income inequality

Income inequality has been negatively correlated with several measures of health, including depression. Some authors assert the claim that this correlation is causal, but this claim is disputed by others (Detollenaere et al., 2018). While the causal effect is relatively modest, reducing the GINI coefficient below 0.3 could avert 1.5 million deaths according to Pickett et al in 2014. Furthermore reducing income inequality should decrease the amount of physical and mental illness as well as violence, and therefore government expenditure on these subjects could be invested elsewhere. It should however be noted that income inequality is most likely to affect health indirectly, through changing the social standing of a person (Pickett & Wilkinson, 2015). A further exploratory study on the effects of income inequality on health in Europe has been conducted in 2018 and found that several parameters of health, self-rated health, life expectancy and mental wellbeing were all negatively correlated with an increase in income inequality (Detollenaere et al., 2018). In Hungary specifically an association between life expectancy and income inequality was found, accounting for a difference between the highest and lowest income groups of 4.6 years for females and 6.9 years for males. The majority of the difference in life expectancy between the highest and lowest income groups is attributed to avoidable causes of death (Bíró et al., 2021). Despite these correlations and while a direct relationship between income inequality and health is hypothesized, causation cannot be confirmed as there has not yet been a longitudinal study concerning the effects of income inequality on the health of a population (Detollenaere et al., 2018).

The consequences of an inequal society are not only expressed through health. In countries with high averages incomes, income inequality slows economic growth. Specifically GDP per capita is influenced negatively by higher levels of income inequality

(Brueckner & Lederman, 2018). A further study conducted involving a statistical analysis on income inequality in 40 countries in the OECD corroborated the result that income inequality had a significant negative relationship with GDP growth (Kim, 2015). Increasing income inequality can furthermore cause macroeconomic instability, the concentration of power and increase the risk of economic crises (IMF, 2015). Besides nationwide effects increased income inequality can increase inequality of opportunity and entrench socio-economic outcomes. The entrenchment of outcomes can in turn cause disillusionment and political instability (IMF, 2015). The consequences of a highly unequal distribution of income are thus numerous, including health, social, economic and political disadvantages.

The global trend

Income inequality has increased among most of the advanced and emerging market economies, with global income inequality across individuals reaching a Gini coefficient of 0.70 in 2013 (IMF, 2015). Furthermore a large increase of income inequality over the last 200 years on the world economy has been identified as one the most worrying features of economic development by van Zanden et al in 2014. The process of globalization, which accelerated over this period, is causing a strong increase in within-country income inequality since 1980, after a period of relative stagnation in the trajectory of income inequality (van Zanden et al., 2013). The current trend of global income inequality is uncertain, and dependent upon how economic growth will be distributed over society. Projections made by Alvaredo et al in 2018 predicted an increase in the top 1% income share and a slight decrease in the income share of the bottom 50% (Alvaredo et al., 2018). This would indicate a future rise in income inequality, and while it remains important to acknowledge the assumptions that are made with projections, an increase in income inequality would have harmful consequences for the societies in which it occurs.

Problem analysis

Before 1989, Hungary’s income inequality was relatively low and stable compared to its comparable capitalistic countries (Bandelj & Mahutga, 2010). After the fall of the communist government the economy was transitioned, with the government abandoning efforts to keep income inequality low (Bandelj & Mahutga, 2010). Thereafter income inequality within Hungary rose sharply, until its trajectory stagnated and slowly started to increase again by 2000. Hungary has since experienced a relatively mild increase in income inequality. Income inequality globally has risen fast and projections predict a further rise in income inequality by 2050. This increase in income inequality is undesirable for states, as increased levels of income inequality can lead to several negative economic and political consequences. Furthermore increased income inequality can lead to negative effects on health, with income inequality even associated with a decrease in life expectancy. By studying the recent economy of Hungary through its macroeconomic variables this paper aims to aid policy makers in identifying the areas of the economy that can be influenced to reduce income inequality. This study will thereafter aim to identify the variables which have had a disproportionate impact on income inequality within Hungary when compared to similar states, in order to see if this global problem could be viably treated by local solutions. In order to answer these questions this study has created two research questions.

Research question 1: *Can the identified macroeconomic variables explain the trajectory of income inequality in Hungary since 1995?*

Research question 2: *Which variables have had a disproportionate impact on income inequality in Hungary after the fall of communism compared to other eastern European states?*

Scope and method

In order to answer the research questions this study will conduct several multiple linear regressions involving macroeconomic variables, which will be chosen through a review of literature. A multiple linear regression is a statistical test that aims to determine the strength, significance and direction of the association between a dependent variable and multiple independent variables. The analysis will be conducted through macroeconomic variables, which are variables that describe aspects of the economy at large. This contrasts with microeconomics, which describes the economic decisions of an individual. This study will focus upon Hungary since 1996, and will include Bulgaria, Czechia, Slovakia, Romania and Poland for comparison. Several significant associations between macroeconomic variables and income inequality have been found by this paper, under which a positive relationship between capital openness and income inequality.

Literature Review

The literature review in this paper will be dedicated to analyse the relationships found between macroeconomic variables and measures of income inequality in relevant literature.

From the connections found throughout literature a directed acyclic graph or DAG will be created, which will visualize a web of relationships between variables. This DAG will allow this study to create a the formula that will be the basis of the linear regression model. From past research it is apparent that there is no clear consensus on the relationship between income inequality and other macroeconomic variables (Furceri & Ostry, 2019). Furthermore, if relationships between different variables are found they are often contested on the ground of statistical methods, origin of data or country specific factors that do not allow for externalization to other states (Deyshappriya, 2017). However, it is nonetheless important to analyse previous literature, as the associations found will be the basis for this study.

Unemployment and Inflation

Research done for the Asian Development Bank by Deyshappriya in 2017 focussed on the impact of macroeconomic variables on income inequality in the Asian region and found unemployment and inflation to be statistically significant, even when including political and demographic variables in the model (Deyshappriya, 2017). Unemployment was furthermore identified as a key driver of income inequality, as those who are most likely to lose their occupations are those in the bottom income shares (Alvaredo et al., 2018). These findings were corroborated by Furceri and Ostry in 2019, whom underscored unemployment and globalization as key drivers of national income inequality (Furceri & Ostry, 2019). However, when using income shares, Jäntti et al found little evidence of a relationship between unemployment and inflation and income inequality. The study conducted took place in the United Kingdom, and analysed five different income groups through a regression model (Jäntti & Jenkins, 2009).

GDP, Economic Growth and Investment

Deyshappriya in 2017 furthermore identifies GDP and trade flows as a significant predictor of income inequality. According to Deyshappriya et al, the association of GDP with income inequality follows a parabolic relationship, as described by the Kuznets curve (Deyshappriya, 2017). The Kuznets curve is a heavily debated topic in literature, and not an accepted hypothesis by every scholar. The Kuznets curve explains the proposed parabolic relationship between income inequality and economic growth by arguing that in poorer countries economic growth causes capital accumulation, which in turn increases income inequality. In richer countries however GDP growth decreases income inequality due to the state funding measures to distribute economic gains more evenly (Furceri & Ostry, 2019) (Deyshappriya, 2017). Not all studies find the proposed parabolic relationship mentioned above, but rather a positive relationship between GDP and income inequality, regardless of the development level of the state (Bandelj & Mahutga, 2010). Nevertheless it is clear that income inequality and GDP growth has an association which is worth including in the model, as not everyone can profit equally from the gains made through economic growth.

Bandelj and Mahutga in 2010 found, while researching the transition from socialist planned to capitalist free market economies, that the socialist states that allowed greater inflow of Foreign Direct Investment or FDI, had significantly higher rates of income inequality (Bandelj & Mahutga, 2010). However, Georgantopoulos and Tsamis in 2011, found that FDI reduces income inequality, as FDI boosts economic growth as well as increases government revenue, which could in turn be used on poverty alleviation programs (Georgantopoulos & Tsamis, 2011) (Aiyar & Ebeke, 2020). Foreign direct investment and investment share to GPD was furthermore found to have a positive correlation with several income inequality variables (Deyshappriya, 2017)(Jäntti & Jenkins, 2009). As foreign direct investment was severely limited during the communist government it is worth looking into the recent effects, especially since most studies support the hypothesis that investment positively influences income inequality.

Government Expenditure and Education

Government Expenditure has a complex and debated relation with income inequality, as government expenditure on infrastructure reduces income inequality in Latin American countries, which have significant levels of income inequality. However, this denotes a specific area of government expenditure, and it occurred in a state that already has a high level of income inequality (Deyshappriya, 2017). Furthermore, government expenditure has been found to increase income inequality in European countries by Roventini et al in 2012, while Sarel et al found no significant relationship at all in 1997 (Maestri & Roventini, 2012) (Sarel, 1997). A third proposed relation with income inequality and government expenditure stipulates that in the short term government expenditure reduces income inequality, while increasing it in the long term (Deyshappriya, 2017). As government expenditure can influence investment, and in turn economic growth, while simultaneously influencing income inequality it is an important variable to include in the statistical analysis of this study.

Government expenditure is furthermore very broad, and the relationships found by this study could unearth the combined effect of all government measures through capital.

Education is a variable that does not fit the macroeconomic scope this study is based upon. Nevertheless the effect of education on income inequality is large and bi-directional (Anyanwu et al., 2016) (IMF, 2015). A person with an education has the possibility to apply for occupations that have increased pay, but education is not accessible for everyone. Secondary education has been associated with a decrease in income inequality, as the majority of population can access a secondary degree. Tertiary education is less accessible, and thus increases income inequality (Anyanwu et al., 2016).

Capital openness and Trade

The negative relationship between the total sum of trade and income inequality is well documented and accepted by scholars (Georgantopoulos & Tsamis, 2011). The hypothesis postulates that trade should mean the more efficient use of goods and services, and therefore those should be more affordable than domestic goods and services. Besides, it should lead to increased globalization, which should increase domestic production efficiency in order to compete. As such the bottom income shares benefit relatively more than the top income share groups (Georgantopoulos & Tsamis, 2011). A study conducted by Furceri and Ostry in 2019 found that trade and financial globalization had asymmetric effects, whereby trade lowered income inequality, while increased financial globalization was associated with higher income inequality (Furceri & Ostry, 2019).

Directed Acyclic Graph

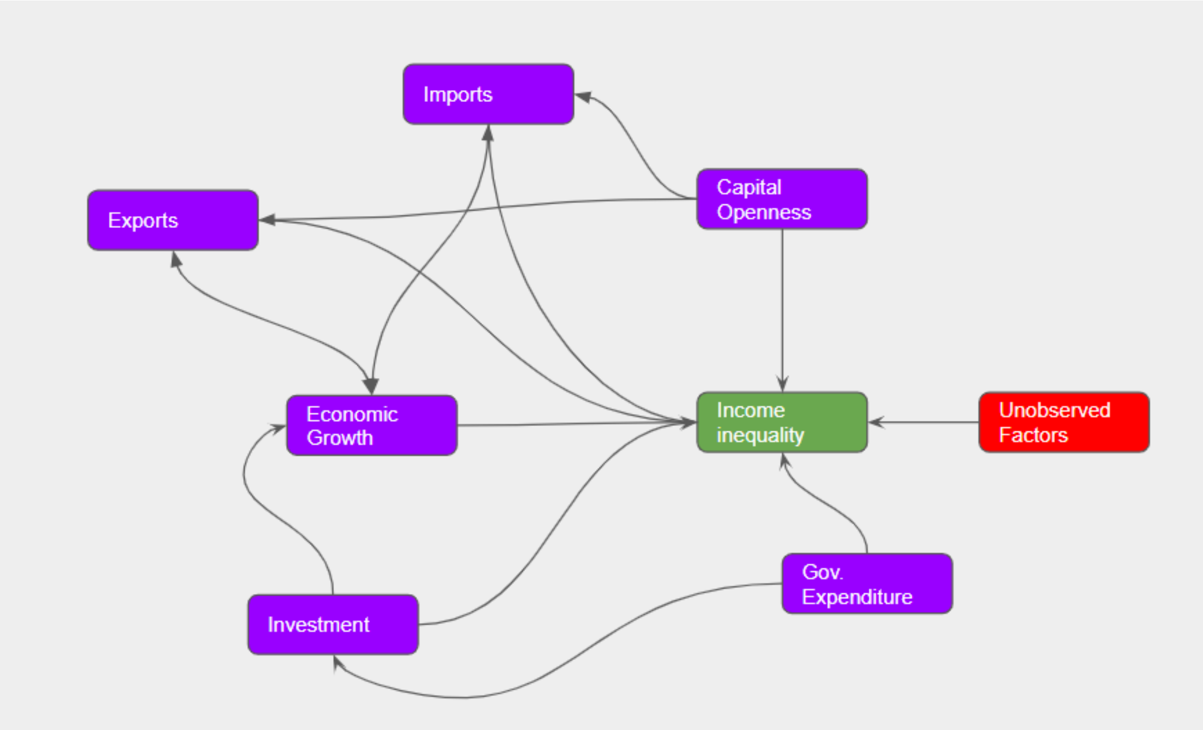

Figure 2: The Directed Acyclic Graph based upon the literature review that will guide the statistical analysis

Based upon the findings of this literature review a directed acyclic graph was produced. This graph helps visualize the relationships between several macroeconomic variables and the dependent variable of interest, in this case income inequality. The macroeconomic variables that have been chosen to be analysed in the linear regression are: government expenditure, economic growth, imports, exports, investment as percentage of GDP and capital openness. These variables have all been supported by literature and have an appropriate amount of observations. The macroeconomic variables that were not included despite being associated with income inequality are: inflation and unemployment. These will not be included in the analysis because the literature is contradicting, and studies which have the most similarities to the background of this study have not found a statistically significant association.

Methodology

The methodology section for this study is subdivided into several sections, which will each describe an essential part of the set-up and structure of this study. The aim of this section is to justify the process the author undertook, while allowing for reproduction of the results found.

Methodological Approach & Scope

Scope. This research has the primary aim to explain the changes in income inequality in Hungary through other relevant macroeconomic variables found in literature. Hungary was chosen for this study as the trajectory of its income inequality through time is worth studying. Hungary had a relatively mild transition from communism to capitalism compared to other states and has shown a relatively stable level of income inequality after the transition had occurred (Bandelj & Mahutga, 2010). The identification of the relevant macroeconomic variables that determine income inequality in Hungary could be the bedrock that future policy to reduce inequalities in other societies can be based on.

The other states included for the comparison with Hungary are Poland, Czechia, Slovakia, Romania and Bulgaria. These states have been identified on the grounds that they share a relatively comparable recent history. The aforementioned states are all eastern European states that experienced communism until 1989-1990, without being a part of the Soviet Union, and are present day members of the European Union. A requirement upon entering the European Union is to have a functioning market economy, and thus all states that are included in this study successfully transitioned from a planned economy to a market economy (Bandelj & Mahutga, 2010). The states thus have a comparable recent economic history, allowing for comparison. The timeframe chosen for the statistical analysis is in part based upon the availability of data, and part on historical developments. The timeframe chosen is from the economic transition until the COVID-19 pandemic. The COVID-19 pandemic was excluded from the timeframe as the measures taken to combat the pandemic were not uniform, and the impacts on the states would skew the data, making the results unreliable. Not all aforementioned states have recorded economic data on all of the aforementioned years, especially in the period following revolutions, and therefore the years used over the different states can differ per state. Table 1 (below) illustrates the timeframe used in the statistical analysis per state that is involved.

State	Hungary	Bulgaria	Romania	Czechia	Slovakia	Poland
Timeframe	1996-2019	2001-2019	2001-2019	1997-2019	1998-2019	1997-2019

Table 1: Shows the Timeframe used in the statistical Analysis per state involved.

As this paper has aims to explain income inequality through other macroeconomic variables, it will therefore solely include macroeconomic variables in its statistical analysis.

Justification of Methodological Approach. Explaining the impact of macroeconomic factors on income inequality in several states over several years requires both a literature review and a statistical analysis. As such the research shall be quantitative, containing no interviews. Qualitative research would have yielded subjective datapoints, while this study focusses on the relationship between variables nationwide, which need several years of observations to accurately identify relationships. Therefore a qualitative research approach would not adequately answer the research question. The literature review is necessary in order to identify the variables that are related to income inequality as well as explain the differences that occur both between the impact of the macroeconomic variables as well as the difference between states. The methodological approach is thus aimed at answering the research questions through a combination of statistical testing and literature review, concluding with a discussion on the results of the statistical analysis supported by relevant literature. This approach best suits the research question as it allows the study to undertake a precise analysis on a large number of datapoints, which have been retrieved from the same origin.

Included Variables. The variables that are used in this paper are visible in the table below.

Variable	**Abbreviation**	**Source**
Gini-Coefficient	GINI	World Inequality Database
Bottom 50% income share	B50	World Inequality Database
Top 1% income share	T01	World Inequality Database
Government Expenditure	GGX	IMF World Economic Outlook (WEO) Dataset
Economic Growth	NGDP	IMF WEO
Imports	TMG_RPCH	IMF WEO
Exports	TX_RPCH	IMF WEO
Investment	NID_NGDP	IMF WEO
Capital Openness	KAOPEN	Chinn-Ito Index

Table 2: Variables, Abbreviation of Variable in statistical testing, Origin of Variable

The variables indicating income inequality were the Gini coefficient, bottom 50 percent income share and top 1 percent income share. The variables were all computed with pre-tax data, otherwise the redistributive effects of the tax system would be included in the statistical analysis, and that is beyond the scope of this paper. The Gini coefficient was used for nationwide analysis, while the income share variables were used to determine disproportionate impact of a macroeconomic variable within a state upon a specific layer of society. As such, separate multiple linear regressions were formulated for all states, over all three variables.

The macroeconomic variables that have been included in the analysis have been identified as having a relationship with income inequality in previous relevant research. The justification is seen in the '*Literature Review*' section of this study.

Research Design

Use of R. The statistical software 'R' was used in the statistical analysis. Besides personal familiarness with the software, the customizability of R was a definitive reason to choose for the software. R allows for customization through the import of libraries, which allow different testing, complementary to the basic statistical tests that R provides. The code can furthermore be easily shared and reproduced, increasing consistency. The intend is to increase the ease of reproduction.

Method of literature review. The literature review was conducted using 'Google Scholar'. As this research required various topics of literature to be read the search terms used varied. Search terms included, but were not limited to 'Income inequality AND trends AND Europe', 'Income inequality AND macroeconomics', 'Income inequality AND regression', 'Income inequality AND Hungary' and 'Income inequality AND consequences'. The papers were then identified through tests of relevance, were the amount of citations, the year the study was published, the organisation involved if relevant, and the information provided in the abstract, introduction and conclusion of each paper. The selection that was then made was read and analysed for the literature review and discussion of this paper.

Data Collection

Origin of the datasets. The variables on income inequality are retrieved from the 'World Inequality Database' or WID. The WID aims to provide precise data on inequality in order to allow comparisons between states and over time periods. The WID distinguishes the data it provides from others due to a different origin of the variables on income inequality. Besides the household surveys that are traditionally used in the computation of inequality variables the WID uses national accounts, survey data, fiscal data and wealth rankings. The

WID furthermore makes explicit when the data has limitations, and where those limitations have their origin (WID.WORLD, n.d.). The WID was chosen for this study due to its scientific origin and precise computation of inequality variables across different states and time periods.

The dataset used for retrieval of data for the macroeconomic variables besides capital openness is the IMF World Economic Outlook dataset by the International Monetary Fund (IMF, n.d.-b). This dataset is a compiled dataset of the bi-annually released IMF World Economic Outlook. The dataset has been compiled and certified by Datahub.io. The dataset includes all states, and contains a large number of macroeconomic variables. Datahub.io is run by Datopian, which is an organisation that aims to publish certified data openly. The organisation has several partners, including The World Bank, the US Government and the OECD (Datopia, n.d.). When a dataset has been certified Datopian assures its quality and has audited its sources. The data in the 'World Economic Outlook' has been gathered through surveys conducted by IMF staff, the results of which have been analysed by economists to distinguish global economic developments. The IMF is a UN organization that seeks to encourage financial stability and economic growth, among others (IMF, n.d.-a). The data that is gathered is used by governments, NGO's and the UN to compile country specific forecasts and advice.

Capital openness, as measured through the variable KAOPEN, was retrieved from the Chinn-Ito index dataset. The dataset was created by Chinn and Ito, and contains the data on capital openness from 1970 until 2019 for 182 countries. This variable specifically was included because of its wide coverage across states and time, as well as its transparency regarding the origin of the data. The data used to compute the variable is based on the IMF Annual Report on Exchange Arrangements and Exchange Restrictions (Chinn & Ito, 2006).

Computation, validity and reliability of data. The Gini coefficient is used to measure the amount of income inequality in a society. It is calculated through the ratio of the area between the line of equality and the Lorenz curve, which plots the proportion of total income to population. It is thus a cumulative measurement of the inequality in a state compared to a state with perfect equality . The Gini coefficient does not measure absolute poverty, states with a high Gini coefficient can still be prosperous. The variables that represent income share have been constructed by the World Inequality Database, by the process laid out above in '*Origin of the datasets*'. The variable ranks the population based upon income, and then divides the population in to the relevant groups. The share of the size of the resulting group compared to total population size is the variable used (WID.WORLD, n.d.). Government expenditure (GGX), is computed as general government total expenditure in billions national currency. Economic growth (NGDP_RPCH) has been computed as annual percent change of gross domestic product in constant prices. Imports (TM_RPCH) and exports (TX_RPCH) have been computed as percent change of volume of imports and exports. The imports and exports were retrieved from the Export and Import Price Index Manual. Total investment (NID_NGDP) as a percentage of GDP was expressed as the ratio of total investment and GDP in current local currency. It was furthermore measured through the total value of gross fixed capital formation and changes in inventories and acquisitions of sectors. The abovementioned variables are all in use by the International Monetary Fund (Datopia, n.d.). Capital openness is a separate index and measures a country's degree of capital account openness based upon data by the IMF on a country's tabulation of restrictions and cross-border financial transactions (Chinn & Ito, 2006).

Data Analysis

Justification of Multiple Linear Regression. This study makes use of multiple linear regression models in R. Multiple linear regression models were chosen as this paper aims to construct a macroeconomic model wherein the impact, directionality and strength of association are measured. Multiple linear regression allows for all of these, but comes with several assumptions. These assumptions in a general linear regression model are linearity, homoscedasticity, normality, independence and endogeneity. To test whether the assumptions made in the models were justified several measures were undertaken. For homoscedasticity the visualization of the variance of datapoints was used through a scale-location plot. Likewise for normality, which was visualized using a normal q-q plot. Linearity was tested twofold, both through a residuals vs fitted values plot and a Ramsay reset test. The Ramsay reset test was furthermore used to assess whether the variables were endogenous and the model misspecified. The independence of the independent variables was avoided as much as possible through analysis of the computation of each variable. It was nevertheless inevitable that some correlation between variables has occurred during linear regression, as macroeconomic variables are often interrelated (Furceri & Ostry, 2019). As our models do not violate their assumptions the multiple linear regression model was a valid choice to determine the nature of the relationships between the chosen macroeconomic variables and income inequality.

Method of Data Analysis. The code used in this paper can be found in appendix A. Hereafter is described the process through which the statistical analysis was conducted. The table below showcases the steps taken, and the order they were taken in.

Step:	**Action done:**
1	Import of the WID, IMF WEO and Chinn-Ito datasets
2	Loading of R libraries: Tidyverse, ggplot2, dplyr, tidyr, broom, ggfortify, readxl, jsonlite and stats
3	Exploratory statistics: creation of two plots visualizing the trend in Gini-coefficient and bottom 50 percent income share across all states
4	Extracting Hungarian income inequality variables; Gini-coefficient, bottom 50 percent and top 1 percent income share by year as data frames.
5	Extracting Hungarian macroeconomic variables; The data was extracted as.numeric. The data was furthermore extracted for t -1. The data was created as a numeric list of values in R.
6	Statistical testing: 3 Multiple Linear Regressions were conducted for Hungary. With the Gini coefficient, bottom 50 percent income share and top 1 percent income share as the dependent variables.
7	Through the autoplot() function the models were visualized in residuals vs fitted, normal q-q and scale-location plots in order to detect non-linearity, non-normality and heteroscedasticity.
8	Extraction of Romanian, Bulgarian, Polish, Czech and Slovak macroeconomic and income inequality variables, through the same method as Hungary.
9	Statistical testing: 3 Multiple Linear Regressions were conducted for each of Romania, Bulgaria, Poland, Czechia and Slovakia. The dependent variables were: Gini Coefficient, Bottom 50 percent income share and Top 1 percent income share.
10	Through the autoplot() function all regression models developed were visualized in order to detect non-linearity, non-normality and heteroscedasticity.
11	Ramsay Reset Tests were conducted on all regression models developed, to rule

	out non-linearity and misspecification.

Table 3: Order of operations conducted during the statistical analysis

For the statistical analysis Rmarkdown was used. After importing the WID, IMF WEO and Chinn-Ito datasets, either manually or through code, the following libraries were loaded: Tidyverse, ggplot2, dplyr, tidyr, broom, ggfortify, readxl, jsonlite and stats. Consequently exploratory statistics was conducted, whereby plots of the Gini-coefficient and bottom 50% income share were created for all states. As Hungary is the main focus of this study, the variables for Hungary were created first. This was done by extracting the data from the respective datasets with the appropriate number of observations, in order for the amount of observations per variable to be consistent per state. The macroeconomic variables were created as a list of values in 'R', while the income inequality variables, Gini-coefficient, top 1 and bottom 50 percent income share, were created as sperate data frames per variable. The individual data frames that contained the separated income inequality data also included the associated years as basis for an index. The inequality data was furthermore collected as a separate data frame as it will serve as the dependent variable in the multiple linear regression. The macroeconomic variables were retrieved for t – 1 compared to the inequality variables. This makes sure that the statistical model tests the explanatory power of the macroeconomic variables for the income inequality in the up following year. This removes the possibility of co-directionality. After the data had been extracted three separate multiple linear regressions were ran, with the macroeconomic data as independent variables. The following formulas were used to construct the models:

1) The multiple linear regression model for Gini Coefficient:

$$GINI_t = \beta_0 + \beta_1 GGX_t + \beta_2 NGDP\_RPCH_t + \beta_3 TM\_RPCH_t + \beta_4 TX\_RPCH_t + \beta_5 NID\_NGDP_t + \beta_6 KAOPEN_t + \varepsilon_t$$

2) The multiple linear regression model for bottom 50% income share:

$$B50_t = \beta_0 + \beta_1 GGX_t + \beta_2 NGDP\_RPCH_t + \beta_3 TM\_RPCH_t + \beta_4 TX\_RPCH_t + \beta_5 NID\_NGDP_t + \beta_6 KAOPEN_t + \varepsilon_t$$

3) The multiple linear regression model for top 1% income share:

$$T01_t = \beta_0 + \beta_1 GGX_t + \beta_2 NGDP\_RPCH_t + \beta_3 TM\_RPCH_t + \beta_4 TX\_RPCH_t + \beta_5 NID\_NGDP_t + \beta_6 KAOPEN_t + \varepsilon_t$$

The function autoplot() was then used to establish whether the conditions for linear regression were met. As the conditions were met the statistical analysis for the other states was then conducted. It followed the exact order of operations as 3Hungary's statistical analysis did. This was done in order for all model regressions to be computed in the exact same manner. The function autoplot() was thereafter used to establish whether the conditions for linear regression were met for all 15 new multiple linear regressions. The final step undertaken in this analysis were Ramsay Reset Tests, or RRT, which determine whether the models are misspecified and the variables are endogenous. In this paper the fitted values of the models tested in the Ramsay reset test were multiplied to the power of 2 and 3. The model formula used in the Ramsay reset test is as follows, wherein the $\hat{Y}$ represents the prediction gathered from the linear regression model, while the $\hat{u}$ represents the residuals of the same original regression:

$$\hat{u} = a_0 + a_1\hat{Y} + a_2\hat{Y}^2 + a_3\hat{Y}^3 + \in$$

If a_1, a_2 and a_3 are not significantly associated with $\hat{u}$, than the model is correctly specified. If the test does find significant association then the results of the model cannot be used for interpretation, as the model is either non-linear or misspecified.

Results

To maintain clarity across this research paper the results section has been subdivided between the several states tested. All results of the regression models that are not visible in this section can be found in appendix B. The results of the Ramsay reset tests can be found in Appendix C.

Hungary

Linear regression. This paper found, through a multiple linear regression model that included income inequality data from 1996 until, and including, 2019 a significant relationship between the Gini coefficient and capital openness ($p < 0.001$). Specifically a 0.027 (± 0.003) increase was found for every 1 increase in capital openness. The other macroeconomic variables did not have a significant relationship with the Gini coefficient in Hungary.

```
Call:
lm(formula = HUNTime_and_Y$HUN_GINI ~ HUN_GGX + HUN_NGDP + HUN_TMGRPCH +
    HUN_TXRPCH + HUN_NIDNGDP + HUN_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0169951 -0.0043361 -0.0001175  0.0051210
       Max
 0.0206267

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.3242365  0.0876916   3.697
HUN_GGX     -0.0002165  0.0017498  -0.124
HUN_NGDP    -0.0001102  0.0014722  -0.075
HUN_TMGRPCH -0.0005805  0.0006884  -0.843
HUN_TXRPCH   0.0004564  0.0005155   0.885
HUN_NIDNGDP  0.0021634  0.0012762   1.695
HUN_KAOPEN   0.0273812  0.0033182   8.252
            Pr(>|t|)
(Intercept)  0.00179 **
HUN_GGX      0.90298
HUN_NGDP     0.94121
HUN_TMGRPCH  0.41081
HUN_TXRPCH   0.38832
HUN_NIDNGDP  0.10826
HUN_KAOPEN  2.39e-07 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01002 on 17 degrees of freedom
Multiple R-squared:  0.9164,    Adjusted R-squared:  0.8869
F-statistic: 31.05 on 6 and 17 DF,  p-value: 2.95e-08
```

Figure 3: Results of the regression model on Hungary's Gini Coefficient.

The linear regression model on bottom 50% income share presented similar results. A significant relationship between bottom 50% income share and capital openness ($p < 0.001$) was found. A 0.016 (± 0.002) decrease in the bottom 50% income share was found for every 1 increase in capital openness. The other macroeconomic variables did not have a significant relationship with bottom 50% income share in Hungary.

```
Call:
lm(formula = HUNTime_and_B$HUN_B50 ~ HUN_GGX + HUN_NGDP + HUN_TMGRPCH +
    HUN_TXRPCH + HUN_NIDNGDP + HUN_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0142799 -0.0025270  0.0000528  0.0036764
       Max
 0.0121640

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.2478790  0.0637602   3.888
HUN_GGX      0.0006892  0.0012722   0.542
HUN_NGDP    -0.0000185  0.0010704  -0.017
HUN_TMGRPCH  0.0001411  0.0005005   0.282
HUN_TXRPCH  -0.0001080  0.0003748  -0.288
HUN_NIDNGDP -0.0005634  0.0009279  -0.607
HUN_KAOPEN  -0.0162800  0.0024126  -6.748
            Pr(>|t|)
(Intercept)  0.00118 **
HUN_GGX      0.59505
HUN_NGDP     0.98641
HUN_TMGRPCH  0.78140
HUN_TXRPCH   0.77672
HUN_NIDNGDP  0.55172
HUN_KAOPEN  3.41e-06 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.007289 on 17 degrees of freedom
Multiple R-squared:  0.8864,    Adjusted R-squared:  0.8463
F-statistic: 22.11 on 6 and 17 DF,  p-value: 3.767e-07
```

Figure 4: Results Hungarian regression model on Bottom 50% income share

The linear regression model on top 1% income share showed a significant relationship between top 1% income share and capital openness (p < 0.001). An 0.011 (± 0.002) increase in the top 1% income share was found for every 1 increase in capital openness. The other macroeconomic variables did not have a significant relationship with top 1% income share in Hungary.

```
Call:
lm(formula = HUNTime_and_T$HUN_T01 ~ HUN_GGX + HUN_NGDP + HUN_TMGRPCH +
    HUN_TXRPCH + HUN_NIDNGDP + HUN_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0113918 -0.0026331  0.0001781  0.0029273
       Max
 0.0099816

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.1173686  0.0472564   2.484
HUN_GGX     -0.0010259  0.0009429  -1.088
HUN_NGDP     0.0010965  0.0007934   1.382
HUN_TMGRPCH -0.0006889  0.0003710  -1.857
HUN_TXRPCH   0.0003230  0.0002778   1.163
HUN_NIDNGDP  0.0009473  0.0006877   1.377
HUN_KAOPEN   0.0116653  0.0017881   6.524
            Pr(>|t|)
(Intercept)   0.0237 *
HUN_GGX       0.2918
HUN_NGDP      0.1848
HUN_TMGRPCH   0.0807 .
HUN_TXRPCH    0.2610
HUN_NIDNGDP   0.1862
HUN_KAOPEN   5.2e-06 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005402 on 17 degrees of freedom
Multiple R-squared:  0.8739,	Adjusted R-squared:  0.8293
F-statistic: 19.63 on 6 and 17 DF,  p-value: 8.937e-07
```

Figure 5: Results of Hungarian regression model on Top 1% income share

Assumptions addressed. The residuals vs fitted values plot based upon the Gini regression model yielded support for linearity, as the values barely deviate from the X-axis. Furthermore normality was established through the Normal Q-Q plot. The scale-location plot provided support of homoscedasticity through equally spread of datapoints. The Ramsay reset test yielded no significant associations, and therefore the regression model on Hungary's Gini coefficient was not found to be misspecified.

Support for linearity, normality and non-misspecification for the models concerning top 1% and bottom 50% incomes share was established through a residuals vs fitted values plot, normal Q-Q plot and Ramsay reset test. Through a scale-location plot the models were found to be homoscedastic.

Bulgaria

Linear regression. This study found, through multiple linear regression that included income inequality data from 2001 until 2019, a significant relationship between government expenditure and the Gini coefficient ($p < 0.05$). For every 1 increase in government expenditure a 0.010 (± 0.004) increase in the Gini coefficient. The other macroeconomic variables did not have a significant relationship with the Gini coefficient in Hungary.

```
Call:
lm(formula = BGRTime_and_Y$BGR_GINI ~ BGR_GGX + BGR_NGDP + BGR_TMGRPCH +
    BGR_TXRPCH + BGR_NIDNGDP + BGR_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.026263 -0.011029 -0.004409  0.011000
      Max
 0.040718

Coefficients:
             Estimate Std. Error t value
(Intercept)  0.128326   0.178857   0.717
BGR_GGX      0.010419   0.004378   2.380
BGR_NGDP    -0.001401   0.004167  -0.336
BGR_TMGRPCH  0.001756   0.001211   1.450
BGR_TXRPCH  -0.001546   0.001270  -1.217
BGR_NIDNGDP -0.000720   0.001755  -0.410
BGR_KAOPEN   0.008475   0.005035   1.683
            Pr(>|t|)
(Intercept)   0.4868
BGR_GGX       0.0348 *
BGR_NGDP      0.7426
BGR_TMGRPCH   0.1727
BGR_TXRPCH    0.2471
BGR_NIDNGDP   0.6889
BGR_KAOPEN    0.1181
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02269 on 12 degrees of freedom
Multiple R-squared:  0.5919,    Adjusted R-squared:  0.3878
F-statistic: 2.901 on 6 and 12 DF,  p-value: 0.05502
```

Figure 6: Results of Bulgarian regression model on Gini Coefficient

The linear regression model concerning the bottom 50% income share and the top 1% income share in Bulgaria did not yield any significant relationships.

Assumptions addressed. The Bulgarian Gini coefficient model was deemed to be linear, homoscedastic and normally divided through the use of the autoplot() functions, which provided the residuals vs fitted, normal q-q and scale-location plot used to address the assumptions. Furthermore the Ramsay reset test did not show any significant associations and as such the model is correctly specified.

The top 1% income share and bottom 50% income share were found to be linear, divided according to a normal distribution and, through the Ramsay reset test, correctly specified. The regression model on bottom 50% income share and top 1% income share were found to be homoscedastic through a scale-location plot.

Czechia

Linear regression. The linear regression on the Gini coefficient of Czechia did not yield any significant relationships between the Gini coefficient and the macroeconomic variables. The linear regression on the bottom 50% income share did however yield two significant relationships. A significant relationship was found between government expenditure and bottom 50% income share ($p < 0.01$). For every 1 increase in government expenditure an increase of 0.002 (± 0.001) was found in the bottom 50% income share. The linear regression involving the top 1% income share yielded no significant relationships.

```
Call:
lm(formula = CZETime_and_B$HUN_B50 ~ CZE_GGX + CZE_NGDP + CZE_TMGRPCH +
    CZE_TXRPCH + CZE_NIDNGDP + CZE_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.008559 -0.003472  0.001530  0.003536
      Max
 0.006573

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.1463208  0.0274803   5.325
CZE_GGX      0.0018853  0.0005826   3.236
CZE_NGDP    -0.0011777  0.0008147  -1.446
CZE_TMGRPCH  0.0007039  0.0005233   1.345
CZE_TXRPCH  -0.0002587  0.0004025  -0.643
CZE_NIDNGDP  0.0012960  0.0006809   1.903
CZE_KAOPEN  -0.0121261  0.0016854  -7.195
            Pr(>|t|)
(Intercept) 6.84e-05 ***
CZE_GGX      0.00517 **
CZE_NGDP     0.16759
CZE_TMGRPCH  0.19731
CZE_TXRPCH   0.52950
CZE_NIDNGDP  0.07515 .
CZE_KAOPEN  2.13e-06 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005111 on 16 degrees of freedom
Multiple R-squared:  0.9308,    Adjusted R-squared:  0.9049
F-statistic: 35.87 on 6 and 16 DF,  p-value: 2.083e-08
```

Figure 7: Results of regression model on Czechia's bottom 50% income share

Assumptions addressed. The three regression models of Czechia were, through residual vs fitted, normal q-q and scale-location plots deemed to be linear, homoscedastic and normally divided. Ramsay reset tests were conducted to rule out misspecification, which granted no significant relationships. Therefore the models are correctly specified.

Poland

Linear regression. The linear regression with the Gini coefficient in Poland yielded two significant relationships. The volume of imports ($p < 0.05$) and capital openness ($p < 0.001$) were found to have a significant relationship with the Gini coefficient. For every 1 increase in volume of imports the Gini coefficient decreased by 0.002 (± 0.001) and for every 1 increase in capital openness the Gini coefficient increased by 0.021 (± 0.004). The other macroeconomic variables were not found to have a significant relationship with the Gini coefficient in Poland.

```
Call:
lm(formula = POLTime_and_Y$POL_GINI ~ POL_GGX + POL_NGDP + POL_TMGRPCH +
    POL_TXRPCH + POL_NIDNGDP + POL_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.013574 -0.007384 -0.003243  0.005647
      Max
 0.016750

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.5225205  0.1004901   5.200
POL_GGX     -0.0006914  0.0017313  -0.399
POL_NGDP     0.0054029  0.0027775   1.945
POL_TMGRPCH -0.0017664  0.0006997  -2.525
POL_TXRPCH   0.0012569  0.0007167   1.754
POL_NIDNGDP -0.0018725  0.0018399  -1.018
POL_KAOPEN   0.0208999  0.0041843   4.995
            Pr(>|t|)
(Intercept) 8.77e-05 ***
POL_GGX     0.694936
POL_NGDP    0.069538 .
POL_TMGRPCH 0.022531 *
POL_TXRPCH  0.098608 .
POL_NIDNGDP 0.323963
POL_KAOPEN  0.000132 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01069 on 16 degrees of freedom
Multiple R-squared:  0.8158,    Adjusted R-squared:  0.7467
F-statistic: 11.81 on 6 and 16 DF,  p-value: 4.174e-05
```

Figure 8: Results of regression model on Poland's Gini Coefficient

The linear regression concerning the bottom 50% income share in Poland yielded similar results, as volume of imports ($p < 0.05$) and capital openness ($p < 0.001$) were found to have a significant relationship. For every 1 increase in volume of imports the bottom 50% income share increased by 0.001 (± 0.000). For every 1 increase in capital openness the bottom 50% income share decreased by 0.010 (± 0.002). The other macroeconomic variables were not found to have a significant relationship with bottom 50% income share.

```
Call:
lm(formula = POLTime_and_T$POL_T01 ~ POL_GGX + POL_NGDP + POL_TMGRPCH +
    POL_TXRPCH + POL_NIDNGDP + POL_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.022086 -0.006398 -0.003148  0.007108
      Max
 0.020518

Coefficients:
              Estimate Std. Error t value
(Intercept)  1.374e-01  1.103e-01   1.245
POL_GGX     -9.842e-05  1.901e-03  -0.052
POL_NGDP     6.181e-03  3.049e-03   2.027
POL_TMGRPCH -1.029e-03  7.682e-04  -1.339
POL_TXRPCH   7.189e-04  7.869e-04   0.914
POL_NIDNGDP -9.686e-04  2.020e-03  -0.479
POL_KAOPEN   2.120e-02  4.594e-03   4.616
            Pr(>|t|)
(Intercept) 0.231060
POL_GGX     0.959347
POL_NGDP    0.059679 .
POL_TMGRPCH 0.199233
POL_TXRPCH  0.374473
POL_NIDNGDP 0.638085
POL_KAOPEN  0.000286 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01174 on 16 degrees of freedom
Multiple R-squared:  0.7329,    Adjusted R-squared:  0.6328
F-statistic: 7.319 on 6 and 16 DF,  p-value: 0.000678
```

Figure 9: Results on the regression on Poland's top 1% income share

The linear regression where the top 1% income share was the dependent variable yielded a significant relationship between capital openness and top 1% income share ($p < 0.001$). An increase of 1 in capital openness yielded a 0.021 (± 0.005) increase in the top 1% income share in Poland. The other macroeconomic variables were not found to have a significant relationship with top 1% income share in Poland.

Assumptions addressed. The top 1% income share and Gini coefficient model for Poland were found to be normally divided, homoscedastic and linear. The Ramsay reset tests furthermore indicated that the models were correctly specified. The bottom 50% income share model was found to be non-linear through the residuals vs fitted plot, which was corroborated with a significant Ramsay Reset test, indicating misspecification of the model.

```
Call:
lm(formula = POL_residualsb ~ POL_yhatb + POL_yhatb2 + POL_yhatb3)

Residuals:
       Min         1Q     Median         3Q
-0.0061500 -0.0024958  0.0002596  0.0028083
       Max
 0.0044993

Coefficients:
             Estimate Std. Error t value
(Intercept)    13.690      6.298   2.174
POL_yhatb    -190.598     90.904  -2.097
POL_yhatb2    881.883    436.803   2.019
POL_yhatb3  -1355.986    698.682  -1.941
            Pr(>|t|)
(Intercept)   0.0426 *
POL_yhatb     0.0496 *
POL_yhatb2    0.0578 .
POL_yhatb3    0.0673 .
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.003245 on 19 degrees of freedom
Multiple R-squared:  0.6006,	Adjusted R-squared:  0.5375
F-statistic: 9.523 on 3 and 19 DF,  p-value: 0.0004719
```

Figure 10: Results of the Ramsay Reset test ran on Poland's bottom 50% income share regression

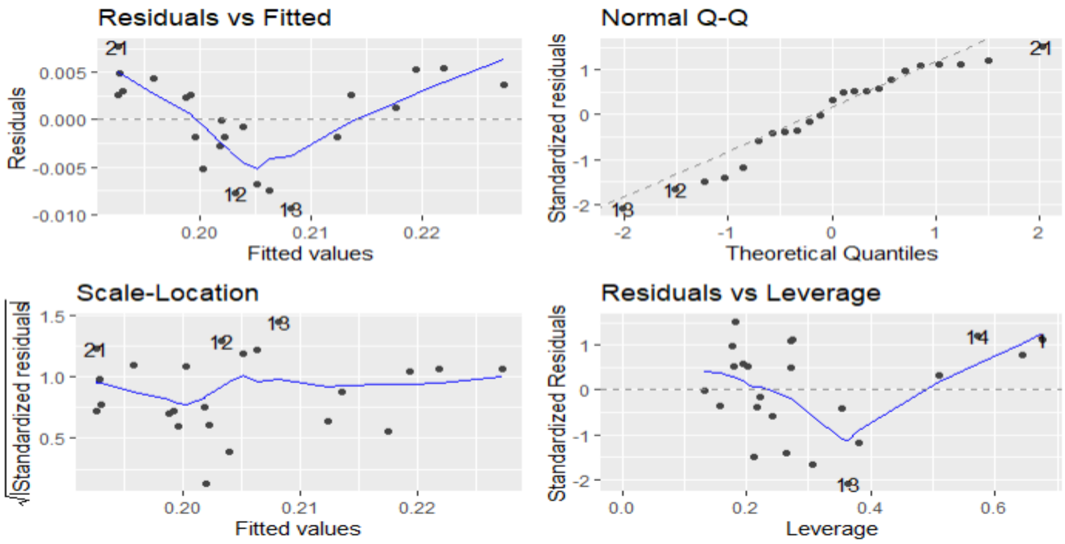

Figure 11: Assumption plots on Poland's bottom 50% income share. Non-Linearity is visible in the Residuals vs Fitted plot

Romania

Linear regression. The linear regression model ran for Romania's Gini coefficient found three significant relationships. Government expenditure ($p < 0.05$), investment as percentage of GDP ($p < 0.01$) and capital openness ($p < 0.01$) were found to have a significant relationship with the Gini coefficient. An increase of 1 in government expenditure was associated with a 0.008 (± 0.003) decrease in Gini coefficient, while investment and capital openness were associated with increases of 0.007 (± 0.002) and 0.011 (± 0.003) respectively. The other macroeconomic variables were not found to have a significant relationship with the Gini coefficient in Romania.

```
Call:
lm(formula = ROUTime_and_Y$ROU_GINI ~ ROU_GGX + ROU_NGDP + ROU_TMGRPCH +
    ROU_TXRPCH + ROU_NIDNGDP + ROU_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0165587 -0.0034871 -0.0003285  0.0031001
       Max
 0.0207678

Coefficients:
              Estimate Std. Error t value
(Intercept)  5.938e-01  7.127e-02   8.331
ROU_GGX     -7.950e-03  3.238e-03  -2.455
ROU_NGDP    -9.903e-05  1.690e-03  -0.059
ROU_TMGRPCH -3.904e-04  4.757e-04  -0.821
ROU_TXRPCH   1.421e-03  8.052e-04   1.765
ROU_NIDNGDP  7.055e-03  2.219e-03   3.180
ROU_KAOPEN   1.132e-02  3.035e-03   3.729
            Pr(>|t|)
(Intercept) 2.48e-06 ***
ROU_GGX      0.03030 *
ROU_NGDP     0.95425
ROU_TMGRPCH  0.42781
ROU_TXRPCH   0.10297
ROU_NIDNGDP  0.00792 **
ROU_KAOPEN   0.00288 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01114 on 12 degrees of freedom
Multiple R-squared:  0.865,	Adjusted R-squared:  0.7975
F-statistic: 12.81 on 6 and 12 DF,  p-value: 0.0001327
```

Figure 12: Results of the regression ran on Romania's Gini Coefficient

The linear regression based upon the bottom 50% of income share yielded similar results, with government expenditure ($p < 0.05$), investment ($p < 0.05$) and capital openness ($p < 0.001$) being identified as significant variables. An increase of 1 in government expenditure was associated with an increase in the bottom 50% income share by 0.004 (± 0.002), while investment as percentage of GPD and capital openness were found to decrease bottom 50% income share by 0.003 (± 0.002) and 0.008 (± 0.002) respectively. The other macroeconomic variables were not found to be significant.

```
Call:
lm(formula = ROUTime_and_B$ROU_B50 ~ ROU_GGX + ROU_NGDP + ROU_TMGRPCH +
    ROU_TXRPCH + ROU_NIDNGDP + ROU_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0079361 -0.0022151  0.0004261  0.0020434
       Max
 0.0088950

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.1150793  0.0371369   3.099
ROU_GGX      0.0040876  0.0016872   2.423
ROU_NGDP     0.0001823  0.0008809   0.207
ROU_TMGRPCH  0.0001428  0.0002479   0.576
ROU_TXRPCH  -0.0006288  0.0004196  -1.499
ROU_NIDNGDP -0.0028756  0.0011560  -2.488
ROU_KAOPEN  -0.0078729  0.0015816  -4.978
            Pr(>|t|)
(Intercept) 0.009211 **
ROU_GGX     0.032153 *
ROU_NGDP    0.839555
ROU_TMGRPCH 0.575137
ROU_TXRPCH  0.159810
ROU_NIDNGDP 0.028561 *
ROU_KAOPEN  0.000321 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005807 on 12 degrees of freedom
Multiple R-squared:  0.8707,	Adjusted R-squared:  0.8061
F-statistic: 13.47 on 6 and 12 DF,  p-value: 0.0001034
```

Figure 13: Results of the regression on Romania's bottom 50% income share

The linear regression with the top 1% income share of Romania yielded 1 significant relationship. A 1 increase in investment as percentage of GDP was found to increase top 1% income share by 0.006 (± 0.002) at $p < 0.05$. The other macroeconomic variables were not found to be significant.

```
Call:
lm(formula = ROUTime_and_T$ROU_T01 ~ ROU_GGX + ROU_NGDP + ROU_TMGRPCH +
    ROU_TXRPCH + ROU_NIDNGDP + ROU_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0111384 -0.0063188 -0.0004683  0.0037285
       Max
 0.0261453

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.1082327  0.0703253   1.539
ROU_GGX     -0.0036294  0.0031949  -1.136
ROU_NGDP     0.0010431  0.0016681   0.625
ROU_TMGRPCH -0.0005916  0.0004694  -1.260
ROU_TXRPCH   0.0012141  0.0007946   1.528
ROU_NIDNGDP  0.0058698  0.0021891   2.681
ROU_KAOPEN   0.0042834  0.0029951   1.430
            Pr(>|t|)
(Intercept)    0.150
ROU_GGX        0.278
ROU_NGDP       0.543
ROU_TMGRPCH    0.231
ROU_TXRPCH     0.152
ROU_NIDNGDP    0.020 *
ROU_KAOPEN     0.178
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.011 on 12 degrees of freedom
Multiple R-squared:  0.7839,	Adjusted R-squared:  0.6758
F-statistic: 7.254 on 6 and 12 DF,  p-value: 0.001896
```

Figure 14: Results of the regression on Romania's top 1% income share

Assumptions addressed. The three models created on Romania were all found to be normally distributed, linear and homoscedastic. The Ramsay reset tests conducted on the three models were found to have no significant relationships and as such the models were correctly specified.

Slovakia

Linear regression. The linear regression ran over the Gini coefficient, bottom 50% income share and top 1% income share in Slovakia yielded no significant relationships with macroeconomic variables. The only notable relationship to be found was economic growth as percentage of GDP, which only barely is not significantly related to top 1% income share at $p = 0.054$.

Assumptions addressed. The models created to fit the income inequality data in Slovakia were found to be linear, homoscedastic and normally distributed. Through the Ramsay reset test misspecification was addressed. The models based upon the Gini coefficient and the top 1% income share were found to be correctly specified. The model based upon bottom 50% income share was however misspecified, showing significant associations with all independent variables.

```
Call:
lm(formula = SVK_residualsb ~ SVK_yhatb + SVK_yhatb2 + SVK_yhatb3)

Residuals:
       Min         1Q     Median         3Q 
-0.0073532 -0.0043764  0.0003658  0.0041817 
       Max 
 0.0095367 

Coefficients:
            Estimate Std. Error t value
(Intercept)    324.3      113.1   2.868
SVK_yhatb    -4129.5     1435.9  -2.876
SVK_yhatb2   17522.1     6077.0   2.883
SVK_yhatb3  -24773.5     8570.5  -2.891
            Pr(>|t|)   
(Intercept)  0.01022 * 
SVK_yhatb    0.01005 * 
SVK_yhatb2   0.00989 **
SVK_yhatb3   0.00974 **
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0054 on 18 degrees of freedom
Multiple R-squared:  0.3409,	Adjusted R-squared:  0.2311 
F-statistic: 3.104 on 3 and 18 DF,  p-value: 0.05261
```

Figure 15: Results of the Ramsay Reset test ran on Slovakia's bottom 50% income share regression.

Discussion

Capital openness

Through multiple linear regression capital openness was found to have a positive significant relationship with the Gini coefficient in Hungary. Its effect is however rather limited. Even though the dataset used by this study offers the Gini coefficient with 12 decimals, in practice these decimals are never used, and the Gini coefficient is visualized through a percentage. The impact of an increase of 1 in capital openness on the Gini coefficient in Hungary is thus only 2.7 (± 0.3) %. The association is however significant and corroborated by the linear regressions over income share. The linear regression including bottom 50% income share returned a negative relationship with capital openness, while the model with top 1% income share returned a positive relationship. The estimate of the impact of capital openness on income share was however limited, only accounting for a decrease of 1.6 (± 0.2) % in bottom 50% and an increase of 1.2 (± 0.2) % in the top 1% share of income. The association found in Hungary’s regression model regarding capital association is corroborated by the regression models of other states. Capital openness was related in the same direction in all regression models if it was found to be significantly associated with income inequality. Capital openness was always, if significant, positively related with the Gini coefficient. For Hungary, Poland and Romania capital openness was thus found to increase income inequality. This finding was further supported by both a negative significant relationship of capital openness with the bottom 50% income share in these aforementioned states and a positive significant relationship with the top 1% income share.

The variable capital openness was computed as a combination of the tabulation of restrictions on cross border financial transactions, and therefore is a measure of the ease of international financial flows (Chinn & Ito, 2006). A study conducted on the relationship between income inequality and globalization furthermore found that capital openness was a driving factor in income inequality in the EU over the period of 1995 – 2009 (Asteriou et al., 2014). A further paper, although naming technological change the main driver behind income inequality instead of globalization, came to this conclusion only due to the offsetting effect of trade, which decreases income inequality. The other factor in globalization, capital openness was found to exacerbate it (Jaumotte et al., 2008). The finding that a state with a higher degree of capital openness would have a higher level of income inequality is therefore supported by literature.

This effect of capital openness could be caused by the nature of the variable, those in the higher income shares profit more from the ability to move capital internationally, as they have the capital to move. Those in the lower income shares do not have the capital to profit off of capital openness in the same manner as the richer income shares.

Government Expenditure

Government expenditure was found to have both positive and negative relationships with income inequality across different countries. In Bulgaria the model found government expenditure to increase the Gini coefficient by 1 (± 0.43) %, while in Romania an increase in government expenditure was found to decrease income inequality by 0.8 (± 0.3) %. In Czechia government expenditure was not significantly associated with the Gini coefficient, but government expenditure did increase the income share of the bottom 50% by 0.2 (± 0.1) %. This paper argues that the difference in direction of impact can be explained by the computation of the variable itself. As government expenditure is a general variable it measures all expenditure. Governments are free to spend state capital on measures combatting income inequality or on measures (indirectly) accelerating it. The statistical models presented in this paper can therefore be used to gouge the impact of government efforts to limit income inequality. The states where a significant association was missing between government expenditure and income inequality, in this paper Hungary, Poland and Slovakia, can be explained through the same argument. The government efforts to combat income inequality in these states might be ineffective and therefore no significant relationship was found.

As seen during the literature review, government expenditure’s relationship to income inequality is heavily debated. In 1997 no significant relationship was found, while in 2012 in an analysis of income inequality in European countries was conducted, and found a positive relationship between income inequality and government expenditure (Maestri & Roventini, 2012) (Sarel, 1997). The hypothesis that government expenditure decreases income inequality in the short term and increases it in the long term could not be tested through the analysis done in this study (Deyshappriya, 2017). The lack of clear consensus on the relationship between income inequality and government expenditure does not have to disprove the theory for its varying impact stipulated by this study above. As perceptions on how the distribution of income in a state should be are shaped by the collective attitudes of the population of a state, and as those attitudes in socialist states favour a more egalitarian society, and as the attitudes only change slowly over time, it could be reasoned that western European democratic states invest less capital or have fewer effective methods to combat income inequality than democratic eastern European states (Gijsberts, 2002).

Investment in Romania

In the regression models for Romania two positive associations were found between investment as a percentage of GDP and income inequality. A 1 increase in investment as a percentage of GDP was estimated to increase the Gini coefficient by 0.7 (± 0.2) % and the income share of the top 1% by 0.6 (± 0.2) %. The level of significance was however not as high as measured by previously mentioned variables. The abovementioned associations were only significant at $p < 0.01$ for Gini coefficient and $p < 0.05$ for the top 1% income share. From these results we can thus reason that total investment in Romania increases income inequality through increasing the income share of the top 1%. The findings are corroborated by literature, which finds investment to be a major predictor of income inequality in states during the transition from a communist economic system towards a free market economy. (Bandelj & Mahutga, 2010) (Georgantopoulos & Tsamis, 2011) (Aiyar & Ebeke, 2020). Foreign direct investment and investment as a share of GDP were even after the transition found to have positive correlations with several inequality variables (Deyshappriya, 2017)

(Jäntti & Jenkins, 2009). Romania is an outlier in this study as it is the only state to have a significant relationship between income inequality and investment, but the findings are corroborated by a consensus in academic literature.

Hungary's macroeconomic results

Hungary's regression models provided no significant associations besides the relationship between income inequality and capital openness. The macroeconomic variables included in this study were identified through literature review, and a such a significant relationship between the variables and income inequality has been pointed out in other studies. Nevertheless the possibility remains that these macroeconomic variables are unable to explain income inequality in Hungary. Besides this possibility the lack of associations in the regression models could also be due to a limited amount of observation, or an incorrect setup of the directed acyclic graph.

Limitations

Failed Ramsay Reset Tests

Two of the Ramsay Reset tests conducted, which test for misspecification and linearity in the model, had results which were significantly related. The two regression models which failed the tests were Poland's and Slovakia's bottom 50% income share regression. The polish regression probably failed the Ramsay Reset test due to non-linearity. The residuals vs fitted values plot of the Polish bottom income share regression clearly shows a parabolic line, which cannot be interpreted differently. The failure of the Slovakian bottom income share regression model to pass the Ramsay reset test is harder to explain. The residuals vs fitted values plot does show some irregularity, but the variance throughout the plot remains linear. Therefore this study hypothesizes that the Slovakian bottom 50% regression model was misspecified. The results of both abovementioned regressions have not been interpreted as they are unusable.

Socio-cultural aspects

The scope of this paper limited it to macroeconomic variables and therefore no socio-cultural variables could be tested, while these are of importance to income inequality (Bandelj & Mahutga, 2010). Furthermore economies are never purely reasonable, as behavioural aspects play a large role in the everyday choices of people. The social exclusion of minorities could therefore not be separately determined. The influence of this exclusion might therefore be present within the current results of the models, skewing them from the impact the macroeconomic variable would otherwise have. Besides, the level of income inequality that is accepted by the population depends upon the collective attitudes of that state, and those attitudes can change over time (Gijsberts, 2002). A further variable that was not included in the regression models was education. As it lies outside the scope of this paper it was not possible to include it, but if the models were to be improved, to describe income inequality more fittingly, then education would be a variable to include.

Exclusion of variables

The size of the shadow economy of a state was found to be strongly positively associated with income inequality by (Berdiev & Saunoris, 2018). The variable was thus deemed to be viable to include in the statistical regression models of this study. However, when collecting the data for this study, it was found that the amount of observations present would reduce the temporal scope of this paper to just the years between 1996 and 2004. The explaining power and significance of the models to draw conclusions from would then have been undermined. Therefore the trade-off on including the size of the shadow economy as a variable was not deemed worth it.

Consensus in modern relevant literature was hard to identify, as almost all papers analysed in the literature review used different methods of data collection and analysis. Several of the papers were attempting to find proof for the Kuznets curve, while others actively discarded the theory. This study has aimed to analyse the most relevant literature for the research questions mentioned in this paper, but decisions had to be made on the inclusion and exclusion on certain macroeconomic variables, such as unemployment and inflation.

Lack of observations

The countries that were within this models scope to run model regressions for were relatively limited in the amount of observations compared to western European states. The data for the states within this study was limited because there are hardly any observations from the time when the states had communist governments. Furthermore the states had different amounts of observations among themselves, and therefore the ability to compare the states with each other might have decreased. The hypothesized impact of this is small however, as the difference in observations is limited to 5 years.

Conclusion

To conclude this paper this study shall answer the research questions postulated after the introduction, and will then give recommendations for future research.

The first research question tested by this study was as follows: *Can the identified macroeconomic variables explain the trajectory of income inequality in Hungary since 1995?*

The identified macroeconomic variables were unable to explain the trajectory of income inequality in Hungary since 1995. The only macroeconomic variable that succeeded in explaining part of the increase in income inequality in Hungary since 1995 was capital openness, which was found to increase the Gini coefficient by 2.7 (± 0.3) % for every 1 increase in the capital openness index. The other variables did not register as statistically significant and therefore did not explain income inequality in Hungary in the multiple linear regression model created by this study.

The second research question posed by this study was: *Which variables have had a disproportionate impact on income inequality in Hungary after the fall of communism compared to other eastern European states?*

The other states that were assessed in this study displayed varying results, with government expenditure being significantly related to income inequality in Romania, Czechia and Bulgaria. The hypothesis given for this disproportionate impact of government expenditure on separate states in this paper is that the variables measures the collective sum of all effects of government expenditure on income inequality, and is therefore a good measure of overall effectiveness of the state to combat income inequality through capital. The other variables of disproportionate impact were volume of imports in Poland, which was negatively associated with income inequality and total investment as percentage of GDP in Romania which was estimated to increase income inequality. The different findings across different states highlight the importance of local solutions to the global problem of income inequality. Similar aspects might affect different states differently, and as such tailor made policies are the most effective tool to combat the global economic issues of today.

Future Research

Several recommendations have been constructed by this study for future research. The first recommendation for future research is to test the significance of capital openness against states different than those tested in this paper. Furthermore the models should test the relationship of capital openness and growth of income for the bottom income shares in several states. Thereby an assessment can be made of the merits of capital openness, and if it is desirable by states or should be avoided.

Further suggestions include the improvement of the models created by this study, namely the involvement of more socio-cultural variables supported by academic literature. The models should furthermore be expanded with additional observations where possible. Besides, the models should be tested for the best fit. A model would thus be created which is the closest to identifying the complete trajectory of income inequality in a state. Literature review can then be conducted to identify the relevant policies which contributed the most towards the level of income inequality in the states used. Policy advice can then be constructed in order to improve the ability of the state to find solutions tailored to the specifics of their culture and economy.

References

Aiyar, S., & Ebeke, C. (2020). Inequality of opportunity, inequality of income and economic growth. *World Development*, *136*, 105115. https://doi.org/10.1016/j.worlddev.2020.105115

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2018). The Elephant Curve of Global Inequality and Growth. *AEA Papers and Proceedings*, *108*, 103–108. https://doi.org/10.1257/pandp.20181073

Anyanwu, J. C., Erhijakpor, A. E., & Obi, E. (2016). Empirical analysis of the key drivers of income inequality in West Africa. *African Development Review*, *28*(1), 18–38. https://doi.org/10.1111/1467-8268.12164

Asteriou, D., Dimelis, S., & Moudatsou, A. (2014). Globalization and income inequality: A panel data econometric approach for the EU27 countries. *Economic Modelling*, *36*, 592–599. https://doi.org/10.1016/j.econmod.2013.09.051

Bandelj, N., & Mahutga, M. C. (2010). How Socio-Economic Change Shapes Income Inequality in Post-Socialist Europe. *Social Forces*, *88*(5), 2133–2161. https://doi.org/10.1353/sof.2010.0042

Berdiev, A. N., & Saunoris, J. W. (2018). On the Relationship Between Income Inequality and the Shadow Economy. *Eastern Economic Journal*, *45*(2), 224–249. https://doi.org/10.1057/s41302-018-0120-y

Bíró, A., Hajdu, T., Kertesi, G., & Prinz, D. (2021). Life expectancy inequalities in Hungary over 25 years: The role of avoidable deaths. *Population Studies*, *75*(3), 443–455. https://doi.org/10.1080/00324728.2021.1877332

Brueckner, M., & Lederman, D. (2018). Inequality and economic growth: the role of initial income. *Journal of Economic Growth*, *23*(3), 341–366. https://doi.org/10.1007/s10887-018-9156-4

Chinn, M. D., & Ito, H. (2006). What matters for financial development? Capital controls, institutions, and interactions. *Journal of Development Economics*, *81*(1), 163–192. https://doi.org/10.1016/j.jdeveco.2005.05.010

Datopia. (n.d.). *Datahub WEO dataset*. Datahub.Io. Retrieved June 10, 2022, from https://datahub.io/core/imf-weo#data

Detollenaere, J., Desmarest, A. S., Boeckxstaens, P., & Willems, S. (2018). The link between income inequality and health in Europe, adding strength dimensions of primary care to the equation. *Social Science & Medicine*, *201*, 103–110. https://doi.org/10.1016/j.socscimed.2018.01.041

Deyshappriya, N. P. R. (2017). Impact of macroeconomic factors on income inequality and income distribution in asian countries. *ADBI Working Paper 696*. https://ecommons.cornell.edu/bitstream/handle/1813/87207/ADB_Impact_of_macroeconomic_factors.pdf?sequence=1&isAllowed=y

Furceri, D., & Ostry, J. D. (2019). OUP accepted manuscript. *Oxford Review Of Economic Policy*, *35*(3). https://doi.org/10.1093/oxrep/grz014

Georgantopoulos, A. G., & Tsamis, A. (2011). The impact of globalization on income distribution: the case of Hungary. *Research Journal of International Studıes*, *21*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2063022

Gijsberts, M. (2002). The legitimation of income inequality in state-socialist and market societies. *Acta Sociologica*, *45*(4), 269–285. https://journals.sagepub.com/doi/abs/10.1177/000169930204500402?casa_token=n0awMwxdJP4AAAAA:bljC_Tu-aCs86AwcLVLhggDEvup29Dlyt1waRj9rmrAI_Nk4qiV0oHo1vPvc7XlTcN53ryV_To7H

IMF. (n.d.-a). *IMF World Economic Outlook*. Https://Www.Imf.Org/En/Publications/WEO. Retrieved June 10, 2022, from https://www.imf.org/en/Publications/WEO

IMF. (n.d.-b). *IMF World Economic Outlook Databases*. Retrieved June 10, 2022, from https://www.imf.org/en/Publications/SPROLLS/world-economic-outlook-databases#sort=%40imfdate%20descending

IMF. (2015). *Causes and Consequences of Income Inequality; Staff Discussion Notes* (No. A001). https://www.elibrary.imf.org/view/journals/006/2015/013/article-A001-en.xml

Jäntti, M., & Jenkins, S. P. (2009). The impact of macroeconomic conditions on income inequality. *The Journal of Economic Inequality*, *8*(2), 221–240. https://doi.org/10.1007/s10888-009-9113-8

Jaumotte, F., Papageorgiou, C., & Lall, S. (2008). Rising Income Inequality: Technology, or Trade and Financial Globalization? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1175363

Kim, J. H. (2015). A Study on the Effect of Financial Inclusion on the Relationship Between Income Inequality and Economic Growth. *Emerging Markets Finance and Trade*, *52*(2), 498–512. https://doi.org/10.1080/1540496x.2016.1110467

Maestri, V., & Roventini, A. (2012). *Inequality and macroeconomic factors: A time-series analysis for a set of OECD countries.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2181399

Mihaylova, S. (2015). Foreign direct investment and income inequality in Central and Eastern Europe. *Theoretical & Applied Economics*, *22*(2), 23–42. http://store.ectap.ro/articole/1080.pdf

Pickett, K. E., & Wilkinson, R. G. (2015). Income inequality and health: A causal review. *Social Science & Medicine*, *128*, 316–326. https://doi.org/10.1016/j.socscimed.2014.12.031

Sarel, M. M. (1997). *How macroeconomic factors affect income distribution: The cross-country evidence*. International Monetary Fund. https://books.google.com/books?hl=nl&lr=&id=K4cYEAAAQBAJ&oi=fnd&pg=PA3&dq=sarel+et+al+1997&ots=XLcKxC_Dc5&sig=OijO2RaWrlPjdlw7j_ed4AVslCk

van Zanden, J. L., Baten, J., Foldvari, P., & van Leeuwen, B. (2013). The Changing Shape of Global Inequality 1820–2000; Exploring a New Dataset. *Review of Income and Wealth*, *60*(2), 279–297. https://doi.org/10.1111/roiw.12014

WID.WORLD. (n.d.). *World Inequality Database*. WID. Retrieved June 10, 2022, from https://wid.world/data/

ŽÍdek, L. (2014). Evaluation of economic transformation in Hungary. *Review of Economic Perspectives*, *14*(1), 55–88. https://www.econstor.eu/handle/10419/179804

Appendix A: R code

````
```{r setup, include=FALSE}

knitr::opts_chunk$set(echo = TRUE)

```

#0 Getting Packages

```{r}

#install.packages('tidyverse')

#install.packages('ggplot2')

#install.packages('tidyr')

#install.packages('broom')

install.packages('ggfortify')

```

#0 Loading Packages

```{r}

library(tidyverse)

library(ggplot2)

library(dplyr)

library(tidyr)

library(broom)

library(ggfortify)

library(readxl)

library(stats)

```

#1 Downloading Macroeconomic Dataset
````

````
```{r}
install.packages("jsonlite", repos="https://cran.rstudio.com/")
library("jsonlite")

json_file <- 'https://datahub.io/core/imf-weo/datapackage.json'
json_data <- fromJSON(paste(readLines(json_file), collapse=""))

# get list of all resources:
print(json_data$resources$name)

# print all tabular data(if exists any)
for(i in 1:length(json_data$resources$datahub$type)){
  if(json_data$resources$datahub$type[i]=='derived/csv'){
    path_to_file = json_data$resources$path[i]
    data <- read.csv(url(path_to_file))
    print(data)
  }
}
```
#1 Reading Datasets
```{r}
kaopen_2019 <- read_excel("C:\\Users\\ruben\\Downloads\\kaopen_2019.xlsx")
#df_ineq <- read.csv2('WID_data_ineq.csv')
Polity_V <- read_excel("C:\\Users\\ruben\\Downloads\\Polity V.xls")
````
````

```

#2 Subsetting Countries

```{r}
df <- subset(data, data$Country == 'HUN' | data$Country == 'ROU' | data$Country == 'CZE' | data$Country == 'SVK' | data$Country == 'POL' | data$Country == 'BGR')
```

#2 Making Exploratory visualisations

1) GINI Graph

2) Bottom 50% incomeshare graph

```{r}
#detach(df_ineq)
attach(df_ineq)
plot(HUN_GINI, type = 'o', col = 'red', ylab = 'GINI-Scores', xlab = 'Years since 1985', ylim = c(0.1, 0.75), main = 'Comparing GINI scores')
   lines(ROU_GINI, type = 'l', col = 'blue')
   lines(CZE_GINI, type = 'l', col = 'green')
   lines(SVK_GINI, type = 'l', col = 'yellow')
   lines(POL_GINI, type = 'l', col = 'black')
   lines(BGR_GINI, type = 'l', col = 'purple')
   legend(0, 0.75, legend = c('Hungary', 'Romania', 'Czechia', 'Slovakia', 'Poland', 'Bulgaria'), col=c('red', 'blue', 'green', 'yellow', 'black', 'purple'), lty = 1:1, cex = 0.45)

````
```

```{r}
attach(df_ineq)
plot(HUN_B50, type = 'o', col = 'red', ylab = '% Incomeshare', xlab = 'Years since 1985', ylim = c(0.1, 0.4), main = 'Bottom 50% income share')
    lines(ROU_B50, type = 'l', col = 'blue')
    lines(CZE_B50, type = 'l', col = 'green')
    lines(SVK_B50, type = 'l', col = 'yellow')
    lines(POL_B50, type = 'l', col = 'black')
    lines(BGR_B50, type = 'l', col = 'purple')
    legend(25, 0.4, legend = c('Hungary', 'Romania', 'Czechia', 'Slovakia', 'Poland', 'Bulgaria'), col=c('red', 'blue', 'green', 'yellow', 'black', 'purple'), lty = 1:1, cex = 0.45)
```


Comparing Inequality to Gov expenditure (GGX)
```{r}
plot(df_ineq$Year[df_ineq$Year > 1994], df_ineq$HUN_GINI[df_ineq$Year > 1994 & df_ineq$Year < 2020], type = 'o', col = 'red', ylab = 'GINI-Score', xlab = 'Years', ylim = c(0.25, 0.5), main = 'GINI & GGX')

par(new=TRUE)
plot(df_hun$Year[df_hun$Indicator == 'GGX'], df_hun$Value[df_hun$Indicator == 'GGX'], xlab = '', ylab = '', ylim = c(2000, 22000), axes = FALSE, type = 'b', col = 'black')
````

~~~~
  mtext('Government Expenditure', side = 4, col = 'black', line = 4)
  axis(4, ylim=c(2000, 22000), col = 'black', col.axis = 'black', las=1)
```


## Getting Dependent Variables
```{r Data_Org, include=TRUE}
#Creating Variables
## HUN Bottom 50% incomeshare
HUNTime_and_B1 = df_ineq[,c(1,5)]
HUNTime_and_B <- subset(HUNTime_and_B1, Year >= 1996)


##HUN Top 1% incomeshare
HUNTime_and_T1 = df_ineq[,c(1,4)]
HUNTime_and_T <- subset(HUNTime_and_T1, Year >= 1996)


## HUN GINI
HUNTime_and_Y1 = df_ineq[,c(1,2)]
HUNTime_and_Y <- subset(HUNTime_and_Y1, Year >= 1996)
```
## Getting Independent Variables
```{r Data_org, include=TRUE}
## GGX / Gov Expenses
HUN_GGX = as.numeric(df[df$Country == 'HUN' & df$Indicator == 'GGX_NGDP' & df$Year >= 1995 & df$Year <= 2018,]$Value)
~~~~

NGDP / Eco Growth

HUN_NGDP = as.numeric(df[df$Country == 'HUN' & df$Indicator == 'NGDP_RPCH' & df$Year >= 1995 & df$Year <= 2018,]$Value)

TMG_RPCH / Imports Volume

HUN_TMGRPCH = as.numeric(df[df$Country == 'HUN' & df$Indicator == 'TM_RPCH' & df$Year >= 1995 & df$Year <= 2018,]$Value)

TX_RPCH / Exports Volume

HUN_TXRPCH = as.numeric(df[df$Country == 'HUN' & df$Indicator == 'TX_RPCH' & df$Year >= 1995 & df$Year <= 2018,]$Value)

NID_NGDP / Investment

HUN_NIDNGDP = as.numeric(df[df$Country == 'HUN' & df$Indicator == 'NID_NGDP' & df$Year >= 1995 & df$Year <= 2018,]$Value)

Kaopen / Capital openness

HUN_KAOPEN = as.numeric(kaopen_2019[kaopen_2019$ccode == 'HUN' & kaopen_2019$year >= 1995 & kaopen_2019$year <= 2018,]$kaopen)

```

## Linear regression analysis hungary

```{r}

# GINI

HUN_lm = lm(HUNTime_and_Y$HUN_GINI ~ HUN_GGX + HUN_NGDP + HUN_TMGRPCH + HUN_TXRPCH + HUN_NIDNGDP + HUN_KAOPEN)

# Bottom 50

HUN_lmb = lm(HUNTime_and_B$HUN_B50 ~ HUN_GGX + HUN_NGDP + HUN_TMGRPCH + HUN_TXRPCH + HUN_NIDNGDP + HUN_KAOPEN)

# Top 01

~~~
HUN_lmt = lm(HUNTime_and_T$HUN_T01 ~ HUN_GGX + HUN_NGDP + HUN_TMGRPCH + HUN_TXRPCH + HUN_NIDNGDP + HUN_KAOPEN)



summary(HUN_lm)

summary(HUN_lm_b50)

summary(HUN_lm_t1)

```

HUNGARY Visualizing

```{r}

autoplot(HUN_lm)

```

ROMANIA, BULGARIA, POLAND, CZECHIA, SLOVAKIA VARIABLES

Bulgaria

```{r}

## BGR GINI

BGRTime_and_Y1 = df_ineq[,c(1,27)]

BGRTime_and_Y <- subset(BGRTime_and_Y1, Year >= 2001)

# BGR B50

BGRTime_and_B1 = df_ineq[,c(1,30)]

BGRTime_and_B <- subset(BGRTime_and_B1, Year >= 2001)

# BGR T01

BGRTime_and_T1 = df_ineq[,c(1,29)]

BGRTime_and_T <- subset(BGRTime_and_T1, Year >= 2001)
~~~

## GGX / Gov Expenses

BGR_GGX = as.numeric(df[df$Country == 'BGR' & df$Indicator == 'GGX_NGDP' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## NGDP / Eco Growth

BGR_NGDP = as.numeric(df[df$Country == 'BGR' & df$Indicator == 'NGDP_RPCH' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## TMG_RPCH / Imports Volume

BGR_TMGRPCH = as.numeric(df[df$Country == 'BGR' & df$Indicator == 'TM_RPCH' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## TX_RPCH / Exports Volume

BGR_TXRPCH = as.numeric(df[df$Country == 'BGR' & df$Indicator == 'TX_RPCH' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## NID_NGDP / Investment

BGR_NIDNGDP = as.numeric(df[df$Country == 'BGR' & df$Indicator == 'NID_NGDP' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## Kaopen / Capital openness

BGR_KAOPEN = as.numeric(kaopen_2019[kaopen_2019$ccode == 'BGR' & kaopen_2019$year >= 2000 & kaopen_2019$year <= 2018,]$kaopen)

```

Czechia

```{r}

## CZE GINI

## cze Bottom 50% incomeshare

CZETime_and_B1 = df_ineq[,c(1,15)]

CZETime_and_B <- subset(HUNTime_and_B1, Year >= 1997)

```
##cze Top 1% incomeshare

CZETime_and_T1 = df_ineq[,c(1,14)]

CZETime_and_T <- subset(CZETime_and_T1, Year >= 1997)

GINI

CZETime_and_Y1 = df_ineq[,c(1,12)]

CZETime_and_Y <- subset(CZETime_and_Y1, Year >= 1997)

GGX / Gov Expenses

CZE_GGX = as.numeric(df[df$Country == 'CZE' & df$Indicator == 'GGX_NGDP' & df$Year >= 1996 & df$Year <= 2018,]$Value)

NGDP / Eco Growth

CZE_NGDP = as.numeric(df[df$Country == 'CZE' & df$Indicator == 'NGDP_RPCH' & df$Year >= 1996 & df$Year <= 2018,]$Value)

TMG_RPCH / Imports Volume

CZE_TMGRPCH = as.numeric(df[df$Country == 'CZE' & df$Indicator == 'TM_RPCH' & df$Year >= 1996 & df$Year <= 2018,]$Value)

TX_RPCH / Exports Volume

CZE_TXRPCH = as.numeric(df[df$Country == 'CZE' & df$Indicator == 'TX_RPCH' & df$Year >= 1996 & df$Year <= 2018,]$Value)

NID_NGDP / Investment

CZE_NIDNGDP = as.numeric(df[df$Country == 'CZE' & df$Indicator == 'NID_NGDP' & df$Year >= 1996 & df$Year <= 2018,]$Value)

Kaopen / Capital openness

CZE_KAOPEN = as.numeric(kaopen_2019[kaopen_2019$ccode == 'CZE' & kaopen_2019$year >= 1996 & kaopen_2019$year <= 2018,]$kaopen)
```

```

# Poland

```{r}

## POL GINI

POLTime_and_Y1 = df_ineq[,c(1,22)]

POLTime_and_Y <- subset(POLTime_and_Y1, Year >= 1997)

# BGR B50

POLTime_and_B1 = df_ineq[,c(1,25)]

POLTime_and_B <- subset(POLTime_and_B1, Year >= 1997)

# BGR T01

POLTime_and_T1 = df_ineq[,c(1,24)]

POLTime_and_T <- subset(POLTime_and_T1, Year >= 1997)

## GGX / Gov Expenses

POL_GGX = as.numeric(df[df$Country == 'POL' & df$Indicator == 'GGX_NGDP' & df$Year >= 1996 & df$Year <= 2018,]$Value)

## NGDP / Eco Growth

POL_NGDP = as.numeric(df[df$Country == 'POL' & df$Indicator == 'NGDP_RPCH' & df$Year >= 1996 & df$Year <= 2018,]$Value)

## TMG_RPCH / Imports Volume

POL_TMGRPCH = as.numeric(df[df$Country == 'POL' & df$Indicator == 'TM_RPCH' & df$Year >= 1996 & df$Year <= 2018,]$Value)

## TX_RPCH / Exports Volume

POL_TXRPCH = as.numeric(df[df$Country == 'POL' & df$Indicator == 'TX_RPCH' & df$Year >= 1996 & df$Year <= 2018,]$Value)

## NID_NGDP / Investment

POL_NIDNGDP = as.numeric(df[df$Country == 'POL' & df$Indicator == 'NID_NGDP' & df$Year >= 1996 & df$Year <= 2018,]$Value)

## Kaopen / Capital openness

POL_KAOPEN = as.numeric(kaopen_2019[kaopen_2019$ccode == 'POL' & kaopen_2019$year >= 1996 & kaopen_2019$year <= 2018,]$kaopen)

```

Slovakia

```{r}

## SVK GINI

SVKTime_and_Y1 = df_ineq[,c(1,17)]

SVKTime_and_Y <- subset(SVKTime_and_Y1, Year >= 1998)

#B50

SVKTime_and_B1 = df_ineq[,c(1,20)]

SVKTime_and_B <- subset(SVKTime_and_B1, Year >= 1998)

#To1

SVKTime_and_T1 = df_ineq[,c(1,19)]

SVKTime_and_T <- subset(SVKTime_and_T1, Year >= 1998)

## GGX / Gov Expenses

SVK_GGX = as.numeric(df[df$Country == 'SVK' & df$Indicator == 'GGX_NGDP' & df$Year >= 1997 & df$Year <= 2018,]$Value)

## NGDP / Eco Growth

SVK_NGDP = as.numeric(df[df$Country == 'SVK' & df$Indicator == 'NGDP_RPCH' & df$Year >= 1997 & df$Year <= 2018,]$Value)

## TMG_RPCH / Imports Volume

SVK_TMGRPCH = as.numeric(df[df$Country == 'SVK' & df$Indicator == 'TM_RPCH' & df$Year >= 1997 & df$Year <= 2018,]$Value)

## TX_RPCH / Exports Volume

SVK_TXRPCH = as.numeric(df[df$Country == 'SVK' & df$Indicator == 'TX_RPCH' & df$Year >= 1997 & df$Year <= 2018,]$Value)

## NID_NGDP / Investment

SVK_NIDNGDP = as.numeric(df[df$Country == 'SVK' & df$Indicator == 'NID_NGDP' & df$Year >= 1997 & df$Year <= 2018,]$Value)

## Kaopen / Capital openness

SVK_KAOPEN = as.numeric(kaopen_2019[kaopen_2019$ccode == 'SVK' & kaopen_2019$year >= 1997 & kaopen_2019$year <= 2018,]$kaopen)

```

Romania

```{r}

## ROU GINI

ROUTime_and_Y1 = df_ineq[,c(1,7)]

ROUTime_and_Y <- subset(ROUTime_and_Y1, Year >= 2001)

#B50

ROUTime_and_B1 = df_ineq[,c(1,10)]

ROUTime_and_B <- subset(ROUTime_and_B1, Year >= 2001)

#To1

ROUTime_and_T1 = df_ineq[,c(1,9)]

ROUTime_and_T <- subset(ROUTime_and_T1, Year >= 2001)

## GGX / Gov Expenses

ROU_GGX = as.numeric(df[df$Country == 'ROU' & df$Indicator == 'GGX_NGDP' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## NGDP / Eco Growth

ROU_NGDP = as.numeric(df[df$Country == 'ROU' & df$Indicator == 'NGDP_RPCH' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## TMG_RPCH / Imports Volume

ROU_TMGRPCH = as.numeric(df[df$Country == 'ROU' & df$Indicator == 'TM_RPCH' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## TX_RPCH / Exports Volume

ROU_TXRPCH = as.numeric(df[df$Country == 'ROU' & df$Indicator == 'TX_RPCH' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## NID_NGDP / Investment

ROU_NIDNGDP = as.numeric(df[df$Country == 'ROU' & df$Indicator == 'NID_NGDP' & df$Year >= 2000 & df$Year <= 2018,]$Value)

## Kaopen / Capital openness

ROU_KAOPEN = as.numeric(kaopen_2019[kaopen_2019$ccode == 'ROM' & kaopen_2019$year >= 2000 & kaopen_2019$year <= 2018,]$kaopen)

```

TESTING LM

BGR

```{r}

# GINI

BGR_lm = lm(BGRTime_and_Y$BGR_GINI ~ BGR_GGX + BGR_NGDP + BGR_TMGRPCH + BGR_TXRPCH + BGR_NIDNGDP + BGR_KAOPEN) summary(BGR_lm)

# B50

BGR_lmb = lm(BGRTime_and_B$BGR_B50 ~ BGR_GGX + BGR_NGDP + BGR_TMGRPCH + BGR_TXRPCH + BGR_NIDNGDP + BGR_KAOPEN)

summary(BGR_lmb)

# T01

BGR_lmt = lm(BGRTime_and_T$BGR_T01 ~ BGR_GGX + BGR_NGDP + BGR_TMGRPCH + BGR_TXRPCH + BGR_NIDNGDP + BGR_KAOPEN)

summary(BGR_lmt)

```

CZE lm

```{r}

#Gini

CZE_lm = lm(CZETime_and_Y$CZE_GINI ~ CZE_GGX + CZE_NGDP + CZE_TMGRPCH + CZE_TXRPCH + CZE_NIDNGDP)

summary(CZE_lm)

# B50

CZE_lmb = lm(CZETime_and_B$HUN_B50 ~ CZE_GGX + CZE_NGDP + CZE_TMGRPCH + CZE_TXRPCH + CZE_NIDNGDP + CZE_KAOPEN)

summary(CZE_lmb)

# T01

CZE_lmt = lm(CZETime_and_T$CZE_T01 ~ CZE_GGX + CZE_NGDP + CZE_TMGRPCH + CZE_TXRPCH + CZE_NIDNGDP + CZE_KAOPEN)

summary(CZE_lmt)

```
```

POL LM

```{r}

# GINI

POL_lm = lm(POLTime_and_Y$POL_GINI ~ POL_GGX + POL_NGDP + POL_TMGRPCH + POL_TXRPCH + POL_NIDNGDP + POL_KAOPEN)

summary(POL_lm)

# B50

POL_lmb = lm(POLTime_and_B$POL_B50 ~ POL_GGX + POL_NGDP + POL_TMGRPCH + POL_TXRPCH + POL_NIDNGDP + POL_KAOPEN)

summary(POL_lmb)

# T01

POL_lmt = lm(POLTime_and_T$POL_T01 ~ POL_GGX + POL_NGDP + POL_TMGRPCH + POL_TXRPCH + POL_NIDNGDP + POL_KAOPEN)

summary(POL_lmt)

```

ROU LM

```{r}

# GINI

ROU_lm = lm(ROUTime_and_Y$ROU_GINI ~ ROU_GGX + ROU_NGDP + ROU_TMGRPCH + ROU_TXRPCH + ROU_NIDNGDP + ROU_KAOPEN)

summary(ROU_lm)

# B50

ROU_lmb = lm(ROUTime_and_B$ROU_B50 ~ ROU_GGX + ROU_NGDP + ROU_TMGRPCH + ROU_TXRPCH + ROU_NIDNGDP + ROU_KAOPEN)

summary(ROU_lmb)
```

T01

ROU_lmt = lm(ROUTime_and_T$ROU_T01 ~ ROU_GGX + ROU_NGDP + ROU_TMGRPCH + ROU_TXRPCH + ROU_NIDNGDP + ROU_KAOPEN)

summary(ROU_lmt)

```

# SVK LM

```{r}

# GINI

SVK_lm = lm(SVKTime_and_Y$SVK_GINI ~ SVK_GGX + SVK_NGDP + SVK_TMGRPCH + SVK_TXRPCH + SVK_NIDNGDP + SVK_KAOPEN)

summary(SVK_lm)

# B50

SVK_lmb = lm(SVKTime_and_B$SVK_B50 ~ SVK_GGX + SVK_NGDP + SVK_TMGRPCH + SVK_TXRPCH + SVK_NIDNGDP + SVK_KAOPEN)

summary(SVK_lmb)

# T01

SVK_lmt = lm(SVKTime_and_T$SVK_T01 ~ SVK_GGX + SVK_NGDP + SVK_TMGRPCH + SVK_TXRPCH + SVK_NIDNGDP + SVK_KAOPEN)

summary(SVK_lmt)

```

#VISUALIZING

```{r}

# GINI

autoplot(HUN_lm)

autoplot(BGR_lm)

autoplot(CZE_lm)

autoplot(POL_lm)

autoplot(ROU_lm)

autoplot(SVK_lm)

```

Visualizing B50

```{r}

autoplot(HUN_lmb)

autoplot(BGR_lmb)

autoplot(CZE_lmb)

autoplot(POL_lmb)

autoplot(ROU_lmb)

autoplot(SVK_lmb)

```

Visualizing T01

```{r}

autoplot(HUN_lmt)

autoplot(BGR_lmt)

autoplot(CZE_lmt)

autoplot(POL_lmt)

autoplot(ROU_lmt)

autoplot(SVK_lmt)

```
```

```
#### RAMSAY TEST Hungary

```{r Data_Org, include=TRUE}

## Gini

HUN_yhat = fitted.values(HUN_lm)

HUN_yhat2 = HUN_yhat^2

HUN_yhat3 = HUN_yhat^3

HUN_residuals = residuals(HUN_lm)

HUN_RRT = lm(HUN_residuals ~ HUN_yhat + HUN_yhat2 + HUN_yhat3)

summary(HUN_RRT)

# H0 = Error term not related to independent variables, thus Ramsay test = not significant


# T01

HUN_yhatt = fitted.values(HUN_lmt)

HUN_yhatt2 = HUN_yhatt^2

HUN_yhatt3 = HUN_yhatt^3

HUN_residualst = residuals(HUN_lmt)

HUN_RRT_t = lm(HUN_residualst ~ HUN_yhatt + HUN_yhatt2 + HUN_yhatt3)

summary(HUN_RRT_t)


# B50

HUN_yhatb = fitted.values(HUN_lmb)

HUN_yhatb2 = HUN_yhatb^2

HUN_yhatb3 = HUN_yhatb^3

HUN_residualsb = residuals(HUN_lmb)
```
```

````
HUN_RRT_b = lm(HUN_residualsb ~ HUN_yhatb + HUN_yhatb2 + HUN_yhatb3)
summary(HUN_RRT_b)

```
# RRT BGR
```{r Data_Org, include=TRUE}
# GINI
BGR_yhat = fitted.values(BGR_lm)
BGR_yhat2 = BGR_yhat^2
BGR_yhat3 = BGR_yhat^3
BGR_residuals = residuals(BGR_lm)
BGR_RRT = lm(BGR_residuals ~ BGR_yhat + BGR_yhat2 + BGR_yhat3)
summary(BGR_RRT)

# T01
BGR_yhatt = fitted.values(BGR_lmt)
BGR_yhatt2 = BGR_yhatt^2
BGR_yhatt3 = BGR_yhatt^3
BGR_residualst = residuals(BGR_lmt)
BGR_RRT_t = lm(BGR_residualst ~ BGR_yhatt + BGR_yhatt2 + BGR_yhatt3)
summary(BGR_RRT_t)

# B50
BGR_yhatb = fitted.values(BGR_lmb)
````

~~~
BGR_yhatb2 = BGR_yhatb^2

BGR_yhatb3 = BGR_yhatb^3

BGR_residualsb = residuals(BGR_lmb)

BGR_RRT_b = lm(BGR_residualsb ~ BGR_yhatb + BGR_yhatb2 + BGR_yhatb3)

summary(BGR_RRT_b)

```

# RRT czE

```{r Data_Org, include=TRUE}

# GINI

CZE_yhat = fitted.values(CZE_lm)

CZE_yhat2 = CZE_yhat^2

CZE_yhat3 = CZE_yhat^3

CZE_residuals = residuals(CZE_lm)

CZE_RRT = lm(CZE_residuals ~ CZE_yhat + CZE_yhat2 + CZE_yhat3)

summary(CZE_RRT)


# T01

CZE_yhatt = fitted.values(CZE_lmt)

CZE_yhatt2 = CZE_yhatt^2

CZE_yhatt3 = CZE_yhatt^3

CZE_residualst = residuals(CZE_lmt)

CZE_RRT_t = lm(CZE_residualst ~ CZE_yhatt + CZE_yhatt2 + CZE_yhatt3)

summary(CZE_RRT_t)
~~~

````
# B50
CZE_yhatb = fitted.values(CZE_lmb)
CZE_yhatb2 = CZE_yhatb^2
CZE_yhatb3 = CZE_yhatb^3
CZE_residualsb = residuals(CZE_lmb)
CZE_RRT_b = lm(CZE_residualsb ~ CZE_yhatb + CZE_yhatb2 + CZE_yhatb3)
summary(CZE_RRT_b)
```
# RRT POL
```{r Data_Org, include=TRUE}
# GINI
POL_yhat = fitted.values(POL_lm)
POL_yhat2 = POL_yhat^2
POL_yhat3 = POL_yhat^3
POL_residuals = residuals(POL_lm)
POL_RRT = lm(POL_residuals ~ POL_yhat + POL_yhat2 + POL_yhat3)
summary(POL_RRT)

# T01
POL_yhatt = fitted.values(POL_lmt)
POL_yhatt2 = POL_yhatt^2
POL_yhatt3 = POL_yhatt^3
POL_residualst = residuals(POL_lmt)
POL_RRT_t = lm(POL_residualst ~ POL_yhatt + POL_yhatt2 + POL_yhatt3)
````

summary(POL_RRT_t)

B50

POL_yhatb = fitted.values(POL_lmb)

POL_yhatb2 = POL_yhatb^2

POL_yhatb3 = POL_yhatb^3

POL_residualsb = residuals(POL_lmb)

POL_RRT_b = lm(POL_residualsb ~ POL_yhatb + POL_yhatb2 + POL_yhatb3)

summary(POL_RRT_b)

```

# RRT ROU

```{r Data_Org, include=TRUE}

# GINI

ROU_yhat = fitted.values(ROU_lm)

ROU_yhat2 = ROU_yhat^2

ROU_yhat3 = ROU_yhat^3

ROU_residuals = residuals(ROU_lm)

ROU_RRT = lm(ROU_residuals ~ ROU_yhat + ROU_yhat2 + ROU_yhat3)

summary(ROU_RRT)

# T01

ROU_yhatt = fitted.values(ROU_lmt)

ROU_yhatt2 = ROU_yhatt^2

ROU_yhatt3 = ROU_yhatt^3

~~~
ROU_residualst = residuals(ROU_lmt)

ROU_RRT_t = lm(ROU_residualst ~ ROU_yhatt + ROU_yhatt2 + ROU_yhatt3)

summary(ROU_RRT_t)


# B50

ROU_yhatb = fitted.values(ROU_lmb)

ROU_yhatb2 = ROU_yhatb^2

ROU_yhatb3 = ROU_yhatb^3

ROU_residualsb = residuals(ROU_lmb)

ROU_RRT_b = lm(ROU_residualsb ~ ROU_yhatb + ROU_yhatb2 + ROU_yhatb3)

summary(ROU_RRT_b)

```

RRT SVK

```{r Data_Org, include=TRUE}

# GINI

SVK_yhat = fitted.values(SVK_lm)

SVK_yhat2 = SVK_yhat^2

SVK_yhat3 = SVK_yhat^3

SVK_residuals = residuals(SVK_lm)

SVK_RRT = lm(SVK_residuals ~ SVK_yhat + SVK_yhat2 + SVK_yhat3)

summary(SVK_RRT)


# T01

SVK_yhatt = fitted.values(SVK_lmt)
~~~

```
SVK_yhatt2 = SVK_yhatt^2

SVK_yhatt3 = SVK_yhatt^3

SVK_residualst = residuals(SVK_lmt)

SVK_RRT_t = lm(SVK_residualst ~ SVK_yhatt + SVK_yhatt2 + SVK_yhatt3)

summary(SVK_RRT_t)

B50

SVK_yhatb = fitted.values(SVK_lmb)

SVK_yhatb2 = SVK_yhatb^2

SVK_yhatb3 = SVK_yhatb^3

SVK_residualsb = residuals(SVK_lmb)

SVK_RRT_b = lm(SVK_residualsb ~ SVK_yhatb + SVK_yhatb2 + SVK_yhatb3)

summary(SVK_RRT_b)
```

Appendix B: All regression models

```
Call:
lm(formula = HUNTime_and_Y$HUN_GINI ~ HUN_GGX + HUN_NGDP + HUN_TMGRPCH +
    HUN_TXRPCH + HUN_NIDNGDP + HUN_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0169951 -0.0043361 -0.0001175  0.0051210
       Max
 0.0206267

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.3242365  0.0876916   3.697
HUN_GGX     -0.0002165  0.0017498  -0.124
HUN_NGDP    -0.0001102  0.0014722  -0.075
HUN_TMGRPCH -0.0005805  0.0006884  -0.843
HUN_TXRPCH   0.0004564  0.0005155   0.885
HUN_NIDNGDP  0.0021634  0.0012762   1.695
HUN_KAOPEN   0.0273812  0.0033182   8.252
            Pr(>|t|)
(Intercept)  0.00179 **
HUN_GGX      0.90298
HUN_NGDP     0.94121
HUN_TMGRPCH  0.41081
HUN_TXRPCH   0.38832
HUN_NIDNGDP  0.10826
HUN_KAOPEN  2.39e-07 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01002 on 17 degrees of freedom
Multiple R-squared:  0.9164,    Adjusted R-squared:  0.8869
F-statistic: 31.05 on 6 and 17 DF,  p-value: 2.95e-08


Call:
lm(formula = HUNTime_and_B$HUN_B50 ~ HUN_GGX + HUN_NGDP + HUN_TMGRPCH +
    HUN_TXRPCH + HUN_NIDNGDP + HUN_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0142799 -0.0025270  0.0000528  0.0036764
       Max
 0.0121640

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.2478790  0.0637602   3.888
HUN_GGX      0.0006892  0.0012722   0.542
HUN_NGDP    -0.0000185  0.0010704  -0.017
HUN_TMGRPCH  0.0001411  0.0005005   0.282
HUN_TXRPCH  -0.0001080  0.0003748  -0.288
HUN_NIDNGDP -0.0005634  0.0009279  -0.607
HUN_KAOPEN  -0.0162800  0.0024126  -6.748
            Pr(>|t|)
(Intercept)  0.00118 **
HUN_GGX      0.59505
HUN_NGDP     0.98641
HUN_TMGRPCH  0.78140
HUN_TXRPCH   0.77672
HUN_NIDNGDP  0.55172
HUN_KAOPEN  3.41e-06 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.007289 on 17 degrees of freedom
Multiple R-squared:  0.8864,    Adjusted R-squared:  0.8463
F-statistic: 22.11 on 6 and 17 DF,  p-value: 3.767e-07


Call:
lm(formula = HUNTime_and_T$HUN_T01 ~ HUN_GGX + HUN_NGDP + HUN_TMGRPCH +
    HUN_TXRPCH + HUN_NIDNGDP + HUN_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0113918 -0.0026331  0.0001781  0.0029273
```

Figure 16: Regressions on hungary, from top to bottom: Gini, Bottom 50%, Top1%

```
Call:
lm(formula = BGRTime_and_Y$BGR_GINI ~ BGR_GGX + BGR_NGDP + BGR_TMGRPCH +
    BGR_TXRPCH + BGR_NIDNGDP + BGR_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.026263 -0.011029 -0.004409  0.011000
      Max
 0.040718

Coefficients:
             Estimate Std. Error t value
(Intercept)  0.128326   0.178857   0.717
BGR_GGX      0.010419   0.004378   2.380
BGR_NGDP    -0.001401   0.004167  -0.336
BGR_TMGRPCH  0.001756   0.001211   1.450
BGR_TXRPCH  -0.001546   0.001270  -1.217
BGR_NIDNGDP -0.000720   0.001755  -0.410
BGR_KAOPEN   0.008475   0.005035   1.683
            Pr(>|t|)
(Intercept)   0.4868
BGR_GGX       0.0348 *
BGR_NGDP      0.7426
BGR_TMGRPCH   0.1727
BGR_TXRPCH    0.2471
BGR_NIDNGDP   0.6889
BGR_KAOPEN    0.1181
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02269 on 12 degrees of freedom
Multiple R-squared:  0.5919,    Adjusted R-squared:  0.3878
F-statistic: 2.901 on 6 and 12 DF,  p-value: 0.05502


Call:
lm(formula = BGRTime_and_B$BGR_B50 ~ BGR_GGX + BGR_NGDP + BGR_TMGRPCH +
    BGR_TXRPCH + BGR_NIDNGDP + BGR_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0172906 -0.0058324  0.0007841  0.0056005
       Max
 0.0143685

Coefficients:
              Estimate Std. Error t value
(Intercept)  3.410e-01  8.287e-02   4.115
BGR_GGX     -4.397e-03  2.029e-03  -2.167
BGR_NGDP     5.334e-05  1.930e-03   0.028
BGR_TMGRPCH -7.404e-04  5.610e-04  -1.320
BGR_TXRPCH   7.802e-04  5.886e-04   1.326
BGR_NIDNGDP  2.707e-04  8.132e-04   0.333
BGR_KAOPEN  -4.029e-03  2.333e-03  -1.727
            Pr(>|t|)
(Intercept)  0.00143 **
BGR_GGX      0.05103 .
BGR_NGDP     0.97841
BGR_TMGRPCH  0.21157
BGR_TXRPCH   0.20967
BGR_NIDNGDP  0.74500
BGR_KAOPEN   0.10975
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01051 on 12 degrees of freedom
Multiple R-squared:  0.5425,    Adjusted R-squared:  0.3137
F-statistic: 2.371 on 6 and 12 DF,  p-value: 0.09573


Call:
lm(formula = BGRTime_and_T$BGR_T01 ~ BGR_GGX + BGR_NGDP + BGR_TMGRPCH +
    BGR_TXRPCH + BGR_NIDNGDP + BGR_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.028506 -0.013144 -0.006195  0.009191
```

```
Call:
lm(formula = BGRTime_and_T$BGR_T01 ~ BGR_GGX + BGR_NGDP + BGR_TMGRPCH +
    BGR_TXRPCH + BGR_NIDNGDP + BGR_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.028506 -0.013144 -0.006195  0.009191
      Max
 0.038595

Coefficients:
              Estimate Std. Error t value
(Intercept) -0.1707321  0.1734208  -0.984
BGR_GGX      0.0092365  0.0042453   2.176
BGR_NGDP    -0.0027300  0.0040399  -0.676
BGR_TMGRPCH  0.0014739  0.0011741   1.255
BGR_TXRPCH  -0.0009942  0.0012317  -0.807
BGR_NIDNGDP -0.0012471  0.0017018  -0.733
BGR_KAOPEN   0.0075907  0.0048816   1.555
            Pr(>|t|)
(Intercept)   0.3443
BGR_GGX       0.0503 .
BGR_NGDP      0.5120
BGR_TMGRPCH   0.2332
BGR_TXRPCH    0.4353
BGR_NIDNGDP   0.4778
BGR_KAOPEN    0.1459
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.022 on 12 degrees of freedom
Multiple R-squared:  0.6195,    Adjusted R-squared:  0.4292
F-statistic: 3.256 on 6 and 12 DF,  p-value: 0.03878
```

Figure 17: Bulgarian regressions, Top to bottom: Gini, Bottom 50%, Top1%

```
Call:
lm(formula = CZETime_and_Y$CZE_GINI ~ CZE_GGX + CZE_NGDP + CZE_TMGRPCH +
    CZE_TXRPCH + CZE_NIDNGDP)

Residuals:
       Min         1Q     Median         3Q
-0.0089079 -0.0060645 -0.0000172  0.0022302
       Max
 0.0209516

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.3864185  0.0403415   9.579
CZE_GGX     -0.0003974  0.0009128  -0.435
CZE_NGDP     0.0007173  0.0011254   0.637
CZE_TMGRPCH -0.0015002  0.0008393  -1.787
CZE_TXRPCH   0.0012377  0.0006472   1.912
CZE_NIDNGDP  0.0005525  0.0007512   0.735
            Pr(>|t|)
(Intercept)  2.9e-08 ***
CZE_GGX       0.6688
CZE_NGDP      0.5324
CZE_TMGRPCH   0.0917 .
CZE_TXRPCH    0.0728 .
CZE_NIDNGDP   0.4721
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.008228 on 17 degrees of freedom
Multiple R-squared:  0.1975,    Adjusted R-squared:  -0.03857
F-statistic: 0.8366 on 5 and 17 DF,  p-value: 0.5418


Call:
lm(formula = CZETime_and_B$HUN_B50 ~ CZE_GGX + CZE_NGDP + CZE_TMGRPCH +
    CZE_TXRPCH + CZE_NIDNGDP + CZE_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.008559 -0.003472  0.001530  0.003536
      Max
 0.006573

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.1463208  0.0274803   5.325
CZE_GGX      0.0018853  0.0005826   3.236
CZE_NGDP    -0.0011777  0.0008147  -1.446
CZE_TMGRPCH  0.0007039  0.0005233   1.345
CZE_TXRPCH  -0.0002587  0.0004025  -0.643
CZE_NIDNGDP  0.0012960  0.0006809   1.903
CZE_KAOPEN  -0.0121261  0.0016854  -7.195
            Pr(>|t|)
(Intercept) 6.84e-05 ***
CZE_GGX      0.00517 **
CZE_NGDP     0.16759
CZE_TMGRPCH  0.19731
CZE_TXRPCH   0.52950
CZE_NIDNGDP  0.07515 .
CZE_KAOPEN  2.13e-06 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005111 on 16 degrees of freedom
Multiple R-squared:  0.9308,    Adjusted R-squared:  0.9049
F-statistic: 35.87 on 6 and 16 DF,  p-value: 2.083e-08
```

Figure 18: Czech regression models, top to bottom, Gini , Bottom 50%

```
Call:
lm(formula = CZETime_and_T$CZE_T01 ~ CZE_GGX + CZE_NGDP + CZE_TMGRPCH +
    CZE_TXRPCH + CZE_NIDNGDP + CZE_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0084337 -0.0043358  0.0007635  0.0016283
       Max
 0.0198254

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.0858505  0.0382577   2.244
CZE_GGX     -0.0008730  0.0008111  -1.076
CZE_NGDP    -0.0007279  0.0011342  -0.642
CZE_TMGRPCH -0.0006315  0.0007285  -0.867
CZE_TXRPCH   0.0008408  0.0005604   1.501
CZE_NIDNGDP  0.0018779  0.0009480   1.981
CZE_KAOPEN   0.0028856  0.0023464   1.230
            Pr(>|t|)
(Intercept)   0.0393 *
CZE_GGX       0.2977
CZE_NGDP      0.5301
CZE_TMGRPCH   0.3988
CZE_TXRPCH    0.1529
CZE_NIDNGDP   0.0651 .
CZE_KAOPEN    0.2365
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.007115 on 16 degrees of freedom
Multiple R-squared:  0.2993,    Adjusted R-squared:  0.03658
F-statistic: 1.139 on 6 and 16 DF,  p-value: 0.3846
```

Figure 19: Czech regression model on Top1%

```
Call:
lm(formula = POLTime_and_Y$POL_GINI ~ POL_GGX + POL_NGDP + POL_TMGRPCH +
    POL_TXRPCH + POL_NIDNGDP + POL_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.013574 -0.007384 -0.003243  0.005647
      Max
 0.016750

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.5225205  0.1004901   5.200
POL_GGX     -0.0006914  0.0017313  -0.399
POL_NGDP     0.0054029  0.0027775   1.945
POL_TMGRPCH -0.0017664  0.0006997  -2.525
POL_TXRPCH   0.0012569  0.0007167   1.754
POL_NIDNGDP -0.0018725  0.0018399  -1.018
POL_KAOPEN   0.0208999  0.0041843   4.995
            Pr(>|t|)
(Intercept) 8.77e-05 ***
POL_GGX     0.694936
POL_NGDP    0.069538 .
POL_TMGRPCH 0.022531 *
POL_TXRPCH  0.098608 .
POL_NIDNGDP 0.323963
POL_KAOPEN  0.000132 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01069 on 16 degrees of freedom
Multiple R-squared:  0.8158,	Adjusted R-squared:  0.7467
F-statistic: 11.81 on 6 and 16 DF,  p-value: 4.174e-05


Call:
lm(formula = POLTime_and_B$POL_B50 ~ POL_GGX + POL_NGDP + POL_TMGRPCH +
    POL_TXRPCH + POL_NIDNGDP + POL_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.009302 -0.002322  0.001234  0.003294
      Max
 0.007646

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.1736984  0.0525862   3.303
POL_GGX      0.0003450  0.0009060   0.381
POL_NGDP    -0.0027768  0.0014535  -1.910
POL_TMGRPCH  0.0009699  0.0003661   2.649
POL_TXRPCH  -0.0006897  0.0003751  -1.839
POL_NIDNGDP  0.0010525  0.0009628   1.093
POL_KAOPEN  -0.0103530  0.0021896  -4.728
            Pr(>|t|)
(Intercept) 0.004490 **
POL_GGX     0.708390
POL_NGDP    0.074165 .
POL_TMGRPCH 0.017503 *
POL_TXRPCH  0.084571 .
POL_NIDNGDP 0.290506
POL_KAOPEN  0.000227 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005595 on 16 degrees of freedom
Multiple R-squared:  0.8093,	Adjusted R-squared:  0.7378
F-statistic: 11.32 on 6 and 16 DF,  p-value: 5.435e-05
```

Figure 20: Polish regression models for GINi (top) and Bottom 50% (bottom)

```
Call:
lm(formula = POLTime_and_T$POL_T01 ~ POL_GGX + POL_NGDP + POL_TMGRPCH +
    POL_TXRPCH + POL_NIDNGDP + POL_KAOPEN)

Residuals:
      Min         1Q    Median        3Q
-0.022086 -0.006398 -0.003148  0.007108
      Max
 0.020518

Coefficients:
              Estimate Std. Error t value
(Intercept)  1.374e-01  1.103e-01   1.245
POL_GGX     -9.842e-05  1.901e-03  -0.052
POL_NGDP     6.181e-03  3.049e-03   2.027
POL_TMGRPCH -1.029e-03  7.682e-04  -1.339
POL_TXRPCH   7.189e-04  7.869e-04   0.914
POL_NIDNGDP -9.686e-04  2.020e-03  -0.479
POL_KAOPEN   2.120e-02  4.594e-03   4.616
            Pr(>|t|)
(Intercept) 0.231060
POL_GGX     0.959347
POL_NGDP    0.059679 .
POL_TMGRPCH 0.199233
POL_TXRPCH  0.374473
POL_NIDNGDP 0.638085
POL_KAOPEN  0.000286 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01174 on 16 degrees of freedom
Multiple R-squared:  0.7329,	Adjusted R-squared:  0.6328
F-statistic: 7.319 on 6 and 16 DF,  p-value: 0.000678
```

Figure 21: Polish Top1% regression model

```
Call:
lm(formula = ROUTime_and_Y$ROU_GINI ~ ROU_GGX + ROU_NGDP + ROU_TMGRPCH +
    ROU_TXRPCH + ROU_NIDNGDP + ROU_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0165587 -0.0034871 -0.0003285  0.0031001
       Max
 0.0207678

Coefficients:
              Estimate Std. Error t value
(Intercept)  5.938e-01  7.127e-02   8.331
ROU_GGX     -7.950e-03  3.238e-03  -2.455
ROU_NGDP    -9.903e-05  1.690e-03  -0.059
ROU_TMGRPCH -3.904e-04  4.757e-04  -0.821
ROU_TXRPCH   1.421e-03  8.052e-04   1.765
ROU_NIDNGDP  7.055e-03  2.219e-03   3.180
ROU_KAOPEN   1.132e-02  3.035e-03   3.729
            Pr(>|t|)
(Intercept) 2.48e-06 ***
ROU_GGX      0.03030 *
ROU_NGDP     0.95425
ROU_TMGRPCH  0.42781
ROU_TXRPCH   0.10297
ROU_NIDNGDP  0.00792 **
ROU_KAOPEN   0.00288 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01114 on 12 degrees of freedom
Multiple R-squared:  0.865,     Adjusted R-squared:  0.7975
F-statistic: 12.81 on 6 and 12 DF,  p-value: 0.0001327


Call:
lm(formula = ROUTime_and_B$ROU_B50 ~ ROU_GGX + ROU_NGDP + ROU_TMGRPCH +
    ROU_TXRPCH + ROU_NIDNGDP + ROU_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0079361 -0.0022151  0.0004261  0.0020434
       Max
 0.0088950

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.1150793  0.0371369   3.099
ROU_GGX      0.0040876  0.0016872   2.423
ROU_NGDP     0.0001823  0.0008809   0.207
ROU_TMGRPCH  0.0001428  0.0002479   0.576
ROU_TXRPCH  -0.0006288  0.0004196  -1.499
ROU_NIDNGDP -0.0028756  0.0011560  -2.488
ROU_KAOPEN  -0.0078729  0.0015816  -4.978
            Pr(>|t|)
(Intercept) 0.009211 **
ROU_GGX     0.032153 *
ROU_NGDP    0.839555
ROU_TMGRPCH 0.575137
ROU_TXRPCH  0.159810
ROU_NIDNGDP 0.028561 *
ROU_KAOPEN  0.000321 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005807 on 12 degrees of freedom
Multiple R-squared:  0.8707,    Adjusted R-squared:  0.8061
F-statistic: 13.47 on 6 and 12 DF,  p-value: 0.0001034
```

Figure 22: Romanian Regression models for GINI, (top) and bottom 50% (bottom)

```
Call:
lm(formula = ROUTime_and_T$ROU_T01 ~ ROU_GGX + ROU_NGDP + ROU_TMGRPCH +
    ROU_TXRPCH + ROU_NIDNGDP + ROU_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0111384 -0.0063188 -0.0004683  0.0037285
       Max
 0.0261453

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.1082327  0.0703253   1.539
ROU_GGX     -0.0036294  0.0031949  -1.136
ROU_NGDP     0.0010431  0.0016681   0.625
ROU_TMGRPCH -0.0005916  0.0004694  -1.260
ROU_TXRPCH   0.0012141  0.0007946   1.528
ROU_NIDNGDP  0.0058698  0.0021891   2.681
ROU_KAOPEN   0.0042834  0.0029951   1.430
            Pr(>|t|)
(Intercept)    0.150
ROU_GGX        0.278
ROU_NGDP       0.543
ROU_TMGRPCH    0.231
ROU_TXRPCH     0.152
ROU_NIDNGDP    0.020 *
ROU_KAOPEN     0.178
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.011 on 12 degrees of freedom
Multiple R-squared:  0.7839,	Adjusted R-squared:  0.6758
F-statistic: 7.254 on 6 and 12 DF,  p-value: 0.001896
```

Figure 23: Romanian regression model on top 1%

```
Call:
lm(formula = SVKTime_and_Y$SVK_GINI ~ SVK_GGX + SVK_NGDP + SVK_TMGRPCH +
    SVK_TXRPCH + SVK_NIDNGDP + SVK_KAOPEN)

Residuals:
       Min         1Q     Median         3Q
-0.0201208 -0.0101150  0.0000113  0.0117545
       Max
 0.0201764

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.4665098  0.0919737   5.072
SVK_GGX     -0.0023513  0.0017115  -1.374
SVK_NGDP    -0.0041765  0.0022491  -1.857
SVK_TMGRPCH  0.0002893  0.0007827   0.370
SVK_TXRPCH   0.0008567  0.0008274   1.035
SVK_NIDNGDP  0.0013835  0.0015618   0.886
SVK_KAOPEN   0.0020729  0.0085754   0.242
            Pr(>|t|)
(Intercept) 0.000138 ***
SVK_GGX     0.189679
SVK_NGDP    0.083066 .
SVK_TMGRPCH 0.716797
SVK_TXRPCH  0.316851
SVK_NIDNGDP 0.389701
SVK_KAOPEN  0.812270
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0148 on 15 degrees of freedom
Multiple R-squared:  0.3034,	Adjusted R-squared:  0.02481
F-statistic: 1.089 on 6 and 15 DF,  p-value: 0.4124


Call:
lm(formula = SVKTime_and_B$SVK_B50 ~ SVK_GGX + SVK_NGDP + SVK_TMGRPCH +
    SVK_TXRPCH + SVK_NIDNGDP + SVK_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.012443 -0.003838  0.000324  0.004077
      Max
 0.009793

Coefficients:
              Estimate Std. Error t value
(Intercept)  1.858e-01  4.528e-02   4.103
SVK_GGX      1.418e-03  8.426e-04   1.683
SVK_NGDP     1.588e-03  1.107e-03   1.435
SVK_TMGRPCH -9.346e-05  3.853e-04  -0.243
SVK_TXRPCH  -4.492e-04  4.073e-04  -1.103
SVK_NIDNGDP -3.278e-04  7.689e-04  -0.426
SVK_KAOPEN   2.167e-03  4.222e-03   0.513
            Pr(>|t|)
(Intercept)  0.00094 ***
SVK_GGX      0.11311
SVK_NGDP     0.17194
SVK_TMGRPCH  0.81163
SVK_TXRPCH   0.28749
SVK_NIDNGDP  0.67590
SVK_KAOPEN   0.61524
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.007287 on 15 degrees of freedom
Multiple R-squared:  0.3456,	Adjusted R-squared:  0.08379
F-statistic:  1.32 on 6 and 15 DF,  p-value: 0.3077
```

Figure 24: Slovakian regression, Gini (top), bottom 50% (bottom)

```
Call:
lm(formula = SVKTime_and_T$SVK_T01 ~ SVK_GGX + SVK_NGDP + SVK_TMGRPCH +
    SVK_TXRPCH + SVK_NIDNGDP + SVK_KAOPEN)

Residuals:
      Min        1Q    Median        3Q
-0.021039 -0.007868  0.002335  0.007628
      Max
 0.016600

Coefficients:
              Estimate Std. Error t value
(Intercept)  0.1640462  0.0769810   2.131
SVK_GGX     -0.0018653  0.0014325  -1.302
SVK_NGDP    -0.0039344  0.0018825  -2.090
SVK_TMGRPCH  0.0004686  0.0006551   0.715
SVK_TXRPCH   0.0003042  0.0006925   0.439
SVK_NIDNGDP  0.0004550  0.0013072   0.348
SVK_KAOPEN   0.0031808  0.0071776   0.443
            Pr(>|t|)
(Intercept)   0.0500 .
SVK_GGX       0.2125
SVK_NGDP      0.0541 .
SVK_TMGRPCH   0.4854
SVK_TXRPCH    0.6668
SVK_NIDNGDP   0.7326
SVK_KAOPEN    0.6640
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01239 on 15 degrees of freedom
Multiple R-squared:  0.4229,	Adjusted R-squared:  0.192
F-statistic: 1.832 on 6 and 15 DF,  p-value: 0.1599
```

Figure 25: Slovakian regression model on top1%

APPENDIX C: Ramsay Reset tests

```
Call:
lm(formula = HUN_residuals ~ HUN_yhat + HUN_yhat2 + HUN_yhat3)

Residuals:
      Min        1Q    Median        3Q
-0.015687 -0.005400  0.001340  0.005121
      Max
 0.019606

Coefficients:
            Estimate Std. Error t value
(Intercept)   -3.257      5.828  -0.559
HUN_yhat      25.472     45.494   0.560
HUN_yhat2    -66.189    118.099  -0.560
HUN_yhat3     57.140    101.945   0.560
            Pr(>|t|)
(Intercept)    0.583
HUN_yhat       0.582
HUN_yhat2      0.581
HUN_yhat3      0.581

Residual standard error: 0.00917 on 20 degrees of freedom
Multiple R-squared:  0.01547,	Adjusted R-squared:  -0.1322
F-statistic: 0.1047 on 3 and 20 DF,  p-value: 0.9563


Call:
lm(formula = HUN_residualst ~ HUN_yhatt + HUN_yhatt2 + HUN_yhatt3)

Residuals:
       Min         1Q     Median         3Q
-0.0106675 -0.0015899  0.0001633  0.0034848
       Max
 0.0089631

Coefficients:
             Estimate Std. Error t value
(Intercept)   -0.6708     0.7206  -0.931
HUN_yhatt     20.5569    21.3951   0.961
HUN_yhatt2  -208.1220   210.3090  -0.990
HUN_yhatt3   696.1194   685.0815   1.016
            Pr(>|t|)
(Intercept)    0.363
HUN_yhatt      0.348
HUN_yhatt2     0.334
HUN_yhatt3     0.322

Residual standard error: 0.004824 on 20 degrees of freedom
Multiple R-squared:  0.06197,	Adjusted R-squared:  -0.07874
F-statistic: 0.4404 on 3 and 20 DF,  p-value: 0.7266


Call:
lm(formula = HUN_residualsb ~ HUN_yhatb + HUN_yhatb2 + HUN_yhatb3)

Residuals:
      Min        1Q    Median        3Q
-0.013281 -0.003699 -0.001092  0.004284
      Max
 0.010462

Coefficients:
            Estimate Std. Error t value
(Intercept)   -7.547      6.032  -1.251
HUN_yhatb     87.673     70.724   1.240
HUN_yhatb2  -338.281    275.678  -1.227
HUN_yhatb3   433.594    357.328   1.213
            Pr(>|t|)
(Intercept)    0.225
HUN_yhatb      0.229
HUN_yhatb2     0.234
HUN_yhatb3     0.239

Residual standard error: 0.006435 on 20 degrees of freedom
Multiple R-squared:  0.08288,	Adjusted R-squared:  -0.05469
F-statistic: 0.6025 on 3 and 20 DF,  p-value: 0.621
```

Figure 26: Hungarian RRT on (GINI, Top 1 and Bottom 50)

```
Call:
lm(formula = BGR_residuals ~ BGR_yhat + BGR_yhat2 + BGR_yhat3)

Residuals:
      Min        1Q    Median        3Q 
-0.025563 -0.010548  0.000187  0.009887 
      Max 
 0.037104 

Coefficients:
            Estimate Std. Error t value
(Intercept)   119.78      68.46   1.749
BGR_yhat     -726.83     417.52  -1.741
BGR_yhat2    1468.47     847.91   1.732
BGR_yhat3    -987.80     573.43  -1.723
            Pr(>|t|)
(Intercept)    0.101
BGR_yhat       0.102
BGR_yhat2      0.104
BGR_yhat3      0.106

Residual standard error: 0.01826 on 15 degrees of freedom
Multiple R-squared:  0.1899,	Adjusted R-squared:  0.02783 
F-statistic: 1.172 on 3 and 15 DF,  p-value: 0.3534


Call:
lm(formula = BGR_residualst ~ BGR_yhatt + BGR_yhatt2 + BGR_yhatt3)

Residuals:
      Min        1Q    Median        3Q 
-0.025857 -0.014236  0.002214  0.006617 
      Max 
 0.032871 

Coefficients:
            Estimate Std. Error t value
(Intercept)    2.065      1.578   1.309
BGR_yhatt    -47.406     36.185  -1.310
BGR_yhatt2   356.870    272.495   1.310
BGR_yhatt3  -881.615    674.454  -1.307
            Pr(>|t|)
(Intercept)    0.210
BGR_yhatt      0.210
BGR_yhatt2     0.210
BGR_yhatt3     0.211

Residual standard error: 0.01864 on 15 degrees of freedom
Multiple R-squared:  0.1027,	Adjusted R-squared:  -0.07677 
F-statistic: 0.5722 on 3 and 15 DF,  p-value: 0.6419


Call:
lm(formula = BGR_residualsb ~ BGR_yhatb + BGR_yhatb2 + BGR_yhatb3)

Residuals:
       Min         1Q     Median         3Q 
-0.0163694 -0.0055697  0.0006542  0.0048481 
       Max 
 0.0141976 

Coefficients:
            Estimate Std. Error t value
(Intercept)     23.8       21.6   1.102
BGR_yhatb     -389.8      350.0  -1.114
BGR_yhatb2    2124.5     1888.6   1.125
BGR_yhatb3   -3853.0     3392.2  -1.136
            Pr(>|t|)
(Intercept)    0.288
BGR_yhatb      0.283
BGR_yhatb2     0.278
BGR_yhatb3     0.274

Residual standard error: 0.008865 on 15 degrees of freedom
Multiple R-squared:  0.1108,	Adjusted R-squared:  -0.06701 
F-statistic: 0.6232 on 3 and 15 DF,  p-value: 0.6109
```

Figure 27: Bulgarian RRT: Gini, Top1, Bottom 50

```
Call:
lm(formula = CZE_residuals ~ CZE_yhat + CZE_yhat2 + CZE_yhat3)

Residuals:
       Min         1Q     Median         3Q 
-0.0094547 -0.0047390 -0.0003313  0.0020145 
       Max 
 0.0205581 

Coefficients:
            Estimate Std. Error t value
(Intercept)   -228.4     1284.8  -0.178
CZE_yhat      1757.6     9976.1   0.176
CZE_yhat2    -4507.3    25819.8  -0.175
CZE_yhat3     3852.5    22273.7   0.173
            Pr(>|t|)
(Intercept)    0.861
CZE_yhat       0.862
CZE_yhat2      0.863
CZE_yhat3      0.865

Residual standard error: 0.007737 on 19 degrees of freedom
Multiple R-squared:  0.0116,	Adjusted R-squared:  -0.1445 
F-statistic: 0.07432 on 3 and 19 DF,  p-value: 0.9731


Call:
lm(formula = CZE_residualst ~ CZE_yhatt + CZE_yhatt2 + CZE_yhatt3)

Residuals:
       Min         1Q     Median         3Q 
-0.0089846 -0.0031805 -0.0007797  0.0028221 
       Max 
 0.0167013 

Coefficients:
             Estimate Std. Error t value
(Intercept)     28.48      37.94   0.751
CZE_yhatt     -750.46    1033.89  -0.726
CZE_yhatt2    6579.66    9385.91   0.701
CZE_yhatt3  -19192.45   28384.63  -0.676
            Pr(>|t|)
(Intercept)    0.462
CZE_yhatt      0.477
CZE_yhatt2     0.492
CZE_yhatt3     0.507

Residual standard error: 0.005601 on 19 degrees of freedom
Multiple R-squared:  0.2642,	Adjusted R-squared:  0.148 
F-statistic: 2.274 on 3 and 19 DF,  p-value: 0.1129


Call:
lm(formula = CZE_residualsb ~ CZE_yhatb + CZE_yhatb2 + CZE_yhatb3)

Residuals:
      Min        1Q    Median        3Q 
-0.006130 -0.003440  0.000379  0.003723 
      Max 
 0.005760 

Coefficients:
            Estimate Std. Error t value
(Intercept)   -8.813      5.938  -1.484
CZE_yhatb    109.013     72.166   1.511
CZE_yhatb2  -448.402    291.789  -1.537
CZE_yhatb3   613.201    392.480   1.562
            Pr(>|t|)
(Intercept)    0.154
CZE_yhatb      0.147
CZE_yhatb2     0.141
CZE_yhatb3     0.135

Residual standard error: 0.004201 on 19 degrees of freedom
Multiple R-squared:  0.1977,	Adjusted R-squared:  0.07105 
F-statistic: 1.561 on 3 and 19 DF,  p-value: 0.2317
```

Figure 28: CZE RRT: Gini, Top1, Bottom 50

```
Call:
lm(formula = POL_residuals ~ POL_yhat + POL_yhat2 + POL_yhat3)

Residuals:
      Min        1Q    Median        3Q 
-0.009882 -0.004395  0.000547  0.003114 
      Max 
 0.010128 

Coefficients:
            Estimate Std. Error t value
(Intercept)    26.31      17.82   1.476
POL_yhat     -180.13     116.67  -1.544
POL_yhat2     409.80     254.30   1.611
POL_yhat3    -309.81     184.56  -1.679
            Pr(>|t|)
(Intercept)    0.156
POL_yhat       0.139
POL_yhat2      0.124
POL_yhat3      0.110

Residual standard error: 0.006386 on 19 degrees of freedom
Multiple R-squared:  0.5764,	Adjusted R-squared:  0.5096 
F-statistic: 8.619 on 3 and 19 DF,  p-value: 0.00081


Call:
lm(formula = POL_residualst ~ POL_yhatt + POL_yhatt2 + POL_yhatt3)

Residuals:
       Min         1Q     Median         3Q 
-0.0232058 -0.0052324  0.0007486  0.0049401 
       Max 
 0.0174575 

Coefficients:
            Estimate Std. Error t value
(Intercept)    1.423      1.472   0.967
POL_yhatt    -35.584     34.741  -1.024
POL_yhatt2   292.767    270.631   1.082
POL_yhatt3  -792.439    696.223  -1.138
            Pr(>|t|)
(Intercept)    0.346
POL_yhatt      0.319
POL_yhatt2     0.293
POL_yhatt3     0.269

Residual standard error: 0.009605 on 19 degrees of freedom
Multiple R-squared:  0.2051,	Adjusted R-squared:  0.07954 
F-statistic: 1.634 on 3 and 19 DF,  p-value: 0.215


Call:
lm(formula = POL_residualsb ~ POL_yhatb + POL_yhatb2 + POL_yhatb3)

Residuals:
       Min         1Q     Median         3Q 
-0.0061500 -0.0024958  0.0002596  0.0028083 
       Max 
 0.0044993 

Coefficients:
             Estimate Std. Error t value
(Intercept)    13.690      6.298   2.174
POL_yhatb    -190.598     90.904  -2.097
POL_yhatb2    881.883    436.803   2.019
POL_yhatb3  -1355.986    698.682  -1.941
            Pr(>|t|)  
(Intercept)   0.0426 *
POL_yhatb     0.0496 *
POL_yhatb2    0.0578 .
POL_yhatb3    0.0673 .
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.003245 on 19 degrees of freedom
Multiple R-squared:  0.6006,	Adjusted R-squared:  0.5375 
F-statistic: 9.523 on 3 and 19 DF,  p-value: 0.0004719
```

Figure 29: Polish RRT: GINI, Top1, Bottom 50

```
Call:
lm(formula = ROU_residuals ~ ROU_yhat + ROU_yhat2 + ROU_yhat3)

Residuals:
       Min         1Q     Median         3Q 
-0.0161623 -0.0034658 -0.0005152  0.0027080 
       Max 
 0.0209097 

Coefficients:
            Estimate Std. Error t value
(Intercept)   -6.164     17.473  -0.353
ROU_yhat      36.356    101.987   0.356
ROU_yhat2    -71.299    197.995  -0.360
ROU_yhat3     46.498    127.865   0.364
            Pr(>|t|)
(Intercept)    0.729
ROU_yhat       0.726
ROU_yhat2      0.724
ROU_yhat3      0.721

Residual standard error: 0.009907 on 15 degrees of freedom
Multiple R-squared:  0.01229,	Adjusted R-squared:  -0.1853 
F-statistic: 0.06221 on 3 and 15 DF,  p-value: 0.979


Call:
lm(formula = ROU_residualst ~ ROU_yhatt + ROU_yhatt2 + ROU_yhatt3)

Residuals:
       Min         1Q     Median         3Q 
-0.0110718 -0.0062861 -0.0008422  0.0037176 
       Max 
 0.0266455 

Coefficients:
            Estimate Std. Error t value
(Intercept)   0.1397     0.6859   0.204
ROU_yhatt    -2.8336    14.3951  -0.197
ROU_yhatt2   18.8262    99.1865   0.190
ROU_yhatt3  -40.9929   224.4176  -0.183
            Pr(>|t|)
(Intercept)    0.841
ROU_yhatt      0.847
ROU_yhatt2     0.852
ROU_yhatt3     0.858

Residual standard error: 0.009817 on 15 degrees of freedom
Multiple R-squared:  0.003837,	Adjusted R-squared:  -0.1954 
F-statistic: 0.01926 on 3 and 15 DF,  p-value: 0.9962


Call:
lm(formula = ROU_residualsb ~ ROU_yhatb + ROU_yhatb2 + ROU_yhatb3)

Residuals:
      Min        1Q    Median        3Q 
-0.008270 -0.001716  0.001313  0.002634 
      Max 
 0.008102 

Coefficients:
            Estimate Std. Error t value
(Intercept)   -1.732      2.883  -0.601
ROU_yhatb     29.590     50.159   0.590
ROU_yhatb2  -167.666    289.641  -0.579
ROU_yhatb3   315.073    554.972   0.568
            Pr(>|t|)
(Intercept)    0.557
ROU_yhatb      0.564
ROU_yhatb2     0.571
ROU_yhatb3     0.579

Residual standard error: 0.005097 on 15 degrees of freedom
Multiple R-squared:  0.03716,	Adjusted R-squared:  -0.1554 
F-statistic: 0.193 on 3 and 15 DF,  p-value: 0.8996
```

Figure 30: Romanian RRT; GINI, Top1, Bottom 50

```
Call:
lm(formula = SVK_residuals ~ SVK_yhat + SVK_yhat2 + SVK_yhat3)

Residuals:
      Min        1Q    Median        3Q
-0.020340 -0.008652  0.002003  0.008353
      Max
 0.022788

Coefficients:
            Estimate Std. Error t value
(Intercept)    305.8      197.3   1.550
SVK_yhat     -2290.0     1477.0  -1.550
SVK_yhat2     5713.9     3684.7   1.551
SVK_yhat3    -4750.5     3063.2  -1.551
            Pr(>|t|)
(Intercept)    0.139
SVK_yhat       0.138
SVK_yhat2      0.138
SVK_yhat3      0.138

Residual standard error: 0.01269 on 18 degrees of freedom
Multiple R-squared:  0.1179,    Adjusted R-squared:  -0.02914
F-statistic: 0.8018 on 3 and 18 DF,  p-value: 0.509


Call:
lm(formula = SVK_residualst ~ SVK_yhatt + SVK_yhatt2 + SVK_yhatt3)

Residuals:
      Min        1Q    Median        3Q
-0.018673 -0.007913  0.002237  0.005261
      Max
 0.018778

Coefficients:
             Estimate Std. Error t value
(Intercept)    -4.137      4.105  -1.008
SVK_yhatt     140.711    139.563   1.008
SVK_yhatt2  -1583.900   1570.788  -1.008
SVK_yhatt3   5903.517   5856.357   1.008
            Pr(>|t|)
(Intercept)    0.327
SVK_yhatt      0.327
SVK_yhatt2     0.327
SVK_yhatt3     0.327

Residual standard error: 0.011 on 18 degrees of freedom
Multiple R-squared:  0.05347,   Adjusted R-squared:  -0.1043
F-statistic: 0.3389 on 3 and 18 DF,  p-value: 0.7974


Call:
lm(formula = SVK_residualsb ~ SVK_yhatb + SVK_yhatb2 + SVK_yhatb3)

Residuals:
       Min         1Q     Median         3Q
-0.0073532 -0.0043764  0.0003658  0.0041817
       Max
 0.0095367

Coefficients:
            Estimate Std. Error t value
(Intercept)    324.3      113.1   2.868
SVK_yhatb    -4129.5     1435.9  -2.876
SVK_yhatb2   17522.1     6077.0   2.883
SVK_yhatb3  -24773.5     8570.5  -2.891
            Pr(>|t|)
(Intercept)  0.01022 *
SVK_yhatb    0.01005 *
SVK_yhatb2   0.00989 **
SVK_yhatb3   0.00974 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0054 on 18 degrees of freedom
Multiple R-squared:  0.3409,    Adjusted R-squared:  0.2311
F-statistic: 3.104 on 3 and 18 DF,  p-value: 0.05261
```

Figure 31: SVK RRT: GINI, top 1, Bottom 50