UNIVERSIDADE DA CORUÑA

Faculty of Economics and Business Administration

Bachelor's Dissertation

Income Inequality and Economic Growth

Author: José Javier Caloca Martínez

Tutor: Fernando Bruna

Bachelor’s Degree in Economics

Year 2020

Bacherlor’s dissertation presented at the Faculty of Economics and Business Administration of the University of A Coruña for obtaining the Bacherlor’s degree in Economics

Acknowledgements

This dissertation became a reality with the support and help of many individuals. I want to especially thank Prof. Dr. Manuel Alberto Gómez Suárez, Prof. Dr. Gustavo Rego and the dean´s administration team for giving me the opportunity to demonstrate to myself that I could graduate a year in advance.

First and foremost, I would like to express my gratitude to my dissertation tutor Prof. Dr. Fernando Bruna for guiding me, providing me with academic support and constant supervision, and also for his patience which enabled the completion of this endeavour.

Also, I would like to express my heartfelt gratitude to my mother Ruth Martínez and the rest of my family, for all the support and the encouragement received during the writing of this dissertation.

Last but not least, I am deeply indebted with Isabella de Oliveira for her invaluable help in preparing this dissertation.

Abstract

The effects of income inequality on economic growth and its relationship have been widely studied in recent decades. Most of the studies using cross-sectional data demonstrate a negative relationship between income inequality and economic growth.

In this paper, we will study this relationship in the context of an extended neoclassical growth model of Mankiw, Romer and Weil (1992), following the baseline empirical model as described by De Dominicis et al. (2008). Using cross-sectional data we make estimations on a world sample of 94 countries for the period 1985 – 2017, which is divided into subsamples according to their level of income (high, middle and low). In order to do this, using the programming software R we built a data bank from the World Bank Database, Penn World Tables 9.1 and the Standardized World Income Inequality Database 8.2, and performed the whole econometric analysis of this paper.

Some of our explanatory variables, such as the level of investment and initial income per capita resulted to be highly significant. This result is robust in all our subsamples and consistent with the empirical literature. However, regarding the relationship of income inequality and economic growth, we obtain a positive and statistically significant relationship between income inequality and economic growth in low-income countries, and a negative but statistically insignificant relationship for high and middle-income countries.

Keywords: Income inequality, economic growth, Gini coefficient, cross-sectional data, long-term effects.

Number of words: 14.784

Resumen

Los efectos de la desigualdad de ingresos en el crecimiento económico y su relación han sido ampliamente estudiados en las últimas décadas. La mayoría de los estudios que utilizan datos transversales demuestran una relación negativa entre la desigualdad de ingresos y el crecimiento económico.

En este artículo, se estudia esta relación en el contexto de un modelo de crecimiento neoclásico extendido de Mankiw, Romer y Weil (1992), siguiendo el modelo empírico de referencia descrito por De Dominicis et al. (2008) Utilizando datos transversales, se realizan estimaciones en una muestra mundial de 94 países para el período 1985 - 2017, que se divide en submuestras según su nivel de ingresos (alto, medio y bajo). Para hacer esto, utilizando el software de programación R, se ha creado un banco de datos a partir de la Base de datos del Banco Mundial, las Penn World Tables 9.1 y the Standardized World Income Inequality Database 8.2, asi como el análisis econométrico completo de este estudio.

Algunas de nuestras variables explicativas, como el nivel de inversión y el ingreso inicial per cápita resultaron ser altamente significativas. Este resultado es robusto en todas nuestras submuestras y consistente con la literatura empírica. Sin embargo, con respecto a la relación entre la desigualdad de ingresos y el crecimiento económico, obtenemos una relación positiva y estadísticamente significativa entre la desigualdad de ingresos y el crecimiento económico en los países de bajos ingresos, y una relación negativa pero estadísticamente insignificante para los países de ingresos altos y medianos.

Palabras clave: Desigualdad de ingresos, coeficiente de Gini, datos transversales, efectos a largo plazo.

Número de palabras: 14.784

Table of Contents

List of Figures

List of Tables

Introduction

Economic growth has been widely used as a measure of countries' economic health and progress, and the effects of other phenomena on economic growth is an important issue in macroeconomics. In the past few decades, income inequality has been increasing substantially worldwide Saez (2020). Policymakers have shown an enormous interest to assess the effects that income inequality may have on economic growth in order to provide the best solutions during crises (Piketty, 2015). However, a debate has arisen when explaining the channels in which income inequality affects economic growth providing a large number of theories with ambiguous predictions.

Income inequality is said to be detrimental to growth due to distortions made by governments through redistributive policies and high-income tax to the rich (Perotti, 1996; Alesina and Rodrik 1994). In addition, inefficient state bureaucracy and institutions impact on economic growth and this problem is exacerbated by an increase in income inequality (Acemoglu, 2007; Acemoglu et al., 2011). Moreover, Galor and Zang (1997) and Aghion et al. (1999) proposed that income inequality affects negatively on economic growth by hampering the access to education of the less favoured due to imperfect capital markets and affecting human capital formation in a country. Additionally, income inequality leads to political instability due to the increase of social problems and therefore it impacts negatively on economic growth as there will not be incentives to invest (Alesina and Perotti, 1996). Nevertheless, other economists claim that income inequality promotes growth as it aids savings to grow among the rich and they can afford large and expensive investments (Kaldor, 1957). In addition, Galor and Tsiddon (1997) claimed that income inequality promotes R&D, for which Foellmi and Zweimuller (2008) suggested that income inequality promotes an increase in technology which leads to economic growth.

The empirical evidence also provides mixed results. Initial papers in the 1990s are based on the estimation of cross-section growth regressions inspired by the growth model of Mankiw, Romer and Weil (1992) in which the variable inequality is added to a set of control explanatory variables. On the basis of this approach, studies such as

Alesina and Rodrik (1994), Deininger and Squire (1998), Persson and Tabellini (1991) among others, provided a piece of robust evidence for a negative and statistically significant relationship between income inequality and economic growth. Nevertheless, more recent studies using panel data models have found evidence of a positive and strong relationship between income inequality and economic growth (Székely and Hilgert 1999; Forbes, 2000; Panizza 2002; Castelló 2004). In this sense, there are mixed results theoretically and empirically speaking as no general consensus has emerged so far.

The aim of this paper is to analyse the effects of income distribution on economic growth in a cross-country setting. In order to do that, we run a cross-sectional model based on the traditional empirical literature with updated data that covers 94 countries for the time span from 1985 to 2017.

Our model uses data from three different databases: The World Bank database (Arel-Bundock, 2019), the Penn World Tables 9.1 (Feenstra et al., 2015) and the Standardized World Income Inequality Database 8.2 (Solt, 2019). These three databases have been merged using programming tools in the software R. In a first stage we imported the databases in R. Secondly, we tidied up the data with the Tidyverse package (Wickham et al., 2019) and a set of programming functions, so it can be manipulated and transformed as the literature suggests. Thirdly, by the usage of our empirical model, we were able to estimate and visualise results in R. In this regard, it should be noted that part of the effort of this dissertation, was dedicated to learning this professional programming language for data science. Although learning how to programming in R has a steep learning curve, it provided the necessary tools for our model estimations, tests and data visualisation for large datasets such as the ones used in this paper.

For our sample, countries are selected based on data availability for the chosen years. In this sense, for the purpose of providing a broader analysis, our world sample will be disaggregated into three subsamples in which countries will be classified by income level: high, middle and low. We estimate our model using the method of Ordinary Least Square (OLS), and we test the hypotheses of this method in order to find potential limitations on estimates. To carry out this analysis, cross-sectional regressions will be run for the aforementioned period in all samples. The first regression will attempt to replicate the extended growth model of Mankiw, Romer and Weil (1992). Subsequently, a second regression will be run for the same variables plus income inequality variable. Furthermore, a third regression will be run adding control variables. Finally, a fourth regression is dedicated to a robustness analysis of the model.

We find consistent results in our first regression in all samples, our results are expected and aligned with the obtained by Mankiw, Romer and Weil (1992). It reveals that initial input per capita and the level of investment statistically impacted on economic growth. Regarding the outstanding regression outputs, from one side, we find a positive relationship between income inequality and economic growth in low-income countries, to the other, there's a negative but statistically insignificant relationship in middle and high-income countries.

The rest of the paper is structured as follows: Chapter 1 will be dedicated to reviewing the theoretical and empirical literature on income inequality and economic growth. Chapter 2 presents the methodology, data, programming tools used in R and the econometric model used in this paper. Chapter 3 provides descriptive statistics for all samples used for the analysis. In chapter 4 a global cross-country analysis is based on the estimates of our model. Chapters 5 will present the second stage of the analysis, performing an individual analysis on the selected groups by income level. The empirical results of this research will be analysed and discussed in chapter 6. Chapter 7 offers the conclusions of this research and will present the main suggestions for future research in this field. Finally, the programming code and packages used in R, for the estimation of our model are explained in Appendix B.

1. Theoretical framework and empirical literature

1.1 Theoretical background of income inequality and economic growth

Over the years, part of the literature in this field has attempted to measure the impact of income inequality on economic growth. Some studies suggest that inequality negatively affects growth, however, others disagree, considering inequality to be a *conditio sine qua non* for the economic growth (De Dominicis et al., 2008)

Early investigations showed a positive relationship between income inequality and economic growth. In this regard, Lewis (1954) concluded that the majority of entrepreneurs tend to save much more of their earnings than other groups in the economy, therefore, the existence of inequality is positively correltated with the generating of savings among the rich. Lewis proposed a relationship between the growth of the GDP per capita and the saving rate of a country, which infers that with an increase in the saving rate of a country, the amount of investment will increase which can lead to economic growth.

Kuznets (1955) laid the groundwork for future research on inequality and started the debate on whether economic growth affects the level of inequality. Kuznets defined the “long swing” as a pattern that countries follow during the transition from a rural to an industrial economy. He claimed that inequality increases as the average income increases along with economic growth, reaching a peak where it starts to decline as the average income keeps increasing along with economic growth, giving an inverted U-shaped curve as a result. Several studies have been debating Kuznets´s ideas over the years. It can be said that there are two groups, one which agrees with Kuznets´s ideas that economic growth affects income inequality, whilst a second group suggest that income inequality affects economic growth, an opposite argument to that which Kuznets proposed (Barro, 1991). The approach of the literature in this paper will focus on this second group of studies that empirically studied the effects of income inequality on economic growth.

Later on, Kaldor (1957) studied the proportion of marginal propensity to save in different countries and realised that those countries with high-income inequality and higher marginal propensity to save will grow faster. He concluded that redistribution policies to reduce income inequality through progressive taxes will decrease the amount of disposable income amongst the rich and, thereby, their capacity to save and invest in a country, leading to a decrease in economic growth.

Evidently, the rich play an essential role in the creation of wealth and economic growth. Although the existence of these groups generates inequality, governments must face a challenge through redistributive policies in order to make societies more equal and reduce the collateral effects of income inequality. In response to this, Okun et al. (2015) first introduced the metaphor in his first book in 1975 to refer to the transfers that the governments make from the rich to the poor (redistributive policies) as a “leaky bucket” because there’s a loss of resources in the redistribution process. Additionally, Okun states that goverment transfers discourage people from making any addtional contributions to the economy, which reduces its efficiency and leads to a decrease in economic growth.

1.2 Theories regarding income inequality and economic growth

More recent theories have arisen exposing the existence of a positive relationship between income inequality and economic growth, stating that large investments require a huge proportion of savings, and as efficiency in saving money amongst the rich is higher, it is better for the wealth to be concentrated amongst a small percentage of the population, so they can afford these large investments but with a lower marginal return in comparison to the less favoured (Aghion, 1999). It is important to mention that there’s a positive relationship between technological change and economic growth. To this, there’s historical evidence that time periods with wide income inequality, reported a greater amount of technological inventions, and these time periods were also caracterised by having an increase in investment, thereby an increase of the economic growth is conceived (Galor and Tsiddon 1997).

Some studies propose an approach linked to the taxation system and how the government finances its consumption. Li and Zou (1998) used an econometric model proposed by Alesina and Rodrik (1994) making some corrections to the methodology and using a panel dataset. They based this empirical study on cross-sectional and panel data, and their main findings concentrate on the ambiguity between income inequality and economic growth and their correlation, although sometimes it can be negative or positive. For all the aforementioned theories, there’s not a direct correlation between income inequality and economic growth but a kind of relationship relating economic growth to other economic factors or variables, such as the rate of savings, redistribution policies, credit availability, gross fixed capital formation, fertility rate, school enrolment, technological change, etc.

As previously mentioned, income inequality and economic growth can also be negatively correlated, this focuses on the second line of arguments of this relationship. Income inequality is perceived as an obstacle in the development and economic growth of countries. De Ferranti et al. (2004) Explains that inequality can be harmful as it severely affects the income of the people, increasing poverty. Also, it limits the access to credit and it shrinks all possible investment and opportunities that poorer individuals could take. Moreover, it also reduces the possibilities of some to access education, limiting potential contributions that talented people could make to society and the economy. Additionally, Ferranti mentions the distributional conflicts that may arise when the economy is facing a downturn due to an adverse shock, he highlights that inequality provokes an increase in crime and violence and institutions become weaker in the protection of property rights. Hence, as inequality increases, economic growth tends to decrease.

According to the effects that inequality generates on growth explained by De Ferranti et al. (2004), there are four main approaches based on theoretical models supporting these effects. In this way, the first approach is supported on Perotti (1996) which proposes a model which exposes the effects of fiscal policies when goverments create distortions in the economy by redistributing resources to the less favoured, these distortions are created when progressive taxes are applied and there’s high goverment spending which in some cases both variables are negatively correlated with economic growth. In addition, Alesina and Rodrik (1994) found evidence of a negative correlation between income inequality and growth when adding to the regression model a control variable (government spending), they have reported that in the presence of high-income inequality in developed countries, high-income taxes are imposed with the aim of making society more equal, and in this manner, it generates a decrease in economic growth. Essentially, policies that only focus on the increase of economic growth, are optimal for governments that take care of the capitalist solely. An example of this is in the United

States where Panizza (2002) evidenced a negative relationship between income inequality and economic growth using a cross-state panel.

On the other hand, the second approach refers to the results obtained by Alesina and Perotti (1996) which focuses on the problems of socio-political instability, inferring that inequality leads to an increase in crime, protests and coup d'etats. This increases the level of uncertainty in a country, reducing investmens and discouraging capital accumulation which also affects economic growth negatively.

The third approach referes to the existing inverse correlation between birthrates and disposable income in households, Galor and Zang (1997) developed this model by analysing the effects of fertility on disposable income, they found that countries with a high fertility rate present a high level of inequality as there's less disposable income to invest in education which in turn, negatively impacts growth.

Finally, the last approach was developed by Aghion et al. (1999) focuses on the performance of capital markets, as it is a good indicator caofusality of income inequality. Imperfect capital markets influence the ability of the poor to invest, this group, in particular, is characterised by having a high return when investing in human capital, if capital markets do not allocate resources efficiently, then, it will stop the poor from investing in education, health and from running businesses. As a consequence, this lack of investment in human capital will negatively affect economic growth.

Stiglitz (2012) has shown how the lower classes in the United States have seen their income to shrink and stagnate. The American system is based on a model that efficiently makes money flow from the lower and middle class to the upper class but not the other way around. Stiglitz discusses the idea of the cost of inequality, he claims that as inequality increases, the inequality of opportunities declines. For instance, intelligent but poor kids are less likely to finish their university studies than rich children with bad results. He suggested that income inequality increases poverty and compromises the household's consumption, slowing down economic growth.

In general, most of the studies using cross-sectional data tend to suggest a negative relationship between income inequality and economic growth. On the contrary, those which use panel data tend to have a positive relationship. Although, it has been previously mentioned that the relationship between these variables is ambiguous, thus, problems regarding the existing disparities within econometric results may arise due to different methods of measuring income inequality (Panizza, 2002).

This dispute has been addressed with a different approach by Galor and Moav (2004), they proposed a unified theory of inequality and growth. Unlike the previously mentioned theories, Galor and Moav propose one that explained economic growth according to the degree of development of a country. In phase one, for countries which haven't finished their process of industrialisation, the main engine of growth is the accumulation of physical capital rather than human capital, and this accumulation can be achieved thanks to income disparity among individuals. As previously explained, rich people tend to save and invest more and only they could afford the large costs of large investments. In phase two, the main engine of growth is the accumulation of human capital, in order to achieve this, individuals in a country have to invest in education but capital markets' imperfections can impede this, consequently, government's policies have to be directed towards the reduction of income inequality in order to favour the access to education among the poor.

Galor and Moav (2004) also inferred that nowadays less developed countries, unlike more developed ones may also require the accumulation of human capital in the first stage rather than solely accumulating physical capital. Moreover, the beneficial role of income inequality upon economic growth for less developed countries can be conditioned by the inflows of international capital as they might reduce the encouragement for the rich to accumulate physical capital and enhance growth.

The following table summarises the results obtained by different empirical studies in the field of income inequality and economic growth. It can be appreciated that most of the studies conducted in the field using cross-country models, depict a positive relationship.

Table 1. Findings of the main authors by type of data used in the model

	CROSS-COUNTRY DATA	PANEL DATA
POSITIVE EFFECT	Bleaney and Nishiyama (2004) Castelló (2004) Castelló and Domenech (2002) de la Croix and Doepke (2003) Li and Zou (1998) Partridge (2005) Schipper and Hoogeveen (2005)	Banerjee and Duflo (2003) Castelló (2004) Deininger and Olinto (1998) Forbes (2000) Iradian (2005) Partridge (2005) Szekely and Hilgert (1999) Panizza (2002)
NEGATIVE EFFECT	Alesina and Rodrik (1994) Benjamin et al. (2006) Clarke (1995) Deininger and Squire (1998) Galor and Zang (1997) Gylfason and Zoega (2003) Keefer and Knack (2002) Kenworthy (2004) Khoo and Dennis (1999) Knell (1999) Knowles (2005) Larrain and Vergara (1997) Mbabazi et al. (2001) Panizza (2002) Persson and Tabellini (1991) Rehme (2002) Tanninen (1999) Zhu (2001)	Barro (2000) Benjamin et al. (2006) Litschig (2005) Mbabazi et al. (2001) Odedokun and Round (2001) Panizza (2002) Voitchovsky (2005)

Source: Own elaboration based on De Dominicis et al. (2008).

1.3 Theoretical framework: the Mankiw, Romer and Weil model

In this section we will present the growth model that inspired many studies and researches in the empirical literature. The basic growth model was proposed by Solow (1956) and focuses on the role of saving to increase capital accumulation which leads to economic growth. It also allows for studying the effects of population growth and technical progress. This model assumes: no public sector, two productions factors (labour and capital), full employment of the production factors, amount of production factors and their productivity are exogenous variables, closed economy, competitive good and factor market and unique product (GDP). Using a Cobb-Douglas production function with constant returns to scale in which only 2 inputs and an output level Y were considered.

$$Y = AK^{\alpha}L^{1-\alpha} \tag{1}$$

Solow assumed a constat level of technology represented by A and labour supply L which grows at rate n. Solow explained a model in which the investment in the economy comes from a fraction of the total output, and the accumulated stock of physical capital K depreciates at a constant level of δ. This model explains that in an economy as more capital is used, there's less output due to the diminishing returns of capital. Therefore, if there's less output there's also less investment as it is a fraction of the output. Finally, Solow shows that at a certain point, the level of depreciation equals the level of investment and it is called the steady state. At the steady state level of capital and output, there's zero growth. Makiw, Romer and Weil (1992) extended this neoclassical model by adding human capital as a third production input.

$$Y = AK^{\alpha}H^{\beta}L^{1-\alpha-\beta} \tag{2}$$

This extended model includes a stock level of human capital H, α and β represent the output elasticity with respected to K and H (physical and human capital) respectively. L and A grow exogeneously at n and g rates respectively: $L = L_0e^{nt}$ and $A = A_0e^{gt}$. Mankiw, Romer and Weil (MRW), for the steady state of this model, they supposed s_k and s_h as the fraction of the output dedicated to invest in physical and human capital respectively. Thus, human and physical capital can be represented in effective units of labour, in $y = \frac{Y}{AL}$, $k = \frac{K}{AL}$ and $h = \frac{H}{AL}$ as follows:

$$\dot{k} = s_k y_t - (n + g + \delta)k_t \tag{3.1}$$

$$\dot{h} = s_h y_t - (n + g + \delta)h_t \tag{3.2}$$

In the above system of equations, the existence of dimishing returns to scale implies that $\alpha + \beta < 1$. This system of equations can be solved using the following values of k^* and h^* as follows:

$$k^* = \left(\frac{s_k^{1-\beta} \; s_h^{\beta}}{n + g + \delta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{4.1}$$

$$h^* = \left(\frac{s_k^{\alpha} \; s_h^{1-\alpha}}{n + g + \delta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{4.2}$$

Substituting in the production function (2) the values of k^* and h^* that represent the steady state for the corresponding level of physical and human capital, and taking logs we, can achieve the steady state output in intensive form as a function of investments in human capital s_h, or as a function of the stock of human capital h*. For this paper, in particular, we will be using a structural model taking into account the stock of human capital, with proxied data that will be explained in chapter 2 following Cingano (2014).

$$\ln y^* = \ln A_0 + gt - \frac{\alpha+\beta}{1-a-\beta}\ln(n+g+\delta) + \frac{\alpha}{1-\alpha-\beta}\ln s_k + \frac{\beta}{1-\alpha-\beta}\ln s_h \tag{5}$$

In the above equation, y^* denotes the output per capita in efficiency units in the steady state $y^* = \frac{y}{LA}$. Makiw, Romer and Weil (1992) detailed that the choice between the structural model (4.1) or (4.2) will depend on the availability of data. Hence, the mathematical analysis will be limited to the previous equation. In order to represent the convergence of per capita income to the steady-state (transitional dynamics), let y^* stand for the steady state output in efficiency units and y_t its value in time *t*:

$$\frac{\partial \ln y}{\partial t} = \lambda(\ln \gamma^* - \ln y) \tag{6}$$

Conventionally, in the literature $\lambda = (n+g+\delta)(1-\alpha-\beta)$ represents the rate of convergence. Assuming that $\alpha + \beta < 1$, then we can infer that $\ln y$ gets closer to $\ln \gamma^*$ in an exponential trend. According to the principle of transitional dynamics that predicts a higher growth when using the first units of capital. This can be demonstrated as:

$$\ln y_t - \ln y_0 = \left(1 - \mathrm{e}^{-\lambda t}\right)\ln y^* - \left(1 - \mathrm{e}^{-\lambda t}\right)\ln y_0 \tag{7}$$

Finally, y^* can be substituted from (5) and we have the final equation which was fundamental for the empiral studies of economic growth and laid the groundwork for the estimations of the effects of income inequality on economic growth:

$$\begin{aligned}\ln y_t - \ln y_0 = {} & \left(1 - \mathrm{e}^{-\lambda t}\right)\frac{\alpha}{1-\alpha-\beta}\ln s_k + \left(1 - \mathrm{e}^{-\lambda t}\right)\frac{\beta}{1-\alpha-\beta}\ln s_h \\ & - \left(1 - \mathrm{e}^{-\lambda t}\right)\frac{\alpha+\beta}{1-a-\beta}\ln(n+g+\delta) - \left(1 - \mathrm{e}^{-\lambda t}\right)\ln y_0\end{aligned} \tag{8}$$

The last equation shows that growth is a function of the initial level of income and some other determinants of the steady-state, such as population growth, depreciation rate of the physical capital, technological growth, investment in education and investment in physical capital.

1.4 Measurement issues about income inequality

In order to capture the dispersion of the income distribution, there many inequality measures. Haughton and Khandker (2009); Atuesta et al. (2018); among others. They have remarked that income inequality estimators and indexes should ideally possess the following characteristics:

- Mean independence: If the variable (income) is multiplied for all the individuals by the same scalar, the level of inequality should not change. This implies that inequality can be identified in a relative way, taking as reference the average level of the variable of interest.
- Population size independence: If there's a change in the size of the population, ceteris paribus, the inequality should not change. This allows the results to be compared to others with different population sizes.
- Symmetry: If two individuals swap positions in the income distribution, it should not affect the level of inequality. This feature focuses only on the level of income without taking into consideration the relevance of other characteristics of the individuals.
- Pigou-Dalton principle: Different weights should be applied according to position in the income distribution. The main characterisitc of an inequality index.
- Transfer sensitivity: When there's a progressive transfer between two individuals or households, without changing their position and separated by the same distance in the income distribution, the inequality level will be reduced for the poorest individual or household.
- Decomposability: The index should be additive or decomposable across subgroups. This means that the index should be the sum of the inequality of all subgroups.
- Statistical testability: There should be confidence intervals that can test the significance of the changes in index over time.

There are many ways to measure income inequality, a general way of measuring inequality is to divide the income distribution in quantiles and compute the accumulation of income in these segments, measures of statistical dispersion can also be spotted such as the squared coefficient of variation and the relative mean deviation (Atuesta et al., 2018; Martin-Legendre, 2018). Finally, there are indexes that present an efficient way to calculate income inequality, such as the Gini coefficient, Hoover Index, Theil Index, etc. According to table 2 Theil index, Atkinson class of measures and the Mean log deviation are the best estimators for income inequality.

Table 2. Comparison of the main income inequality estimators.

	Squared coefficient of variation	Relative mean deviation	Gini coefficient	Hoover Index	Theil index	Mean Log Deviation	Atkinson class of measures
	$SCV = \left(\frac{\sigma}{\mu}\right)^2$	$RMD = \frac{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i - \bar{x}\rvert}{\lvert \bar{x}\rvert}$	$G = 1 - \sum_{i=1}^{N}(x_i - x_{i-1})(y_i - y_{i-1})$	$H = \frac{1}{2}\sum_{i=1}^{N}\left\lvert \frac{E_i}{E_{TOTAL}} - \frac{A_i}{A_{TOTAL}} \right\rvert$	$E(\alpha) = \frac{1}{N(\alpha^2 - \alpha)}\sum_{i=1}^{N}\left[\left(\frac{x_i}{\mu}\right)^{(\alpha)} - 1\right]$	$T = \frac{1}{N}\Sigma\frac{x_i}{\mu}\ln\left(\frac{x_i}{\mu}\right)$	$A(\varepsilon) = 1 - \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_1}{\bar{x}}\right)^{1-\varepsilon}\right]^{\left(\frac{1}{1-\varepsilon}\right)}$
Mean independence	X	X	X	X	X	X	X
Population size independence	X	X	X	X	X	X	X
Symmetry	X	X	X	X	X	X	X
Pigou-Dalton principle	X	X	X	X	X	X	X
Transfer sensivility	-	-	X	X	X	X	X
Decomposability	-	-	-	-	X	X	X
Statistical testability	X	X	X	X	X	X	X

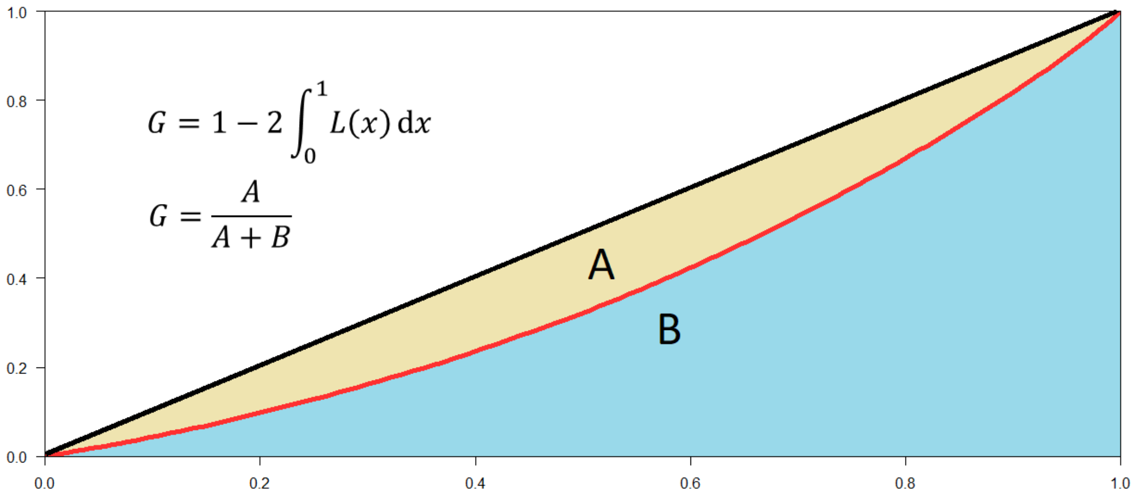
Source: Own elaboration based on Martin-Legendre (2018), Haughton and Khandker (2009) and Atuesta et al. (2018).

1.5 Gini coefficient and Lorenz curve

For the sake of this study, the Gini coefficient is going to be used as a representative measure for income inequality. Today, the Gini coefficient is a widely used measure of income inequality due to the popularity it gained in the past century, and it is the main indicator for inequality used in the empirical literature. However, according to table 2, there are many other indexes which assess a more complete overview of income inequality. In order to move forward and explain the Gini coefficient, it is important to understand the Lorenz curve which is defined as a graphical representation of income distribution, the horizontal axis shows the percentage (portion) of the total population, whereas the vertical axis represents the portion of total income accumulated by the percentage of the total population. In this case, a Lorenz curve of 45º represents the line of total equality. Normally, the Lorenz curve is divided into 10 deciles or five quintiles, and it gives a graphical understanding of income inequality. From the Lorenz curve, the

Gini coefficient can be obtained which is the ratio of the area between the 45º line of perfect equality and the resulting Lorenz curve for the distribution in analysis, and the area of the triangle below the 45 º (Fellman, 2012).

Figure 1. Representation of the Lorenz Curve



Source: Own elaboration using random values for the income distribution.

In other words, the Lorenz curve as a cumulative frequency curve compares the distribution of income with the uniform distribution represented by the 45 degree line. Due to its functionality in terms of comparing regions, over time, regardless of the number of individuals (magnitude of the population) and its ease of interpretation, as it takes values between 0 (perfect equality) and 1 (perfect inequality) the Gini coefficient (index) became the most popular measure of inequality (Martin-Legendre, 2018).

2. Methodology and data

2.1 Methodology

As previously stated, the aim of this paper is to analyse the impact of income inequality on economic growth using a sample of 94 countries selected by the availability of data for the period of 1985 - 2017. Based on the theoretical framework in section 1.3, and emphasising in equation (8) resulting from the extension of Solow's model (1956) by Mankiw, Romer and Weil (1992), the empirical literature since the beginning of the 1990s, has extended this equation by adding control variables that could explain economic growth such as the results presented by Aghion, (2009). Generally, the way to assess the impact of income inequality on economic growth is through a linear relationship. This model can be summarised according to De Dominicis et al. (2008) as follows:

$$\frac{\left(\ln Y_{i,t} - \ln y_{i,t-\tau}\right)}{\tau} = a_0 \ln y_{i,t-\tau} + \alpha_1 \phi_{i,t-\tau} + X_{i_1 t-\tau}\beta + \varepsilon_{i,t} \tag{9}$$

Where *Y(i,t)* is the growth rate of GDP per capita, in country *i* and time *t,* τ is the time span of this model, ϕ stands for the inequality measure to be used which is generally the Gini coefficient and X stands for all the control variables that can be used to explain the economic growth, finally, ε is the error term. This model is widely used when running cross-sectional regressions and it generally reports a negative effect of income inequality to economic growth according to the empirical evidence of Alesina and Rodrik (1994); Panizza (2002); among others that have also used this model. Is important to mention that this model has been criticised when using cross-country data due to omitted, time-invariant variables that can also affect economic growth (Bouincha and Karim, 2018).

Our first model to consider in this analysis is built on Mankiw, Romer and Weil (1992) page 426. In this model, the dependent variable is the log of the difference of GDP per capita in time t,0 which is an approximation of the discrete percentage change in the GDP per capita. Instead, we will be using the average growth rate in 1985-2017 to explain economic growth as follows:

$$\frac{(\ln Y_{2017} - \ln y_{1985})}{2017 - 1985} = \beta_0 + \beta_1 \ln y_{1985} + \beta_2 \ln(n + g + \delta)_{1985-2017} + \beta_3 \ln(S_k)_{1985-2017} + \beta_4 \ln(S_h)_{1985-2017} + \varepsilon_{i,t} \quad (10)$$

In which y_{1985} refers to the initial GDP per capita, S_h stands for the average years of secondary schooling as proxy for human capital and S_k is the gross capital formation as a share of the GDP as proxy for physical capital. Lastly, n stands for population growth and $g + \delta$ which is the exogenous rate of technological progress and the depreciation of the physical capital respectively, is assumed to be 0,05 by Mankiw, Romer and Weil (1992).

Our second model to consider in this analysis is based on the empirical literature. We follow equation (9) described by De Dominicis et al. (2008) and we adapt this model in our sample of 94 countries for the time span from 1985-2017, adding the inequality variable and a set of control variables.

$$\frac{(\ln Y_{2017} - \ln y_{1985})}{2017 - 1985} = \beta_0 + \beta_1 \ln y_{1985} + \beta_2 G_{1985} + \beta X_{1985-2017} + \varepsilon_{i,t} \quad (11)$$

In this case, G_{1985} stands for the intial inequality represented by the Gini coefficient. All the previous determinants of economic growth of equation (10) as S_k, S_h and $n + g + \delta$ are contained in $X_{1985-2017}$ as a matrix of control variables. Additionally, trade openness and price of investment will be also contained in this matrix as relevant control variables according to their usage in the recent studies of Breunig and Majeed (2020); Brueckner and Lederman (2018); Aiyar and Ebeke (2019); among others. Moreover, all the independent variables will be expressed in logs, except for the Gini coefficient which will be expressed in percentage, as in Castelló and Domenech (2002), Berg et al. (2018), Breunig and Majeed (2020) among others. In addition, a robustness analysis will be applied to equation (11) in which some independent variables may no longer be considered if they generate a risk of multicollinearity. We will use the Variance Inflation Factor corresponding to equation (11) as a selection criterion.

Lastly, this analysis is going to be implemented in a general sample of 94 countries, then, these countries will be clustered as per the income level classification of countries according to the World Bank in 2017. In this sense, countries will be classified into 3 income groups: low, middle and high income countries. For each subsample (group) we will apply a regression analysis from equations (10) and (11), plus the corresponding robustness analysis.

The empirical literature has shown a negative effect of income inequality on economic growth (negative relationship) when using cross-sectional models. Thereupon, the hypotheses for this paper are the following:

H0: Income inequality and economic growth have a negative relationship

H1: Income inequality and economic growth have either a positive or no relationship.

The following table reflects the expected impact of the variables in equation (1) and (2) on economic growth:

Table 3. Expected effects of the variables used in the models on economic growth

Variable	Expected sign
Intial GDP per capita	Negative
Gini coefficient	Negative
Average years of secondary schooling	Positive
Gross capital formation	Positive
$n+g+\delta$	Negative
Trade openness	Positive
Price of investment	Negative

2.2 Data

Our dataset is composed of a collection of data for the time span of 1985 - 2017 using a sample of 94 countries. Our data is collected in R using a serie of packages that provide a programmatic access to the databases of the World Bank and the Penn World Table 9.1. In addition, as there's no a package that allows the access to the Standardized World Income Inequality Database (SWIID) 8.2. I have developed a formula in R which enables the access to the data and loads that database in the environment of the software. Lastly, the tidyverse package by Wickham et al. (2019) was mainly used for data wrangling.

In equations (10) and (11) the dependent variable is represented by economic growth. When measuring economic growth, the best indicator is the growth rate of the real gross domestic product (real GDP) (Henderson et al., 2012; Barro, 1996). In order to assess cross-country comparisons, it is necessary for the Real GDP to be presented in the same currency (normally USD) and the same prices. The Real GDP (PPP) per capita, measures the average price-adjusted production per person in a country in order to be compared with many other countries in the world (Callen, 2008; Carbonari, 2011). Therefore, we are interested in considering the growth rate of the Real GDP per capita (PPP) as an indicator to measure economic growth.

For the dependent variable “economic growth”, represented by the average growth rate and the construction of the independent variable “initial income per capita”, represented by the GDP per capita at chained PPPs (2011 USD). We have extracted the data using the R package PWT9 by Zeileis A (2019), this package enables the access to the Penn World Tables 9.1 (Feenstra et al., 2015) in order to acquire data for the Expenditure-side Real GDP at chained PPPs (in millions 2011 USD) and divided it by the population in millions in order to get the Gross Domestic Product per capita (GDPpc) at chained PPPs (2011 USD), using this value for the representation of initial income and also to calculate the average growth rate in accordance with equation (10) and (11). Additionally, from the same database, we have acquired relevant data for the representation of the “price of investment” used as a control variable, proxied as the price level of capital formation (price level of USA GDPo in 2011 = 1) and averaged for the whole time span.

For the outstanding independent variables, we can highlight the usage of the R package WDI (Arel-Bundock, 2019) and Wstats (Piburn, 2018) that enables the access to the World Delovepment Indicators database from the World Bank in order to get the following data for the period 1985-2017:

- Human Capital: represented by the average years of secondary schooling measured as the total average years of secondary education completed by people over the age of 15.
- Physical Capital: represented by the gross fixed capital formation (% of GDP) measured as the weight on GDP of improvements (fences, ditches, drains, and so on); plant, machinery, and equipment purchases; and the construction of roads, railways, and other similar things, including schools, offices, hospitals, private residences, and commercial and industrial buildings.
- Human Capital Depreciation ($\boldsymbol{n}$): population growth (% annual) measured as the exponential rate of growth of the midyear population from year t-1 to t, expressed as a percentage.
- Trade Openness: trade as a percentage of GDP measured as the sum of exports and imports of goods and services measured as a share of GDP.

Our income inequality variable is taken from the SWIID version 8.2 from Solt (2019) which is widely used in the literature on income inequality and economic growth, as it covers a great number of countries and a large time span. In this dataset, we have chosen the net Gini coeficient as a representative which is measured as the Gini coefficient of the income distribution after the payment of taxes. In order to capture as many countries as possible, for the initial period 1985, we have calculated the average of the net Gini coefficient for the period 1985-1990. Since the data, unfortunately, in many cases is not published for 1985. This technique is very common in the economic growth literature of Cingano (2014); Li and Zou (1998); Barro (1996); among others.

2.3 Empiral strategy

In order to offer a quality quantitative study, I have decided to use the software R which is an open-source language and statistical environment useful for statistical computing and graphics (R Core Team, 2020). Therefore, is possible to take advantage of big data and R's supplies of statistical and graphical techniques when conducting empirical studies in the field of social science (Foster et al., 2018).

R is a very powerful tool, however, as it is a programming language, time is required to familiarised oneself with the software, therefore, learning how to use the software from scratch is challenging at the very beginning. For this reason, R has been chosen as the software to perform the statistical and econometrical analysis of the empirical cross-sectional model. Using R, we have created a data bank which combines the databases of Penn world tables, World Bank and the Standardized World Income Inequality Database, these databases contain relevant data and information needed to run our model the literature and provide the accuracy required for estimations. Furthermore, R has a series of packages that allow one to run many econometric tests which provide the level of accuracy in estimations of the model. This can show us how reliable our estimations are, and help us spot problems when estimating and suggest what could be solved by future researchers in this field which might follow the same methodology and model.

In Appendix B we can find the code in R. The first section of this code is dedicated to merging 2 databases using the packages WDI and PWT9 that allow the access to the World Bank and Penn World Tables 9.1 databases respectively. In this section, we selected the variables and changed the data structure so it can be wrangled easily with the tidyverse package. Section 2 is dedicated to loading the SWIID 8.2 database which unlike the aforementioned databases, this one doesn't have a package that allows direct access. However, I have developed a formula that downloads the SWIID database from Solt's GitHub repository and loads it in the environment of R. In section 3 we merge the three databases. Lastly, in section 4 we can find the data wrangling process in which the data is filtered as per the availability of data of countries for the period 1985-2017 and transformed according to the specifications in the methodology for the model.

3. Descriptive analysis

3.1 Descriptive analysis of the general sample

Table 4 provides some descriptive statistics of the dataset. As seen below, the average GDP per capita growth for the whole period is 2% for all the 94 countries in the general sample. Moreover, the net Gini coefficient provides evidence of high-income inequality in 1985 as it is close to 37%, otherwise speaking, 63% of the population share all the income in these countries (in different proportions) and the outstanding 37% gets nothing. In addition, 25% of the sample have a Gini coefficient averaging 43.5%, which is considerably high.

Human capital represents high disparities among countries in the world, reporting a mean of 2.8 years. Regarding Investment as a share of the GDP, we can highlight its importance in many economies as it represents at least 11.6 % of the GDP for the country with the lowest value. In average, $\boldsymbol{n+g+\delta}$ has reported a positive but small value for the vast majority of countries.

Table 4. Descriptive statistics for the whole sample

Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Average GDP per capita growth	94	0.02	0.01	-0.02	0.02	0.03	0.1
GDP per capita 1985	94	10,869	9,063	806	3,179	17,153	36,223
Net Gini 1985	94	36.7	10.0	17.8	28.7	43.5	58.5
Human capital	94	2.8	1.3	0.3	1.7	3.8	6.1
Investment	94	22.4	4.4	11.6	20.2	24.5	36.9
$n+g+\delta$	94	0.1	0.01	0.04	0.1	0.1	0.1
Trade openness	94	81.6	52.7	21.8	50.5	99.8	350.6
Price of investment	94	0.6	0.7	0.3	0.5	0.7	7.2

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

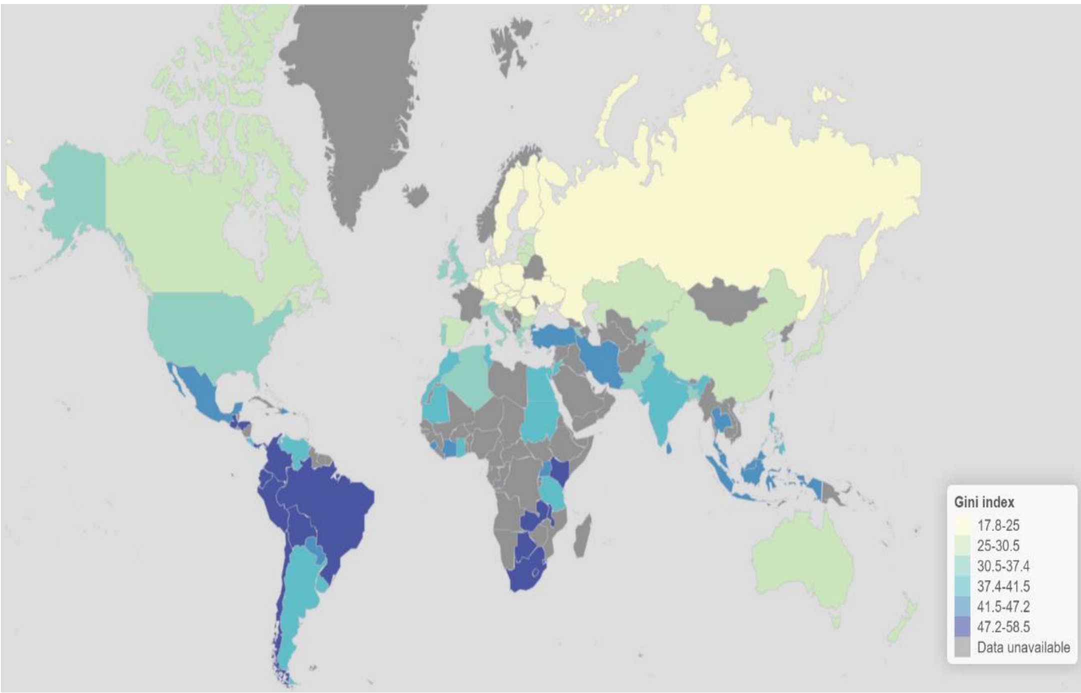
Table 5. Correlation Matrix for the whole sample (logged data)

	Initial GDPpc	Net Gini coefficient	Human Capital	Investment	n+g+δ	Trade openess	Price of investment
GDP pc 1985	1						
Net Gini 1985	-0.640	1					
Human Capital	0.790	-0.560	1				
Investment	0.160	-0.210	0.230	1			
$n + g + \delta$	-0.510	0.600	-0.510	-0.130	1		
Trade openess	0.320	-0.190	0.300	0.240	-0.200	1	
Price of investment	0.480	-0.360	0.330	-0.140	-0.390	0.060	1

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

The above table shows a strong correlation between initial GDPpc and years of schooling and this could be problematic for the outcome of the regression, therefore, this variable will not be taken into account for the robustness analysis of the model. Additionally, there's a moderate correlation between inequality and the initial GDPpc, between $\boldsymbol{n + g + \delta}$ and the net Gini coefficient and between $\boldsymbol{n + g + \delta}$ and years of schooling. Prior to assessing the results of the model, it is important to understand how evident the data of income inequality is, and its evolution for our sample in figure 2.

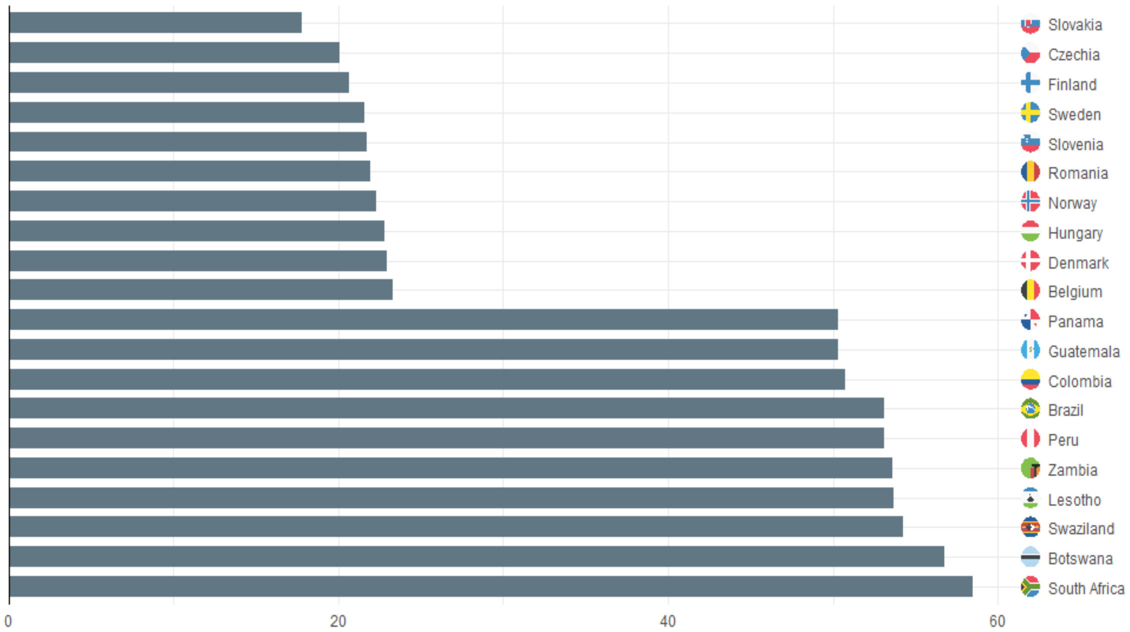
Figure 2. World map of Inequality in sampled countries 1985-1990



Source: own elaboration with data from SWIID 8.2 database

This sample shows high levels of inequality among low-income countries, specifically those in South America, Africa and south-east Asia. When analysing the distribution of income for these countries (GDP per capita) we have compared the values of the tails in this income distribution, to get an idea of how great the disparities are among the sampled countries. Figure 3 shows the 10 most egalitarian countries and the 10 most unequal countries in the sample of 94 countries.

Figure 3. Gini coefficient in 1985: Top and bottom countries.

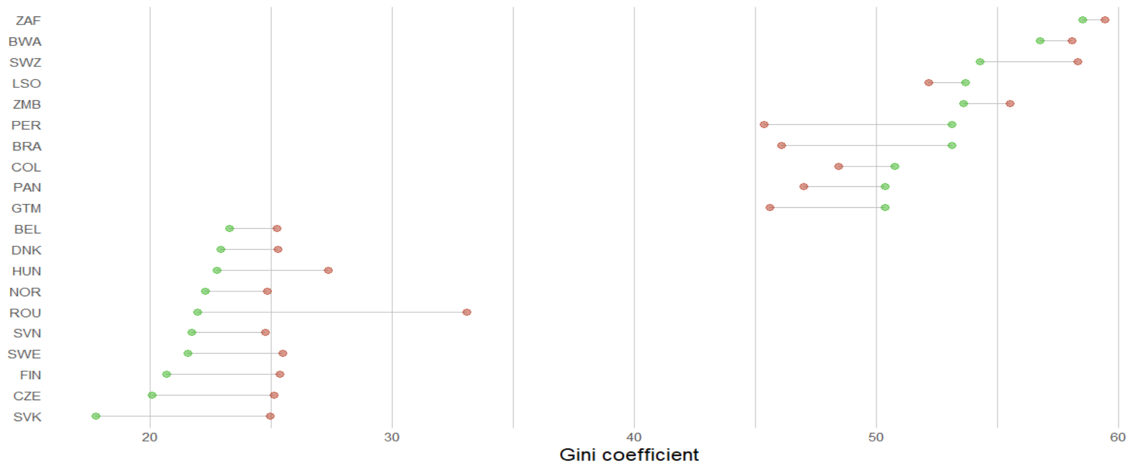


Source: own elaboration with data from SWIID 8.2 database

The above graph shows the inequality gap that exits among countries in the sample. Those with a lower level of inequality were considered to be high-income countries in 2017. Another relevant characteristic is that these high-income countries are all European (see table 13). Some of these countries reporting a low level of inequality in 1985 belonged to the soviet union, and due to political ideological reasons, strong redistributive policies were applied to the population. Lets also focus on Scandinavian countries which managed to lower the level of inequality, due to the effectiveness of the welfare state, this conception can be also be applied to Belgium and Austria. On the other hand, countries reporting a higher level of inequality are those in Latin America and in Africa, many of these countries are politically unstable and few have also suffered problems systematic violence.

The literature stresses the increasing inequality in developed countries in recent decades, we can confirm this issue by averaging the change in the income inequality between the intial period of 1985 and the final period of 2017. The following graphs measure the magnitude of change of the Gini coefficient for the countries in figure 4. The green dots refer to the initial Gini coefficient in 1985-1990, and the red dots represent the final Gini coefficient in 2012-2017.

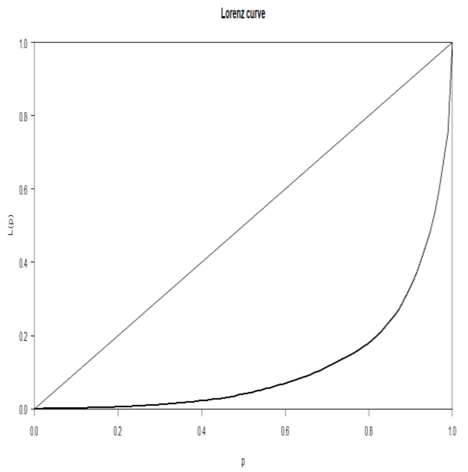
Figure 4. Change in the Gini coefficient 1985-2017: Top and bottom countries.



Source: own elaboration with data from SWIID 8.2 database

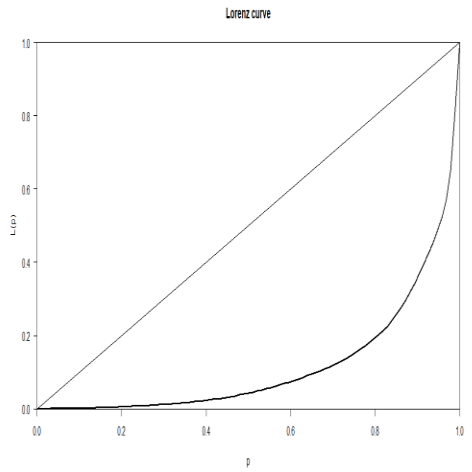
The above graph confirms the evidence in the literature, reflecting the increasing level of inequality in richer countries and a small reduction in income inequality in poor countries, positioning these countries in a margin of high income inequality and squashing the tails of the initial distribution. In order to assess the level of inequality among these countries (spatial inequality), the Gini coefficient for this distribution was calculated, accounting the Real GDP per capita as the income for each of the 94 countries in the sample, reflecting the following level of inequality:

Figure 5. Lorenz curve for the general sample in 1985.



Gini coefficient = 0.773

Figure 6. Lorenz curve for the general sample in 2017



Gini coefficient = 0.764

Source: own elaboration with data from SWIID 8.2 database

As a representative sample, we can infer that income inequality is extremely high worldwide and it has not changed significantly within past few decades. This sample reflects a reduction in inequality of 0.9% in a time span of 32 years. To this, the World Inequality Report by Piketty, et al. (2018) refers that Asian growth contributed to reduce inequality between countries over the past decades, especially, from 1980 - 2018.

3.2 Descriptive analysis of all subsamples

As described in section 3.2 the general sample is clustered into 3 subsamples. Countries can be found in listed in tables 10, 12 and 14 for high, middle and low-income countries respectively. Table 6 provides some descriptive statistics of the subsamples.

Table 6. Descriptive statistics for all sub samples

Descriptive statistics: High-income countries. N=41

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Average GDP per capita growth	0.02	0.01	-0.002	0.02	0.03	0.05
GDP per capita 1985	18,909	7,703	4,998	13,328	22,936	36,223
Net Gini 1985	29.8	7.5	17.8	24.1	32.2	50.4
Human capital	3.6	0.8	2.1	3.0	4.2	5.3
Investment	22.8	3.4	15.8	21.1	24.1	32.2
$n + g + \delta$	0.1	0.01	0.04	0.1	0.1	0.1
Trade openness	97.6	68.7	23.9	57.3	120.8	350.6
Price of investment	0.7	0.1	0.5	0.6	0.8	1.0

Descriptive statistics: Middle-income countries. N=23

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Average GDP per capita growth	0.03	0.01	0.01	0.02	0.04	0.1
GDP per capita 1985	6,622	3,154	2,535	4,182	8,450	14,974
Net Gini 1985	42.4	9.9	22.0	38.0	49.2	58.5
Human capital	2.5	1.0	0.8	1.8	2.9	5.5
Investment	23.0	4.7	15.2	20.1	24.1	36.9
$n + g + \delta$	0.1	0.01	0.04	0.1	0.1	0.1
Trade openness	71.5	34.5	21.8	46.4	95.1	165.5
Price of investment	0.8	1.4	0.3	0.4	0.5	7.2

Descriptive statistics: Low-income countries. N=30

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Average GDP per capita growth	0.02	0.02	-0.02	0.01	0.03	0.1
GDP per capita 1985	3,137	2,547	806	1,300	3,608	10,798
Net Gini 1985	41.8	6.8	25.0	38.7	46.5	54.3
Human capital	1.8	1.5	0.3	1.0	2.1	6.1
Investment	21.6	5.3	11.6	17.7	25.8	32.4
$n + g + \delta$	0.1	0.01	0.05	0.1	0.1	0.1
Trade openness	67.5	29.3	24.9	46.6	87.1	137.4
Price of investment	0.5	0.2	0.3	0.3	0.5	1.2

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

We can observe that middle and low-income countries have in average the largest income inequality, also in the middle-income subsample, there's the country with the largest income inequality, reaching up a Gini coefficient of 58.5. On the other hand, high-income countries have the lowest income inequality. There's a remarkable difference among groups in terms of income per capita.

In terms of population growth, this value is lower in high and middle-income countries as the minimum value is 0.04% and as we stated before, $g + \delta$ is said to be 0.05 according to Mankiw, Romer and Weil (1992). For $n + g + \delta$ a value of 0.04% expresses a value of $n = -0.01\%$. However, for low-income countries, the minimum value of population growth is 0% inferring that there are countries growing at a higher rate.

In average, middle-income countries reported the higher average growth rate, nonetheless, middle and low-income countries (developing countries) have countries which grew their average GDP per capita growth at a rate of 10% contrary to what high-income countries have reported. In page 18, it was mentioned that countries get to the steady state when there's zero growth due to the dimishing returns of capital accumulation. High-income countries slowly are getting to the steady state as the new units of capital report less growth as in low and middle-income countries. Nonetheless, we can state that high-income countries are better off in terms of macroeconomic variables such as human capital, investment, etc.

Table 7 depicts the correlation matrices for all subsamples. In terms of correlation, variables are not correlating as in the general sample, however, for high-income countries there's a moderate correlation between human capital and initial GDP pc, as well as for price of investment and initial GDP pc. For middle-income countries, there's also a moderate correlation between depreciation and Gini and between the price of investment and depreciation. Lastly, for low-income countries, there's a strong correlation between human capital and initial GDP pc and a moderate correlation between depreciation and human capital.

Table 7. Correlation matrices for all sub samples (logged data)

Correlation Matrix: High-income countries. N=41

	gdp85	gini	schooling	I2GDP	popgrowth	trade	priceinv
GDP pc 1985	1						
Net Gini 1985	-0.440	1					
Human Capital	0.660	-0.260	1				
Investment	0.040	-0.240	0.210	1			
$n+g+\delta$	0.170	0.420	0.120	0.050	1		
Trade openess	0.080	0.030	-0.070	0.330	0.260	1	
Price of investment	0.630	-0.250	0.550	-0.130	0.090	-0.280	1

Correlation Matrix: Middle-income countries. N=23

	gdp85	gini	schooling	I2GDP	popgrowth	trade	priceinv
GDP pc 1985	1						
Net Gini 1985	-0.390	1					
Human Capital	0.560	-0.530	1				
Investment	-0.280	-0.240	0.080	1			
$n+g+\delta$	-0.450	0.630	-0.370	-0.050	1		
Trade openess	0.010	-0.170	0.310	0.240	0.110	1	
Price of investment	0.200	-0.280	0.210	-0.130	-0.460	0.150	1

Correlation Matrix: Low-income countries. N=30

	gdp85	gini	schooling	I2GDP	popgrowth	trade	priceinv
GDP pc 1985	1						
Net Gini 1985	-0.490	1					
Human Capital	0.690	-0.580	1				
Investment	0.120	0.010	0.020	1			
$n+g+\delta$	-0.590	0.360	-0.590	-0.190	1		
Trade openess	0.450	0.260	0.380	0.140	-0.340	1	
Price of investment	-0.140	0.030	-0.140	-0.360	0.240	-0.120	1

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 database

4. Estimation for the world sample

4.1 Model estimation

In this section, the estimations of the models (10) and (11) from section 2.1 for the sample of 94 countries will be presented in table 8. These models follow the OLS method and in order to assess the reliability of our econometric models, some tests must be run as a way to detect any potential violations of the OLS assumptions that might lead to biased estimates.

Table 8. Regression output for all the world's sampled countries

	Dependent variable:			
	Average Real GDP per capita growth rate (PPP)			
	(1)	(2)	(3)	(4)
log(GDPpc 1985)	-0.011**	-0.011**	-0.011**	-0.008*
	(0.005)	(0.005)	(0.006)	(0.004)
Gini coefficient 1985		0.0001	0.0001	0.0001
		(0.0002)	(0.0002)	(0.0002)
log(Investment)	0.065***	0.066***	0.071***	0.073***
	(0.015)	(0.015)	(0.016)	(0.016)
log($n+g+\delta$)	-0.010	-0.015	-0.012	-0.017
	(0.023)	(0.025)	(0.025)	(0.025)
log(Human capital)	0.007	0.007	0.008	
	(0.008)	(0.008)	(0.008)	
log(trade Opennes)			-0.004	-0.004
			(0.006)	(0.006)
log(Price of investment)			0.006	0.005
			(0.009)	(0.009)
Constant	-0.033	-0.047	-0.040	-0.060
	(0.033)	(0.043)	(0.044)	(0.039)
Observations	94	94	94	94
R^2	0.222	0.225	0.234	0.226
Adjusted R^2	0.188	0.181	0.172	0.172
Residual Std. Error	0.012 (df = 89)	0.012 (df = 88)	0.012 (df = 86)	0.012 (df = 87)
F Statistic	6.365*** (df = 4; 89)	5.104*** (df = 5; 88)	3.761*** (df = 7; 86)	4.227*** (df = 6; 87)
Note:				*p<0.1; **p<0.05; ***p<0.01

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

4.2 Model analysis

In regards to the regression output of the model when using a sample of 94 countries. Column (1) replicates the extended neoclassical growth model of Mankiw, Romer and Weil (1992), using the same variables, however, the amount of variation in the average Real GDP per capita growth rate (economic growth) explained by the variables investment, human capital and level of depreciation is not relevant as the model has an adjusted R^2 of 0.188, meaning that these variables only explain 18.8% of the variation in economic growth. The variable investment is significant at the 99% level of condifence, and the level of initial income has a statistical significance of 95%. As both variables are in logs, we can infer that a 1 unit increase in the natural log of the gross fixed capital formation increases the average Real GDP per capita growth rate by 0.065, on average, holding all other variables constant. The increase of 1 unit of the natural log of the initial GDP per capita which will also reduce economic growth by -0.011 on average, holding all other variables constant, demonstrating convergence of income per capita in the sample.

Column (2) takes column (1) and adds the variable inequality, which will also be present in columns (3) and (4). We find that there’s a lack of significance in explaining the variation in economic growth due to the low adjusted R^2. Also, the level of inequality can not be representative in explaining the dependent variable as its p-value is high and its coefficient is 0,0001 hardly impacting economic growth in this model. However, as a representative variable, the level of investment and initial income can be highlighted, such as in colum (1). These variables affect them equally but with a proportion of 0.066 and -0.011 respectively.

Column (3) is the reference in the literature of income inequality and economic growth, which was presented by De Dominicis et al. (2008) and takes the initial model of MRW, adding the variable of inequality and some control variables. In this model, by adding more variables, is expected to see an increase in R^2 as the degrees of freedom decrease. However, this didn’t occur, the adjusted R^2 in fact decreased, meaning that all variables in the model are not significant, nor enough to explain the variation in economic growth. Subsequently, the initial income and the level of investment are representative by having p-values less than 0,05 which affects economic growth as in columns (1) and (2) but in different proportions, more specifically, in -0,011 and 0,071 respectively.

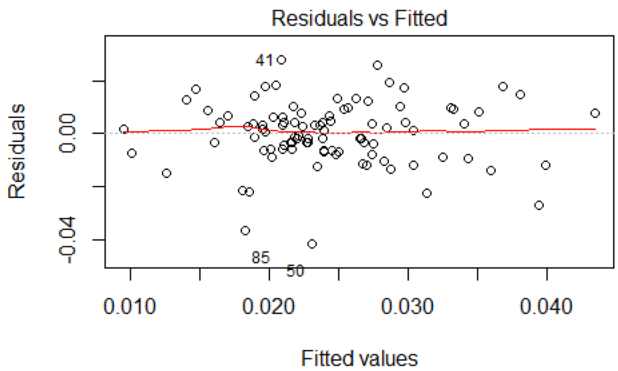
Finally, Column (4) stands for the robustness analysis of column (3), by doing so, we remove the variable of human capital as it might generate multicollinearity problems due to a VIF value close to 4. Moreover, this column has the same pattern of column (1), (2) and (3) in which there's a slightly adjusted R^2. Nonetheless, in this equation, the only representative variable is the level of investment in which we can infer that a 1 unit increase in the natural log of the gross fixed capital formation increases the average Real GDP per capita growth rate by 0.073, on average, holding all other variables constant.

In general, the variables investment and initial GDP per capita depict homogeneous results along the 4 columns. Regarding income inequality it shows a positive gradient close to zero but statistically insignificant.

In order to assess the reliability of the estimates in table 8, the following tests will discuss 4 common assumptions that leads OLS regressions to present biased estimates and weak predictions. All following tests from chapter 4 and 5 will be run for model (11) presented in section 3.1 according to the literature in income inequality and economic growth.

Linearity of the data: The residuals vs. Fitted plot (Figure 7) show a line close to the horizontal axis without any specific pattern. We can assume that there's a linear relationship between the dependent and independent variables.

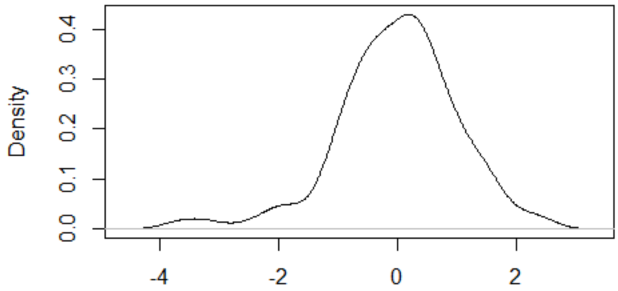
Figure 7. Residuals vs fitted values plot for the general sample



Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

Normality of residuals: We performed a Shapiro-Wilk normality test in order to assess the normality in the residuals, we got the following result W = 0.96366 with a p-value = 0.0103 with 95% confidence, rejecting the null hypothesis of normality of residuals. According to this, there isn´t a normal distribution of residuals and it might affect the result of the standard errors of the OLS estimates, making them less reliable. This result might be due to the existence of outliers, however, we can infer that most of the residuals follow a normal distribution as the density curve of residuals in figure 8 shows, so we can continue with the analysis.

Figure 8. Density curve of residuals for the general sample



Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

Constant variance of residuals (homoscedasticity): We performed a Breusch-Pagan test in order to evaluate the variance of the residuals and to check whether they were homoscedastic. This test computes a score test of the hypothesis of constant error variance (homoscedasticity) against the alternative heteroskedasticity. As a result, we got a Chisquare = 0.3577012, Df = 1, p-value = 0.54979 with 95% confidence. Rejecting the alternative hypothesis and assuming homoscedasticity in the residuals.

No multicollinearity: to check whether two or more variables are strongly correlated we have shown in table 5 the correlation matrix, showing the different Pearson correlation coefficients of the variables in presence of the other. Table 9 reflects the variance inflation factor (VIF) for the equations, this table shows no signs of multicollinearity in the regression, with 4 being the tolerance VIF level (Hair et al., 2010). Variables with VIF values close to 4 are taken off in the robustness analysis. Initial income becomes moderately positive correlated with human capital; hence, human capital is taken off for the robustness analysis (4) in table 9.

Table 9. Variance Inflation Factor for the whole sample

	(1)	(2)	(3)	(4)
log(GDPpc 1985)	2,82	3,19	3,64	2,10
Gini 1985		2,07	2,08	2,08
log(Investment)	1,05	1,07	1,18	1,17
log($n + g + \delta$)	1,40	1,69	1,76	1,70
log(Human capital)	2,88	2,88	2,92	
log(Trade openness)			1,17	1,17
log(Price of investment)			1,47	1,45

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

5. Model estimation by income subsamples

5.1 High-Income countries

This subsample is composed of 41 countries with high income listed in table 10 according to the country classification income level of the World Bank (2017) in which the income threshold is a GNI per capita of 12.235 converted to current U.S. dollars using the World Bank Atlas method.

Table 10. List of high-income countries in the sample

GNI/Capita	Country	GNI/Capita	Country	GNI/Capita	Country
13120	Argentina	44680	Finland	14710	Latvia
51600	Australia	38330	France	47110	Netherlands
45120	Austria	41370	United Kingdom	76210	Norway
42720	Belgium	18340	Greece	38470	New Zealand
15000	Barbados	46420	Hong Kong SAR China	13260	Panama
42960	Canada	12640	Croatia	12730	Poland
81120	Switzerland	13080	Hungary	20040	Portugal
13290	Chile	53050	Ireland	54200	Singapore
24580	Cyprus	37420	Israel	16650	Slovakia
17970	Czechia	31340	Italy	22090	Slovenia
43640	Germany	38470	Japan	52850	Sweden
56340	Denmark	28380	South Korea	14900	Uruguay
27040	Spain	15240	Lithuania	59030	United States
18690	Estonia	66380	Luxembourg		

Source: own elaboration with data from World Bank database

5.1.1. Model estimation

In this section, the estimations of the model for this particular sample are presented in table 11. Additionally, in the following section, the analysis for the model will be presented and discussed with the OLS assumpition.

Table 11. Regression output for high-income sampled countries

	Dependent variable:			
	Average Real GDP per capita growth rate (PPP)			
	(1)	(2)	(3)	(4)
log(GDPpc 1985)	-0.032***	-0.040***	-0.045***	-0.042***
	(0.008)	(0.009)	(0.010)	(0.008)
Gini coefficient 1985		-0.0003	-0.0003	-0.0003
		(0.0002)	(0.0002)	(0.0002)
log(Investment)	0.055***	0.045**	0.035*	0.042**
	(0.019)	(0.020)	(0.020)	(0.019)
log($n+g+\delta$)	0.063***	0.084***	0.077***	0.076***
	(0.022)	(0.026)	(0.025)	(0.025)
log(Human capital)	0.006	0.011	0.013	
	(0.016)	(0.016)	(0.017)	
log(trade Opennes)			0.012**	0.012**
			(0.005)	(0.005)
log(Price of investment)			0.017	0.021
			(0.018)	(0.017)
Constant	0.162***	0.243***	0.249***	0.230***
	(0.047)	(0.072)	(0.069)	(0.064)
Observations	41	41	41	41
R^2	0.524	0.552	0.619	0.612
Adjusted R^2	0.472	0.488	0.538	0.543
Residual Std. Error	0.007 (df = 36)	0.007 (df = 35)	0.007 (df = 33)	0.007 (df = 34)
F Statistic	9.923*** (df = 4; 36)	8.612*** (df = 5; 35)	7.664*** (df = 7; 33)	8.932*** (df = 6; 34)
Note:				*p<0.1; **p<0.05; ***p<0.01 Standard Errors are in brackets

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

5.1.2 Model analysis

Clustered data by income level reveals a better relationship between variables. In this way, for the sample of high-income countries, column (1) becomes more significant in comparison with the previous sample, in this case, the adjusted R^2 is 0.472. There's also evidence of income convergence due to the negative sign of the beta coefficient for the initial GDPpc. Additionally, the representativeness in the model of $n + g + \delta$ and investment can be spotted, affecting the model both positively and negatively. In this regard, investment fulfills expectation according to the initial hypotheses, whereas $n + g + \delta$ does not.

In column (2) there's a slight change in the adjusted R^2 which is 0.488. In this case, the level of inequality in rich countries is not a determinant of economic growth and as determinants of growth, we can underline the same variables that were in column (1) due to their low p-values. By contrast, the average years of secondary schooling are not significant in rich countries, also, this variable in particular has a moderate positive correlation with the initial GDPpc, therefore, it will not be considered for the robustness analysis.

When the model in column (3) is carried out with all the variables, we can see a considerable increase in R^2 which is 0.538, meaning that the variables in this model explain 53.8% of the variation in economic growth. This increase could be due to the representativeness of trade openness in the model, in addition to the previous representative variables in models (1) and (2) which make a considerable improvement in the results of the regression model.

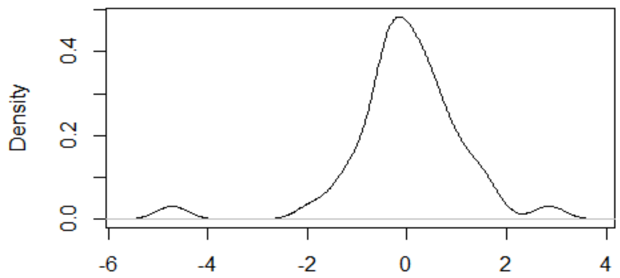
The robustness analysis in column (4), the variables used to explain the variation of economic growth by 54.3% and the representative variables, are all the same as in model (3). Investment and trade openness get positive values, and initial GDPpc gets negative values, according to what was expected in the hypotheses. However, $n + g + \delta$ did not impact growth as expected, although it is significant in this model. We can infer that a 1 unit increase in the natural log of the GDPpc, gross fixed capital formation, $n + g + \delta$, or the trade share in the GDP will cause a variation of -0.042, 0.042, 0.076 and 0.012 respectively, on average, holding all other variables constant.

As in the previous chapter, some tests will be run to our model in order to ensure the quality of the estimates for the empirical model.

Linearity of the data: We assume linearity of the variables according to the Residuals vs fitted values plots in the appendix A (see figure 12).

Normality of residuals: The Shapiro-Wilk normality test in Table 16 shows that there isn´t a normal distribution of residuals. Nevertheless, this might be due to the existence of outliers as most of the residuals seem to follow a normal distribution as figure 9 shows, therefore it is possible to proceed with the analysis.

Figure 9. Density curve of residuals for the sample of high-income countries



Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

Constant variance of residuals (homoscedasticity): The Breusch-Pagan test in table 17 shows homoscedasticity

No multicollinearity: No signs of multicollinearity in the sample (see Table 18) in accordance with the correlation matrix in table 7. A moderate positive relationship between Initial income and average years of schooling can be spotted. As a result, the same treatment will be applied for the last variable as in the previous model.

5.2 Middle-Income countries

This subsample is composed of the 23 countries listed in table 12 with middle income, according to the country classification income level of the World Bank (2017) the income threshold is a GNI/Capita of 3.956 – 12.235 (equivalent US$ in 2017).

Table 12. List of middle-income countries in the sample

GNI/Capita	Country	GNI/Capita	Country
7860	Bulgaria	8040	Kazakhstan
8670	Brazil	8930	Mexico
7020	Botswana	11000	Mauritius
8650	China	9940	Malaysia
5930	Colombia	6060	Peru
11090	Costa Rica	5390	Paraguay
7090	Dominican Republic	10010	Romania
5860	Ecuador	9230	Russia
4060	Guatemala	5950	Thailand
5470	Iran	10900	Turkey
4740	Jamaica	5410	South Africa
4020	Jordan		

Source: own elaboration with data from World Bank database

5.2.1. Model estimation

In this section, the estimations of the model for our middle-income countries sample are presented in table 13.

Table 13. Regression output for middle-income sampled countries

	Dependent variable:			
	Average Real GDP per capita growth rate (PPP)			
	(1)	(2)	(3)	(4)
log(GDPpc 1985)	-0.022**	-0.022**	-0.023**	-0.020*
	(0.009)	(0.010)	(0.010)	(0.010)
Gini coefficient 1985		-0.00002	-0.00002	-0.0001
		(0.0002)	(0.0002)	(0.0002)
log(Investment)	0.090***	0.089***	0.086***	0.091***
	(0.020)	(0.021)	(0.024)	(0.023)
log($n + g + \delta$)	-0.030	-0.028	-0.034	
	(0.023)	(0.030)	(0.038)	
log(Human capital)	-0.011	-0.012	-0.010	-0.011
	(0.010)	(0.011)	(0.012)	(0.012)
log(trade Opennes)			-0.001	-0.004
			(0.008)	(0.008)
log(Price of investment)			-0.003	0.0004
			(0.008)	(0.007)
Constant	-0.041	-0.037	-0.037	-0.003
	(0.051)	(0.064)	(0.071)	(0.059)
Observations	23	23	23	23
R^2	0.724	0.725	0.729	0.714
Adjusted R^2	0.663	0.644	0.602	0.607
Residual Std. Error	0.007 (df = 18)	0.007 (df = 17)	0.007 (df = 15)	0.007 (df = 16)
F Statistic	11.832*** (df = 4; 18)	8.947*** (df = 5; 17)	5.751*** (df = 7; 15)	6.671*** (df = 6; 16)
Note:				*p<0.1; **p<0.05; ***p<0.01 Standard Errors are in brackets

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

5.2.1 Model analysis

For middle-income countries, the explanatory variables in column (1) are more relevant in terms of significance as the adjusted R^2 is higher than the previous samples. The variables in this column explain 66.3% of the variation of economic growth which is a relevant model in order to explain the economic growth. Thus, the level of investment and initial GDPpc play a key role in explaining the dependent variable as they are significant, with a level of 99% and 95% respectively.

In column (2), by adding the variable of inequality the model barely changes, which means that income inequality is insignificant for the model in this sample as it lowers its coefficient with respect to the previous sample. Furthermore, the coefficient of the rest of variables hardly change. For this column, the level of investment and initial GDPpc are also significant in explaining economic growth.

In column (3) the R^2 decreases to 0.602 and this decrease is explained by the unability of the new variables to explain the model. At the same time, the significance of the level of investment and initial GDPpc remain the same. In this column, $n + g + \delta$ demonstrates a moderate correlation with the initial GDPpc and therefore will not be taken into account in the next column.

Finally, in column (4) we can infer that there's a clear sign of convergence in the income of the countries. In addition, The level of investment is a key determinant of growth in this model, affecting it positively, as explained by MRW (as expected in table 3). Unfortunately, not much can be deduced from the rest of the variables due to the lack of statistical significance of the coefficients. It is likely that increasing the dataset of middle-income countries would have improved the outcome of the regression model.

Following with our OLS tests to ensure and validate the hypotheses of OLS models, we perform the same analysis for this sample as in previous sections.

Linearity of the data: According to the Residuals vs fitted values plot in figure 13 a linear model is appropriate for this data.

Normality of residuals: There´s a normal distribution of residuals (see Table 16).

Constant variance of residuals (homoscedasticity): Table 17 shows homoscedasticity.

No multicollinearity: No signs of multicollinearity in the sample (see table 18). Although, there's evidence of a moderate correlation in table 7 between the level of depreciation and net Gini coefficient, as well with the price of investment. Thereupon, the level of depreciation will not be taken into account for the robustness analysis.

5.3 Low-Income countries

This last subsample is composed of the 30 countries with middle income listed in table 14. According to the country classification income level of the World Bank (2017) the income threshold is a GNI/Capita of 3.956 – 12.235 (equivalent US$ in 2017).

Table 14. List of low-income countries in the sample

GNI/Capita	Country	GNI/Capita	Country	GNI/Capita	Country
3950	Armenia	1440	Kenya	1110	Mauritania
1520	Bangladesh	1110	Kyrgyzstan	340	Malawi
3090	Bolivia	3880	Sri Lanka	860	Nepal
1480	Côte d'Ivoire	1250	Lesotho	1500	Pakistan
3920	Algeria	2880	Morocco	3650	Philippines
3040	Egypt	3520	Tunisia	730	Rwanda
1900	Ghana	970	Tanzania	2390	Sudan
2220	Honduras	620	Uganda	520	Sierra Leone
3530	Indonesia	2260	Ukraine	3590	Swaziland
1830	India	1300	Zambia	1000	Tajikistan

Source: own elaboration with data from World Bank database

5.3.1 Model estimation

Our estimations for our middle-income countries sample are presented in table 15.

Table 15. Regression output for low-income sampled countries

	Dependent variable:			
	Average Real GDP per capita growth rate (PPP)			
	(1)	(2)	(3)	(4)
ln(GDPpc 1985)	-0.046***	-0.045***	-0.015	-0.006
	(0.013)	(0.013)	(0.013)	(0.013)
Gini coefficient		0.0002	0.002***	0.001**
		(0.0005)	(0.001)	(0.001)
ln(Investment)	0.054**	0.054**	0.062**	0.057**
	(0.025)	(0.025)	(0.022)	(0.024)
ln(n+g+δ)	-0.043	-0.047	-0.082	-0.112**
	(0.059)	(0.061)	(0.050)	(0.052)
ln(Human capital)	0.015	0.017	0.024**	
	(0.012)	(0.013)	(0.011)	
ln(trade Opennes)			-0.075***	-0.068***
			(0.020)	(0.021)
ln(Price of investment)			-0.003	-0.010
			(0.019)	(0.020)
Constant	0.052	0.036	-0.043	-0.103
	(0.074)	(0.083)	(0.070)	(0.069)
Observations	30	30	30	30
R^2	0.417	0.422	0.653	0.578
Adjusted R^2	0.323	0.302	0.543	0.468
Residual Std. Error	0.015 (df = 25)	0.015 (df = 24)	0.012 (df = 22)	0.013 (df = 23)
F Statistic	4.466*** (df = 4; 25)	3.506** (df = 5; 24)	5.926*** (df = 7; 22)	5.259*** (df = 6; 23)
Note:				*p<0.1; **p<0.05; ***p<0.01 Standard Errors are in brackets

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

5.3.2 Model analysis

For low-income countries in equation (1) when regressed, as in Mankiw, Romer and Weil (1992) there's evidence of a representative impact of initial income per capita and the level of investment, however, this model doesn't explain economic growth with accuracy using these variables, as the adjusted R^2 is 0.314.

In equation (2) when adding the inequality variable to the model, R^2 decreases slightly. Furthermore, the inequality variable to this extent is not significant enough to explain any relevant changes in economic growth. Consequently, the initial income per capita and the level of investment are crucial in explaining economic growth in this sample.

Equation (3) integrates all the variables and shows that income inequality strongly impacts the explanation of economic growth. We can therefore infer that for every additional percentage increase in the Gini coefficient (inequality), the expected value of economic growth increases by 0.002 (or 0.2 percent) on average, holding all other variables constant. The p-value indicates the probability of the coefficients of the variables occurring due to a random change, consequently, the most signficant variables are inequality, investment and trade openess with a 99% level of significance and human capital with a 95% level of significance. Additionally, the adjusted R^2 is quite large to represent at least half of the variation in economic growth this variables.

Finally, equation (4) also reports a significant representation of inequality in explaining economic growth, on the other hand, the level of investment, trade openness and human capital remain equally as significant. Additionally, there's a significant influence of $n + g + \delta$ which explains the variation of economic growth. In this sample, inequality level and trade openness have both a positive and negative sign which does not go along with the expected signs. Moreover, investment and $n + g + \delta$ behaved as expected, reporting a negative sign. Nevertheless, this equation explains only 43.7% of the variation of economic growth, which is not as high as the previous one.

As in our previous analysis, our tests reflect the following information towards the validation of the hypotheses of OLS models.

Linearity of the data: Variables and the dependent variable are linear (see figure 14)

Normality of residuals: There´s a normal distribution of residuals (see table 16).

Constant variance of residuals (homoscedasticity): Table 17 demonstrates that there's no sign for heteroskedasticity.

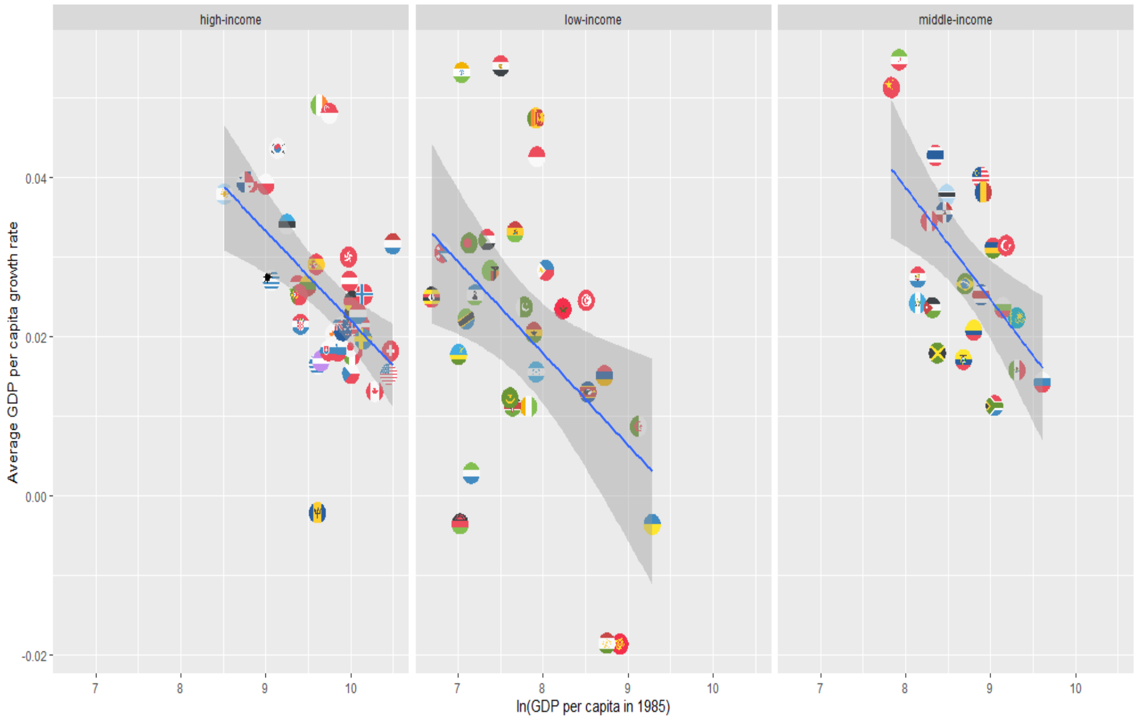
No multicollinearity: No signs of multicollinearity in the sample (see table 18). Also, table 7 shows the existence of a moderate correlation between the average years of schooling and the initial level of income. In this way, the average years of schooling will not be taken into account in the robustness analysis as it could generate problems of multicollinearity.

6. Discussion

This chapter discusses the main findings of our models and compares the outputs from chapter 4 and 5 with respect to the previous empirical literature in this field. In section 2.1 we proposed 2 models that are subject of analysis for our samples. The first model corresponding to the one proposed by Mankiw, Romer and Weil (1992) and the second one is the model used in the empirical literature of income inequality and economic growth for cross-sectional data. The first model is compared to column (1) in all our regression outputs, whereas the second model has been represented and analysed in columns (2), (3) and (4).

Regarding the results of Mankiw, Romer and Weil (1992) they found a strong negative and significant relation between initial income per capita and economic growth for all their samples, this is a strong evidence of convergence, specially for low-income countries. This result is repeted in all our samples following that model and we get honogeneous results indicating convergence in all samples as the literature shows. In order to assess the level of convergence, Mankiw, Romer and Weil (1992) regressed economic growth and initial GDP per capita, and for the general sample they didn't find a strong evidence of convergence, however they did for high-income countries. This model is more representative when the sample is divided and analysed per income level. This explains why the adjusted R^2 in column (1) for all subsample's regression output drastically increased. When regressing economic growth with initial income per capita as in figure 10, we validate the results obtained by Mankiw, Romer and Weil (1992) regarding the convergence of countries.

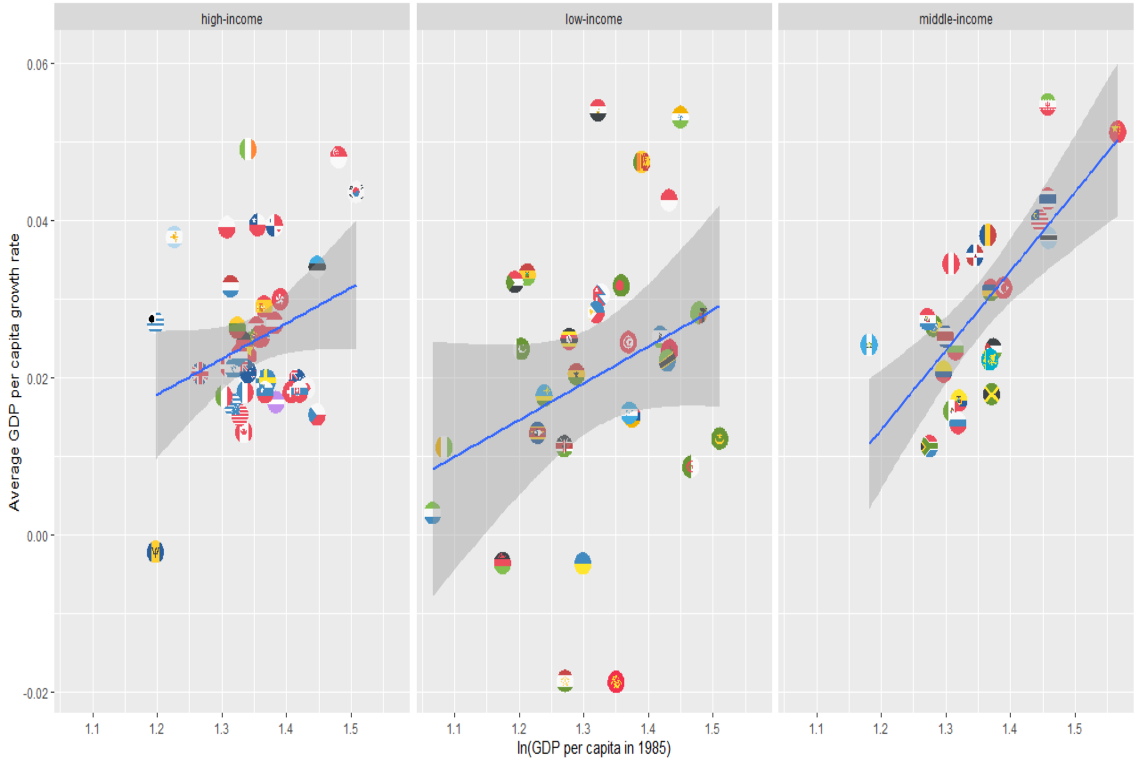
Figure 10. Conditional convergence 1985-2017 in all subsamples



Source: own elaboration with data from PWT 9.1 and WDI database

Mankiw, Romer and Weil (1992) also found a positive coefficient for investment when explaining economic growth. Regarding the impact of investment on economic growth, we can infer that according to our results, it has a greater impact on developing countries (middle-income countries) as it exhibits the bigger gradient (coefficient). In figure 11 we can find countries like China, India, Iran, Thailand and Malaysia which grew economically at a very fast rate in the 1990s and 2000s. In this sample, countries with high investment reported larger economic growth. Solow's model (1956) explains this trend as the "catch-up growth", in which first units of physical capital (investment), generate fast economic growth in poorer countries. On the other hand, the situation seems to be different in developed countries in which most of the countries are concentrated in a similar rank of growth rate, between 1 and 3 per cent, differentiating them from the large growth rates of many of the developing countries and some low-income countries. This means that these countries are getting closer to the theoretical steady-state level as described in section 1.3. According to Solow's model, this is explained by the "cutting edge growth", in which economic growth is not great enough when acquiring more units of physical capital due to the diminishing returns to scale.

Figure 11. Impact of investment on economic growth 1985-2017 in all subsamples



Source: own elaboration with data from PWT 9.1 and WDI database

As for $n + g + \delta$, Mankiw, Romer and Weil (1992) found a negative relationship with economic growth, however, our model proposes a positive and strong statistical relationship for high-income countries and a negative but insignificant relationship for low and middle-income countries. Lastly, we get a positive coefficient for human capital as the literature reflects but not statistically significant.

In general terms, the model (11) proposed in section 2.1 following the empirical literature and represented by equations (2), (3) and (4) in the regression outputs from chapter 4 and 5, did not show a significant relationship between income inequality and economic growth among high and middle-income countries, as well as in the general sample of 94 countries. It does, however, exhibit a strong positive relationship in low-income countries when exposed to the whole set of control variables.

Coefficients of inequality in our output for high and middle-income suggested a negative but statistically insignificant relationship with economic growth. If the coefficients for inequality in the regression output for high and middle-income countries were to be statistically representative, we could infer that these negative results are not robust to the inclusion of more variables in the model. In this way, based on the empirical evidence in all samples we are rejecting the null initial hypothesis in this paper that income inequality negatively affects economic growth.

Therefore, the negative coefficient determined by many of the previous empirical studies using cross-sectional data as in Alesina and Rodrik (1994) stopped being significant to explain income inequality and economic growth when control variables are included in the model as in equation (3) and (4). The outcome of our model is similar to the results of Castelló and Domenech (2002) who also obtained a negative result for their whole sample of 67 countries and only got a positive relationship between income inequality and economic growth when inequality was considered simultaneously with initial income and human capital, using an OLS model with cross-sectional data. Also, in our samples, population growth, when augmented by the rate of technical progress and the depreciation rate of the physical capital $(n + g + \delta)$ exhibited incongruous results. This disparity on the effects of $n + g + \delta$ according to the sample is also frequent in the model of Castelló and Domenech (2002).

Moreover, we can highlight that early cross-section studies analysed the relationship between these variables from the period 1960 – 1985 and the majority of these studies reflected this negative coefficient. However, recent studies that use up to date data and analysed a more recent period as in this paper, tend to present inconsistent results differing from the literature. This paper gets also a similar result as Knowles (2005) which found a negative but insignificant relationship in high and middle-income countries, nevertheless, argues that low-income countries present a negative relationship as in Deininger and Squire (1998).

This model demonstrated a positive and significantly high relationship between investment and economic growth for all samples. Another variable that strongly impacted economic growth for the whole dataset and the consequent subsamples is the initial income per capita, which has negatively impacted ecomic growth (as expected), meaning that the higher the initial income, the lower the economic growth. This results are also present in the model proposed by Mankiw, Romer and Weil (1992) and the vast majority of empirical studies in the field of income inequality and economic growth.

Finally, trade openness demonstrates a positive relationship with growth in high-income countries and a negative one among low-income countries, being more statistically significant in the sample for low-income countries. The sign of this coefficient goes against what was expected. We can infer that and poor countries dependent on trade have suffered the instability of the global crisis more than the rest.

7. Conclusion

The objective of this paper is to assess the impact of income inequality on economic growth during the period 1985-2017. The literature in this field suggests either a positive or negative effect depending on channel and period used in the model. However, these theories have been highly debated and offer ambiguous premises. We use cross-sectional data, which is suitable to find long-run relationships between the explanatory variables and the dependent variable in the chosen period. Our first model was motivated by the extended neoclassical growth model of Mankiw, Romer and Weil (1992), and the second by the empirical literature as described by De Dominicis et al. (2008). Additionally, we have added more relevant variables which have been used in more recent papers consistent with the empirical literature, in order to avoid omitted variable bias. Moreover, we have worked with updated data and performed an econometric analysis of the second model to assess its reliability.

Our main findings suggest that there's a positive significant relationship between income inequality and economic growth for low-income countries in the long run. However, there's a negative but not statistically significant relationship in high and middle-income countries. Additionally, when all countries in the sample are regressed together, there is not a statistically significant relationship between income inequality and economic growth.

In this regard, in our samples, other variables reported higher statistical significance in explaining economic growth among countries than inequality. There's evidence in our model that the level of investment is the most important variable as it was strongly presented in all regression outputs for all subsamples. We have deduced that the level of investment positively correlates with economic growth more closely in poor and developing countries, confirming the neoclassical theory of Solow (1956) which is the baseline of the models used.

Additionally, our models point out that the initial level of income per capita is relevant to explain economic growth. It expresses how countries are converging in time. Those countries with a lower income per capita in 1985 grew faster than those with a higher income per capita, especially in middle-income countries (developing countries) which converged faster and reported higher levels of growth in the analysed period.

Furthermore, our second model reveals the importance of $n + g + \delta$ and trade openness as key indicators to explain economic growth, even if to a lesser extent than the level of investment and initial income per capita. Nonetheless, these variables can vary in sign depending on the sample in which it is regressed, expressing an ambiguous result which is not homogeneous like the level of investment and initial income per capita in all samples.

Personally, this paper supposed a professional boost in quantitative analysis programming. The motivation was high since I intend to do a master on this subject. Additionally, the cost of entry in terms of effort, until all econometric analyses and estimations could be programmed in R as Appendix B reveals, was relatively high.

The main limitation of this paper is the available data for all countries in the world and that OLS regressions might suffer endogeneity problems that might lead to biased estimates. In order to control measurement errors in estimates, it would be convenient to use instrumental variables as they will correct endogeneity problems, although, at this preliminary stage of this study we have only worked with OLS regressions that use cross-sectional data with as many explanatory variables as possible following the empirical literature. As a future extension of this paper, it would suit the usage of another inequality indicator such as the Theil index and to cover (if possible) more countries for the analysis, as well as adding more adequate control variables that can capture countries´ economic, social and institutional differences for each of the homogeneous groups, as suggested by De Dominicis et al. (2008). This paper could be potentially expanded in a second stage by estimating a model with panel data.

Bibliography

Acemoglu, D. (2007). *An Introduction to Modern Economic Growth*. Journal of Economic Theory, Elsevier, vol. 147(2), pages 545-550.

Acemoglu, D., Ticchi, D., & Vindigni, A. (2011). *Emergence and persistence of inefficient states*. Journal of the European Economic Association, 9(2), 177–208.

Aghion, P., Bacchetta, P., Ranciere, R, and Rogoff, K. (2009). *Exchange Rate Volatility and Productivity Growth: The Role of Financial Development.* Journal of Monetary Economics 56 (4): 494-513.

Aghion, P., Caroli, E., and García-Peñalosa, C. (1999). *Inequality and Economic Growth: The Perspective of the New Growth Theories. Journal of Economic LiteratureVol.37(4)*, 1615-1660

Aiyar, S. and Ebeke, C. H. (2019). *Inequality of Opportunity, Inequality of Income and Economic Growth*. International Monetary Fund. IMF Working Papers 19/34.

Alesina, A. and Perotti, R. (1996). *Income distribution, political instability, and investment*. European Economic Review, 81, 5, pp. 1170–89.

Alesina, A. and Rodrik, D. (1994). *Distributive politics and economic growth*. Quarterly Journal of Economics, 109, 2, pp. 465–90.

Atuesta, B., Mancero, X. and Tromben, V. (2018). *Herramientas para el análisis de las desigualdades y del efecto redistributivo de las políticas públicas*. ECLAC documents LC/TS.2018/53

Arel-Bundock, V. (2019). *WDI: World Development Indicators (World Bank)*. R package version 2.6.0. https://CRAN.R-project.org/package=WDI

Barro, R. (1991). *Economic growth in a cross section of countries*. Quarterly Journal of Economics 106 (2), 407–443.

Barro, R. J. (*1996*). *Determinants of Economic Growth: A Cross-Country Empirical Study*. MIT Press Books, The MIT Press, edition 1, volume 1, number 0262522543, April.

Barro, R. J. (2000). *Inequality and growth in a panel of countries*. Journal of Economics Growth 5 (1), 5–32.

Berg, A., Ostry, J. D., Tsangarides, C. G. and Yakhshilikov, Y. (2018). *Redistribution, inequality, and growth: new evidence*. J Econ Growth 23, 259–305.

Bouincha, M, and Karim, M. (2018). *Income Inequality and Economic Growth: An Analysis Using a Panel Data.* International Journal of Economics and Finance. 10. 242.

Breunig, R. and Majeed, O. (2020) *Inequality, poverty and economic growth*. International Economics, Volume 161, Pages 83-99, ISSN 2110-7017

Brueckner, M. and Lederman, D. (2018). *Inequality and economic growth: the role of initial income*. J Econ Growth 23, 341–366.

Callen, T. (2008). *¿Qué es el producto interno bruto?*. Finanzas y desarrollo: quarterly publication of the International Monetary Fund and the World Bank 45(4), 48-49.

Carbonari, L. (2011) *Purchasing power parity (PPP)*. Associazione nazionale per l'enciclopedia della banca e della borsa. Bankpedia Review n.1. pp 93- 100

Castelló, A. (2004). *A reassessment of the relationship between inequality and growth: what human capital inequality data say?* Working Papers. Serie EC 2004-15, Instituto Valenciano de Investigaciones Económicas, S.A. (Ivie).

Castelló, A. and Doménech, R. (2002) *Human Capital Inequality and Economic Growth: Some New Evidence*, *The Economic Journal*, Volume 112, Issue 478, Pages C187–C200

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

De Dominicis, L., Florax, R., and de Groot, H. (2008). *A meta-analysis on the relationship between income inequality and economic growth*. *Scottish Journal of Political Economy*, 55(5), 654– 682.

De Ferranti, D., Ferreira, F. H. G. and Perry, G. E., Walton, M. (2004). *Inequality in Latin America: breaking with history?* (English). World Bank Latin American and Caribbean Studies. Washington, DC: World Bank. Retrieved: May 1st, 2020. from: http://documents.worldbank.org/curated/en/804741468045832887/Inequality-in-Latin-America-breaking-with-history

Deininger, K. and Squire, L. (1996). *A new data set measuring income inequality.* World Bank Economic Review, 10, 3, pp. 563–91.

Feenstra R. C., Inklaar R. and Timmer M. P. (2015). *The Next Generation of the Penn World Table.* American Economic Review*, 105*(10), 3150–3182. Retrieved: March 27th, 2020. From: http://www.ggdc.net/pwt/.

Fellman, J. (2012). *Estimation of Gini coefficients using Lorenz curves*. J. Stat. Econ. Methods. 1.

Foellmi, R., and Oechslin, M. (2008). *Why progressive redistribution can hurt the poor*. Journal of Public Economics, 92(3–4), 738–747.

Forbes, K. (2000). *A Reassessment of the Relationship Between Inequality and Growth*. American Economic Review. 90. 869-887.

Foster, I., Ghani, R., Jarmin, R., Kreuter, F. and Lane, J. (2018) *Big Data and Social Science A Practical Guide to Methods and Tools*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences

Galor, O. and Moav, O. (2004*). From physical to human capital accumulation: inequality and the process of development*. Review of Economic Studies, 71, pp. 1001–26.

Galor, O. and Zang, H. (1997). *Fertility, income distribution, and economic growth: theory and cross-country evidence*. Japan and the World Economy, 9, 2, pp. 197–229. &

Galor, O., and Tsiddon, D. (1997). *Technological Progress, Mobility, and Economic Growth.* The American Economic Review, 87(3), 363-382. Retrieved May 16, 2020, from www.jstor.org/stable/2951350

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2010). *Multivariate Data Analysis.* Seventh Edition. Prentice Hall, Upper Saddle River, New Jersey.

Haughton, J. H. & Khandker, S. R. (2009). *Handbook on Poverty and Inequality*. Washington, D.C.: World Bank Publications.

Henderson, J., Storeygard, A., & Weil, D. (2012). *Measuring Economic Growth from Outer Space.* The American Economic Review, 102(2), 994-1028. Retrieved May 16, 2020, from www.jstor.org/stable/23245442

Kaldor, N. (1957). *A Model of Economic Growth*. *The Economic Journal, 67*(268), 591-624. doi:10.2307/2227704

Knowles, S. (2005) *Inequality and Economic Growth: The Empirical Relationship Reconsidered in the Light of Comparable Data*. The Journal of Development Studies, vol. 41(1): 135-159.

Kuznets, S. (1955). *Economic Growth and Income Inequality*. The American Economic Review, 45(1), 1-28. Retrieved April 12, 2020, from www.jstor.org/stable/1811581

Lewis, W. A. (1954) *Economic Development with Unlimited Supplies of Labor*. The Manchester School 22. 139–91.

Li, H. and Zou, H. (1998). *Income Inequality is not Harmful for Growth: Theory and Evidence*. Review of Development Economics, 2(3): 318-334.

Mankiw, N. G. (2013) *Defending the One Percent*. Journal of Economic Perspectives, 27 (3): 21-34.

Mankiw, N. G., Romer, D. and Weil, D. N. (1992) *A Contribution to the Empirics of Economic Growth.* The Quarterly Journal of Economics*, Volume* 107, Issue 2. Pages 407–437, https://doi.org/10.2307/2118477

Martín-Legendre, J. (2018). *The challenge of measuring poverty and inequality: a comparative analysis of the main indicators.* European Journal of Government and Economics, 7(1), 24-43. https://doi.org/10.17979/ejge.2018.7.1.4331

Okun, A. and Summers, L. (2015). *Equality and Efficiency: The Big Tradeoff*. Brookings Institution Press. Retrieved: May 16th, 2020. From: jstor.org/stable/10.7864/j.ctt13wztjk

Panizza, U. (2002) *Income Inequality and Economic Growth: Evidence from American Data*. Journal of Economic Growth 7, 25–41.

Persson, T. and Tabellini, G. (1991). *Is inequality harmful for growth? Theory and evidence*. CEPR Working Paper no. 581

Perotti, R. (1996). *Growth, Income Distribution, and Democracy: What the Data Say*. Journal of Economic Growth, 1(2), 149-187. Retrieved April 15, 2020, from: www.jstor.org/stable/40215914

Piburn, J. (2018). *wbstats: Programmatic Access to the World Bank API*. Oak Ridge National Laboratory. Oak Ridge, Tennessee. Retrieve: April 2nd, 2020. From: https://www.ornl.gov/division/csed/gist

Piketty, T., and Goldhammer, A. (2015). *The Economics of Inequality*. Cambridge, Massachusetts; London, England: Harvard University Press. Retrieved May 16, 2020, from www.jstor.org/stable/j.ctvjnrtk1

Piketty, T., Saez, E. and Zucman, G. (2018). *World Inequality Report 2018*. Post-Print halshs-01885458, HAL.

R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Saez, E. (2020). *Striking it Richer: The Evolution of Top Incomes in the United States (Updated with 2018 estimates)*. Retrieved: March 25th, 2020. From: eml.berkeley.edu/~saez/saez-UStopincomes-2018.pdf

Solow, R. M. (1956). *A Contribution to the Theory of Economic Growth*. The Quarterly Journal of Economics, Oxford University Press, vol. 70(1), pages 65-94.

Solt, F. (2019). *Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database*. SWIID Version 8.2, November 2019.

Stiglitz, J. (2012). *The Price of Inequality: How Today's Divided Society Endangers Our Future.* New York: W.W. Norton.

Székely, M. and Hilgert, M. (1999) *What's Behind the Inequality We Measure: An Investigation Using Latin American Data*. IDB Working Paper No. 340.

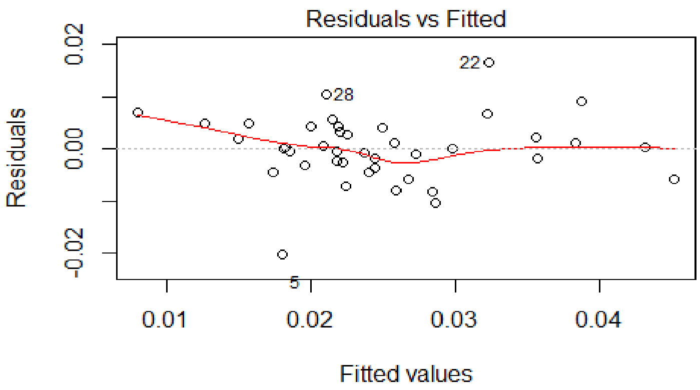
Wickham et al., (2019). *Welcome to the tidyverse*. Journal of Open Source Software, 4(43), 1686.

World Bank (2017) *New country classifications by income level: 2017-2018*. Retrieved: April 15th, 2020. From: https://blogs.worldbank.org/opendata/new-country-classifications-income-level-2017-2018

Zeileis A (2019). *pwt9: Penn World Table (Version 9.x)*. R package version 9.1-0. Retrieved: April 3rd, 2020. From: https://CRAN.R-project.org/package=pwt9.

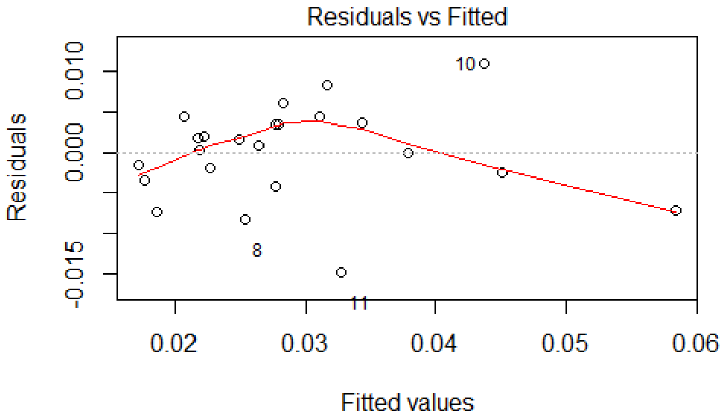
Appendix A

Figure 12. Residuals vs fitted values plot for high-income countries sample



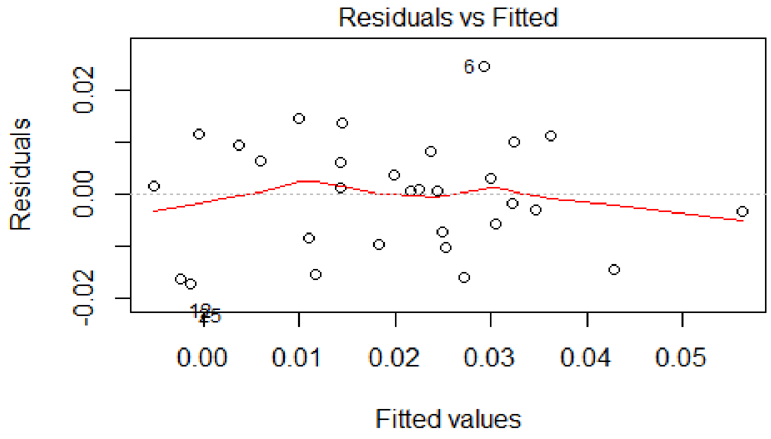
Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

Figure 13. Residuals vs fitted values plot for middle-income countries sample



Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

Figure 14. Residuals vs fitted values plot for low-income countries sample



Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

Table 16. Shapiro-Wilk Normality Test for all sub samples

Sample	W	p-value	Hypothesis testing
High-income countries	0.90318	0.002061	Reject null hypothesis
Middle-income countries	0.97969	0.9009	Accept null hypothesis
Low-income countries	0.96899	0.5118	Accept null hypothesis

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

Table 17. Breusch-Pagan test for all sub samples

Sample	Chisquare	p-value	Hypothesis testing
High-income countries	0.0006929759	0.979	Accept null hypothesis
Middle-income countries	2.300005	0.12937	Accept null hypothesis
Low-income countries	0.3014576	0. 58297	Accept null hypothesis

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

Table 18. Variance Inflation Factor for all sub samples

Variance Inflation Factor analysis					Sample
	(1)	(2)	(3)	(4)	
ln(GDPpc)	1,861759	2,899513	3,262758	2,437865	High-income countries
Gini		2,044949	2,071899	1,929547	
ln(Investment)	1,083458	1,200264	1,454763	1,222101	
ln(n+g+δ)	1,020829	1,343236	1,407706	1,366375	
ln(Human capital)	1,951717	2,068793	2,442584		
ln(Trade openness)			1,348404	1,344142	
ln(Price of investment)			2,380258	2,108492	
	(1)	(2)	(3)	(4)	
ln(GDPpc)	1,780933	1,781152	1,904073	1,714075	Middle-income countries
Gini		2,293119	2,692488	1,455012	
ln(Investment)	1,340588	1,441643	1,645358	1,603631	
ln(n+g+δ)	1,276924	2,131135	4,375859		
ln(Human capital)	1,675688	1,752343	1,919742	1,894179	
ln(Trade openness)			1,268408	1,187041	
ln(Price of investment)			2,504222	1,236003	
	(1)	(2)	(3)	(4)	
ln(GDPpc)	2,139393	2,14166	3,341631	2,940969	Low-income countries
Gini		1,394051	2,739167	2,533441	
ln(Investment)	1,036493	1,036534	1,249367	1,237589	
ln(n+g+δ)	1,612699	1,693189	1,774764	1,691161	
ln(Human capital)	2,26826	2,412538	2,605946		
ln(Trade openness)			2,930742	2,839575	
ln(Price of investment)			1,298289	1,250015	

Source: Own elaboration with data from WDI, SWIID8.2 and PWT9.1 databases

Appendix B

Loading packages:

```
library(tidyverse) # Easily Install and Load the 'Tidyverse'
library(stringr) # Simple, Consistent Wrappers for Common String Operations
library(rworldmap) # Mapping Global Data
library(WDI) # World Development Indicators (World Bank)
library(wbstats) # Programmatic Access to Data and Statistics from the World Bank
library(pwt9) # Penn World Table (Version 9.x) # Penn World Table (Version 9.x)
library(countrycode) # Convert Country Names and Country Codes
library(zoo) # S3 Infrastructure for Regular and Irregular Time Series (Z's
library(readxl) # Read Excel Files
library(writexl) # Export Data Frames to Excel 'xlsx' Format
library(regclass) # Tools for an Introductory Class in Regression and Modeling
library(lmtest) # Testing Linear Regression Models
library(stargazer) # Well-Formatted Regression and Summary Statistics Tables
library(caret) # Classification and Regression Training
library(car) # Companion to Applied Regression
library(ggflags) # Plot flags of the world in ggplot2.
library(normtest) # Tests for Normality
library(annotater) # Annotate Package Load Calls
```

Merge World Bank and Penn World Tables 9.1 databases

```
##### Merge World Bank and PWT databases ####################################################
sink(); rm(list=ls(all=TRUE),envir=globalenv()); cat("\014"); WD <- getwd(); WD
sink( file.path( WD, "MacroBank_out.txt" ), append=FALSE, split=TRUE)

Datos  <- file.path(WD, "Datos")

pathlist <- strsplit(WD, "/", fixed = TRUE)

motherfolder <- paste( pathlist[[1]] [ 1: (length(pathlist[[1]]) -1) ], collapse=.Platform$file.sep)

A <- paste( pathlist[[1]] [ 1: (length(pathlist[[1]]) -2) ], collapse=.Platform$file.sep)

cat('Data Bank')
cat('Author: José Caloca, Universidade da Coruña')
date()

cat('DATA BANK PACKAGES')

cat('Penn World Tables')

data("pwt9.1")
pwt <- pwt9.1
pwt <- pwt %>% select(country, year, isocode, rgdpe, csh_g, pop, pl_i)
rm(pwt9.1)

pwt$country <- as.character(pwt$country)
pwt <- rename(pwt, iso3 = isocode)
pwt$iso3 <- as.character(pwt$iso3)

pwt <- pwt[ , c("iso3", "year", setdiff( colnames(pwt), c("iso3", "year")))]

cat('Wolrd Bank')

wbcache <- wbcache(lang = "en") # english or es spanish  / SLOW

# In order to compare different ginis with the SWIID database
# Gini coefficient: "SI.POV.GINI"
wbsearch(pattern = "gini", fields = c("indicator", "indicatorDesc"), extra = FALSE, cache=wbcache)
# 15121    SI.POV.GINI    GINI index (World Bank estimate)

wbsearch(pattern = "Gross fixed capital formation", fields = c("indicator", "indicatorDesc"), extra =
FALSE, cache=wbcache)
# 9769    NE.GDI.FTOT.ZS    Gross fixed capital formation (% of GDP)
```

```
# Trade openness index: (X+M)/GDP "NE.TRD.GNFS.ZS"
wbsearch(pattern = "Trade", fields = "indicatorDesc", extra = FALSE, cache=wbcache)
# 9533  Trade (% of GDP) exports and imports of goods and services measured as a share of gross domestic
product.

# Years of schooling, secondary "BAR.SEC.SCHL.15UP"
wbsearch(pattern = "secondary schooling", fields = "indicatorDesc", extra = FALSE, cache=wbcache)
# 1359    BAR.SEC.SCHL.15UP     Barro-Lee: Average years of secondary schooling, age 15+, total

# Population growth, as in Mankiew 1992
wbsearch(pattern = "Population growth", fields = "indicatorDesc", extra = FALSE, cache=wbcache)
#15818        SP.POP.GROW     Population growth (annual %)

# GNI per capita, Atlas method (current US$)
wbsearch(pattern = "GNI per capita", fields = "indicatorDesc", extra = FALSE, cache=wbcache)
## 10166    NY.GNP.PCAP.CD           GNI per capita, Atlas method (current US$)

wbvars  <-  c("SI.POV.GINI",  "NE.GDI.FTOT.ZS",  "NE.TRD.GNFS.ZS",  "BAR.SEC.SCHL.15UP",  "SP.POP.GROW",
"NY.GNP.PCAP.CD" )

wbdata <- wb(indicator = wbvars, return_wide = TRUE, lang = "en" ) # We want in wide format, one column
by variable, the same as PWT

wbdata <- rename(wbdata , year = date,
                 iso3          = iso3c,
                 iso2c         = iso2c,
                 giniwb = SI.POV.GINI,
                 I2GDP         = NE.GDI.FTOT.ZS,       # Investment-to-GDP ratio
                 trade2GDP     = NE.TRD.GNFS.ZS,       # Trade openness index: (X+M)/GDP
                 yearSecSchoo   = BAR.SEC.SCHL.15UP,     # 1359     BAR.SEC.SCHL.15UP      Barro-Lee:
Average years of secondary schooling, age 15+, total
                 popgrowth = SP.POP.GROW, # Adolescent fertility rate
                 GNIpc = NY.GNP.PCAP.CD # GNI per capita
)

wbdata$year <-  as.numeric(wbdata$year)
unique(sort(wbdata$year)) # 1960-2018

# Given that I am going to lose a lot of observations, I postpone the cleaning of world bank data


cat('CORRESPONDENCES TO MERGE')

pwt.countries <- unique(pwt[ , c("country", "iso3")])

# 1) PENN & WORLD BANK
# https://wits.worldbank.org/wits/wits/witshelp/content/codes/country_codes.htm
wbdata.countries <- unique(wbdata[ , c("country", "iso3")])
# write.csv( wbdata.countries, file.path(Datos, "wbdata.countries.csv"), row.names = F )

# setdiff(A, B) elements in A that are not present in B
setdiff(pwt$iso3        , wbdata$iso3)
setdiff(wbdata$iso3     , pwt$iso3)
# Codes that are in WORLD BANK but not in pwt: check with both excels
unique(wbdata[ wbdata$iso3 %in% setdiff(wbdata$iso3 , pwt$iso3), c("country", "iso3")])

cat('VARIABLE SELECTION AND MERGE')
unique(pwt$year)            # 1950-2017
unique(wbdata$year)         # 1960-2018


length(unique(pwt$iso3))    # 182 countries
length(pwt$iso3)            # 12376 obs

# all.x = TRUE  means that it keeps all the rows of x, so if those countries are not in y, the variables
of y will have NA
# https://stat.ethz.ch/R-manual/R-devel/library/base/html/merge.html
macro <- merge (pwt, wbdata, by=c("iso3", "year"), all.x=TRUE)
length(unique(macro$iso3))
length(macro$iso3)
names(macro)
macro <- rename(macro, country = country.x)
macro$country.y <- NULL

c('DATA BANK ')

macro <- macro [order(macro$iso3, macro$year), ]
macro <- data.frame(macro)
head(rownames(macro))
rownames(macro)[duplicated(rownames(macro))]
macro[macro$iso3=="MEX", "year"]

macro$iso3      <- as.character(macro$iso3)
```

```
macro$country   <- as.character(macro$country)
macro$year      <- as.numeric(as.character(macro$year))

rm(pathlist, pwt, pwt.countries, wbcache, wbdata, wbdata.countries, A, motherfolder, wbvars, WD)
```

Development of a function that enables the access to the SWIID 8.2 database and loads it in R's environment.

```
##### Load SWIID database #####################################################

url <- "https://github.com/fsolt/swiid/raw/master/data/swiid8_1.rda"

download.file(url, "swiid8_2.rda")

e <- new.env(parent = emptyenv())
load("swiid8_2.rda", envir = e)
out <- eapply(e, function(x) {
})

nms <- load("swiid8_2.rda")
for (nm in nms) {
    x <- get(nm)
}

out <- lapply(lapply(nms, get), function(x) {
})

rm(nm, nms, url, swiid, x, out, e)
```

Merge previous databases with SWIID 8.2 database

```
##### Merge all 3 databases ###################################################

datos <- swiid_summary

datos <- datos %>%
    dplyr::select(country, year, gini_disp, gini_mkt) %>%
    cbind(countrycode(sourcevar = datos$country, origin = "country.name",destination = "iso3c"))  %>%
    rename(iso3 = "countrycode(sourcevar = datos$country, origin = \"country.name\", ")

datos <- datos[!is.na(datos$iso3),]

datos <- filter(datos, datos$year >= 1985 & datos$year <= 1990)

as.factor(datos$iso3)

datos <- aggregate(gini_disp~iso3,datos,mean)
rm(swiid_summary)
```

Data wrangling process

```
##### Data wrangling ##########################################################

muestra <- filter(macro, macro$iso3 %in% datos$iso3 & macro$year >= 1985 & macro$year <= 2017)

schooling <- data.frame(aggregate(yearSecSchoo ~ iso3, muestra, mean))
popgrowth <- data.frame(aggregate(popgrowth ~ iso3, muestra, mean))
I2GDP <- data.frame(aggregate(I2GDP ~ iso3, muestra, mean))
trade <- data.frame(aggregate(trade2GDP ~ iso3, muestra, mean))
rateschooling <- data.frame(aggregate(ratepopsecch ~ iso3, muestra, mean))
priceinv <- data.frame(aggregate(pl_i ~ iso3, muestra, mean))
gov <- data.frame(aggregate(csh_g ~ iso3, muestra, mean))

schooling <- cbind(datos, merge(datos, schooling, by.x = "iso3", by.y = "iso3", all = TRUE))
popgrowth <- cbind(datos, merge(datos, popgrowth, by.x = "iso3", by.y = "iso3", all = TRUE))
I2GDP <- cbind(datos, merge(datos, I2GDP, by.x = "iso3", by.y = "iso3", all = TRUE))
trade <- cbind(datos, merge(datos, trade, by.x = "iso3", by.y = "iso3", all = TRUE))
rateschooling <- cbind(datos, merge(datos, rateschooling, by.x = "iso3", by.y = "iso3", all = TRUE))
priceinv <- cbind(datos, merge(datos, priceinv, by.x = "iso3", by.y = "iso3", all = TRUE))
gov <- cbind(datos, merge(datos, gov, by.x = "iso3", by.y = "iso3", all = TRUE))

datos    <-    data.frame(datos$iso3,   datos$gini_disp,   schooling$yearSecSchoo,   popgrowth$popgrowth,
I2GDP$I2GDP, trade$trade2GDP, rateschooling$ratepopsecch, priceinv$pl_i, gov$csh_g)
names(datos)[1] <- "iso3"
names(datos)[2] <- "gini"
names(datos)[3] <- "schooling"
names(datos)[4] <- "popgrowth"
names(datos)[5] <- "I2GDP"
names(datos)[6] <- "trade"
names(datos)[7] <- "rateschooling"
names(datos)[8] <- "priceinv"
names(datos)[9] <- "gov"
```

```
datos <- datos[!is.na(datos$gini),]
datos <- datos[!is.na(datos$schooling),]
datos <- datos[!is.na(datos$popgrowth),]
datos <- datos[!is.na(datos$I2GDP),]
datos <- datos[!is.na(datos$trade),]
datos <- datos[!is.na(datos$rateschooling),]
datos <- datos[!is.na(datos$priceinv),]
datos <- datos[!is.na(datos$gov),]
```

Data transformation

```
gdp17 <- filter(macro, macro$iso3 %in% datos$iso3 & macro$year == 2017)
gdp17 <- log(gdp17$rgdpe/gdp17$pop)
gdp85 <- filter(macro, macro$iso3 %in% datos$iso3 & macro$year>= 1985 & macro$year <= 1990)
gdp85 <- data.frame(aggregate(rgdpe ~ iso3, gdp85, mean), aggregate(pop ~ iso3, gdp85, mean))
gdp85 <- gdp85$rgdpe/gdp85$pop
growth <- (gdp17 - log(gdp85))/(2017-1985)

datos <- cbind(datos, gdp85, growth)
datos <- datos[!is.na(datos$growth),]
datos <- cbind(datos, datos$gini*datos$gdp85, datos$popgrowth/100 + 0.05)
names(datos)[12] <- "ginixgdp"
names(datos)[13] <- "popgrowth"
datos[4] <- NULL
datos[8] <- NULL

rm(gov, I2GDP, muestra, popgrowth, priceinv, rateschooling, schooling, trade, gdp17, gdp85, growth)
```

Export of descriptive statistics in an academic format with the stargazer package

```
#####Descriptive analisis ##############################################################

##### GENERAL SAMPLE

gni <- filter(macro, macro$iso3 %in% datos$iso3 & macro$year == 2017)
gni <- gni$GNIpc
datos <- cbind(datos, gni)
datos <- datos[!is.na(datos$gni),]

low <- filter(datos, datos$gni < 3955) ## filters low income countries
middle <- filter(datos, datos$gni > 3956 & datos$gni < 12235)
high <- filter(datos, datos$gni > 12235)


# correlation matrix - multicolinearity general sample
vars <- c("gdp85", "gini", "schooling", "I2GDP", "popgrowth", "trade", "priceinv")
r <- datos %>%
  dplyr::select(vars) # correlation matrix

correlation.matrix <- round(cor(r, use= "pairwise.complete.obs"), digits = 2)
stargazer(correlation.matrix,
          title="Correlation Matrix",
          out = "correlation.htm") ## exports correlation matrix

# descriptive stats general sample
vars <- c("growth", "gdp85", "gini", "schooling", "I2GDP", "popgrowth", "trade", "priceinv")
r <- datos %>%
  dplyr::select(vars)

stargazer(r, type = "text",
          title="Descriptive statistics",
          digits=1, out="table1.htm") ## exports descriptive stats

##### LOW INCOME COUNTRIES

# correlation matrix - multicolinearity
vars <- c("gdp85", "gini", "schooling", "I2GDP", "popgrowth", "trade", "priceinv")
r <- low %>% dplyr::select(vars) # correlation matrix

correlation.matrix <- round(cor(r, use= "pairwise.complete.obs"), digits = 2)
stargazer(correlation.matrix, title="Correlation Matrix", out = "correlationlow.htm") ## exports
correlation matrix
vars <- c("growth", "gdp85", "gini", "schooling", "I2GDP", "popgrowth", "trade", "priceinv")
r <- low %>% dplyr::select(vars)

stargazer(r, type = "text", title="Descriptive statistics", digits=1, out="tablelow.htm") ## export
descriptive stats

##### MIDDLE INCOME COUNTRIES

# correlation matrix - multicolinearity
vars <- c("gdp85", "gini", "schooling", "I2GDP", "popgrowth", "trade", "priceinv")
```

```
r <- middle %>% dplyr::select(vars) # correlation matrix

correlation.matrix <- round(cor(r, use= "pairwise.complete.obs"), digits = 2)
stargazer(correlation.matrix, title="Correlation Matrix", out = "correlationmiddle.htm") ## exports
correlation matrix
vars <- c("growth", "gdp85", "gini", "schooling", "I2GDP", "popgrowth", "trade", "priceinv")
r <- middle %>% dplyr::select(vars)

stargazer(r, type = "text", title="Descriptive statistics", digits=1, out="tablemiddle.htm") ##
export descriptive stats

##### HIGH INCOME COUNTRIES

# correlation matrix - multicolinearity
vars <- c("gdp85", "gini", "schooling", "I2GDP", "popgrowth", "trade", "priceinv")
r <- high %>% dplyr::select(vars) # correlation matrix
correlation.matrix <- round(cor(r, use= "pairwise.complete.obs"), digits = 2)
stargazer(correlation.matrix, title="Correlation Matrix", out = "correlationhigh.htm") ## exports
correlation matrix
vars <- c("growth", "gdp85", "gini", "schooling", "I2GDP", "popgrowth", "trade", "priceinv")
r <- high %>% dplyr::select(vars)
stargazer(r, type = "text", title="Descriptive statistics", digits=1, out="tablehigh.htm") ## export
descriptive stats

rm(correlation.matrix, low, middle, high, r, gni, vars)
```

World sample model and OLS assumptions

```
###### Model world ##################################################

macro <- read_excel("macro.xlsx") ## Cargar banco de datos WB y PWT9
datos <- read_excel("datos con log.xlsx")

gni <- filter(macro, macro$iso3 %in% datos$iso3 & macro$year == 2017)
gni <- gni$GNIpc
datos <- cbind(datos, gni)
datos <- datos[!is.na(datos$gni),]

mankiew1 <- lm(data = datos, growth ~ gdp85 + I2GDP + popgrowth + schooling)
VIF(mankiew1)
mankiew2 <- lm(data = datos, growth ~ gdp85 + gini + I2GDP + popgrowth + schooling)
VIF(mankiew2)
mankiew3 <- lm(data = datos, growth ~ gdp85 + gini + I2GDP + popgrowth + schooling + trade + priceinv
)
VIF(mankiew3)
mankiew4 <- lm(data = datos, growth ~ gdp85 + gini + I2GDP + popgrowth + trade + priceinv ) ##
schooling is taken off as it shows a high correlation with initial GDPpc
VIF(mankiew4)

stargazer(mankiew1, mankiew2, mankiew3, mankiew4, type="html",
          dep.var.labels=c("Average Real GDP per capita growth rate (PPP)"),
          out="model.htm")


#### beta convergence for all the countries

beta <- read_excel("beta.xlsx")
growth <- beta$growth
gdp85 <- beta$gdp85
countries <- countrycode(sourcevar = beta$iso3, origin = "iso3c",destination = "iso2c")
countries <- tolower(countries)
class <- beta$classification
df <- data.frame(growth, countries, gdp85, class)

ggplot(df, aes(x=gdp85, y=growth)) +
  geom_flag(mapping = aes(country=countries), size = 6) +
  geom_smooth(method=lm) +
  facet_grid(~ class) +
  scale_size(range = c(0, 7)) +
  labs(x = "ln(GDP per capita in 1985)",
       y = "Average GDP per capita growth rate")

## OLS assumptions MODEL world #################################################################

# Linearity of the data
plot(mankiew3, 1)


# distribution of studentized residuals
sresid <- studres(mankiew3)
shapiro.test(sresid)
hist(sresid)
plot(density(sresid))
```

```
# homoscedasticity

plot(mankiew3, 3)
# non-constant error variance test
ncvTest(mankiew3) ## http://math.furman.edu/~dcs/courses/math47/R/library/car/html/ncv.test.html

# Variation inflation factor
VIF(mankiew3)
```

Subsamples modelling

```
## Model low income #####################################################################

## classification of countries: https://blogs.worldbank.org/opendata/new-country-classifications-
income-level-2017-2018

low <- filter(datos, datos$gni < 3955)

mankiew1 <- lm(data = low, growth ~ gdp85 + I2GDP + popgrowth + schooling)
VIF(mankiew1)
mankiew2 <- lm(data = low, growth ~ gdp85 + gini + I2GDP + popgrowth + schooling)
VIF(mankiew2)
mankiew3 <- lm(data = low, growth ~ gdp85 + gini + I2GDP + popgrowth + schooling + trade + priceinv )
VIF(mankiew3)
mankiew4 <- lm(data = low, growth ~ gdp85 + gini + I2GDP + popgrowth + trade + priceinv ) ##
schooling is taken off as it shows a high correlation with initial GDPpc
VIF(mankiew4)

stargazer(mankiew1, mankiew2, mankiew3, mankiew4, type="html",
          dep.var.labels=c("Average Real GDP per capita growth rate (PPP)"),
          out="model2.htm")

## List of middle countries in the sample:

countries <- data.frame(cbind(low$gni, countrycode(sourcevar = low$iso3, origin = "iso3c",destination
= "country.name")))
write_xlsx(countries, "C:\\Users\\Jose Caloca\\Desktop\\banco de datos\\low countries.xlsx")

## OLS assumptions MODEL low income ##############################################################

# Linearity of the data
plot(mankiew3, 1)


# distribution of studentized residuals
sresid <- studres(mankiew3)
shapiro.test(sresid)
hist(sresid)
plot(density(sresid))


# homoscedasticity

plot(mankiew3, 3)
# non-constant error variance test
ncvTest(mankiew3) ## http://math.furman.edu/~dcs/courses/math47/R/library/car/html/ncv.test.html

# Variation inflation factor
VIF(mankiew3)

## Model middle income #####################################################################

middle <- filter(datos, datos$gni > 3956 & datos$gni < 12235)

mankiew1 <- lm(data = middle, growth ~ gdp85 + I2GDP + popgrowth + schooling)
VIF(mankiew1)
mankiew2 <- lm(data = middle, growth ~ gdp85 + gini + I2GDP + popgrowth + schooling)
VIF(mankiew2)
mankiew3 <- lm(data = middle, growth ~ gdp85 + gini + I2GDP + popgrowth + schooling + trade +
priceinv )
VIF(mankiew3)
mankiew4 <- lm(data = middle, growth ~ gdp85 + gini + I2GDP + schooling + trade + priceinv ) ##
schooling is taken off as it shows a high correlation with initial GDPpc
VIF(mankiew4)

stargazer(mankiew1, mankiew2, mankiew3, mankiew4, type="html",
          dep.var.labels=c("Average Real GDP per capita growth rate (PPP)"),
          out="model3.htm")

## List of middle countries in the sample:

countries <- data.frame(cbind(middle$gni, countrycode(sourcevar = middle$iso3, origin =
"iso3c",destination = "country.name")))
write_xlsx(countries, "C:\\Users\\Jose Caloca\\Desktop\\banco de datos\\middle countries.xlsx")
```

```
## OLS assumptions MODEL middle income ############################################################

# Linearity of the data
plot(mankiew3, 1)

# distribution of studentized residuals
plot(mankiew3, 2)
sresid <- studres(mankiew3)
shapiro.test(sresid)
hist(sresid)
plot(density(sresid))

# homoscedasticity

plot(mankiew3, 3)
# non-constant error variance test
ncvTest(mankiew3) ## http://math.furman.edu/~dcs/courses/math47/R/library/car/html/ncv.test.html

# Variation inflation factor
VIF(mankiew3)

## Model high income #####################################################################
high <- filter(datos, datos$gni > 12235)

mankiew1 <- lm(data = high, growth ~ gdp85 + I2GDP + popgrowth + schooling)
VIF(mankiew1)
mankiew2 <- lm(data = high, growth ~ gdp85 + gini + I2GDP + popgrowth + schooling)
VIF(mankiew2)
mankiew3 <- lm(data = high, growth ~ gdp85 + gini + I2GDP + popgrowth + schooling + trade + priceinv
)
VIF(mankiew3)
mankiew4 <- lm(data = high, growth ~ gdp85 + gini + I2GDP + popgrowth + trade + priceinv ) ##
schooling is taken off as it shows a high correlation with initial GDPpc
VIF(mankiew4)

stargazer(mankiew1, mankiew2, mankiew3, mankiew4, type="html",
          dep.var.labels=c("Average Real GDP per capita growth rate (PPP)"),
          out="model4.htm")

### regression plot growth and gini
summary(lm(data = high, growth ~ gini))
growth <- high$growth
gini <- high$gini
countries <- countrycode(sourcevar = high$iso3, origin = "iso3c",destination = "iso2c")
countries <- tolower(countries)
df <- data.frame(growth, countries, gini)

ggplot(df, aes(x=gini, y=growth, country=countries)) +
  geom_flag() +
  scale_country() +
  geom_abline(intercept = 0.0145600, slope = 0.0003426  , colour = "black", size = 1) +
  scale_size(range = c(0, 7))


## List of rich countries in the sample:

countries <- data.frame(cbind(high$gni, countrycode(sourcevar = high$iso3, origin =
"iso3c",destination = "country.name")))
write_xlsx(countries, "C:\\Users\\Jose Caloca\\Desktop\\banco de datos\\rich countries.xlsx")

## OLS assumptions MODEL high income #####################################################

# Linearity of the data
plot(mankiew3, 1)


# distribution of studentized residuals
plot(mankiew3, 2)
sresid <- studres(mankiew3)
shapiro.test(sresid)
hist(sresid)
plot(density(sresid))

# homoscedasticity

plot(mankiew3, 3)
# non-constant error variance test
ncvTest(mankiew3) ## http://math.furman.edu/~dcs/courses/math47/R/library/car/html/ncv.test.html

# Variation inflation factor
VIF(mankiew3)
```