TÉCNICO LISBOA

New Thermodynamic Measures of Inequality

Dulce de Jesus Costa

Thesis to obtain the Master of Science Degree in

Engineering Physics

Supervisor(s): Prof. Henrique Manuel dos Santos Silveira de Oliveira
Prof. Luis Humberto Viseu Melo

Examination Committee

Chairperson: Prof. Carlos Manuel Dos Santos Rodrigues da Cruz
Supervisor: Prof. Luís Humberto Viseu Melo
Member of the Committee: Prof. António Miguel Areias Dias Amaral

May 2021

”Say what you want to say,
and the let the words fall out.”

Acknowledgments

A few people were fundamental to the successful elaboration of this thesis and I use this section to acknowledge their contribution.

First, I'd like to thank my friends for their support and patience, for always kept me focused and motivated during this process. A special thanks to Rita, Tatiana, Soraia and Sofia for all the help in reviewing the work and encouragement they gave me to go through with this thesis.

I also thank my family for always supporting me and giving me everything I needed to get an education and all the support for finishing this last task in the university phase.

To Ana Banha, thank you so much for show me that you will never give up on me.

Last but not least I thank my supervisor, Henrique Oliveira, for everything he did for me. For all the knowledge transmitted, the ideas and advice he gave me and for always being patient and understanding of my lack of time due to my job. A simple thank you is not enough. Also I thank Professor Luis Melo for not giving up on me, joining to this idea and supporting my work.

Resumo

Existem actualmente diversas medidas de desigualdade utilizadas para analisar e comparar as diferenças de distribuição de rendimentos entre países/populações.

Com o objectivo de compreender as raizes da desigualdade, este trabalho apresenta um novo método complementar ao Índice de Gini. O Gini é uma das medidas mais utilizadas mas tem limitações relativamente às raizes de desiguldade em distribuições que se traduzem num mesmo Indice de Gini. Esta nova métrica é uma medida alternativa de segunda ordem, que tem por base medidas já conhecidas e a variação da entropia. O trabalho irá focar-se na aplicação e comportamento dos novos coeficientes em casos reais, utilizando para isso informação relativa a rendimentos, disponível em bases de dados mundiais (WIID). A desigualdade de distribuições de rendimentos de Portugal, Europa e OCDE ao longo dos anos e entre países serão testadas usando os novos coeficientes, assim como serão apresentados os resultados para países com maior desigualdade.

Estes novos coeficientes (GMais e GMenos) mostram-se vantajosos na análise de desigualdade de distribuições de rendimentos, principalmente no que diz respeito às caudas da distribuição, sendo portanto boas ferramentas auxiliares para a comparação de distribuições com o mesmo coeficiente de Gini.

Quando combinamos o GMais e o GMenos é possivel perceber se a desigualdade se deve a um aumento de ricos, de pobres ou a um aumento/diminuição da classe média. A utilização destes dois novos indicadores possibilita assim a apresentação de resultados de uma forma mais intuitiva, sem a necessidade de complementos gráficos ou extensas tabelas e valores.

Palavras-chave: Medidas de Desigualdade, Rendimentos, Entropia, Coeficiente de Gini, GMais e GMenos, Curvas de Lorenz

Abstract

There are currently several measures of inequality that can be used to analyze and compare differences in income distribution between countries and populations.

This work presents a new method for understanding the inequality source complementing Gini's coefficient. Gini's coefficient is one of the most widely used inequality measures, but fails to resolve between sources of inequality leading to the same value. This new method is a second order measure based on current inequality measures in use and the entropy variation. The work will focus on the application and behavior of the new coefficients in real cases, using information related to yields, available in worldwide databases. The inequality of income distributions from Portugal, Europe and OECD over the years and between countries will be tested using the new coefficients, as well as the results for countries with greater inequality will be presented.

These new coefficients (GPlus and GMinus) prove to be advantageous in the analysis of inequality of income distributions, mainly with regard to the distribution tails, therefore being good auxiliary tools for comparing distributions with the same Gini coefficient. When we combine GPlus and GMinus, it is possible to see if inequality is due to an increase in the rich, the poor or if there is an increase or decrease in the middle class.

The use of these two new indicators makes it possible to present results in an easier and more intuitive way, without the need for graphical supplements or with extensive tables and values.

Keywords: Inequality measures, Income, Entropy, Gini Index, GPlus and GMinus, Lorenz Curves

Contents

List of Tables

List of Figures

Nomenclature

∞ Infinity.

G Gini Index.

$G+$ GPlus coefficient.

$G-$ GMinus coefficient.

GDP Gross domestic product.

K Entropy unit - Boltzmann's constant.

MLF Middle Layer Force.

$S(W)$ Entropy.

x, y, i Normally used to represent income or computational indexes.

Chapter 1

Introduction

In the last couple of decades, throughout the world, social inequalities have grown [1] [2].

There is a need to understand the social processes that sustain inequality and develop social policies that can reduce it, and in order to do that, we need to understand and measure inequality [1] [3].

Inequality is not a self-defining concept, as its definition may depend on economic interpretations as well as ideological and intellectual positions [4]. If we compare total income to a cake, inequality can be defined as the way that a cake is divided among individuals.

Economic attitudes, as well as ideological and intellectual positions, may cause differing views about the size of inequality, its relevance and policies that might be implemented to deal with it. Furthermore, income inequality might be seen as a part of a more general concept of economic inequality, even though income conditions are often a good proxy for economic conditions because income shapes people's living standards and it is generally highly correlated with other well-being indicators (eg: health, education) [2]. Economic inequality can be viewed as part of a dynamic system, where endowments and opportunities lead to differences in economic outcomes (Figure 1.1) [5].

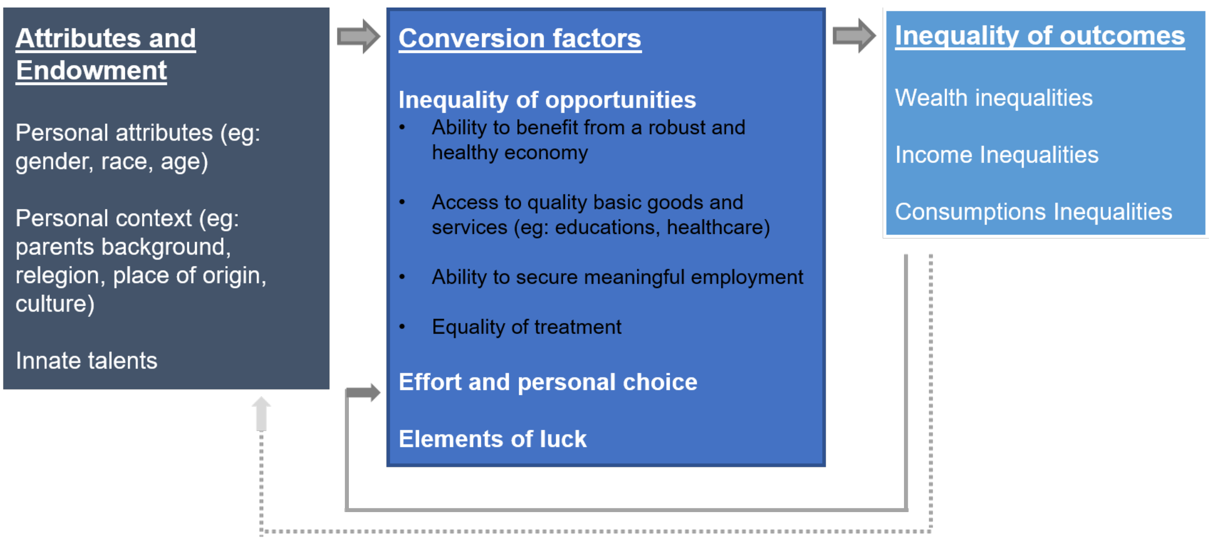

Figure 1.1: Inequalities as a Dynamic System with several sources. Source: McKinsey Global Institute analysis [5].

Inequality is not a new phenomenon, but it has become one of the burning social and economic issues of our time, prompting a polarized debate about causes and solutions and fueling public discontent with governments and other institutions [2]. Studying inequality can break down some barriers, contributing to a more equal society and a healthier world [5].

Recently, there has been a lot of concern about the increase of inequality in income and wealth and as such, the interest in this subject among the leading social scientists across the globe has been renewed [1] [2]. In the economics front, how inequalities affect financial markets, firms and their dynamics, and vice versa is an important area to look into [5]. Many deeper and important issues in society still need attention in terms of inequality research, as per example gender equality regarding income and education. Available data and its further analysis can bring out hidden patterns which may be used to counter those situations [1]. The main handicap is the lack of good quality data, and the abstractness of the issues, which may not always be easily amenable to statistical modeling [6] [7] [2].

Many laws of nature are statistical in origin, a well established fact that applies across most areas of modern physics, which gives statistical physics the status of a prominent and extremely useful discipline [8]. The subject, being developed in a very general framework, makes it applicable to various areas outside the commonly perceived boundaries of physics, as in biology, computer science and information technology. Statistical physics treats its basic entities as particles, while in society the basic constituents are human individuals [9].

Physicists' interests are mostly concentrated on subjects that are amenable to macroscopic or microscopic modeling, where tools of statistical physics prove useful in explaining the emergence of broad distributions. A huge body of literature has been already developed, containing serious attempts to understand socio-economic phenomena, branded under Econophysics and Sociophysics [8] [10] [11] [12]. The physics perspective brings new ideas and an alternative outlook to the traditional approach taken by social scientists, and is seen in the increasing collaborations across disciplines [9].

When we talk about inequality we can easily associate it with a disorder, not only because there are unequal distributions but also because the political opinions are controversial. With tools of statistical physics as a core, the knowledge and techniques from various disciplines like statistics, applied mathematics, information theory and computer science are incorporated for a better understanding of the nature and origin of socio-economic inequalities that shape Humankind.

Measuring inequality is important for answering a wide range of questions. For instance:

- Is the income distribution more equal than what it was in the past?
- Are underdeveloped countries characterized by greater inequality than developed countries?
- Do taxes or other kinds of policy interventions lead to greater equality in the distribution of income or wealth?
- Is the inequality relative to the poor or to the rich?

Since the way inequality is measured also determines how the above questions (among others) are answered, a rigorous discussion of the measurement of inequality is necessary.

It is known that there are no perfect inequality measures [13], and, in order to contribute to this research, this work aims to develop and test a second order measure, using physical entropy concepts allied with the already known measures. The entropic measures are versatile and can deal with the types of curves needed for this task [12]. The goal is that this new measure can be used as a complementary tool to give us more information about the sources of inequality, given the Gini Index fails to resolve between sources of inequality leading to the same value.

In the following chapters the key ideas about entropy - new coefficients basis - will be presented. The current inequality measures in use, their applications and some history and limitations will also be presented. Lorenz Curves and Gini Index will be analysed for some real world cases, showing the importance of having a new complementary measure. The axioms for a good inequality measure will be introduced and applied to the new coefficients created. Results for new coefficients will be showed using toy models and real world cases.

Chapter 2

Background

In the following subsections, an overview of the main components related to this work will be discussed. Starting from a brief explanation about entropy, followed by inequality measures, this section will provide the necessary background to understand the main concepts used to develop the new coefficients and the need to create them.

2.1 Entropy

Several problems from Economics, Financial, or Social areas are analysed using concepts and methods from Physics [8] [10]. The application of statistical physics methods to these subjects promises fresh insights into several problems. One of the examples of a physical concept that is being applied in a range of fields is entropy.

In physics, the entropy of a physical system is proportional to the quantity of energy no longer available to do physical work - the center of the second law of thermodynamics, which states that in an isolated system any activity increases the entropy [14]. In probability theory, the entropy of a random variable measures the uncertainty about the value that might be assumed by the variable [10]. In information theory, the compression entropy of a message (e.g. a computer file) quantifies the information content carried by the message in terms of the best lossless compression rate [15]. In the theory of dynamical systems, entropy quantifies the exponential complexity of a dynamical system or the average flow of information per unit of time. In sociology, entropy is the natural decay of structure (such as law, organization, and convention) in a social system [10]. In common sense, entropy means disorder or chaos.

The entropy concept was first raised in response to a puzzle: the loss of energy. From the XVII century, it was noticed that heat-powered engines such as Thomas Savery's (1698), the Newcomen engine (1712) and the Cugnot steam tricycle (1769) were inefficient, converting less than two percent of the input energy into useful work. The observation that a certain amount of functional energy is always lost to dissipation or friction and is thus not transformed into useful work leads physicists to solve the charade [16].

Around 1850, Rudolf Clausius [17] presented the concept of the thermodynamic system and placed the argument that in any irreversible process a small amount of heat energy δQ is incrementally dissipated across the system boundary. Clausius did several studies about the loss of energy, leading to the development of the second law of thermodynamics and he created the term ”entropy”.

From Clausius' formulas, laws and theorems, several investigators developed new approaches related to entropy. In contrast to Clausius, Boltzmann [14] attempts not to combine macroscopic and microscopic points of view but to determine the macroscopic behavior of the particles. He established the order and disorder possibilities of a molecule arrangement, noting that the most probable state of a gas in a container is such that the molecules do not retain any specific configuration, but rather, are totally random.

Entropy can be associated with system or information disorder with implications for a diverse range of problems involving information, uncertainty, and risk, and it has been used for diverse applications in economic and financial markets [18]. For example, entropy of the distribution of intraday returns is shown to be a good predictor for daily value at risk [19]. Entropy has powerful applications in portfolio selection and asset pricing [20]. In spatial economics and economic geography, entropy is used to measure firm specialization and cluster-based measurement of agglomeration and sector decomposition [21]. Entropy has deep interconnection with resource allocation and our understanding of economic systems, their stability, and complexity [22]. The concept of entropy in the analysis of economic problems was first introduced by Georgescu-Roegen, in his book, ”The Entropy Law and the Economic Process” (1971) [23]. From there, several authors use Entropy to help in Economic problems, creating new models that allow us to understand some economic issues [8]. This work is fully aligned with the necessity of understanding a part of economic inequalities and for that reason, it is a new application of the entropy concept in this field.

One of the most well-known entropy measures is applied in the field of information theory, describing an analogous loss of data in information transmission systems and it was first introduced in the late 1940s by an electrical engineer at Bell Labs called Claude Shannon. Shannon entropy [15] was further developed into a relative measure of entropy by Kullback and Leibler [24].

From Shannon and Boltzman's ideas, and to quantify the species' capacity of increase and pull in some evolutionary extensions and interpretations, Demetrius introduced demographic entropy as a measure of species adaptation and founded the directionality theory [12]. In brief, the concept invokes the fundamental attributes of demographic entropy, as a measure of Darwinian fitness of age-structured populations, to study the changes in genotypic and phenotypic composition as generated by the mutation-selection regime [25]. Demetrius's work is the basis for this research and his idea is described below, together with the entropy theory pillars.

2.1.1 Boltzmann's Entropy

In 1877, Boltzmann [26] for the first time explained what entropy is and why, according to the 2^{nd} Law of Thermodynamics, entropy increases. Boltzmann also showed that there were three contributions to entropy: from the motion of atoms (heat), from the distribution of atoms in space (position), and radiation (photon entropy).

Considering a classical system consisting of N elements (organisms, cells...) classified into n classes (species, states...), with N_i $(i = 1, 2...n)$ the occupation numbers of the i^{th} class, the macrostate of the system is given by the set of occupation numbers A_n=$N_1, N_2...N_n$. The statistical weight/thermodynamical probability of the macrostate A_n is given by equation 2.1:

$$W(A_n) = \frac{N!}{\prod_{i=1}^{n} N_i!}. \tag{2.1}$$

This is the total number of microscopic states compatible with the constraints of the system. The distribution of the elements among the different classes possesses a great deal of disorder and the statistical measure of this disorder is given by Boltzmann entropy [11]:

$$S = K \ln W(A_n), \tag{2.2}$$

where K is a constant and depends on the unit of measurement of entropy. In thermodynamics K is called Boltzmann constant. For large N_i, the Boltzmann entropy (Equation 2.2) reduces to the form in Equation 2.3:

$$S = -KN \sum_{i=1}^{n} p_i \ln(p_i). \tag{2.3}$$

Note, these equations follow the two fundamental properties of thermodynamic entropy (and common sense):

- Axiom 1)
  The entropy $S(W)$ of a system is a positive increasing function of the disorder W, ie:

  $$S(W) \leq S(W + 1). \tag{2.4}$$

- Axiom 2)
  The entropy $S(W)$ is assumed to be an additive function of the disorder W. That is, for any two statistically independent systems with degrees of disorder W1 and W2 respectively, the entropy of the composite system is given by:

  $$S(W1 \cdot W2) = S(W1) + S(W2). \tag{2.5}$$

  This axiom is the statistical analogue of the thermodynamic postulate of the additivity of entropy.

These formulations are the basis for the numerous derivations of entropy developed by several authors, as the case of Shannon entropy described in the next subsection.

2.1.2 Shannon's Entropy

The goal of Shannon was to construct a measure for the amount of information associated with a message.

The Shannon entropy measure S(p), for a probability distribution of a random variable with discrete values defined over $x \in X$, where X is a random variable with "n" possible outcomes and probability $(p_i) = \frac{1}{n}$ for the ith outcome, $1 \leq i \leq n$. The entropy is thus defined as [15]:

$$S(p) = -\sum_{i=1}^{n} p_i log_2(p_i). \tag{2.6}$$

This entropy measure takes on values larger than 0. Its value is zero if $(p_i) = 1$ for an $i \in \{1, 2, ..., n\}$, and is maximum for equally likely outcomes of $(p_i) = \frac{1}{n}$ for all i. The Shannon entropy expression can be rewritten as:

$$S(P) = -\frac{1}{ln(2)} \sum_{i=1}^{n} p_i ln(p_i). \tag{2.7}$$

Note the use of logarithmic base 2 instead of a natural base, as the goal of this entropy measure is to be used in information systems (bits).

2.1.3 Demetrius's entropy

Demetrius (1974) [12] brought out the concept of population entropy H as a summary statistic to define species fitness. In practical terms, the maximization of population entropy H under various constraints yields to distributions of reproduction and survivorship. By considering that l_x stands for the probability that an individual born at age zero survives to age x, and that m_x is the age-specific fecundity of cohort life table, equations and R0 the net reproductive rate, and so the population entropy can be derived as:

$$H = -\int_0^{\infty} q(x) \log q(x) dx, \tag{2.8}$$

where

$$q(x) = \frac{l_x m_x}{R_o}, \tag{2.9}$$

is the probability density function of the age of reproducing individuals. The net reproductive rate (R_0), If $l_x m_x$ is the net maternity function and for fixed $m_x = 1$, then H is given by:

$$H = -\int_0^{\infty} \frac{l_x}{e_0} \log \frac{l_x}{e_0} dx = -\frac{\int_0^{\infty} l_x \log l_x dx}{\int_0^{\infty} l_x dx} + \log e_0. \tag{2.10}$$

where e_0 is the life expectancy:

$$e_0 = \int_0^{\infty} l_x dx. \tag{2.11}$$

If we define the normalized entropy H^* as 2.12 it's possible to define two forms of entropy [27]. The first one, H^*, corresponds to the population in cases where no environmental forces interfere. The second, $H^{'}$, consists of the conditional entropy, which is addressed to a perturbed population related entropy given that unperturbed mortality is known.

$$H^* = -\frac{\int_0^\infty l_x \, log l_x dx}{\int_0^\infty l_x dx}. \tag{2.12}$$

The relation: $\Psi = H^* + H^{'}$ consists of the adaptive value of a given population and can be interpreted as a measure of correlation between the variability of the mortality distribution and environmental variability [12]. Equation 2.10 can be regarded as an analogue of the Boltzmann–Gibbs definition of the entropy of a thermodynamic system. The numerator is analogous to the Shannon–Weaver measure of the amount of information in a message; it measures the variability of the contribution of different age classes to the stationary population. The denominator is the average maternal age, which is a natural measure of generation time.

From an information theoretical standpoint, where entropy is referred to as a measure of information, minimizing H yields minimum uncertainty (high information and population stability), and while maximum H represents maximum uncertainty (low information and population stability).

Concepts of entropy are really important when dealing with statistical data and have the main part in world inequality analysis. In the following sections, several measures of inequality will be presented, as well as the axioms that grant a good measure.

2.2 Income Inequality

The dictionary defines ”inequality” as ”social or economic disparity between people or groups or the condition or an instance of not being equal” [28]. Economic income inequality measures the disparity between a percentage of the population and percentage of income received by that population [29]. When disparity moves up, inequality rises. From the definition, we can have two marginal situations: inequality minimum or equality (each person from the population holds the same amount of income) and inequality maximum (when population income is in the hands of a single person). Income encompasses labor earnings (such as wages, salaries, and bonuses), capital income derived from dividends, interest on savings accounts, rent from real estate, as well as welfare benefits, state pensions, and other government transfers [30]. In addition, it is possible to distinguish between individual versus family income, pre-tax versus after-tax (disposable) income, and labor earnings versus capital income [13]. In more-developed countries, wages and salaries are the major sources of income for most households, while property, including capital gains, is the major source for the richest [1] [30]. Income inequality can be studied within countries, between countries, or across the world's population (Figure 2.1).

Despite popular belief that income inequality can be derived from individual differences in talent and motivation, there are significant structural and cultural causes for inequality, such as discrimination, racism and sexism, gender roles, and family responsibilities, that can lead to segmented labour markets

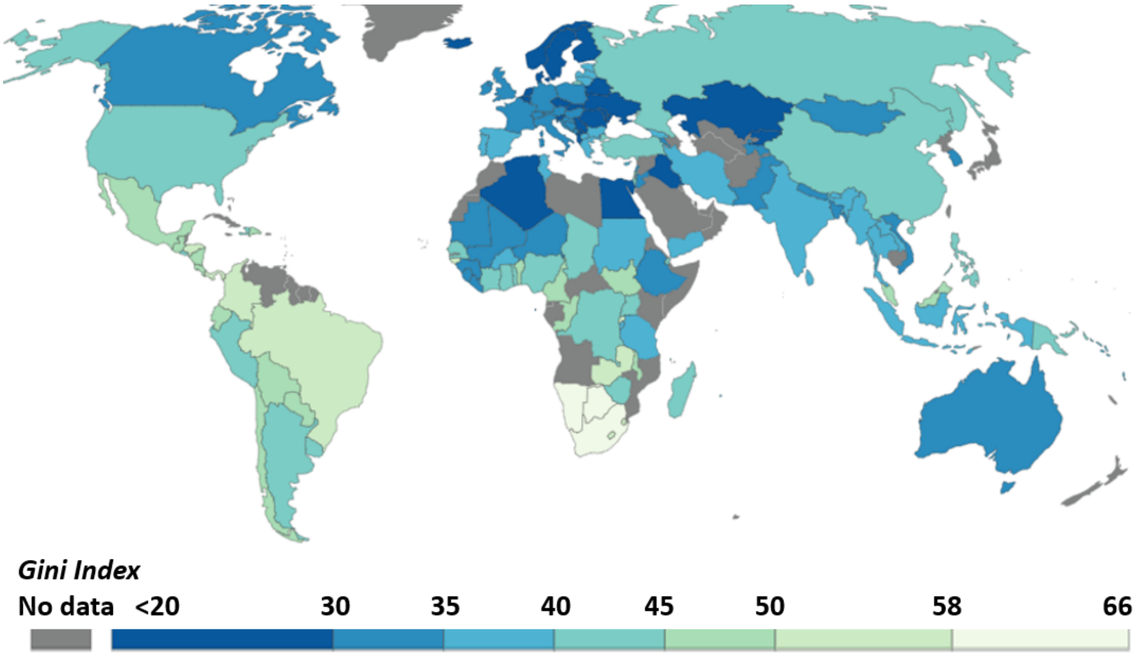

Figure 2.1: World Income Inequality in 2013 - Gini Index. Data Source: World Bank [13].

[1] [30]. In Figure 2.2 we can see the influence of race in income obtained in the USA. Other legal, political, and economic factors also affect income levels independently of individual traits, such as corporate power and the degree of private versus public ownership and control of resources [30] [1].

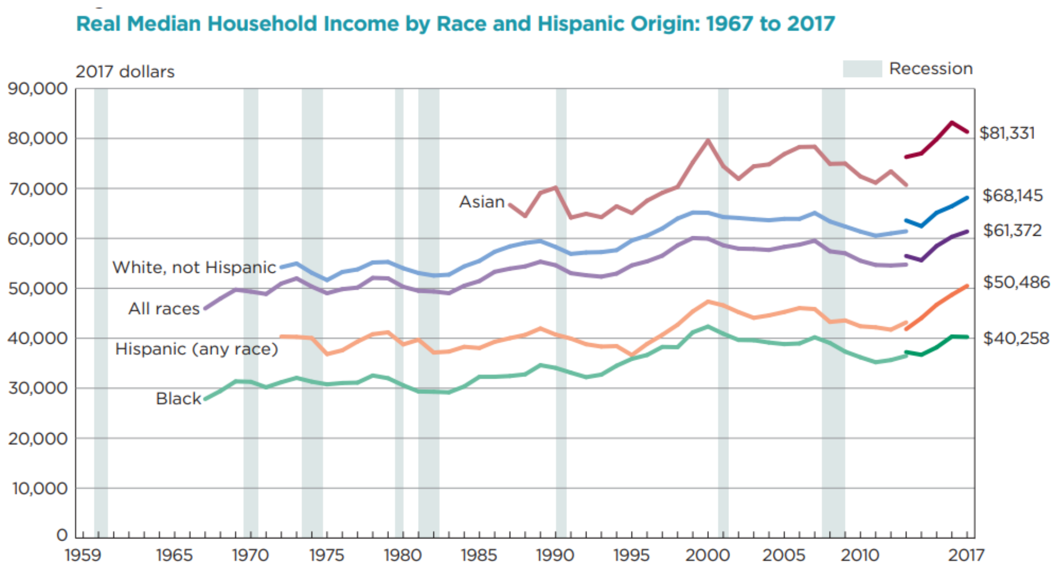

Figure 2.2: Real Median Household Income by Race and Hispanic Origin: 1967 to 2017. Data Source: World Bank [13].

Levels of global inequality remain extreme, with persistent high numbers of people in absolute poverty. According to the World Bank [13], in 2012 nearly 13 percent of the world's population received less than $1.90 per day, and about 35 percent, lived on less than $3.10 per day. Such poverty produces low levels of education, sanitation, nourishment, and medical care and high rates of child labour and exploitation as well as child and infant mortality. The richest 1 percent of the world's population owns more wealth than the rest of the world combined. The assets of the 10 richest billionaires are greater than the GDPs (gross Domestic Product) of most countries, for example, in Figure 2.3, it's seen that the Jeff Bezos (Amazon) and Bill Gates (Microsoft) GDP in 2019 is higher than some European countries.

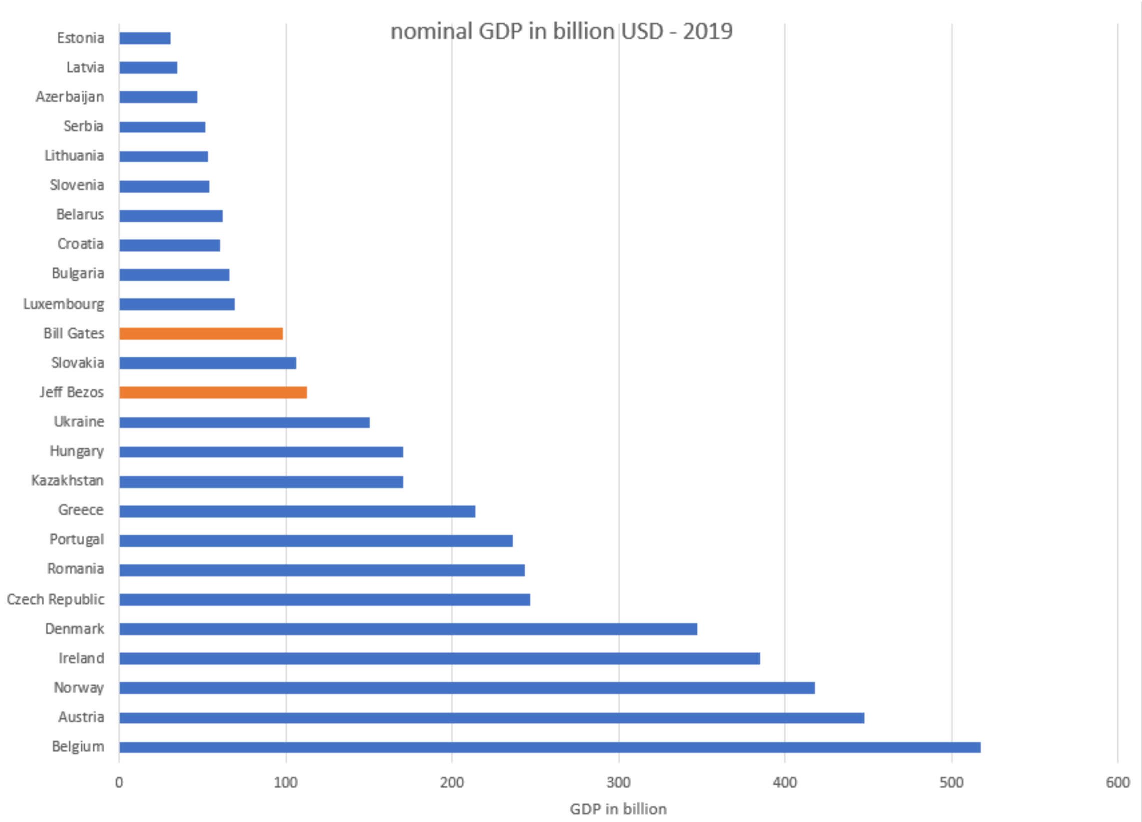

Figure 2.3: Nominal GDP in billion USD in 2019 for several European Countries and two world's richer people. Data Source: WID [31].

Policies to address income inequality can focus narrowly on individual skills, opportunities, and aspirations or may focus more broadly on altering the social, political, and economic structures that create and maintain income inequality [2].

Overall economic inequality is affected by policies that provide public goods, such as health care and education, leaving a larger proportion of individuals' incomes to be spent on other goods. Wealth redistribution through steeper inheritance taxes, promotion of broader ownership (e.g., greater worker ownership), and socialization or redistribution of capital and land equally to all citizens are ways to reduce income inequality indirectly, as they will equalize the unearned income that derives from ownership of wealth. Enforcement of affirmative action and nondiscrimination policies by employers, governments, and educational institutions and policies such as government-subsidized child care that enable people to enter the labour market should also affect income inequality through facilitating greater access to higher-income jobs [2] [1]. Income inequality can be reduced directly by decreasing the incomes of the richest or by increasing the incomes of the poorest. Policies focusing on the latter include increasing employment or wages and transferring income. Because of continued global instability and environmental degradation linked to inequalities in income and other resources, such policy efforts will continue to be critical not only for ethical reasons but also for the sake of national security and global survival [3].

Research on inequality and on the impact of changes in income distribution on economic processes and social conditions has a long history in economics and was prominent in the works of Smith, Ricardo, Mill and Marx [32] [33]. While the topic fell somewhat out of fashion in the last quarter of the twentieth century [34], it has again moved into the centre of political and economic debates. One important reason for the recent revival in interest is new empirical evidence that suggests that high and growing inequality can have adverse effects on macroeconomic stability and can hurt economic growth [33].

Lately, there has emerged a substantial empirical literature about global inequality but this literature has a ”limitation” in that it is focused on the Gini and Theil Indices, which both measure relative differences between incomes (i.e. ratios of incomes to the mean) [3] [35]. Inequality also can be measured in absolute monetary terms, and absolute and relative inequality trends can be quite different. If, for example, the income of the whole population increases by the same percentage, the Gini and Theil coefficients remain constant, even though the absolute income gap increases. The most appropriate measures to estimate the absolute degree of inequality are the Absolute Gini Index and the variance. In this work we will mainly focus on relative inequality.

2.3 Inequality - Distributions and Measures

The issue of inequality in terms of income and wealth is a widely debated subject in economics [35]. Economists and philosophers have pondered over the normative aspects of this problem. More than a century ago, Pareto made extensive studies in Europe and found that wealth distribution follows a power law tail for the richer section of the society, known as the Pareto law [36]. Independently, Gibrat worked on the same problem and he proposed a “law of proportionate effect” [37]. Much later, Champernowne’s systematic approach provided a probabilistic theory to justify Pareto’s claim [10] . Subsequent studies revealed that the distributions of income and wealth possess some globally stable and robust features. In general, the bulk of the distribution of both income and wealth seems to reasonably fit both the log-normal and the Gamma distributions. Economists have a preference for the log-normal distribution, while statisticians and, rather recently, physicists prefer the Gamma distribution for the probability density or Gibbs/ exponential distribution for the corresponding cumulative distribution [38]. Statistical physics tools have helped to formulate these microscopic models, which are simple enough yet rich with socio-economic ingredients [8]. Toy models help in understanding the basic mechanism and bring out the crucial elements that shape the emergent distributions of income and wealth. A variety of models, from zero-intelligence variants to the much complex agent based models incorporating game theory have been proposed and found to be successful in interpreting results. Simple modeling has been found to be effective in understanding how entropy maximization produces distributions dominated by exponential, and also explains the reasons for aggregation at the high range of wealth, including the emergence of the power law Pareto tail [10].

Socio-economic inequalities can be quantified in several ways. The most common measures are absolute, in terms of indices, e.g., Gini, Theil indices [13]. Alternatively one can use a relative measure, by considering the probability distributions of various quantities, whereas the indices can be computed easily from the distributions themselves.

Economists use various metrics for measuring income inequality [35]. The most commonly used measures are the Lorenz curve, the Gini coefficient, decile ratios, the Palma ratio, and the Theil index [39]. These many measures of inequality, when combined, provide nuance and depth to our understanding of how income is distributed. Choosing which measure to use requires understanding the strengths and weaknesses of each, and how they can complement each other to provide a complete picture. We can split the inequality measures in Graphical, Indices and Ratios (Figure 2.4) and an overview of them will be presented in the next subsections [39]. Ratios constitute the most basic inequality measures available as they are simple, direct, easy to understand, and they offer few data and computation challenges. Accordingly, they do not provide as much information as other measures.

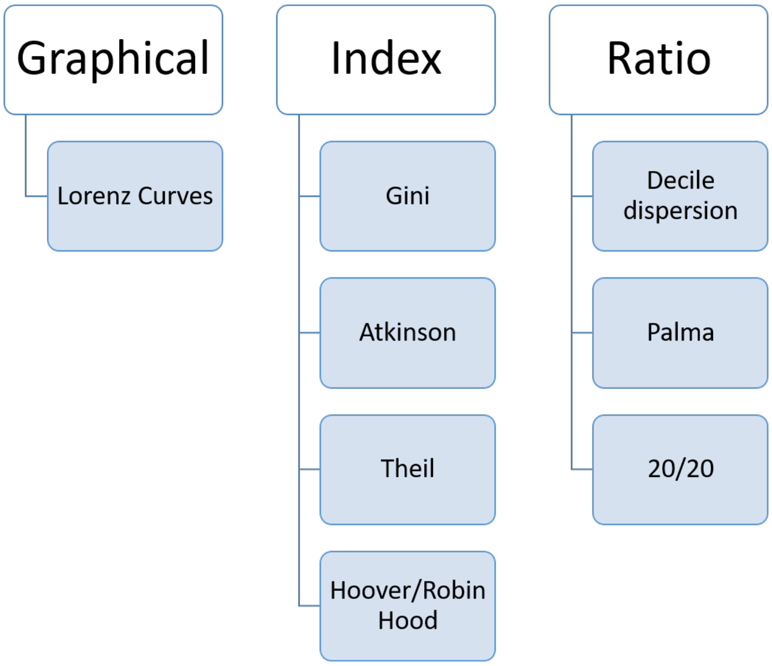

Figure 2.4: Inequality measures examples.

2.3.1 Good Inequality Measures and General Axioms

Axioms, in inequality measurement, are desirable properties of inequality measures. They define the way in which inequality measures should behave. Using axioms may help to choose among inequality indexes. When an inequality index is chosen because it respects some desirable properties, it is said that inequality measurement follows an axiomatic approach [40].

Thon proposes an axiomatization of the Gini coefficient in a general case [41]. He works in a framework where income distributions' aggregate incomes and populations are not necessarily the same. Its axioms are imposed directly over the numerical function representing the inequality index. An axiomatization for the continuous version of the Gini index is provided in Aaberge [42], and it is based on the imposition of axioms to characterize an order over Lorenz curves, viewed as lotteries of the theory of choice under uncertainty. The axiomatization of Weymark [43] presents the Gini coefficient as an element of a parameterized family of inequality indices. His axioms generate social evaluation functions, i.e., representations of social welfare orders, which are transformed into inequality indices through the Atkinson methodology [44]. Below the general axioms for an inequality measure are defined [40] [45].

- **Scale Invariance**

  Scaling the incomes x_i by a constant factor has no effect on the inequality index, so, if the income is measured in other currencies, then the index remains the same. The inequality index I is said to be scale-invariant if $I(x) = I(\lambda x)$ for all $x \in \mathbb{R}^n_+$ and $0 \neq \lambda \in \mathbb{R}$.

- **Symmetry and Anonymity**

  This axiom requires that the inequality index be independent of any characteristic of the individuals other than income x. The inequality index I is symmetric if $I(\theta \cdot x) = I(x)$ for all $\theta \in S_n$ and $x \in \mathbb{R}^n_+$. Also, all permutations of personal labels within a given distribution (i.e. who earns what), should not affect overall inequality.

- **The Pigou-Dalton principle of transfers (PT**)

  This axiom requires the inequality measure to change when income transfers occur among individuals in the income distribution. In particular, Inequality indexes should fall with a progressive transfer (an income transfer from richer to poorer individuals) and Inequality indexes should rise with a regressive transfer (an income transfer from poorer to richer individuals). Let's assume an ordered income distribution $A = \{x_1, x_2, x_3, x_4\}$. If a progressive transfer, T, occurs from x_3 to x_2, the new distribution is given by $B = \{x_1, x_2 + T, x_3 - T, x_4\}$. We say that an Inequality index satisfies the principle of transfers if, in this case, $I(A) > I(B)$ i.e., the second moment gives a lower inequality value.

- **Population Independence**

  The income inequality metric should not depend on whether an economy has a large or small population. An economy with only a few people should not be automatically judged by the metric as being more equal than a large economy with lots of people. This means that the metric should be independent of the level of the population: $I(x \cup x) = I(x)$, where, $(x \cup x)$ is the union of x with a "copy" of itself [46].

- **Decomposability**

  Inequality may occur among different elements of income distribution (e.g., earned income or income from capital) or among different groups of population (e.g. workers, pensioners, agricultural workers, manifacturing workers). In any case, the decomposability axiom requires a consistent relation between overall inequality and its parts. If the original income distribution, A, is composed by n groups, and has an overall inequality $I(A)$, it must be given as $I(A) = I(A)_1 + I(A)_2, ...I(A)_n$, i.e., total inequality must be equal to the sum of the various group inequalities.

2.3.2 Ratios

Ratios constitute the most basic inequality measures available as they are simple, direct, easy to understand, and they offer few data and computation challenges. Accordingly, they do not provide as much information as other measures:

- **Palma ratio**
  It is the ratio of national income shares of the top 10 percent of households to the bottom 40 percent. It is based on economist José Gabriel Palma's empirical observation that difference in the income distribution of different countries (or overtime) is largely the result of changes in the 'tails' of the distribution (the poorest and the richest) as there tends to be relative stability in the share of income that goes to the 'middle' [47].

- **20/20 Ratio**
  It compares the ratio of the average income of the richest 20 percent of the population to the average income of the poorest 20 percent of the population.

- **Decile dispersion ratio (or inter-decile ratio) and Percentiles**
  It is the ratio of the average income of the richest x percent of the population to the average income of the poorest x percent. It expresses the income (or income share) of the rich as a multiple of that of the poor. However, it is vulnerable to extreme values and outliers [40]. Common decile ratios include:

  - D9/D1: ratio of the income of the 10 percent richest to that of the 10 percent poorest;
  - D9/D5: ratio of the income of the 10 percent richest to the income of those at the median of the earnings distribution;
  - D5/D1: ratio of the income of those at the median of the earnings distribution to the 10 percent poorest.

In table 2.1 the income percentage of the top 10% richer in Portugal between 2011 and 2017 is shown. This is also a simple measure of inequality as, if we compare the top 10% in Portugal with another country we can draw several conclusions. In Figure 2.5, a comparison between Portugal and Brazil is presented.

Table 2.1: Pre-tax Portugal income - top 10% share between 2011 and 2017. Data Source: WID [31].

Percentile	**Year**	**Pre-tax national income Top 10% - share Portugal**
p90p100	2011	0.3754
p90p100	2012	0.3668
p90p100	2013	0.3709
p90p100	2014	0.3722
p90p100	2015	0.3663
p90p100	2016	0.3705
p90p100	2017	0.3712

Figure 2.5: Pre-tax income - top 10% share between 2011 and 2015 in Portugal and Brazil. Data Source: WID [31].

2.3.3 Theil's measure and the Generalised Entropy indices

The Generalised Entropy (GE) measures constitute the family of indicators that display additive decomposability as well as anonymity, the population principle, the principle of transfers, and scale invariance [48]. The GE measures depend on a parameter α that expresses the sensitivity of the indicator to different parts of the distribution. The special cases of $\alpha = 1$ and $\alpha = 0$ are known as the Theil index and the mean log deviation, respectively [49]. The case of $\alpha = 2$ is equal to half the squared coefficient of variation. In the case of an empirical distribution with n elements where y_i denotes the income per household i and $\bar{y}$ is the sample average, the Generalised Entropy indices can be expressed as Equation 2.13:

$$GE(\alpha) = \frac{1}{\alpha(\alpha-1)}\frac{1}{n}\sum_{i=1}^{n}((\frac{y_i}{\bar{y}})^{\alpha} - 1). \quad (2.13)$$

The Generalised Entropy indicators are equal to zero in the case of complete equality, i.e. when all individuals have the same wealth/income. A larger value of the index indicates larger inequality in the distribution. The GE is unbounded when considered for a theoretical distribution but is bounded in the case of a finite independent and identically distributed sample.

The special case of Theil's measure ($GE(\alpha = 1)$) was introduced by Theil (1967) as a consequence of Shannon's information theory [50] [48]. In previous chapters we defined general entropy as:

$$S = -k\sum_{i=1}^{n} p_i \log p_i, \quad (2.14)$$

where p_i is the probability of finding member i from a random sample of the population and k is Boltzmann's constant.

When talking about Shannon entropy (information theory), K is equal to 1, and (because of bits) the log base is 2. In Theil index, the natural logarithm was chosen as the base. To convert the Shannon entropy [15] to deal with income per person (y_i), the probability defined in (2.14) needs to be normalized, by dividing by the total population income, and we obtain the Theil's entropy [50]:

$$S_{Theil} = \sum_{i=1}^{n} \frac{y_i}{N\bar{y}} \ln(\frac{N\bar{y}}{y_i}). \tag{2.15}$$

The Theil index is then obtained by the difference between the maximum possible entropy and the observed entropy of the distribution ($S_{Max} - S_{Theil}$). The index is a "negative entropy" as it gets smaller as the disorder gets larger. In this case, maximum entropy is when all incomes are equal, so the average income is equal to a singular income, which leads us to:

$$S_{Max} = \ln N, \tag{2.16}$$

and this lead us to the special GE case, or Theil index (T_T), defined in Equation 2.17 [48].

$$GE(1) = T_T = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} \ln(\frac{y_i}{\bar{y}}). \tag{2.17}$$

Figure 2.6 shows a comparison of Theil and Gini index (that will be described in the following chapters) in Portugal during the last years and it is possible to note the variations are quite equivalent between the two measures. Note however, for some years, the coefficients not behave in line with each other but the variations are so small that it is not possible to draw meaningful conclusions.

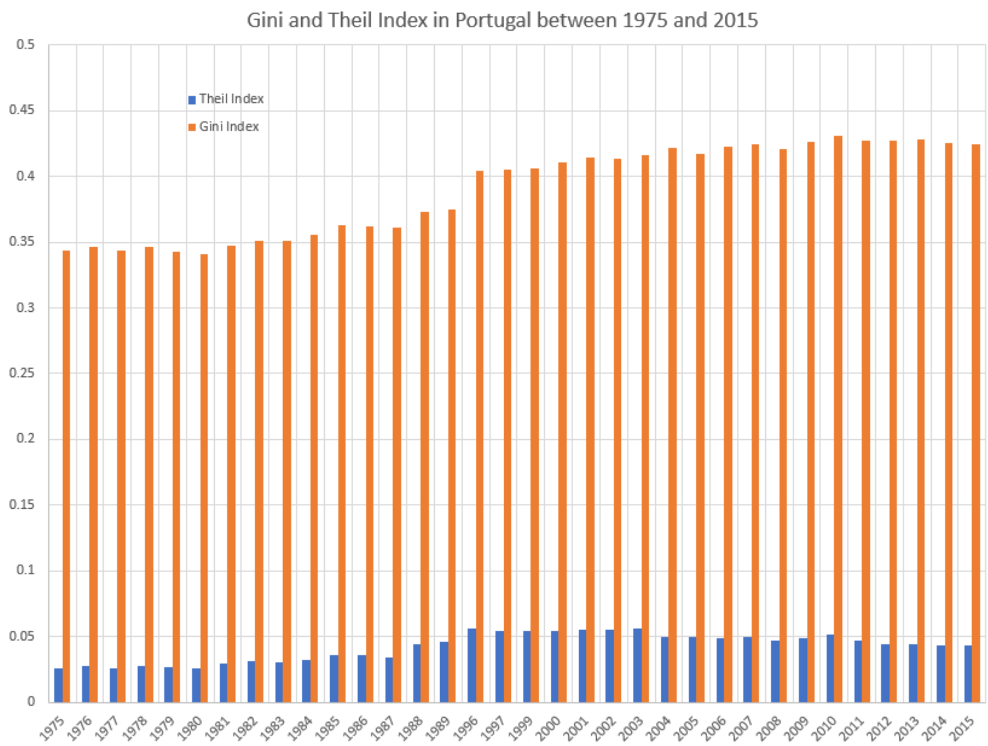

Figure 2.6: Theil and Gini Index in Portugal from 1975 to 2015. Note: lack of data between 1990 and 1995. Data Source - EHII dataset.

2.3.4 Lorenz Curves

The ”Lorenz Curve” is named after Max O. Lorenz, who developed a method to graphically represent the concentration of wealth within a population (1905) [51]. This method orders equal-sized segments by the amount of wealth each segment has, such that a cumulative total is obtained with the addition of each segment [40]. For example, populations are often segmented by quintiles (fifths), and a hypothetical five segments might have the following percentages of the total wealth: 4%, 10%, 15%, 21%, and 50%. The two lowest segments cumulatively account for 14% of the wealth, and the four lowest segments cumulatively account for 50% of the wealth. Graphically, in Figure 2.7 it is showed the curve generated by plotting these points - Lorenz Curve, as well as the ”equality” line.

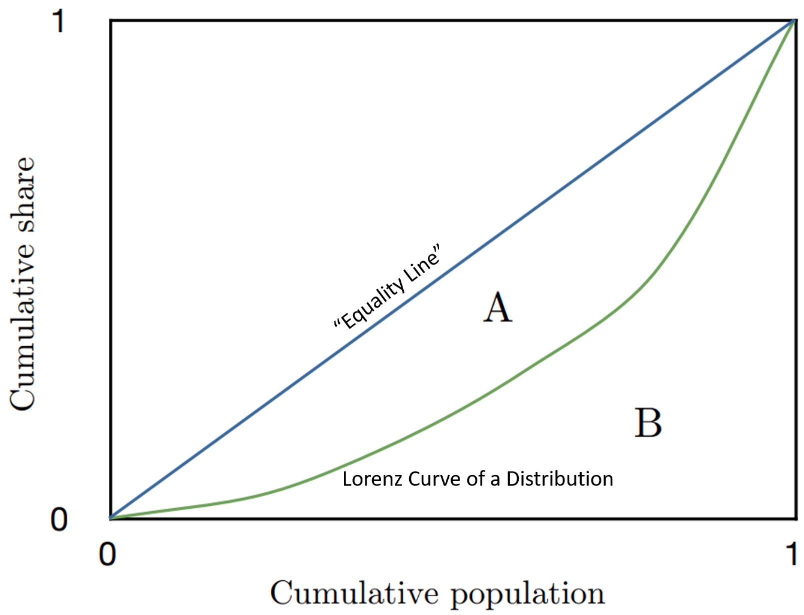

Figure 2.7: Illustration of a Lorenz Curve based on an example distribution and the Equality Line representation.

The $x-axis$ in the Lorenz curve records the cumulative proportion of population ranked by income level. Its range is therefore (0,1). The $y-axis$ records the cumulative proportion of income for a given proportion of population, i.e. the income share calculated by taking the accumulated income of a given share of the population, divided by the total income Y, as per Equation 2.18 [52] [51] [40],

$$L(\frac{k}{P}) = \frac{\sum_{i=1}^{k} y_i}{Y}, \tag{2.18}$$

where:

$k = 1, ..., n$ is the position of each individual in the income distribution;

$i = 1, \ldots, k$ is the position of each individual in the income distribution;

P is the total number of individuals in the distribution;

y_i is the income of the y^{th} individual in the distribution;

$\sum_{i=1}^{k} y_i$ is the accumulated income up to the k^{th} individual.

Lorenz Curve Properties

Lorenz curves have numerous applications and one of the reasons is their specific properties [52].

- **Property 1**

  As the Lorenz Curve is a relation between cumulative proportions, its initial point has coordinates (0,0) and its final point has coordinates (1,1). If each individual had the same income the Lorenz Curve would be equal to the equi-distribution line. Since incomes are not equal and poor individuals own proportionally less income than rich people, the Lorenz Curve will lie below the equi-distribution (equality) line. In fact, in a standard income distribution, the Lorenz Curve is convex.

- **Property 2**

  It is invariant to equi-proportional changes of the original distribution but not to equal absolute changes. When all incomes are scaled by the same percentage factor, the Lorenz Curve does not change. This is so because scaling all incomes by the same percentage will also increase total incomes by the same percentage. However, when all incomes are added or subtracted by the same absolute amount, the Lorenz Curve change. An example of this behaviours is showed in Figure 2.8, using the income distributions data in Table 2.2. The distribution B is equal to distribution A, multiplied by 2, and this leads to the same Lorenz curve. The Lorenz curve related to distribution C is different from the other two, as this is obtained from Income Distribution A, plus an increment of 2000 for each individual.

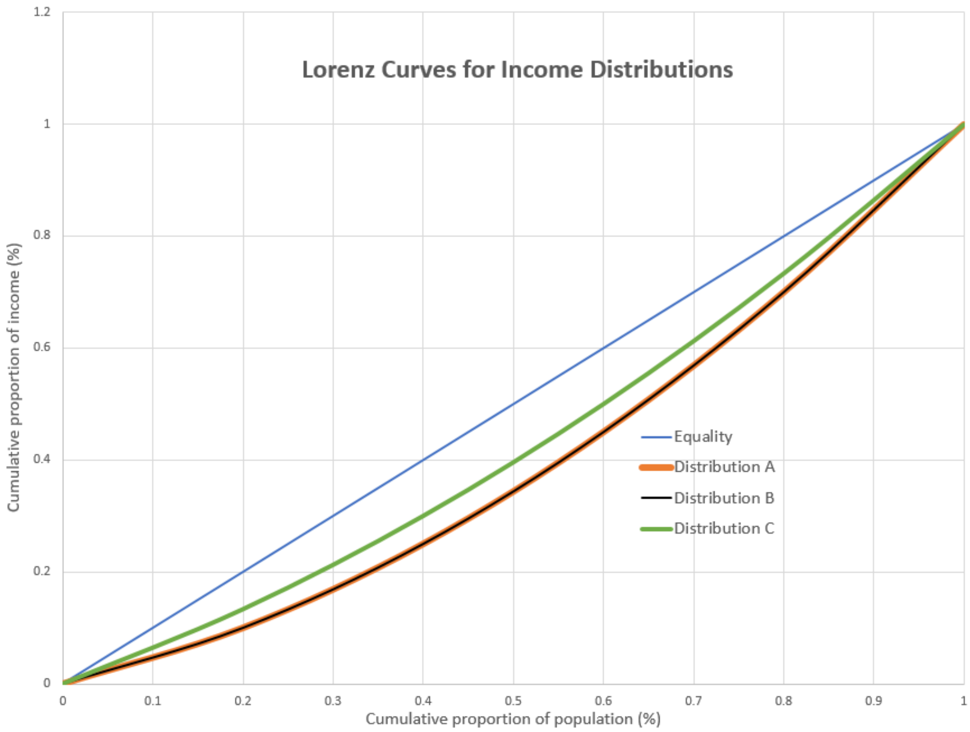

Figure 2.8: Lorenz curves obtained from 3 different income distributions in Table 2.2.

Table 2.2: Data for 3 income distributions used to obtain Lorenz curves in Figure 2.8.

Income Distributions			
ID Person	**A**	**B**	**C**
1	2000	4000	4000
2	3000	6000	5000
3	4000	8000	6000
4	5000	10000	7000
5	6000	12000	8000

Lorenz dominance of one income distribution over another occurs when, for any given cumulative proportion of population p, the Lorenz Curve of a given income distribution is above the Lorenz Curve(s) of the other distribution(s). Given the Lorenz Curve and its properties, the dominating Lorenz Curve implies an income distribution with less inequality. However, there is no guarantee that given two income distributions one would dominate the other. It may be the case that Lorenz Curves intersect. In this case, by considering only Lorenz Curves, nothing can be said about which income distribution has less inequality, and this is one of the Lorenz curve's limitations when measuring inequality.

2.3.5 Hoover Index

The Hoover Index is equivalent to the longest vertical distance between the Lorenz Curve and the 45-degree line representing the perfect equality line [49]. The value of the Hoover index shows an estimate of the share of the total community income that needs to be transferred from the households that are above the mean to the households that are ranked below the mean in order to achieve equality in income distribution. Since this represents the proportion of money needed to be transferred from the rich to the poor to achieve equality, the Hoover Index is also called Robin Hood Index. The relationship can be graphically represented on a graph as per Figure 2.9. This index is also known by Schutz index or Pietra ratio.

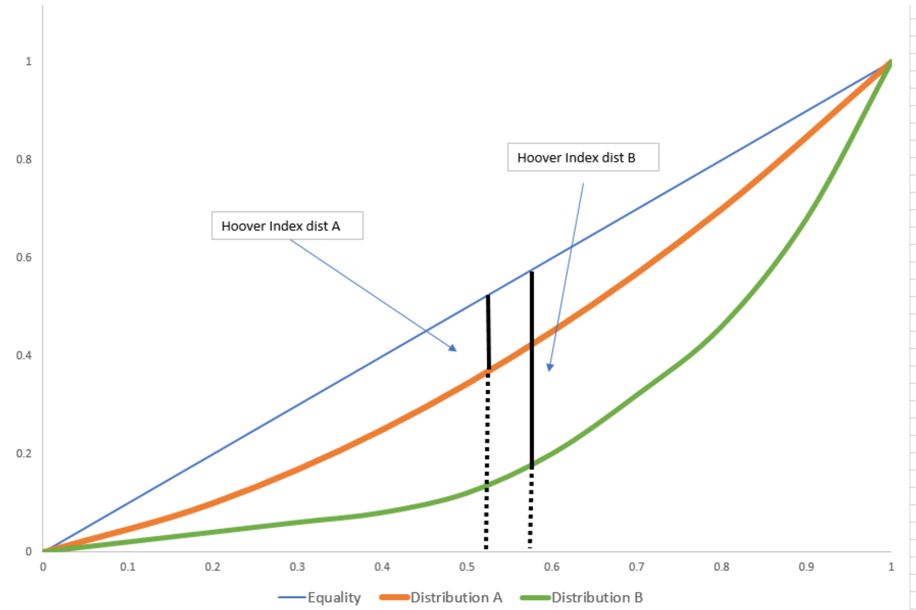

Figure 2.9: Hoover Index is represented by black vertical line between curve and equality line.

2.3.6 Gini Index

Another common measure of inequality is to examine the area between the Line of Equality (y=x) and the Lorenz Curve (Area A in Figure 2.7, representing the deviation from equality), and take its ratio to the total overall area (Areas A+B). This ratio is known as the Gini Coefficient [53]. Gini coefficients range from 0 (no deviation from equality) to 1 (complete deviation from equality).

Initially, the method for calculating the Gini coefficient was proposed differently by its originator Corrado Gini, using a method known as "relative mean differences" [54]. Dalton (1920) [55] attributes Umberto Ricci in *L'indice di variabilita e la curve dei redditi* (1916) as the first to publish a proof connecting the Lorenz Curve with the Gini coefficient, although Dalton says potentially Gini himself offered the proof [54]. Sen and Foster (1997) [53] give the formula in Equation 2.19 for calculating the Gini coefficient when all data points are known and not ordered and also discuss an early analysis of the Gini coefficient.

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2\mu}. \tag{2.19}$$

The absolute difference between each pair of data, $|x_i - x_j|$, where x_i and x_j are members in the set being compared, is calculated and summed for each member; this quantity is divided by twice the number of data points (n) multiplied by the data set mean, μ. A fundamental aspect is that the Gini coefficient is one half of the relative mean difference between all pairs of data [56].

Alternatively, if the mathematical function for the Lorenz Curve ($L(x)$) is known , the area under the curve can be calculated using integration between the limits 0 and 1 (representing from 0% to 100% of the population) as per Equations 2.23 and 2.24 [52]. This can be constructed with the knowledge that the areas represented by A and B total one half of the area in the unit square bounded by the two axes, according to the following equations.

$$G = \frac{A}{A+B}, \tag{2.20}$$

$$A + B = \frac{1}{2}, \tag{2.21}$$

$$G = 1 - 2B, \tag{2.22}$$

$$B = \int_0^1 L(x)dx, \tag{2.23}$$

$$G = 1 - 2\int_0^1 L(x)dx. \tag{2.24}$$

According to Equation 2.25 we define L as the Lorenz curve for the cumulative income distribution F with mean μ:

$$L(u) = \frac{1}{\mu} \int_0^1 F(x)dx, \tag{2.25}$$

where L is an increasing convex function with range $[0,\ 1]$.

The Gini index has been used primarily as a tool for comparison of income distributions among countries or geographical regions. At the same time, it is used as one of the most important indicators to take into account the allocation of public resources. For example, Sen [53] suggests that the comparison of welfare among different countries is not limited to the valuation of GDP per capita, he proposes to weigh the GDP by the degree of equality of the income distribution. Thus, between two countries with the same per capita income, it is considered a better country the one with less inequality in the distribution of resources. The Gini index is also used as one of the factors that explain poverty.

In fact, the Gini index is not so much a measure of inequality but a measure of concentration. To put it another way, the Gini index is a kind of "poverty counter" if we use it to evaluate income inequality [13] [44]. It has a kind of built-in sensitivity to small values. Rejecting this sensitivity it should be considered that, for example vectors $(0,0,1)$ and $(0,1,1)$ have the same degree of inequality. However, the Gini index gives here the values $G(0,0,1) = \frac{2}{3}$ and $G(0,1,1) = \frac{1}{3}$. In this case, the index value is simply the percentage of extremely poor in the population. The elementary characterization of the Gini coefficient presented in the article [44] shows that this property is not an accidental consequence of the adopted method of calculating the value of the coefficient but it is its base condition.

Chapter 3

Gini Index and Lorenz Curves

The most common usage of the Gini coefficient is in economics, to measure income inequality [44]. An advantage of using the Gini coefficient for comparisons is that it is independent of the scale of the attribute being measured. While the USA, Portugal and Mozambique have substantially different economies in terms of size and GDP, if we use the Gini Index, they can be compared against each other in terms of income inequality [35].

Wilkinson and Pickett (2009) took this approach and analyzed the outcomes of a large number of social ills across many countries, in their book The Spirit Level: Why Equality is Better for Everyone [57]. With only the occasional exception, Wilkinson and Pickett found a strong correlation between inequality within a country and unfavorable outcomes for social issues such as alcoholism, crime and incarceration rates, and obesity.

Auger, Zang, and Daniel (2009) [58] cite an inability to calculate the Gini coefficient from available data as an impediment to quantifying inequality in communities in Quebec, Canada, given that the other indicators they used (decile ratio, coefficient of variation, and mean share) showed an inverse relationship between income inequality and mortality rates (due to alcohol, tobacco and suicide).

Other uses of the Gini coefficient can be found in the literature. Several authors use the Gini coefficient, Concentration Curve and the Atkinson Index to analyze global resource consumption. Chakraborty (2001) [59] uses the Gini coefficient to characterize the land distribution in Nepal while analyzing the outcomes of common pool forestry management institutions, and identifies these management institutions as responsible for the distribution of access to forest products. Fum and Hodler (2010) [60] find income inequality among natural resource rich countries increases with a few large ”polarized” ethnic groups, but decreases in countries with many small ethnic groups. To this end, Pérez-Cirera and Lovett (2006) [61] construct a model using Gini to inform government authorities which community forests need greater oversight due to power imbalances.

Outcomes of government policies can be quantified using the Gini coefficient. The Gini coefficient can be used to examine the effects of corruption on economic development. Gupta, Davoodi, and Alonso-Terme (2002) [62] found a clear correlation between government corruption and unequal income distributions, with greater levels of corruption also impeding economic development at the lower income levels.

As the last example, Wang et al. (2011) [63], in a study on the distribution of water supply and demand on the Yellow River in China, developed an integration method as described above based on a step-wise population function of water consumption, instead of income. The authors found that inequality in water consumption had peaked in 2001 and dropped in the following five years, and further investigation was warranted into balancing equality and equity in terms of economic activity.

3.1 Main properties of the Gini Index

As described before, the Gini index is traditionally constructed in two ways, both with roots from statistical measures of dispersion: the discrete version as a standardized average of all income differences between individuals or groups and the continuous version through the Lorenz curve [44]. The Gini coefficient or index is one of the most used indicators of social and economic conditions. In this section, it will be characterized as one function that satisfies the properties of scale invariance, symmetry, proportionality and convexity in similar rankings [44]. An example of the Gini index behaviour in relation to several income inputs can be found in Figure 3.1 [64].

- **G has zero as lower limit for any distribution:**

  When all incomes are equal, the covariance between income levels and the cumulative distribution function is zero, therefore, the Gini Index is zero [64]. As an example, an income distribution for 4 people, each one with a income of 1000 units, gives a Gini coefficient of zero since:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\mu} = \frac{0}{2 \times 16 \times 1000} = 0. \tag{3.1}$$

- **G has $\frac{n-1}{n}$ as upper limit:**

  The limit of this value, for very large populations, is 1. When all incomes are zero except for the last, the last income is also equal to total income. In the limit (i.e. in a continuous framework) the value of the area Z tends towards zero. Therefore, the Gini Index tends towards 1.

  As an example, an income distribution for 4 people, each one with an income of 0 units and one with 1000 units, gives a Gini coefficient of 0.75 since:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\mu} = \frac{6000}{2 \times 16 \times 250} = 0.75, \tag{3.2}$$

  and an income distribution for 5 people, each one with a income of 0 units and one with 1000 units, gives a Gini coefficient of 0.8 since:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\mu} = \frac{8000}{2 \times 25 \times 200} = 0.8, \tag{3.3}$$

  so, $\lim G = 1$.

- **The Gini Index is scale invariant:**
  By multiplying all incomes by a factor λ, the value of the Gini Index G does not change.

  Intuitively, when all incomes are scaled by a common factor, the cumulative distribution of income does not change, as a given fraction of the population still holds the same fraction of total income. The areas under the Lorenz Curve, therefore, do not change. With regard to the covariance formula, the application of a common factor to all incomes makes the covariance and the average income increase by the same factor.

- **The Gini Index G is not translation invariant:**
  By adding/subtracting the same amount of money to all incomes, the Gini Index would increase/decrease accordingly [64] [44].

- **The Gini Index satisfies the principle of transfers:**
  If income is redistributed from relatively richer individuals to relatively poorer individuals, G decreases. The opposite holds true if income is redistributed from relatively poorer to relatively richer individuals. The size of a change in Gini, following a change in any income, depends on the rank of the individuals involved in redistribution and on the sample size. It does not depend on the level of individual incomes involved in redistribution, but it depends on total income. In particular, the Gini Index reacts more to redistribution occurring among individuals who have a greater difference in ranks. The same amount of redistribution, indeed, generates a much lower effect if the two individuals have a close rank [64] [44].

Table 3.1: Illustration of how the main properties of the Gini Index work on a sample of an income distribution. Looking for the last two columns, we note The Gini index reacts less to transfers among individuals with a close rank [64].

	Original	**scale invariance (A)**	**translation invariant (B)**	**principle of transfers (C)**	**principle of transfers (D)**
ID	Original Income Distribution	Original income distribution with all incomes increased by 20 per cent	Original income distribution with all incomes increased by 2,000 units	Original income distribution with a redistribution of 500 units from the richest to the poorest	Original income distribution with a redistribution of 500 units from two individuals with a close rank
1	1000	1200	3000	1500	1000
2	2000	2400	4000	2000	2500
3	3000	3600	5000	3000	2500
4	4000	4800	6000	4000	4000
5	5000	6000	7000	4500	5000
GINI	**0.267**	**0.267**	**0.16**	**0.213**	**0.253**
		No changes	**Decrease**	**Decrease**	**Decrease**

3.2 Applications

Gini coefficients can be used to compare inequality changes between countries or over time (considering the type of data collected). In Figure 3.1 we can observe there has been a generalized downward trend (although levels remain very high) over time in Latin America region, one of the world regions with the highest income inequality, as shown in Figure 3.2.

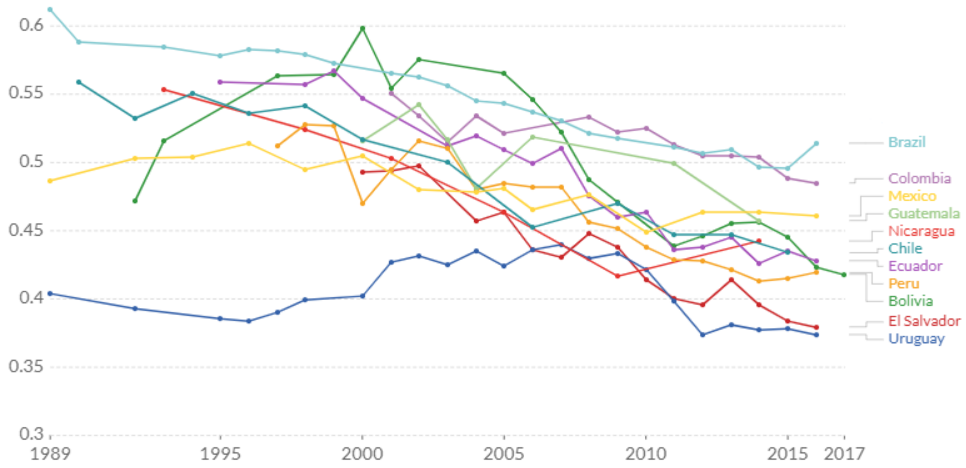

Figure 3.1: Inequality in Latin America from 1989 to 2017. The Gini index measures the distribution of household equivalent income, including zero income. Source: WID [31].

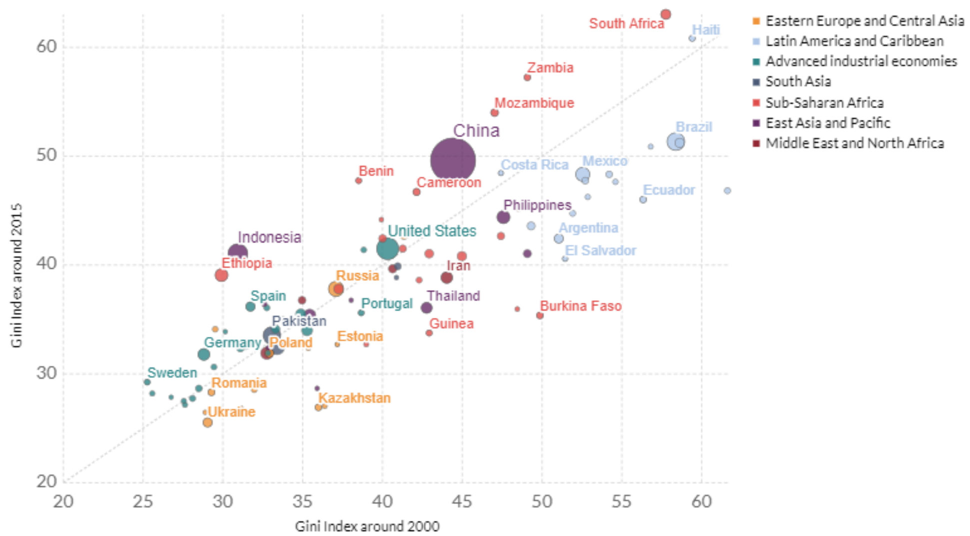

Figure 3.2: Gini Index around 2015 vs. Gini Index around 2000. Estimates are based on household survey data of either incomes or consumption. The circles areas represent population size. Source: WIID [65].

Looking at Figure 3.2, we can see the countries below the "equality" line had an inequality reduction between 2000 and 2015, however, we don't have enough information to understand why and how the income distributions changed. In the following subsection we will explore the limitations of using Gini coefficient as an inequality measure, exploring some examples and extreme cases.

3.3 Limitations and Extreme Cases

The Gini index has been found to have a number of limitations. One can be related to the 'income concept', since it can be defined at the household level weighted by household size, other scales or at the individual level taking into consideration financial holdings or just wage earnings, it can take the net income (after taxes) or not. Each income definition gives a different measure of income and different levels of income inequality [66]. Thus differences in income concepts can lead to differences in measures of income, inequality and the ranking of countries, and for that reason is really important to know the data we are using to deduct the coefficient. In Figure 3.3 it can be seen there's a big difference in Gini coefficient in Portugal after and before taxes.

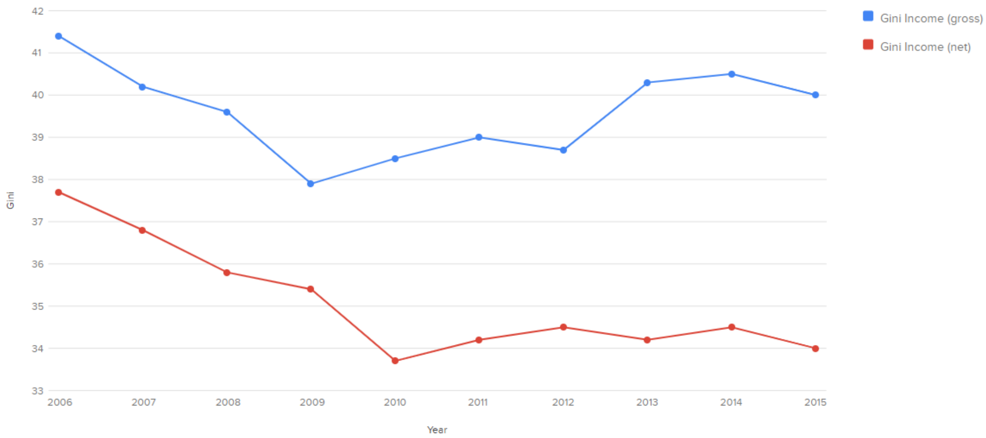

Figure 3.3: Inequality of incomes before (gross) and after (net) taxes and transfers in Portugal based on the Gini Index, from 2006 to 2015. Note: Income before taxes and transfers corresponds to 'market income' (gross wages and salaries + self-employment income + capital and property income). Income after taxes and transfers corresponds to 'disposable income' (disposable income = market income + social security cash transfers + private transfers – income tax). Source: WIID Database [65].

If we rely In Figure 3.4 without knowing the type of data it's being used, we will say Portugal has a higher Gini index (inequality) than Chile, as we are plotting two lines with a different meaning. As described, this limitation can lead to wrong analysis and comparisons.

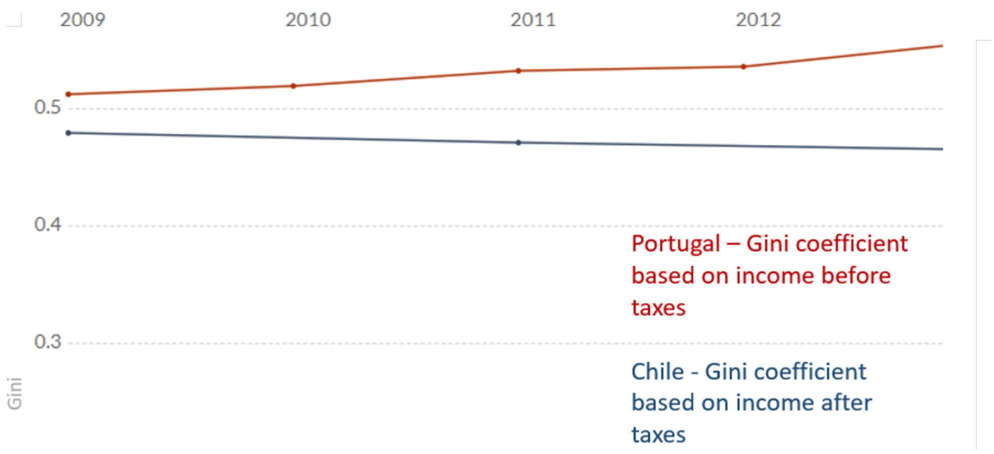

Figure 3.4: Inequality of incomes before (Portugal) and after (Chile) taxes and transfers based on the Gini Index, from 2009 to 2015. Source: OECD Database [67].

Also, a Gini index based on individual incomes is different from a Gini Index based on household incomes for the same country. As a result, the rankings of countries change depending on whether the index is based on household incomes or individual incomes, creating some subjectivity in its use and interpretation.

The Gini index also does not capture social benefits or other interventions aimed at bridging the inequality between rich and poor. Subsidised housing, healthcare, education and social grants for the vulnerable are measures that subsidise household incomes, reducing income inequality to some extent [44].

Population features or demographic changes are not reflected by the Gini index. Countries with high ratios of elderly people whose main sources of income are pensions, or countries with high student ratios are likely to have higher levels of income inequality as measured by the Gini index [66].

Additionally, the Gini index is a relative measure that fails to capture absolute differences in income. It is possible for the Gini index of a country to rise due to increasing income inequality while the number of people living in absolute poverty is actually declining. This is because the Gini index violates the Pareto improvement principle, which says income inequality can increase despite an increase in all incomes in a given society. Thus, although the level of income inequality has increased, the Gini index fails to capture the fact that absolute levels of income have also increased. Similarly, the Gini index could reflect a lower level of income inequality in a scenario where there is a decrease in all incomes in a given society.

As described, the Gini Index is Lorenz-derived and Lorenz consistent. However, it is worth recalling that the ordering provided by Lorenz Curves, in particular by Lorenz dominance, is a partial ordering. As the Gini Index reduces the whole income distribution to a single number, it allows an easier comparison between distributions. The basic difference between the two approaches can be appreciated by considering a case in which Lorenz Curves cross each other.

For example, let us consider the two income distributions reported in Table 3.2.

Table 3.2: Data for Income Distributions A and B.

ID	Income Distribution A	Cumulative % Population	% Income	Cumulative % Income
1	2,000	20%	10%	10%
2	3,000	40%	15%	25%
3	4,000	60%	20%	45%
4	5,000	80%	25%	70%
5	6,000	100%	30%	100%
TOTAL	**20,000**			
GINI	**0.200**			

ID	Income Distribution B	Cumulative % Population	% Income	Cumulative % Income
1	900	20%	5%	5%
2	4,000	40%	20%	25%
3	4,800	60%	24%	49%
4	4,800	80%	24%	73%
5	5,500	100%	28%	100%
TOTAL	**20,000**			
GINI	**0.200**			

Incomes are distributed differently, as we can see in Figure 3.5, but the Gini Index is the same (0.200). These two income distributions give rise to Lorenz Curves in Figure 3.6.

Figure 3.5: Graphic of Income Distributions A and B per person.

Lorenz Curves intersect but the way they intersect is such that the area before the intersection and the area after the intersection are of the same value. This gives rise to the same Gini Index, even in the presence of quite different income distributions.

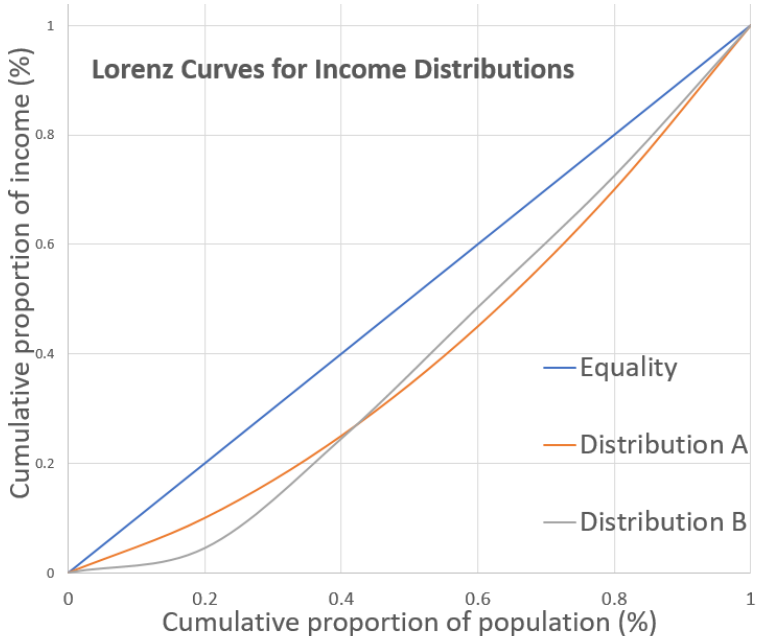

Figure 3.6: Lorenz curves of Distributions A and B from Table 3.2. Although the curves intersect, the Gini coefficient is the same for both distributions.

As per the last examples described, in some cases, the Gini coefficient can be the same for countries with different income distributions, and this is one the limitations of using this measure and the problem we aim to address in this work.

Also, one of the flaws of the Gini index is that, from a single number, we can not see the full picture of what's happening in terms of differences in income distributions and the reason why the index is increasing or decreasing.

Transfers between two middle-income households do affect a higher fraction of the population than other transfers but those transfers do not receive an excessive weight relative to other transfers because the difference in the ranks of donor and recipient is smaller than the corresponding difference in other transfers. Thus, progressive transfers between two households in the middle of the distribution changes the Gini index less than a transfer of the same amount from an upper income household to a lower income household. Similarly, the effect on the Gini index when a household in either tail of the distribution receives an additional increment is larger than when a middle-income household receives it.

This is the main aspect we will try to change, introducing the new coefficients described in the following chapters.

Chapter 4

Numerical Model - New Complementary Coefficients

In order to complement the Gini Coefficient known limitations, two new coefficients were created.

Inspired by the Demetrius idea [12] [27] of measuring the difference between the mortality curves relative to one constant mortality over time (entropy), these 2 new complementary coefficients were developed to give more information about the reasons for inequality. These new inequality measures were named as GPlus (G+) and GMinus (G-), according to their equations format only (Equations 4.1 and 4.2) and they are a second order measure for the Gini index:

$$G+ = 1 + 2\frac{\int_0^1 (1 - F(x)) \times \ln(1 - F(x))\,\mathrm{d}x}{\int_0^1 (1 - F(x))\,\mathrm{d}x}, \tag{4.1}$$

$$G- = -1 - 2\frac{\int_0^1 (F(x)) \times \ln(F(x))\,\mathrm{d}x}{\int_0^1 F(x)\,\mathrm{d}x}, \tag{4.2}$$

where $F(x)$ is a Lorenz curve.

Their main goal is to be used to observe what happens in inequality for distributions with the same Gini index but with different properties, although they can also be used to compare distributions with different Gini Index. GPlus will focus on the poorest side of population, and GMinus on the richest. They are also a complementary tool that helps to analyse Lorenz curves without using graphical support. For example, in cases where two Lorenz curves are practically identical, GPlus and GMinus factors will give us the summarized information that our eyes can not see. Both coefficients were developed in the same way as they were both constructed using curves with cumulative values (Lorenz) and then normalized. Equations 4.2 and 4.1 can be regarded as an analogue of the Boltzmann–Gibbs definition of the entropy of a thermodynamic system. The numerator is analogous to the Shannon–Weaver measure of the amount of information in a message. They are both based on Lorenz curves, being GPlus given by a set of Lorenz dependent terms, $1 - F(x)$, and GMinus by a set of Lorenz curves terms, $F(x)$. In the end, we can say these are entropic measures of inequality that measure the difference between the distributions curves to one constant distribution over time.

4.1 Verification and Validation - Axioms

GPlus and GMinus also respect some of the inequality measures axioms presented in Section 2.3.1 and for that reason, they can be considered as good complementary measures. In Table 4.1 we have summarized the principal axioms and the behaviour of the new measures for different "inputs".

Table 4.1: Results of New coefficients - axioms proof.

	Original	**scale invariance (A)**	**translation invariant (B)**	**principle of transfers (C)**	**principle of transfers (D)**
ID	Original Income Distribution	Original income distribution with all incomes increased by 20 per cent	Original income distribution with all incomes increased by 2,000 units	Original income distribution with a redistribution of 500 units from the richest to the poorest	Original income distribution with a redistribution of 500 units from two individuals with a close rank
1	1000	1200	3000	1500	1000
2	2000	2400	4000	2000	2500
3	3000	3600	5000	3000	2500
4	4000	4800	6000	4000	4000
5	5000	6000	7000	4500	5000

GINI	**0.267**	**0.267**	**0.16**	**0.213**	**0.253**
G+	**0.357**	**0.357**	**0.216**	**0.289**	**0.334**
G-	**0.311**	**0.311**	**0.205**	**0.267**	**0.307**

- **Scale Invariance:**

  Scaling the incomes x_i by a constant factor does not change these inequality coefficients. This can be observed comparing the original distribution results in Table 4.1 with the scale invariance results (distribution A) - there are no changes in both G+ or G- when each income is multiplied by a factor of 1.2.

- **The Pigou-Dalton principle of transfers:**

  This axiom requires the coefficients to change when income transfers occur among individuals in the income distribution. In particular, Inequality indexes should fall with a progressive transfer (an income transfer from richer to poorer individuals) and they should rise with a regressive transfer (an income transfer from poorer to richer individuals). Let's assume the original income distribution in Table 4.1. If a progressive transfer, $T = 500$, occurs from ID3 to ID2, the new distribution is given by last column (distribution D) in Table 4.1. We say that an Inequality index satisfies the principle of transfers if, in this case, $I(original) > I(D)$ and, from Table 4.1: $GPlus(original) = 0.357 > GPlus(D) = 0.334$ and $GMinus(original) = 0.311 > GMinus(D) = 0.307$. The same happens for distribution C, a Robin Hood example, when a transfer of 500 units is done from the richest to the poorest.

- **Not translation invariant:**

  As per the Gini Index, by adding/subtracting the same amount of money to all incomes, the new coefficients would increase or decrease (Distribution B in Table 4.1).

- **Symmetry, Anonymity and Population Independence:**

  Both inequality coefficients are not dependent on any characteristic of the individuals other than income and the permutation of each characteristic among the individuals does not affect the results. Also, they are population independent so the level of the population (few people or many people) also does not impact the results.

4.2 Finding Bounds for GPlus and GMinus

Let's now analyze the extreme values of GMinus. We denote the part of the numerator of G- by N- as per Equation 4.3:

$$N- = \int_0^1 (F(x)) \times \ln(F(x))\,\mathrm{d}x, \tag{4.3}$$

integrating by parts using $\int uv' = uv - \int u'v$, with $u = \ln F(x)$, $u' = \frac{F'(x)}{F(x)}$, $v' = F(x)$ and $v = \int F(x)$, we obtain 4.4:

$$N- = \ln F(x) \int F(x)\mathrm{d}x - \int \frac{F'(x)}{F(x)}\mathrm{d}x \left[\int F(x)\mathrm{d}x\right]. \tag{4.4}$$

For an easier analysis of the extremes let's assume that $F(x)$ is a function with the form x^a, where a is a real number higher than 1. These types of functions are in line with a Lorenz curve as they are always positive, their maximum and minimum are 0 and 1 and they are always convex, i.e, the second derivative is always positive, as shown in Figure 4.1:

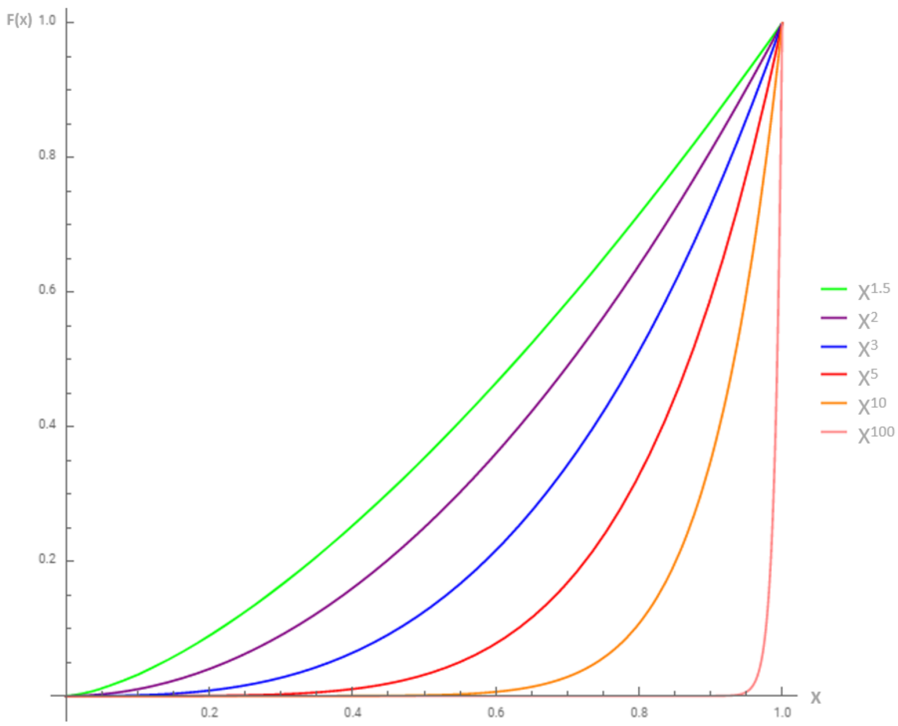

Figure 4.1: Graphical representation of the Convexity of Lorenz curve type functions - $F(x) = x^a$ for different values of a.

Developing the result from Equation 4.4 for $F(x) = x^a$ we obtain the expression in 4.5:

$$\ln x^a \int x^a \mathrm{d}x - \int \frac{(x^a)'}{x^a} \mathrm{d}x \left[\int x^a \mathrm{d}x \right]. \tag{4.5}$$

Solving this in an integration interval between 0 and 1 results in 4.6

$$-\frac{a}{(a+1)^2}. \tag{4.6}$$

Doing the same thing for the denominator and replacing the results in GMinus equations, gives the expression in Equation 4.7:

$$G-_{x^a} = -1 - 2\left(\frac{-\frac{a}{(a+1)^2}}{\frac{1}{(a+1)}}\right) = -1 + 2\frac{a(a+1)}{(a+1)^2} = -1 + 2\frac{a}{(a+1)}. \tag{4.7}$$

As $a \in [1, \infty[$, when doing $\lim_{a \to 1} G-_{x^a}$ and $\lim_{a \to \infty} G-_{x^a}$ it is possible to conclude about the maximum and minimum of the GMinus. Both limits solutions are showed in 4.8.

$$\lim_{a \to 1} G-_{x^a} = 0 \text{ and } \lim_{a \to \infty} G-_{x^a} = 1. \tag{4.8}$$

Now, to complete the same exercise for GPlus from Equation 4.1, the term $1 - F(x)$ will be replaced by functions with the same behaviour, $1 - x^a$, where $a \in]1, \infty[$. This type of functions will behave as $1 - F(x)$, where $F(x)$ is a Lorenz curve, since they will always be positive, with extremes in 0 and 1 (as can be seen in Figure 4.2).

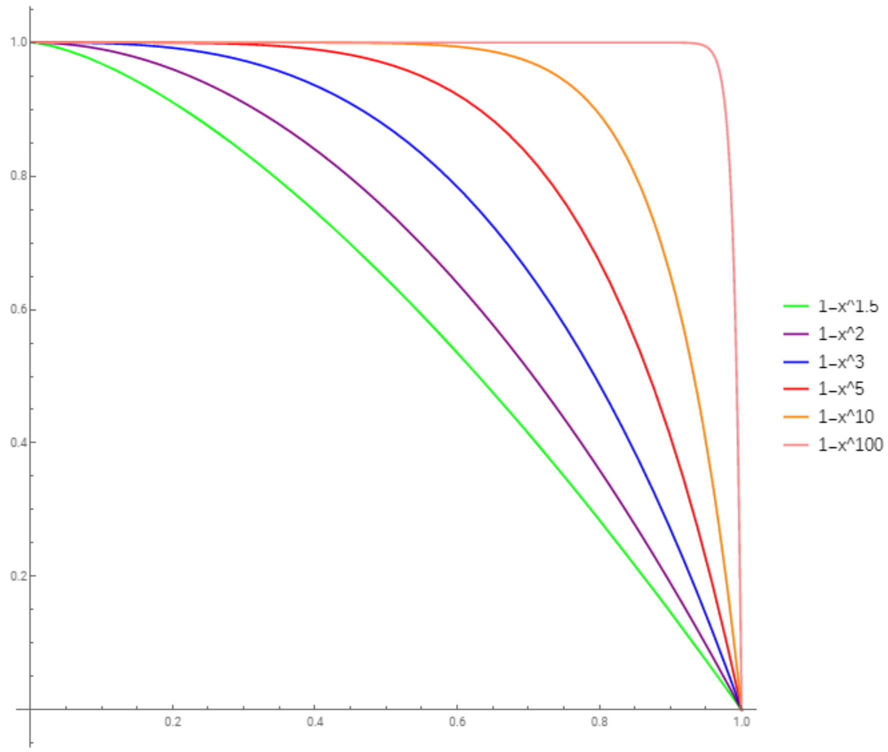

Figure 4.2: Graphical representation of $1 - F(x)$ (curve type functions), given by $1 - x^a$ for different values of a.

Using the same logic, i.e, integrating by parts the numerator of GPlus with the new function inside, doing the same for the denominator and replacing the results into GPlus expressions, it is possible to obtain the final expression (Equation 4.9) to calculate the limits and obtain the extremes of GPlus.

$$G+_{1-x^a} = 1 + 2\frac{a - a(a+1)HarmonicNumber(1/a)}{a(1+a)}, \tag{4.9}$$

where, HarmonicNumber(1)= 1 and HarmonicNumber(0)= 0, by Euler definition, which can be used to simplify the expression for limit calculations [68] [69].

Both limits solutions are showed in 4.10.

$$\lim_{a\to 1} G+_{1-x^a} = 0 \text{ and } \lim_{a\to\infty} G+_{1-x^a} = 1. \tag{4.10}$$

In line with the Gini index properties, GMinus and GPlus also have a possible range of values between 0 and 1.

For the several functions (distributions) in Figure 4.1, it is known that, for higher a, more inequality will exist, since, from the Gini Index definition, the area under the imaginary equality curve will be higher for $F(x) = x^{100}$ than for $F(x) = x^{10}$. In fact, it is also possible to see that the slope of the curve with $a = 100$ is higher than for lower a, which means that there is a big percentage of total income (if we are associating this with an income distribution) for the top decile population (richest).

On Figure 4.3 the evolution of GMinus index for different values of a using the example type of functions - Lorenz consistent - $F(x) = x^a$ is shown. It is possible to conclude that for higher a, GMinus will approach 1, increasing in line with the Gini Index, i.e., more inequality.

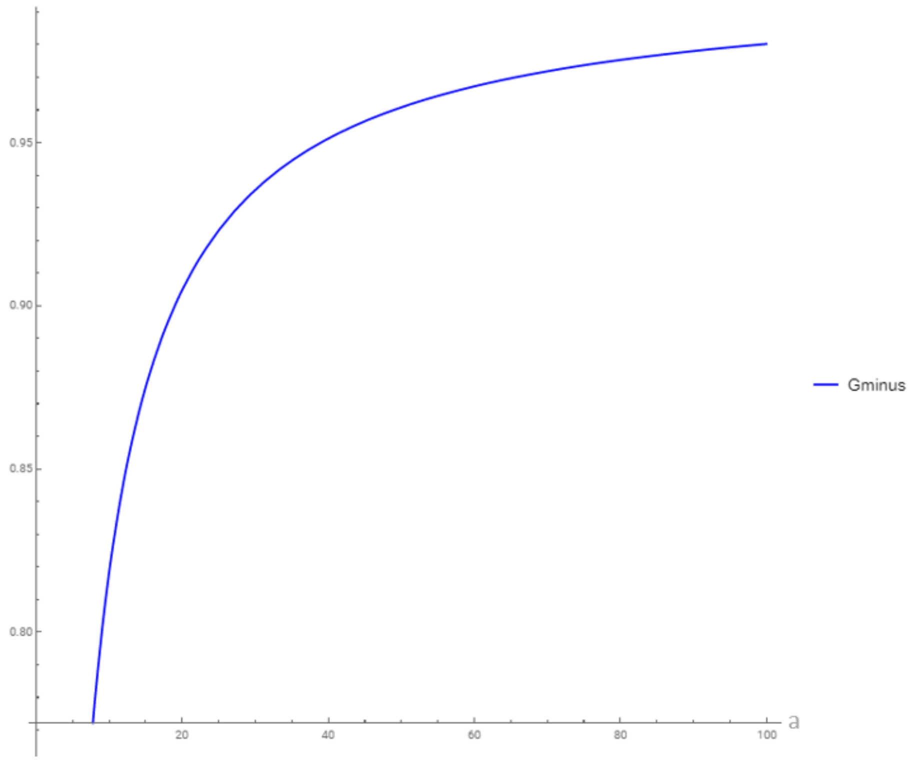

Figure 4.3: Evolution of GMinus index for different values of a in function $= x^a$.

An increase in GPlus is also expected when there is an increase in a, as it can be seen in Figure 4.4. It is possible to verify that GPlus increases more quickly than GMinus in this type of curves and GPlus is always higher than GMinus.

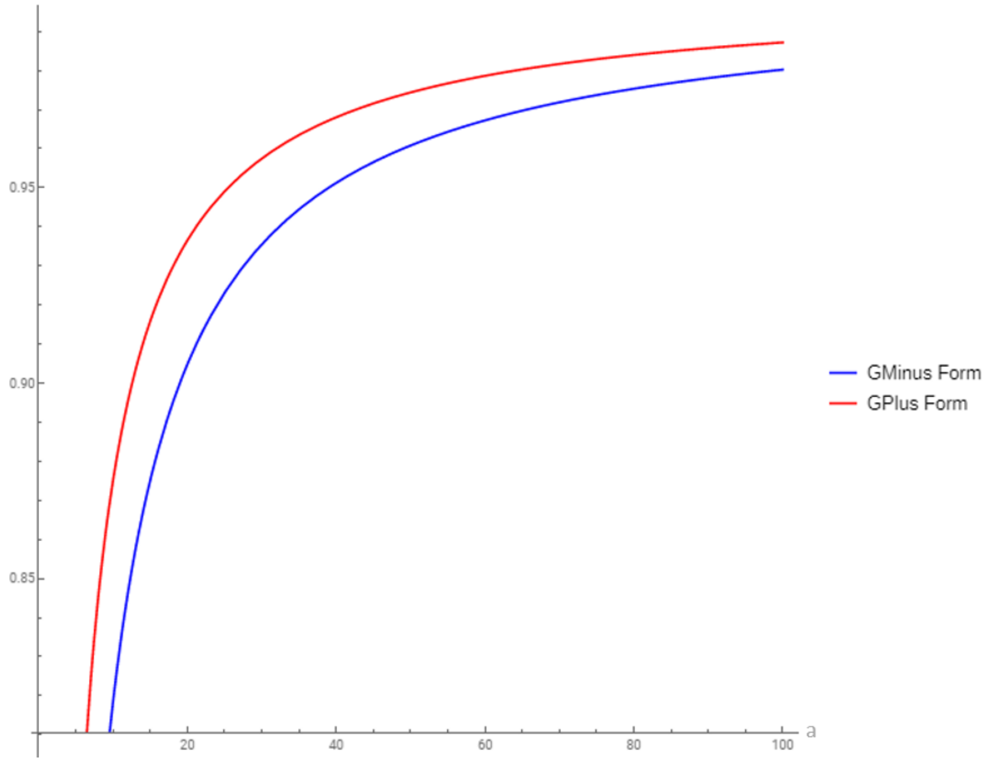

Figure 4.4: Evolution of GPlus and GMinus indexes for different values of a.

The results for boundaries can also be checked using Mathematica software tools for "limit" curves as per Figure 4.5. This software helped us to obtain the results presented in the next sections [69].

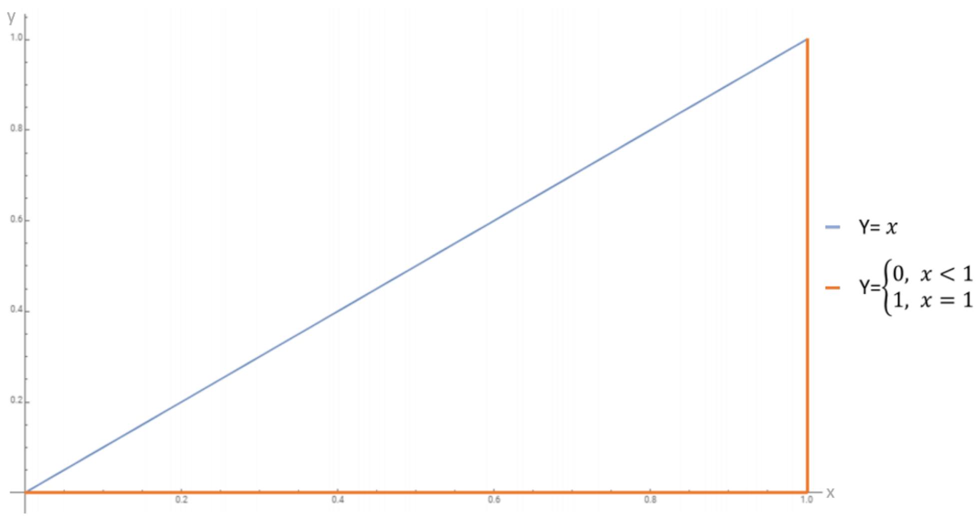

Figure 4.5: Representation of potential limit curves related with inequality (orange) and equality (blue).

The blue line in Figure 4.5 represents the equality line, while the orange line represents an extreme case of inequality (it can be associated with a population where there is one person that has all the income, and the rest of the people don't have any income). The obtained results are presented in Table 4.2.

Table 4.2: Results of new coefficients for limit curves represented in Figure 4.5.

Line	Representing	Gini Index	G+	G-
Blue	Equality	0	0	0
Orange	Inequality	1	1	1

In summary, it was proved that both coefficients will share the same behaviour of the Gini Index with the increase and decrease of income percentages as well as the same boundaries (maximum of 1 and minimum of 0).

It was also showed the behaviour of the coefficients when a transfer occurs from a top earner to a low earner. This is of extreme importance when managing policies as the new factors allow to predict the behaviour of inequality in the distributions tails, adding more value and information to the police makers. Several scenarios can be tested using the new coefficients and the results can be used as an extra input to make important governmental decisions.

These conclusions will be presented in the next sections through real cases and toy models.

Chapter 5

Results

In this chapter GPlus and GMinus coefficients will be applied to toy models and real world cases to prove their efficiency and utility.

Toy models are analogous examples of a real case scenario, but reduced to a simple model, with specific parameters that will be used to illustrate the behaviour of the new coefficients when dealing with small parameter changes. The use of toy models will allow to draw basic conclusions, which can then be applied and verified in the real cases scenarios, where several comparisons between both the same and different countries in different years, and extreme cases will be analyzed.

The cases that are being studied are shown below:

- Toy Models
- Slovakia (2014) vs Germany (2005)
- Germany in 1983 and in 2005
- Portugal - several years
- Europe 2018
- OECD - several years
- High Inequality
- Sensitivities - Deciles vs Quintiles
- Mathematical Sensitivities
- Re-Scale

In order to present the most accurate and compatible data we've only used the WIID dataset information to get the results presented in this chapter. A full description of this dataset (and others used in the last sections) and the relevant extracted information is available in Appendix A. For the majority of the presented results, the data that was used was based on Net income (after taxes) and an equalised scale (i.e, the population unit is considered as individual). This is what should be considered in the following sections as used by default when no other information is given.

5.1 Toy Models

In this section, toy models will be explored in order to assess the advantages of using the new coefficients and to know their limitations.

For the first considerations we've used discrete samples of incomes and, as a first example, we've calculated the Gini index and the new coefficients for several distributions with specific variations. From table 5.1, it is possible to verify that the 3 coefficients are increasing (inequality), from distribution I to K. With the increase of people earning less it is expecting that the Gini index also increases and, as the GPlus and GMinus will (typically) follow the same evolution.

Table 5.1: Results of New coefficients and Gini index for toy models distributions.

	Distribution I	Distribution J	Distribution K
Income 1	10000	10000	10000
Income 2	100000	10000	10000
Income 3	100000	100000	10000
Income 4	100000	100000	100000
Income 5	100000	100000	100000
Gini	0,1700	0,3375	0,4695
G+	0,2950	0,4885	0,5954
G-	0,0395	0,1880	0,5400

As a second example, we've calculated the new coefficients for four different distributions with the same Gini index and the results can be found in Table 5.2.

Table 5.2: Toy models distributions with an equal Gini Index, Quintiles and New coefficients results.

ID	Income Distribution A	Income Distribution B	Income Distribution C	Income Distribution D
1	2000	900	3000	20
2	3000	4000	3000	4970
3	4000	4800	3000	5000
4	5000	4800	3000	5000
5	6000	5500	8000	5010
TOTAL	**20,000**	**20,000**	**20,000**	**20,000**
	Quintiles - Proportion of Total Income			
Q1	10.00%	4.50%	15.00%	0.10%
Q2	15.00%	20.00%	15.00%	24.85%
Q3	20.00%	24.00%	15.00%	25.00%
Q4	25.00%	24.00%	15.00%	25.00%
Q5	30.00%	27.50%	40.00%	25.05%
GINI	**0.200**	**0.200**	**0.200**	**0.200**
G+	**0.269**	**0.303**	**0.204**	**0.333**
G-	**0.250**	**0.134**	**0.383**	**0.005**
Theil	**0.063**	**0.102**	**0.100**	**0.195**
Hoover	**0.150**	**0.155**	**0.200**	**0.199**

In distribution A, all 5 people receive a different amount, in distribution B there's one person with an income of 900 units and the other four with an income $>$ 4000 units; in distribution C, there is one person with an income of 8000 units and other four people with 3000 units, and distribution D is similar to distribution B but one of the people has as income of 20 units, when compared to other people with an income circa 5000 units.

For completeness, the four curves are displayed in Figure 5.1. Comparing B and D distributions and the Gini +/- coefficients leads us to conclude the G- will decrease from B to D (ie, will be more equal in D) as people with higher incomes are closer to each other, and the G+ will increase (more inequality), as the people with lower income are less close to each other.

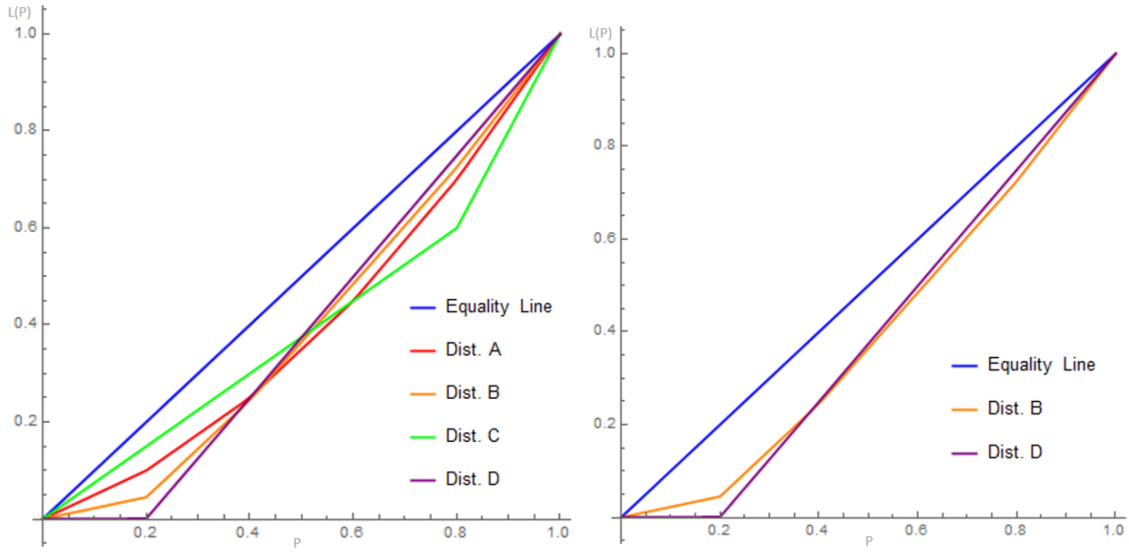

Figure 5.1: Lorenz curves related with distributions in Table 5.2.

Our first conclusion is that GPlus is more sensitive to poor people, and GMinus is more sensible to rich people. According to this statement, when looking only to the quintile percentage information in 5.2 we can assume that the G+ will be higher for distribution D (poor people receiving less percentage of income = 0.10% of total income), followed by Distribution B (4.5%), A (10%) and C (15%), which is proved by the results obtained for the four distributions (G+=0.333, 0.303, 0.263 and 0.204, for D, B, A and C, respectively). The same will happen for the GMinus coefficient: When comparing the percentage of total income for the top 20% richest it is possible to conclude that Distribution C will have the highest G- coefficient (with a percentage of 40% of all the incomes), followed by distribution A (30%), then B and C (with 27.50% and 25.05%). Therefore, it can be said that distribution D has more inequality for the poor than the others, but distribution C has more inequality on the rich side and the balancing of these coefficients going up and down between the distributions is what gives the same Gini index for all. This will also be proved in the section 5.9 using mathematical approaches.

Another observation is that the difference between GMinus and GPlus will decrease for the same Gini Index if the wealth share is higher for middle earners, i.e., if the proportion of total income is higher for middle quintiles in one distribution than in the other one. This means that the difference between GMinus and GPlus will be lower. In Figure 5.2 it is possible to see that the middle earners from distribution D have a higher percentage of total income than in distribution C, so we expect the difference between the GMinus and GPlus to be lower in D. Taking the numbers from table 5.2 for distribution C we will

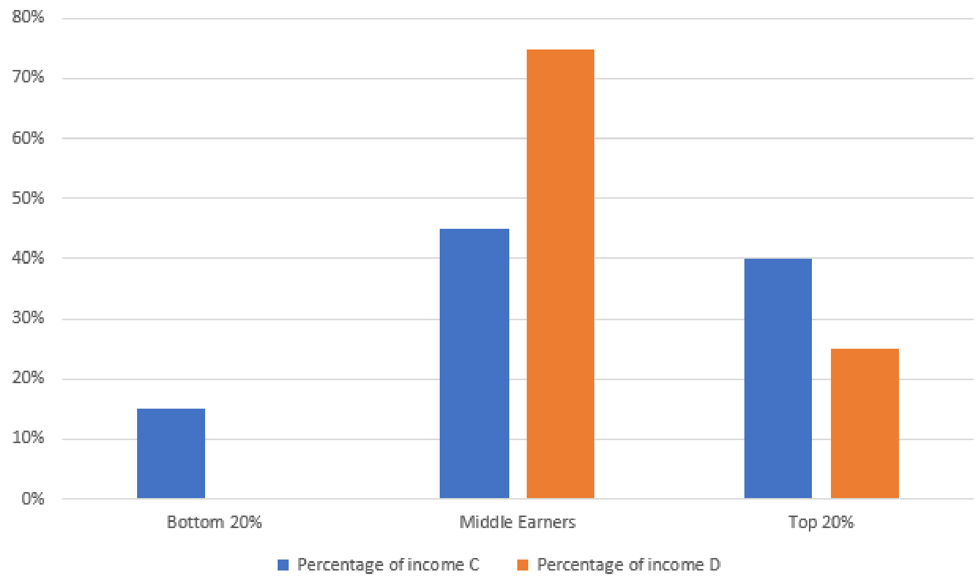

Figure 5.2: Percentage of income for richest, poorest and middle class population for distributions C and D presented in Table 5.2.

have GMinus-GPlus=0.179 while for distribution D the difference is -0.328, which is in line with our expectations (lower in D). According to this, a new term was defined (Middle Layers Force - MLF) that measures the force of middle layers and is given by Equation 5.1. For two given distributions, the one with higher MLF will have higher income percentage for the middle class earners than the other.

$$MLF = 1 - (GMinus - GPlus). \tag{5.1}$$

5.2 Slovakia (2014) and Germany (2005)

As a first test we choose two income distributions with a similar Lorenz curve and the same Gini index. Although we can not easily compare the historical/political facts when using these two countries and years, they have some similarities in their income distributions and that makes them interesting to study.

From Figure 5.3 it is possible to see that the curve is more narrowed for Germany for the bottom population percentage. This means that the income percentage of the poorest 10% in Slovakia is lower than in Germany. From our toy model's conclusions we noted the GPlus coefficient is more sensitive to the poorest people, and for that reason we are expecting the GPlus to be lower in Germany (more equality in the poor side) than in Slovakia. On the other hand, for the richest people, we can make some conclusions about the GMinus index looking to the Income Quintiles from each country in Figure 5.4. The income percentage of the richest 20% in Slovakia is lower than in Germany, so we are expecting the GMinus to be higher for Germany (more inequality). Both results are presented in Table 5.3 and are in line with what was expected.

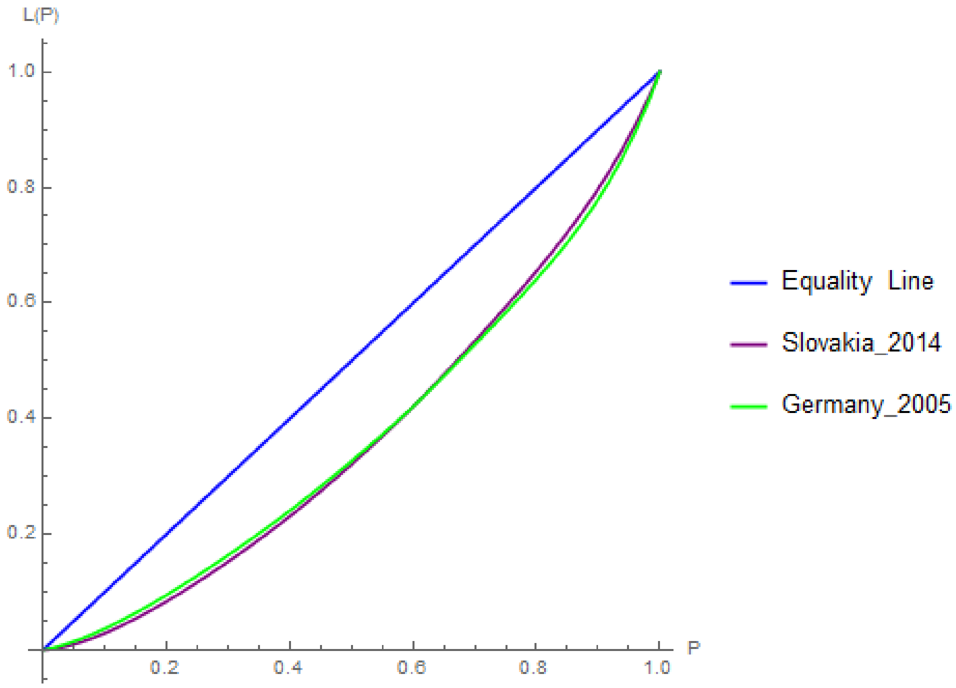

Figure 5.3: Lorenz curves of Slovakia (2014) and Germany (2005) income distributions. Data Source: WIID [65].

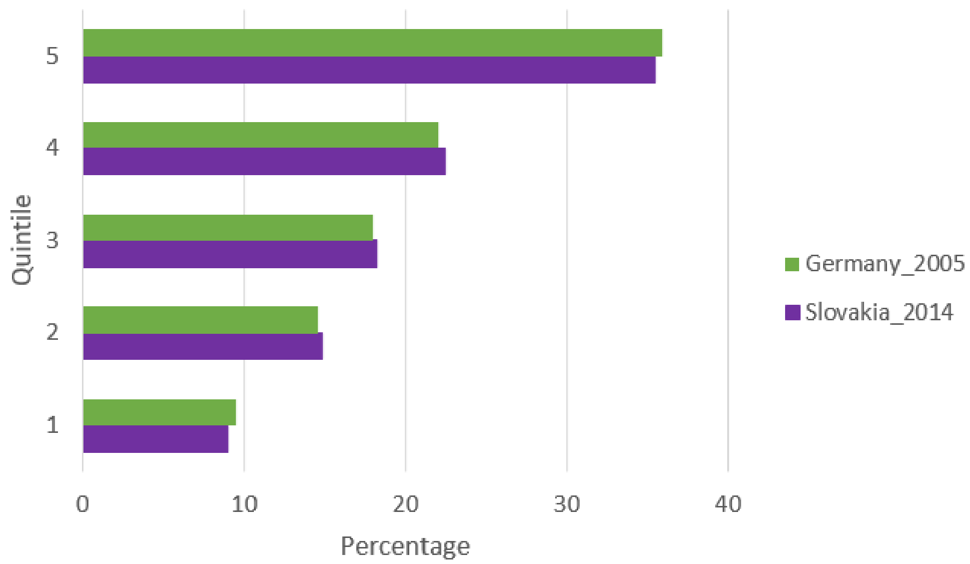

Figure 5.4: Quintiles of Slovakia (2014) and Germany (2005) income distributions. Data Source: WIID [65].

The Palma ratio is also presented in the table and it is the ratio of income shares of the top 10 percent of the population to the bottom 40. It's expected that, if the Palma ratio is higher, then, the GMinus will also be higher in a distribution, since it is more sensible to the bottom quintiles. In Germany (2005) the Palma ratio is 0.92, i.e., the top 10% of the population has 22.1% of total income, while the bottom 40% of the population has approximately the same (24.1%). For that reason, and given that the Palma ratio in Slovakia is lower (0.88), the GMinus is higher in Germany.

Table 5.3: Inequality common Measures and new coefficients results for income distributions in Figure 5.3.

	Slovakia 2014	**Germany 2005**
Gini	0.26	0.26
G+	0.32	0.30
G-	0.34	0.38
Palma	0.88	0.92
MLF	0.98	0.92

5.3 Germany in 1983 and in 2005

One of the main motivations for this work is to understand the reason for inequality and it's interesting to analyse cases within different years for the same country. Given the fact that there are many things that occur in one country that contribute to its income differences, it's interesting to study what happened with the income distributions in Germany between 1983 and 2005, which have the same Gini coefficient. It's important to understand if this is due to an increase of the proportion of total income for the top deciles/quintiles (richer people being richer), if it is due to an inverse option as bottom quintiles having less proportion of income, or if there are no changes between the distributions.

By observing Lorenz curves, some conclusions can already be taken, although it's not easy to consolidate all the information and it can not be interpreted by anyone. GPlus and GMinus are more intuitive and do not require graphical analysis, which makes them a good complement to other measures. In line with the distributions in Figure 5.3, these two different years in Germany have similar Lorenz curves (Figure 5.5) and the same Gini index but using only these two measures won't give us much more information about the differences between these two distributions.

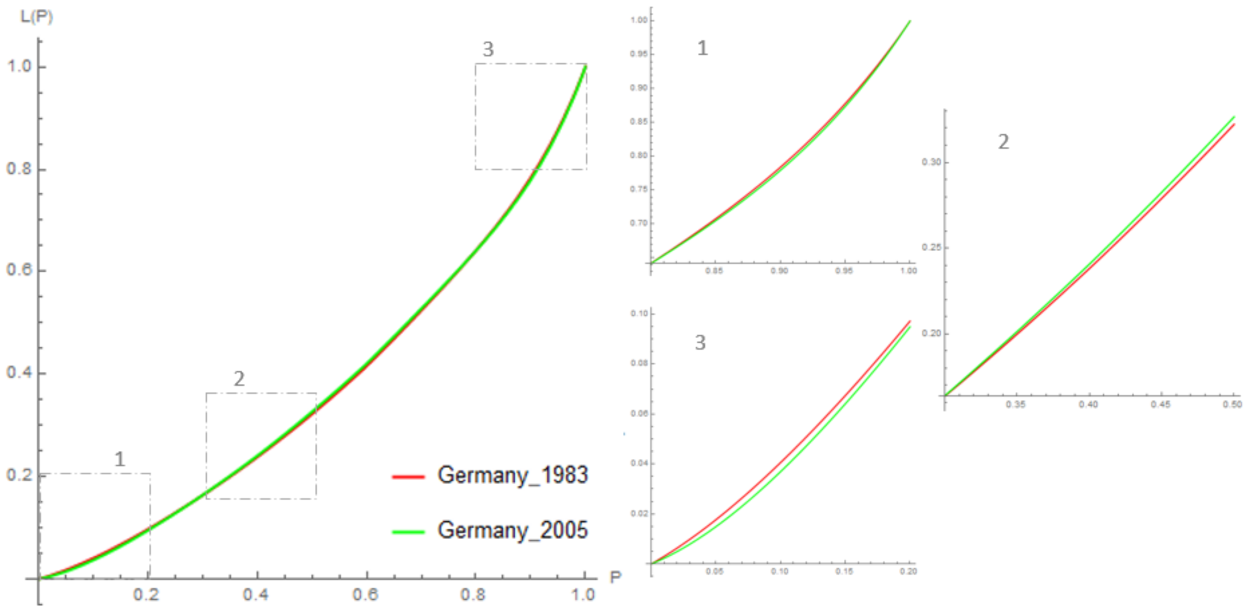

Figure 5.5: Lorenz curves of Germany income distributions in 1983 and 2005. Data Source: WIID [65].

On the contrary, if we use deciles or quintiles of these distributions we can gather more conclusions, although these ratios are not simple to use since we need to compare several values or use a graphical approach to derive conclusions faster. Looking to the information in Figure 5.6, it is possible to say that in 2005 the percentage of income of the richest people is similar to the one in 1983 and the percentage of income of the poorest is lower in 2005.

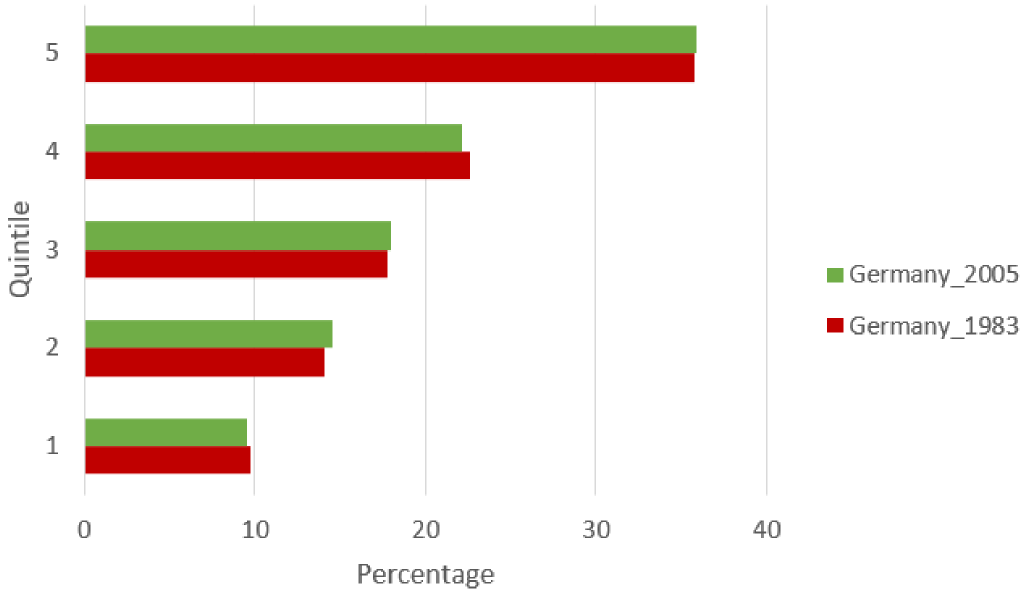

Figure 5.6: Quintiles of Germany income distributions in 1983 and 2005. Data Source: WIID [65].

As we have stated before, the GPlus is more sensitive to the poorest layers and the GMinus to the richest layers and, from the above, we are expecting to have a higher GPlus in 2005 (more inequality), although given the poorest middle layers (second and third quintile) shares, the GPlus will be slightly higher in 1983. The same happens for the richest layers as we are expecting to have a similar GMinus, but the fourth quintile will affect the result, and so, the GMinus in 1983 is higher (more inequality). These results are given in Table 5.4.

Table 5.4: Inequality common Measures and new coefficients results for income distributions in Figure 5.5.

	Germany 1983	**Germany 2005**
Gini	0.26	0.26
G+	0.300	0.302
G-	0.383	0.389
Theil	0.11	0.11
Palma	0.91	0.92
MLF	0.917	0.913

When calculating the difference between GMinus and GPlus for each year, it is possible to derive the Middle Layer Force and it will have a value of 0.917 in 1983 and 0.913 in 2005. This value is useful to understand which of the distributions has more percentage of income in the middle layers population. The higher value will represent the distribution with more income in these layers (Germany in 1983) as it is possible to view on Table 5.5.

Table 5.5: Percentage of total income for each society layer regarding the distributions in Figure 5.5.

	Percentage of income for		
	Bottom 10% of the population	**Medium Layers**	**Top 10% of the Population**
Germany_1983	4.06	74.32	21.62
Germany_2005	3.7	74.1	22.1

5.4 Portugal

From 1980 to 2018 the available data indicates the Gini Index in Portugal has changed over time, hitting a maximum (inequality) in 2005 (0.381) and a minimum in 1990 (0.31). In Figure 5.7 a graphical representation of the Gini index by year is presented. The green bars represent the years chosen to perform the analysis using the new coefficients (1980, 2001, 2005, 2010 and 2018). It is possible to see the Gini index in 1980 and 2018 is the same (0.32), even with the several fluctuations that occurred in the meantime. For that reason these will be interesting case studies.

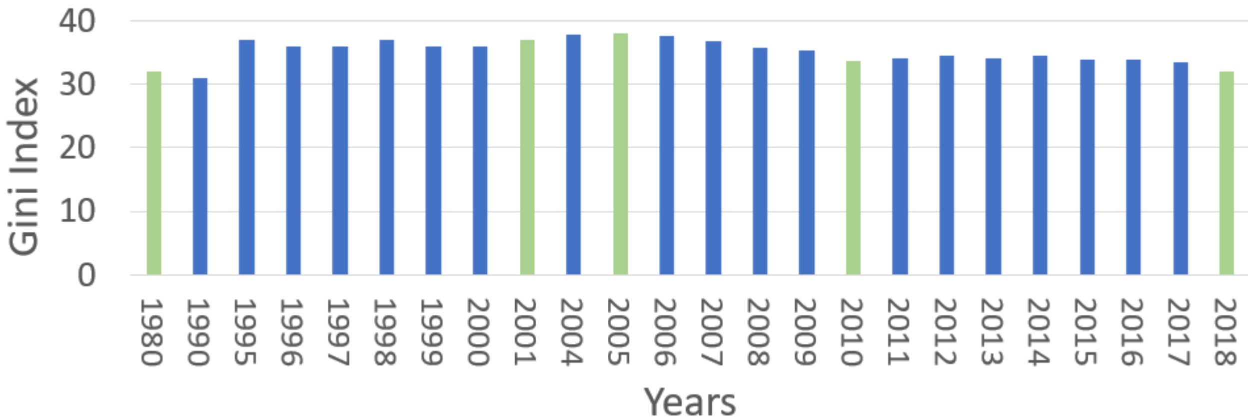

Figure 5.7: Graphical representation of Gini index in Portugal for several years from 1980 to 2018. Data Source: WIID [65]. The green bars represent the years that will be considered for further analysis.

In Figure 5.8 the five Lorenz curves are presented and the first conclusion about this dataset, obtained when looking at the graph, is that the Gini index is higher in 2005 (0.38), as there is more area between the blue curve and the equality line, followed by 2001 (0.36). For the other three distributions the Gini index should be really close and it is not possible to take meaningful conclusions using only the Lorenz curves. The GPlus and GMinus were calculated (Table 5.6) in order to interpret the distribution of inequality.

Table 5.6: Inequality common Measures and new coefficients results for income distributions in Figure 5.8.

	Portugal 1980	**Portugal 2001**	**Portugal 2005**	**Portugal 2010**	**Portugal 2018**
Gini	0.32	0.36	0.38	0.33	0.32
G+	0.37	0.41	0.42	0.38	0.37
G-	0.47	0.60	0.62	0.53	0.49
Theil	0.16	0.23	0.23	0.19	0.17
Palma	1.20	1.61	1.69	1.35	1.23
MLF	0.9	0.81	0.8	0.85	0.88

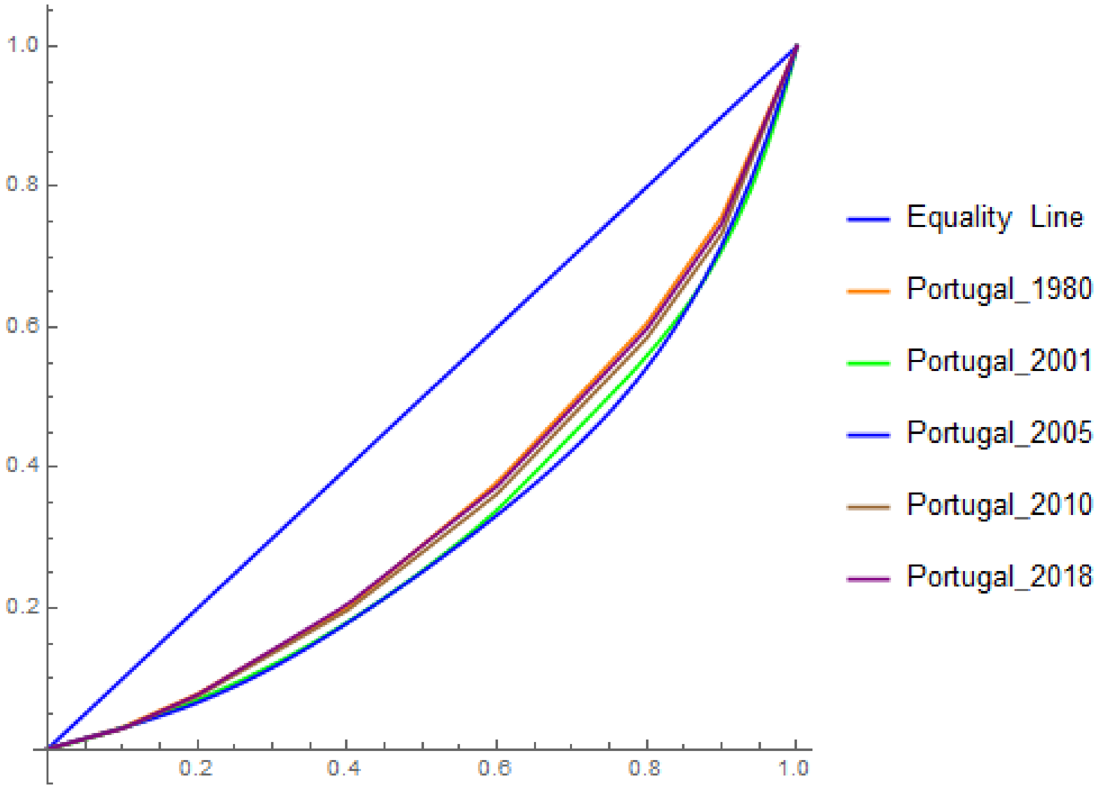

Figure 5.8: Lorenz curves of Portugal income distributions in 1980, 2001, 2005, 2010 and 2018. Data Source: WIID [65].

Looking at the GPlus results for the five distributions it is possible to say GPlus is higher in 2005, which makes sense given the fact that this is the most unequal and probably the one that has more poor people, i.e. the percentage of income for the bottom population (Quintile 1 in Table 5.9) is lower than in others. The GPlus is lower in 1980 and 2018, so these are the distributions with a higher percentage of income for the poorest people. On the other hand, GMinus is higher in 2005 as it has more rich people, i.e. more inequality than the others.

It is interesting to compare the distributions in Figure 5.10 (1980 and 2018) as they have the same Gini Index and it is not possible to draw meaningful conclusions when looking only to the Lorenz curves without zooming on the tails.

The first thing to note is that GPlus is also equal in 1980 and 2018, but GMinus is lower in 1980 (went from 0.47 in 1980 to 0.49 in 2018), which is an indicator that in 1980 there must have been more middle layers (the MLF is higher in 1980 than in 2018 - the middle layers are stronger). This can be observed when looking at the percentage of income by quintiles in Table 5.9, mainly in quintiles 3 and 4, where the percentage is higher in 1980. Also, the fact that GMinus is higher in 2018 - more inequality - can be corroborated by the Palma Ratio. As it is higher in 2018, it will need a higher transfer of income from the top population to the bottom to obtain a more equal distribution. Despite the fact that the Gini Index is equal, it is important to note that with other measures combined, there are some significant differences, mainly in the distributions tails, that lead to income inequality.

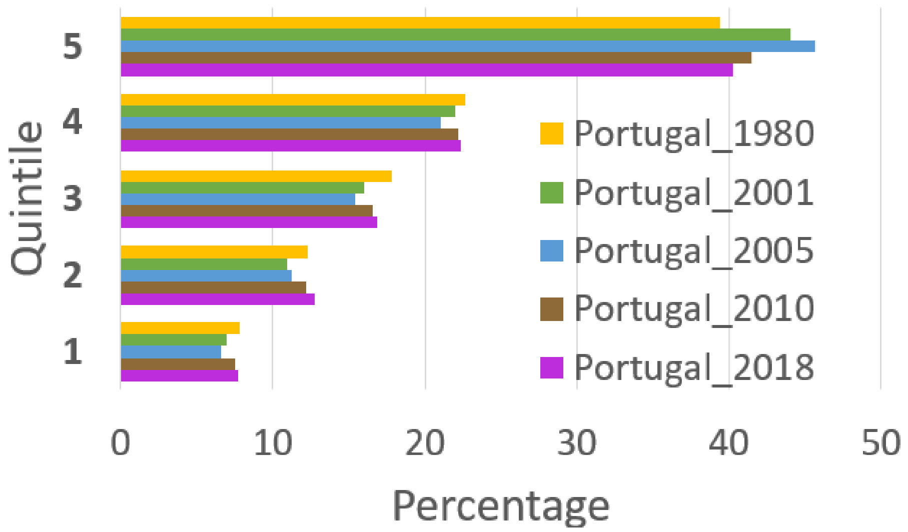

Figure 5.9: Quintiles of Portugal income distributions in 1980, 2001, 2005, 2010 and 2018. Data Source: WIID.

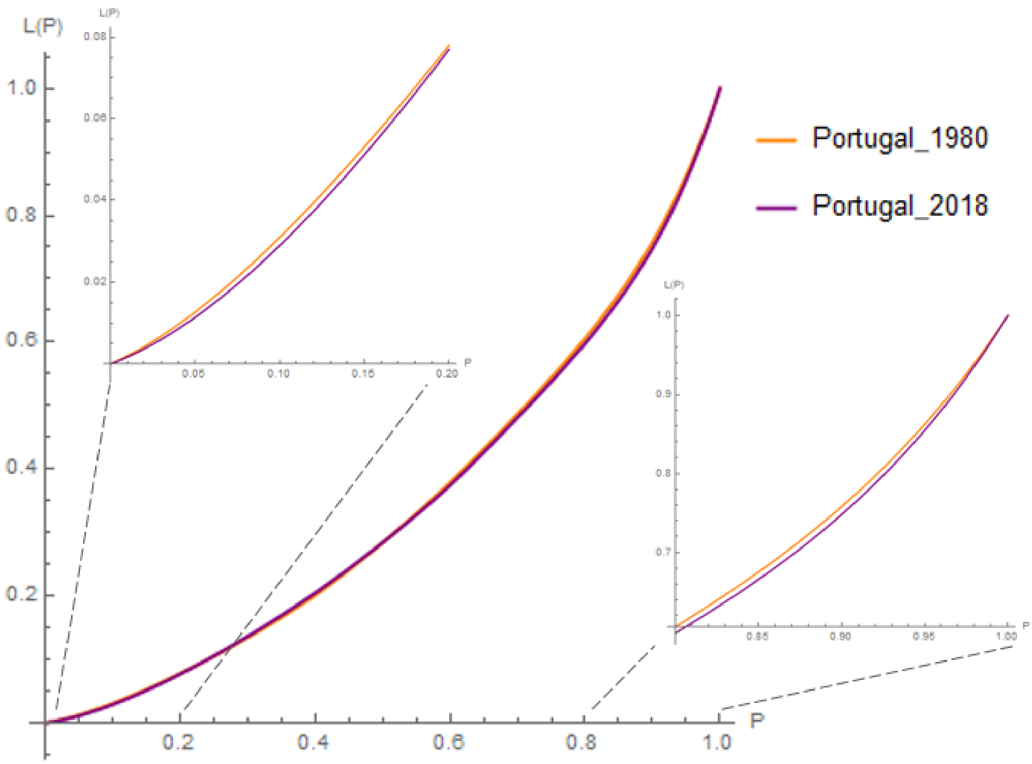

Figure 5.10: Lorenz curves of Portugal income distributions in 1980 and 2018 (part of Figure 5.8). Data Source: WIID [65].

5.5 Europe 2018

Looking to Figure 5.11, European countries are identified by their Gini index in 2018 (where the ones painted with dark blue are where there is more inequality) it is possible to conclude that Bulgaria (40.2) is the country with more inequality, while Slovakia and Slovenia (23.2) have more equality in their income distributions.

One fact about the countries represented in Figure 5.11 is that Bulgaria in 2000 had a Gini Index of 0.20 and Slovakia's index in 2015 was 0.28, meaning the inequality level changed across the years in these countries. From the 35 countries with data for 2018, Portugal is number 28 in an ordered list of countries (from low to high Gini index) which means that it is one of the countries in Europe with more income inequality, with a Gini index of 0.32, being only behind Spain, Italy, Greece, Latvia, Serbia, Lithuania and Bulgaria.

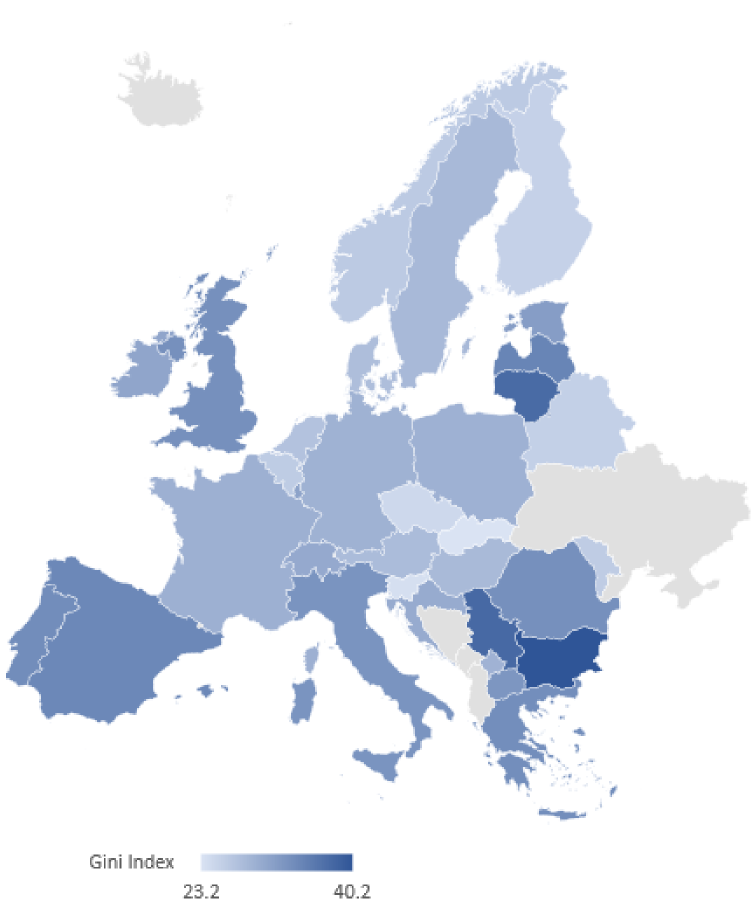

Figure 5.11: Europe inequality (given by Gini index) in 2018 for several European Countries. Lack of data for countries in grey. Data Source: WIID [65].

It is interesting to study Portugal's geographic area and how the countries around behave in terms of income distribution. For that the new coefficients were calculated for some European countries in 2018 (using the most complete and recent data available).

In Table 5.7 the results for 14 countries are presented and the following conclusions were gathered:

- As expected, GPlus and GMinus generally increase when Gini Index increases.
- For countries with the same Gini Index it's possible to compare them on the distribution tails, i.e., if the inequality is due to an increase of top deciles or bottom deciles shares. For example Belgium and Finland have a Gini Index of 0.26 but the bottom population in Belgium has less percentage of total income than in Finland, and the top decile in Finland has more percentage of total income

Table 5.7: Gini Index, GPlus and GMinus coefficients for several European countries in 2018.

	Slovenia	Belgium	Finland	Netherlands	Sweden	Denmark	Poland
G	0.23	0.26	0.26	0.27	0.27	0.28	0.28
G+	0.28	0.31	0.30	0.31	0.32	0.32	0.33
G-	0.33	0.36	0.39	0.39	0.38	0.41	0.41
MLF	0.95	0.95	0.91	0.92	0.94	0.91	0.92

	Germany	Portugal	Greece	Spain	Italy	Lithuania	Bulgaria
G	0.31	0.32	0.32	0.33	0.33	0.37	0.40
G+	0.36	0.37	0.37	0.39	0.39	0.42	0.44
G-	0.46	0.49	0.47	0.47	0.48	0.56	0.65
MLF	0.90	0.88	0.90	0.92	0.91	0.86	0.79

than in Belgium. For that reason, it is possible to say that the inequality in Finland, arise from top earners and Belgium is the country with more medium class.

- When comparing Portugal and Spain, although inequality in Spain is higher (0.33 vs 0.32), the GMinus coefficient is higher in Portugal. This means that the unequal distribution of income for the top population in Portugal has a lot of weight in Portugal's inequality.

- From the list, Portugal is the third country with less percentage of income for the medium class population. This can be seen when comparing the results obtained for Middle Layer Force term. For completeness, the income percentages per society layer are presented in Table 5.8. The country with more percentage of income for middle layer earners is Slovenia and with less percentage is Bulgaria, in line with the expected from MLF results.

Table 5.8: Percentage of the total income for each society's layer regarding the countries in Figure 5.7.

	Percentage of income for		
	Bottom 10% of the population	Medium Layers	Top 10% of the Population
Bulgaria	2.1	66.7	31.2
Lithuania	2.1	70.3	27.6
Portugal	2.9	71.9	25.2
Germany	2.5	72.8	24.7
Greece	2.5	72.9	24.6
Italy	2	72.9	25.1
Denmark	3.3	73.5	23.2
Spain	2.2	73.6	24.2
Finland	4.1	74	21.9
Netherlands	3.5	74.4	22.1
Poland	3.3	74.6	22.1
Sweden	3.3	75.4	21.3
Belgium	3.7	75.8	20.5
Slovenia	4.1	76.3	19.6

5.6 OECD - Organisation for Economic Co-operation and Development

The Organisation for Economic Co-operation and Development is an intergovernmental economic organisation founded in 1961 to stimulate economic progress and world trade. OECD and Key Partners represent about 80% of world trade and investment and so it is an interesting subject to study. OECD is composed of 37 countries, from North and South America to Europe and Asia-Pacific, as listed in Table 5.9, with the respective year of accession [70].

Table 5.9: OECD members and Accession year [70].

COUNTRY	ACCESS YEAR	COUNTRY	ACCESS YEAR	COUNTRY	ACCESS YEAR
AUSTRALIA	1971	HUNGARY	1996	NEW ZEALAND	1973
AUSTRIA	1961	ICELAND	1961	NORWAY	1961
BELGIUM	1961	IRELAND	1961	POLAND	1996
CANADA	1961	ISRAEL	2010	PORTUGAL	1961
CHILE	2010	ITALY	1962	SLOVAK REPUBLIC	2000
COLOMBIA	2020	JAPAN	1964	SLOVENIA	2010
CZECH REPUBLIC	1995	KOREA	1996	SPAIN	1961
DENMARK	1961	LATVIA	2016	SWEDEN	1961
ESTONIA	2010	LITHUANIA	2018	SWITZERLAND	1961
FINLAND	1969	LUXEMBOURG	1961	TURKEY	1961
FRANCE	1961	MEXICO	1994	UNITED KINGDOM	1961
GERMANY	1961	NETHERLANDS	1961	UNITED STATES	1961
GREECE	1961				

New coefficients were calculated using the average of the data available for OECD country members in the last years.

Due to the lack of equivalent data needed for new coefficients calculations (income percentage by quintile or decile) for all the countries and years, some of the countries were not included in all the years and the data used is detailed in Appendix B.

In order to verify the robustness of the chosen data, the Gini Index obtained using the quintile data available, was compared with the OECD average Gini Index reported in OECD data statistics reports [70]. Given it follows the same pattern and has an average difference across the years of only 2%, as represented in Figure 5.12, the quintile data selected was accepted and used to calculate the GPlus and GMinus coefficients.

The new coefficients were computed for all the years between 1999 and 2018, with the exception of 2002 and 2003, as there was not enough data to obtain valid results.

Although some of the OECD countries only joined the organization recently (as Colombia, that joined in 2020), all of them were used (when data is available) to calculate the OECD average quintile data in each year.

In Figure 5.12 the Gini Index in Portugal is also represented, and it is always superior to the OECD average Gini Index between 1999 and 2018, starting to be closer to the OECD average in recent years.

The lowest Gini Index reported in OECD was in 2001 and it was also the year with the lowest GPlus

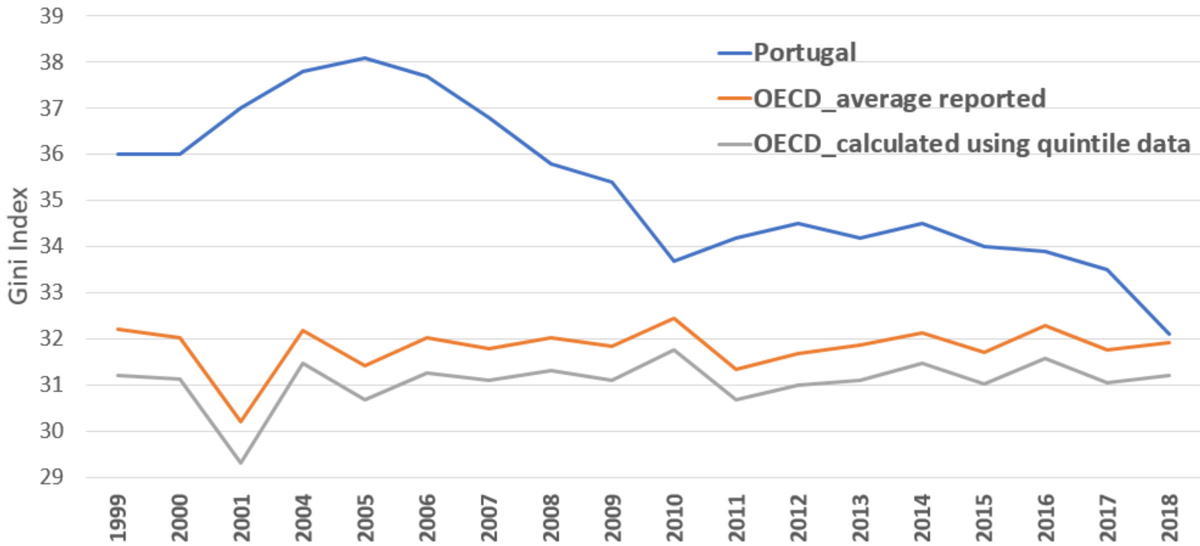

Figure 5.12: Portugal and OECD Average Gini Index between 1999 and 2018 (calculated using decile data and OECD reported).

and GMinus and the highest MLF (higher force of the middle layers), which means that 2001 was the year with less income inequality in the past 20 years. On the other hand, 2010 was the year with more income inequality, also corroborated by the higher GPlus and GMinus. Both 2001 and 2010 Lorenz curves are presented in Figure 5.13. Despite the fact that some higher and lower Gini Index results were obtained, it is clear that in OECD the Gini Index, GPlus and GMinus were almost constant over the last years, with no relevant changes between the years. The results are displayed in Table 5.10.

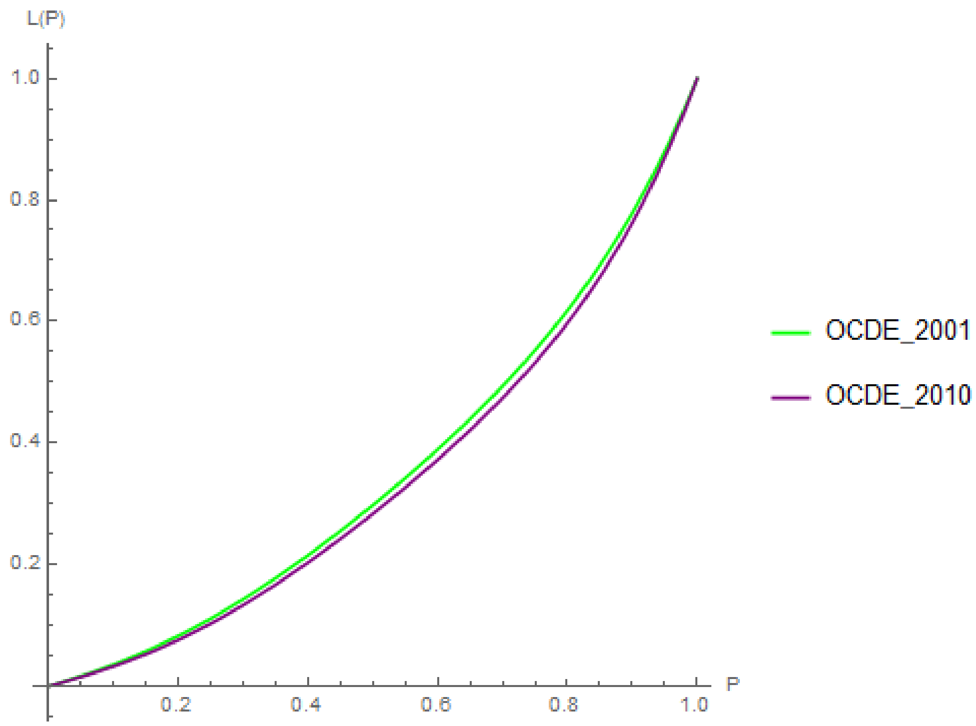

Figure 5.13: Lorenz Curves for OECD Average income distributions in 2001 and 2010.

In Figure 5.14 the Gini Index and MLF are represented for several OECD countries in 2018. As

Table 5.10: Gini Index, GPlus and GMinus coefficients for OECD between 1999 and 2018.

	1999	2000	2001	2004	2005	2006	2007	2008	2009
Gini Index	0.312	0.311	0.293	0.315	0.307	0.313	0.311	0.313	0.311
G+	0.365	0.365	0.349	0.369	0.361	0.367	0.365	0.367	0.365
G-	0.482	0.478	0.436	0.483	0.467	0.479	0.475	0.481	0.474
MLF	0.882	0.886	0.913	0.886	0.894	0.888	0.891	0.886	0.891

	2010	2011	2012	2013	2014	2015	2016	2017	2018
Gini Index	0.318	0.307	0.310	0.311	0.315	0.310	0.316	0.311	0.312
G+	0.373	0.362	0.366	0.367	0.371	0.366	0.372	0.366	0.368
G-	0.483	0.464	0.468	0.473	0.476	0.469	0.477	0.468	0.470
MLF	0.890	0.898	0.898	0.893	0.895	0.896	0.895	0.899	0.897

expected, the countries with higher inequality are the ones with the lower MLF and vice versa. Also, as expected, the South American countries (Mexico and Colombia) are the ones with higher income inequality (above OECD average). The new coefficients results were also computed for these countries (2018) and are presented in Table 5.11.

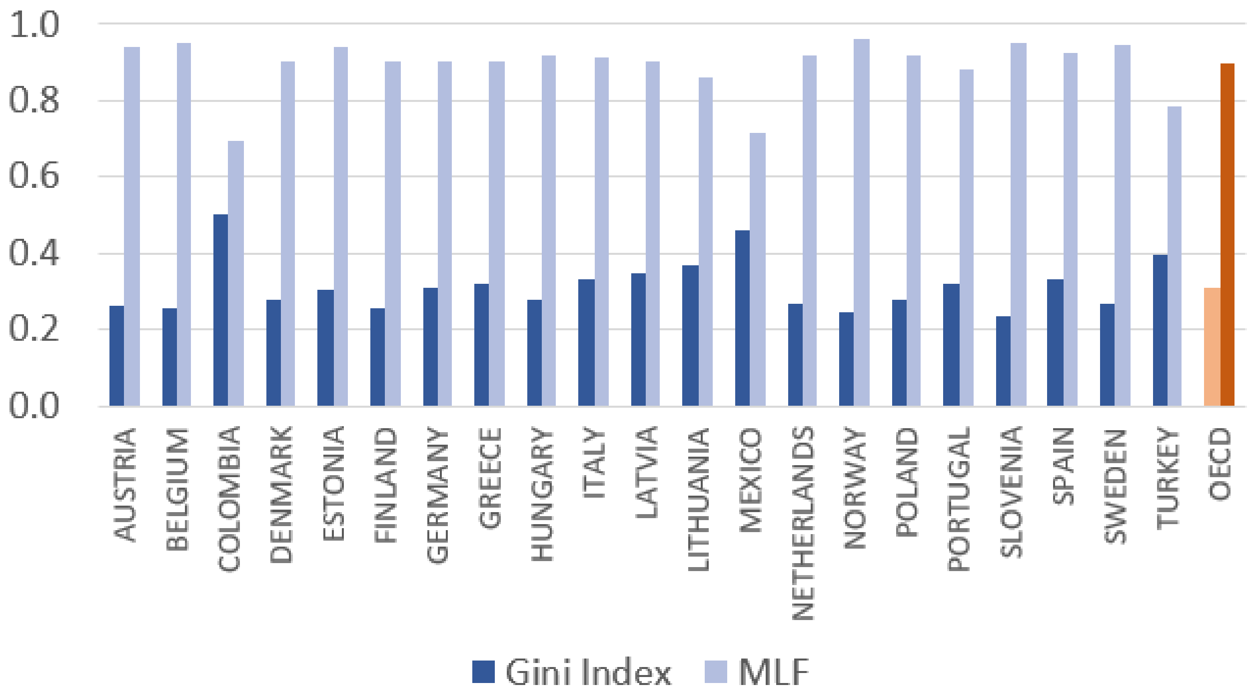

Figure 5.14: Gini Index and Middle Layers Force in 2018 for some OECD countries.

From the 21 countries analysed, 10 of them are over the OECD average for GPlus and 9 are over for GMinus. Estonia is a special case as, despite the fact that its Gini Index and GMinus are under the OECD average, its GPlus is higher, which implies a significant income inequality in the bottom percentage of population.

For countries with a Gini Index under the OECD average, the Middle Layer Force is higher than the OECD MLF but there are 4 more countries with a significant MLF: Greece, Spain, Italy and Latvia. These countries are then well balanced in the middle layer when compared with the OECD average.

The extreme cases in respect of all the coefficients are Slovenia (low inequality) and Colombia (high inequality). All the inequality coefficients in Colombia are more than twice times the ones in Slovenia and the Middle Layers Force in Colombia is really small when compared with the remaining countries. Their Lorenz curves are presented in Figure 5.15. It is also important to consider that, although Colombia is now part of OECD, it was not yet in 2018.

Table 5.11: New coefficients results for OECD countries in 2018. The cells highlighted in blue represent the OECD average results and the countries are ordered by Gini Index.

	Gini Index	G+	G-	MLF
Slovenia	0.234	0.281	0.331	0.950
Norway	0.244	0.298	0.338	0.960
Belgium	0.255	0.307	0.358	0.949
Finland	0.258	0.298	0.393	0.905
Austria	0.264	0.318	0.378	0.940
Netherlands	0.268	0.314	0.394	0.920
Sweden	0.269	0.322	0.378	0.944
Denmark	0.276	0.319	0.414	0.905
Hungary	0.280	0.333	0.415	0.918
Poland	0.280	0.330	0.410	0.920
Estonia	0.303	0.370	0.427	0.943
Germany	0.309	0.358	0.457	0.901
OECD	0.312	0.368	0.470	0.897
Portugal	0.320	0.370	0.490	0.880
Greece	0.321	0.374	0.471	0.904
Spain	0.331	0.393	0.470	0.923
Italy	0.333	0.390	0.476	0.915
Latvia	0.350	0.415	0.513	0.902
Lithuania	0.368	0.424	0.564	0.860
Turkey	0.396	0.449	0.664	0.785
Mexico	0.460	0.516	0.800	0.716
Colombia	0.503	0.564	0.869	0.695

Using the Portugal results from the middle years in the last 20 years and comparing them with the OECD average is also important to verify the inequality evolution in Portugal. The comparison between the new coefficients and MLF are graphically represented in Figures 5.16 and 5.17.

It is possible to observe the Portuguese decrease in income inequality as it is becoming closer to the OECD global trend, both in relation to bottom and top population (GPlus and GMinus). When looking to the MLF evolution, the efforts to break the inequality are also visible, given the strong increase that the MLF had across these years, increasing the percentage of income in the middle layer society.

5.7 High Inequality

Figure 3.2 presented the Gini Index around the world in 2000 and 2015. For some of the countries represented there was an increase in income inequality, while for others a decrease occurred, but, in general, some Latin America and African countries represent the ones with higher income inequality in the world in both periods. We've already explored European countries, the ones with lower inequality in the world and the results for some countries with more income inequality will now be presented. Due to the lack of recent good data to obtain the new coefficients results for African countries, the analysis will

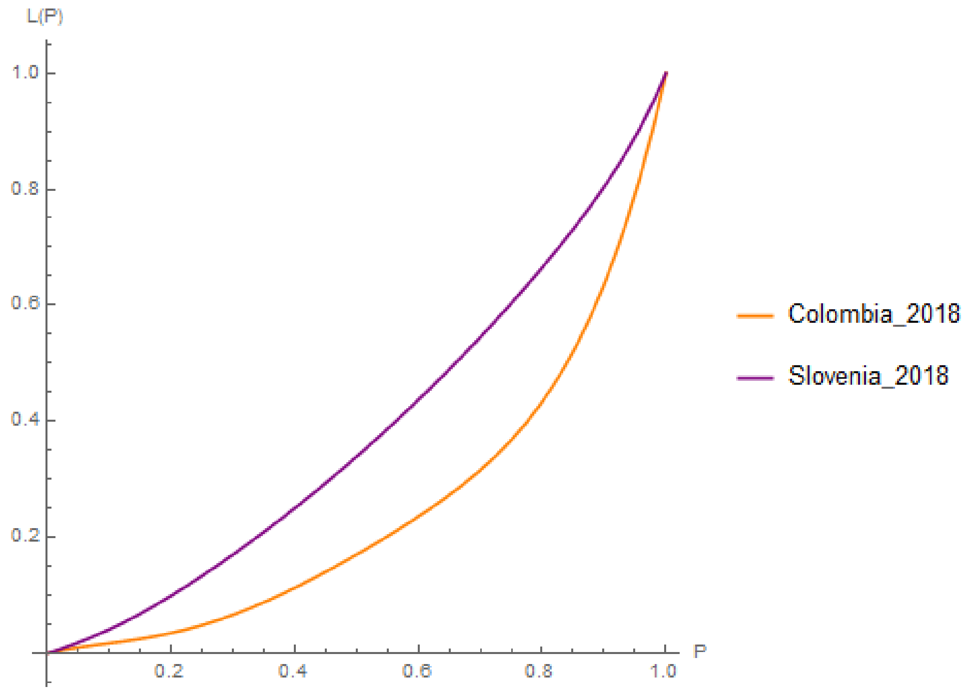

Figure 5.15: Lorenz curves of OECD with extreme results in 2018: Slovenia and Colombia.

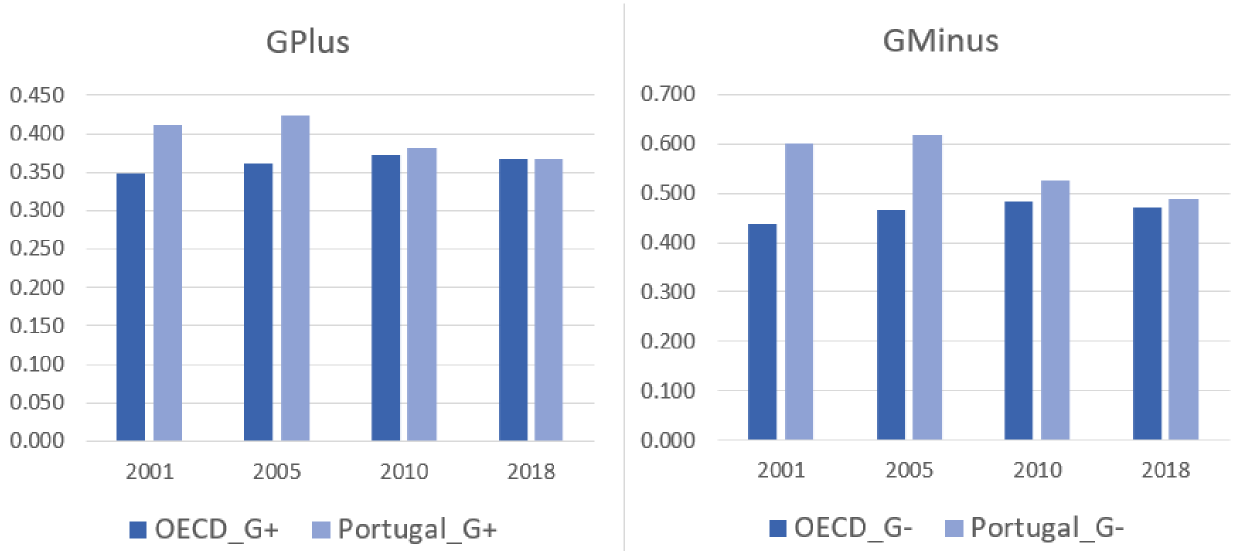

Figure 5.16: GPLus and GMinus in OECD and Portugal in 2001, 2005, 2010 and 2018.

consider four Latin American countries in 2015 and 2017, the ones with higher Gini Index in 2017 and decile information available.

In Figure 5.18 the Gini Index differences between 2015 and 2017 are presented for Brazil, Colombia, Paraguay and Chile. Both Brazil and Paraguay had an increase in their Gini Index between 2015 and 2017, while for Colombia and Chile, it decreased.

For Brazil and Paraguay the % of total income shared by the richest people (top 20%) increased over the 2 years (3.2% and 2.1%, respectively). However, in Brazil, the % of total income for the bottom 20% of the population decreased, from 4.1% to 3.25%, and in Paraguay increased from 4.5% to 5%, as can be observed in Figure 5.19. As expected the GPlus and GMinus results are in line with these changes.

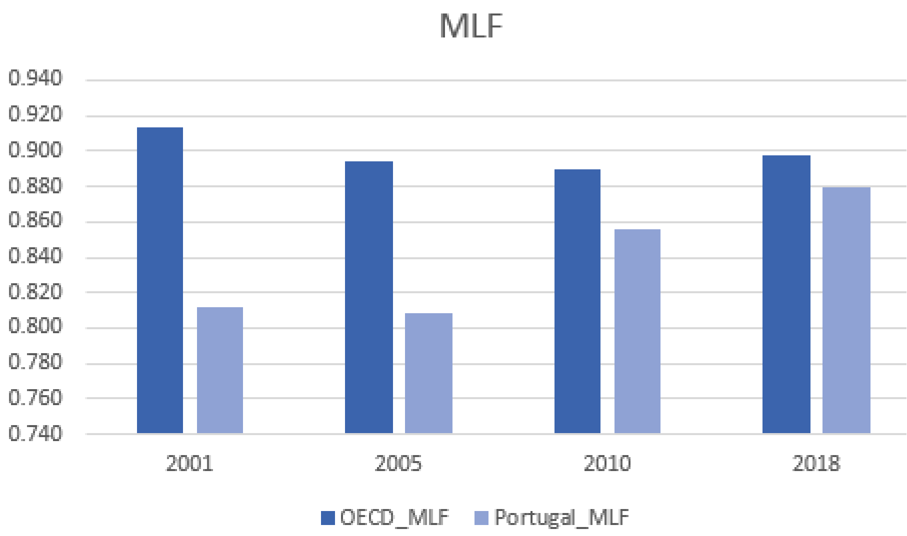

Figure 5.17: MLF in OECD and Portugal in 2001, 2005, 2010 and 2018.

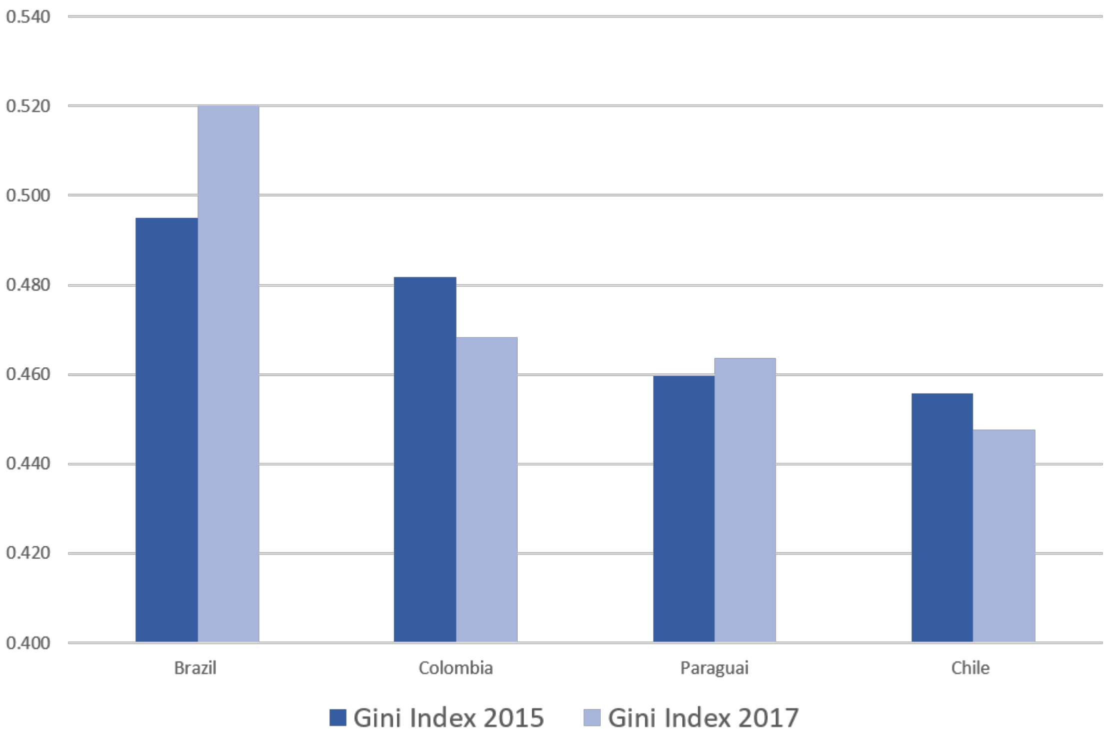

Figure 5.18: Gini Index in 2015 and 2017 for 4 Latin America countries.

For Brazil, between 2015 and 2017 the GPlus and GMinus increased, while for Paraguay, GPlus decreased and GMinus increased, that is consistent with the fact that the bottom 20% of the population in 2017 receives a bigger percentage of total income and consequently there is a small reduction in the lower tail inequality.

In both cases (Brazil and Paraguay) it is possible to conclude from the MLF results presented in Table 5.12 that in 2015 the income shared by the middle class is higher than in 2017, given the MLF in 2015 is higher than in 2017.

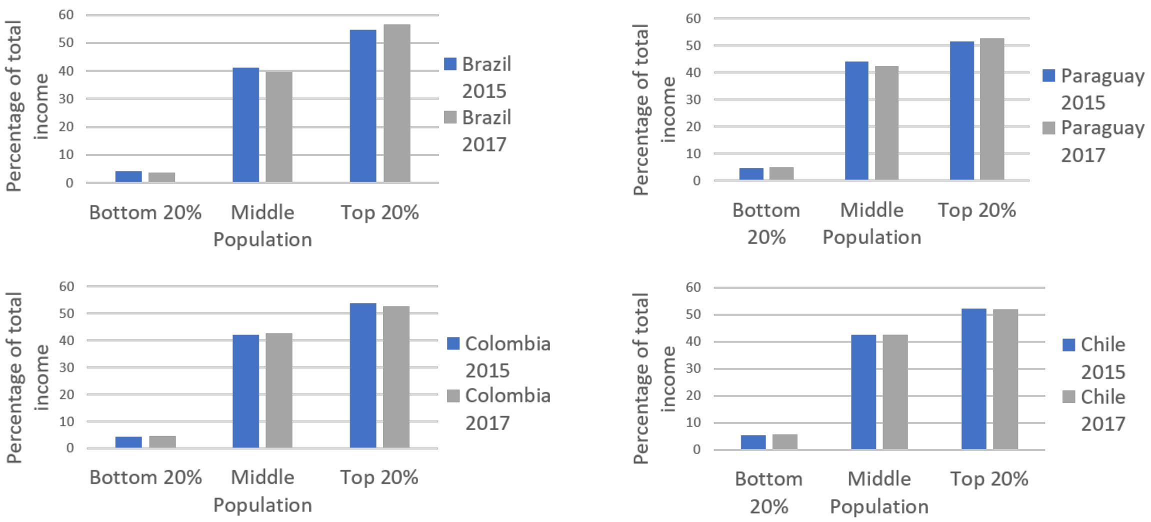

Figure 5.19: Percentage of total income for each society layer regarding the countries in Figure 5.18.

Table 5.12: Gini Index and new coefficients results in 2015 and 2017 for 4 Latin America countries.

2015

	Brazil	**Colombia**	**Paraguay**	**Chile**
G	0.495	0.482	0.460	0.456
G+	0.555	0.528	0.509	0.497
G-	0.857	0.859	0.789	0.824
MLF	0.698	0.669	0.719	0.674

2017

	Brazil	**Colombia**	**Paraguay**	**Chile**
G	0.520	0.468	0.464	0.448
G+	0.560	0.514	0.504	0.486
G-	0.865	0.830	0.848	0.822
MLF	0.694	0.684	0.656	0.664

Colombia and Chile had a decrease in their Gini Index, becoming less unequal, which was achieved with some sharing (transfers) of total income from the richest to the poorest. The top percentage of the population has a lower proportion of total income in 2017 than in 2015, and the bottom population receives more income in 2017, which leads to more income equality in both lower and higher tails of the distributions. This can be confirmed when looking to the new coefficients results, as both GPlus and GMinus decreased from 2015 to 2017 in Colombia and Chile.

In both cases (Colombia and Chile) it is possible to conclude from the MLF results presented in Table 5.12 that in 2015 the income shared by the middle class is lower than in 2017, given the MLF in 2017 is higher than in 2015. Given the results for these four countries, it can be concluded that the force of middle layers is an important factor to consider, when analysing income inequality.

5.8 Sensitivities - Deciles vs Quintiles

In this section we will compare the results of new coefficients for the same income distributions but using decile and quintile data. The goal is to understand if GPlus (the index of the poor) is more sensitive for the bottom 10% of the population or bottom 20% and the behaviour of GMinus for the top shares.

Brazil is one of the countries in the world with a higher inequality index and for that reason it is an interesting subject to study. In Figure 5.20 a graph with the evolution of Gini Index, GPlus and GMinus coefficients in Brazil is shown for several years.

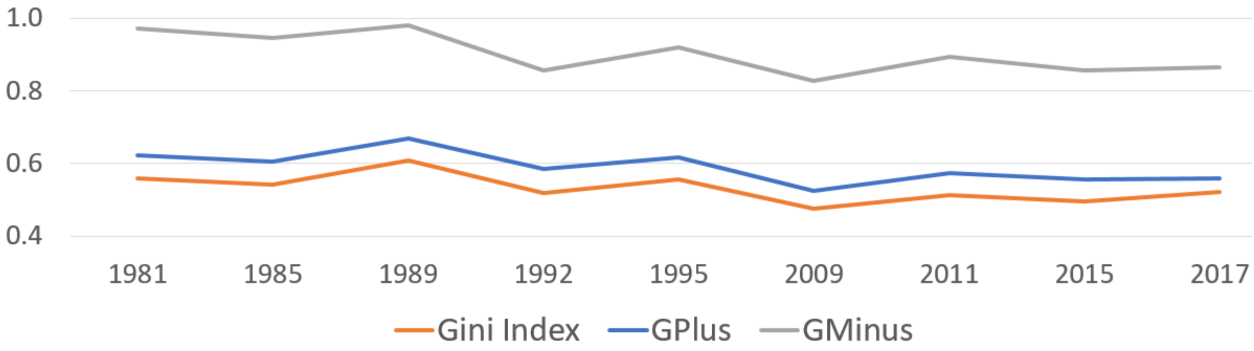

Figure 5.20: Graphical representation of Gini Index and new coefficients for Brazil income distribution between 1981 and 2017 (From Table 5.13).

The first point to note is that the minimum Gini index represented is greater than 40 (Table 5.13), which is higher than the highest Gini Index in Europe (Bulgaria 2018 = 40.2). The second important fact is that, as already flagged, both GPlus and GMinus have the same behaviour has Gini Index, i.e., they increase and decrease when Gini Index increases and decreases. sharing the same extremes.

As already reported, Brazil is one of the countries with more income inequality, and from the results obtained it is possible to say that this inequality is due to the fact that the total income is almost all shared by the top richest people. Also, the force of the middle layers is not too strong in Brazil. Looking at 1992, the year with the highest MLF, 0.728, this is still a low value compared to the European average in 2018 (0.90). Also, the bottom population in Brazil has a low percentage of the total income (associated with high values of GPlus). The average 2017 GPlus in OECD is 0.366, which is low compared to the Brazil result of 0.560.

Table 5.13: Gini Index and new coefficients for Brazil income distribution between 1981 and 2017.

YEAR	1981	1985	1989	1992	1995	2009	2011	2015	2017
G	0.559	0.541	0.608	0.518	0.556	0.475	0.512	0.495	0.520
G+	0.622	0.600	0.668	0.585	0.616	0.525	0.572	0.555	0.560
G-	0.971	0.945	0.980	0.857	0.940	0.828	0.893	0.857	0.865
MLF	0.651	0.655	0.688	0.728	0.676	0.697	0.679	0.698	0.695

The results showed in Figure 5.20 are based on decile data but with an interpolation factor of 7, in order to approximate the points toward a more robust Lorenz curve, and the respective results are presented in Table 5.13.

New coefficients were also calculated using decile and quintile data but with an interpolation function of order 1, i.e., using the specific points (a line instead of a curve). The difference between these two approaches, is that we have more information using decile data instead of quintile data. As already described, to construct a Lorenz curve (F(x), in GPlus and GMinus equations) a set of points are needed, representing the accumulated percentage of income. The more points we have, the more robust is our curve, without the need of using interpolation functions to "detect" the missing points. For that reason, when using only quintile data points we are more distant from the real Lorenz curve for a given country than when using decile ratios, and, the addition of these points into the curve reflects in GPlus and GMinus coefficients results. In Figure 5.21 a representation of accumulated functions using quintile and decile data from Brazil 2017 is presented and it is possible to see that the line using decile data is closer to the Lorenz curve function (obtained using the decile data with an interpolation factor of 7).

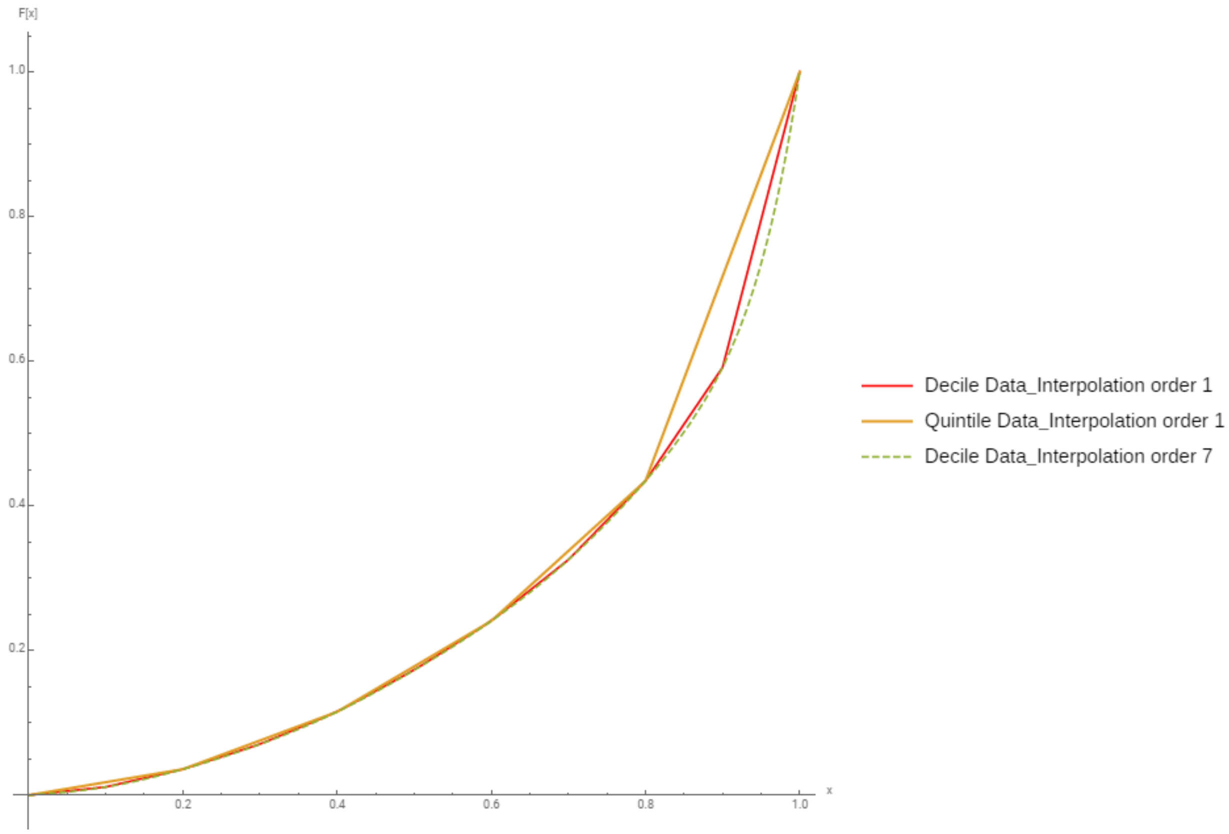

Figure 5.21: Graphical representation of accumulated functions using quintile and decile data from Brazil 2017.

In Table 5.14 the results obtained using some of Brazil deciles and quintiles information are presented, as well as some arbitrary distributions based on Brazil 2017.

The first point to note is that the Gini Index is closer to the "real" Gini Index (from Table 5.13) when using decile data for the three Brazil results presented, reflecting the "gain" in data that is expected. Secondly, it is noted the decile results are higher than the quintile results. This is also a proof of the gain of information we have when using decile data as it reflects the variation that can occur, mainly in the tails of the distributions.

Table 5.14: Gini Index and new coefficients results for Brazil income distribution and arbitrary distribution based on Brazil 2017 data, using quintile and decile information.

	Brazil 1995	Brazil 2009	Brazil 2017	A	B	C	D	E	A2
			Quintile Data Results						
G	0.521	0.438	0.469	0.469	0.469	0.469	0.469	0.469	0.469
G+	0.606	0.519	0.552	0.552	0.552	0.552	0.552	0.552	0.552
G-	0.818	0.676	0.727	0.727	0.727	0.727	0.727	0.727	0.727
			Decile Data Results						
G	0.557	0.465	0.500	0.500	0.500	0.500	0.501	0.500	0.502
G+	0.616	0.525	0.560	0.560	0.560	0.560	0.560	0.560	0.562
G-	0.970	0.785	0.865	0.863	0.862	0.862	0.865	0.855	0.851

In Table 5.14 distributions A/B/C/D/E/A2 are given by small variations of Brazil 2017 distribution (difference in the 1^{st}, 3^{rd}, 5^{th}, 7^{th} and 9^{th} decile data as identified in Figures 5.22 to 5.26). So, the results using quintile data are exactly the same, but the results for decile data changes for each distribution given the small variations on each decile:

- The difference between Brazil 2017 and distribution A is on the 1^{st} decile, as can be observed in Figure 5.22. In distribution A, the 1^{st} decile data was reduced in order to add more disorder in the lower tail of the distribution. In Table 5.14 the results obtained indicate that there is no change for GPlus (the index of the poor) from Brazil 2017 to distribution A results, but there is a small change of GMinus. So, although the variation is in the poor tail side, the GMinus decreases a bit, in line with a small disorder on the rich side (when compared to the poor). This allows to conclude that, despite the fact that GPlus can be associated with the bottom percentage of population and GMinus with the top earners, both coefficients are affected by overall changes in income percentages.

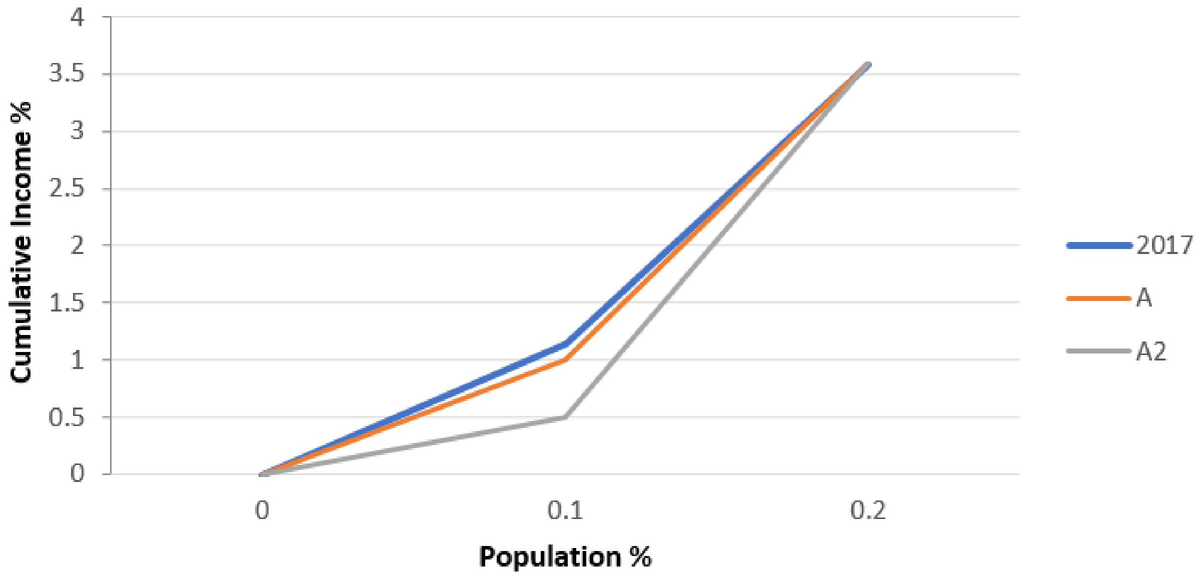

Figure 5.22: Graphical representation of part of the accumulated functions from Brazil 2017 and distributions A and A2.

- From A to B there is a low reduction (from 7.05% to 7%) on the 3^{rd} decile shared percentage (Figure 5.23), which is reflected only on the GMinus result - the inequality on the low side is not affected, but overall, GMinus coefficient can be sensitive to this little change and decrease (less unequal) on the top side. This result is in line with the one obtained from A to A2 distributions changes.

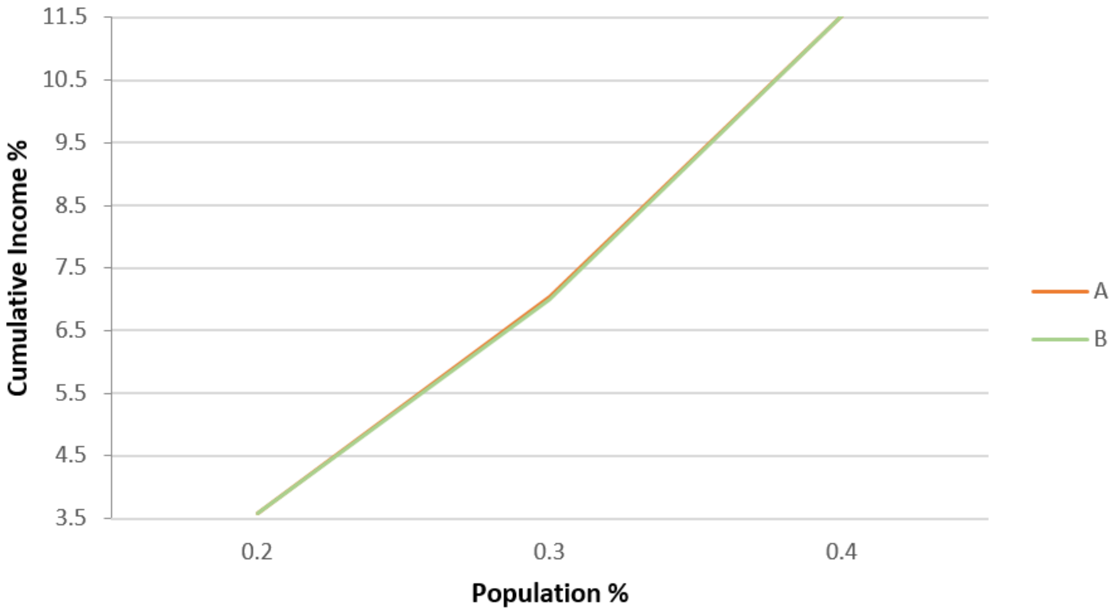

Figure 5.23: Graphical representation of part of the accumulated functions from A and B distributions.

- From B to C there is a small increase (from 17.13% to 17.5%) on the 5^{th} decile shared percentage (Figure 5.24), but, there are no changes in any coefficient - i.e., a small change in the middle of the distribution doesn't seem to affect the new coefficients results. It is also important to note the use of rounded income percentages as well as rounded results which can also have an associated error.

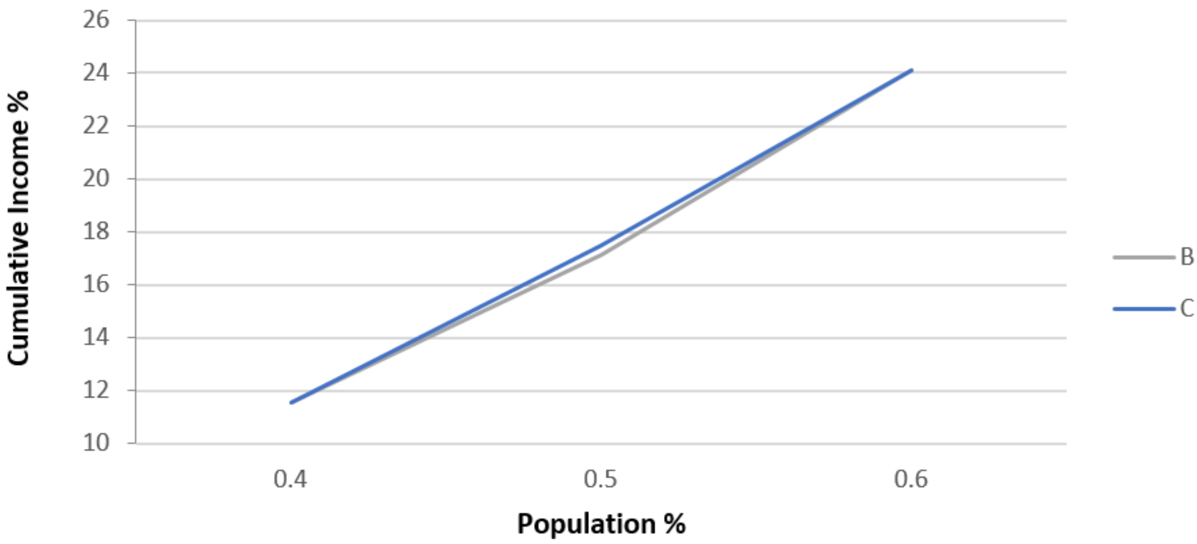

Figure 5.24: Graphical representation of part of the accumulated functions from B and C distributions.

- From C to D there is a small reduction (from 32.55% to 32%) on the 7^{th} decile shared percentage (Figure 5.25), which is reflected on the GMinus result - a decrease in a top decile slightly increases the inequality on the top side of the distribution. GPlus is equal in both distributions but the GMinus moves from 0.862 to 0.865.

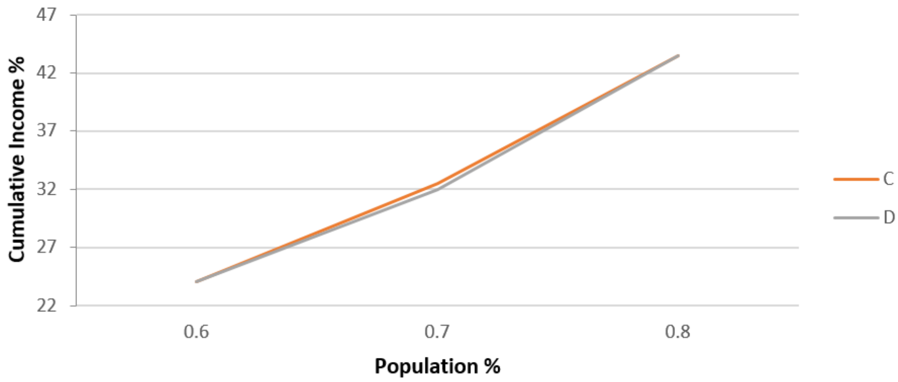

Figure 5.25: Graphical representation of part of the accumulated functions from C and D distributions.

- From D to E there is a small increase (from 59.15% to 60%) on the 9^{th} decile shared percentage (Figure 5.26), which is reflected in the GMinus result - an increase in a top decile slightly decreases the inequality on the top side of the distribution.

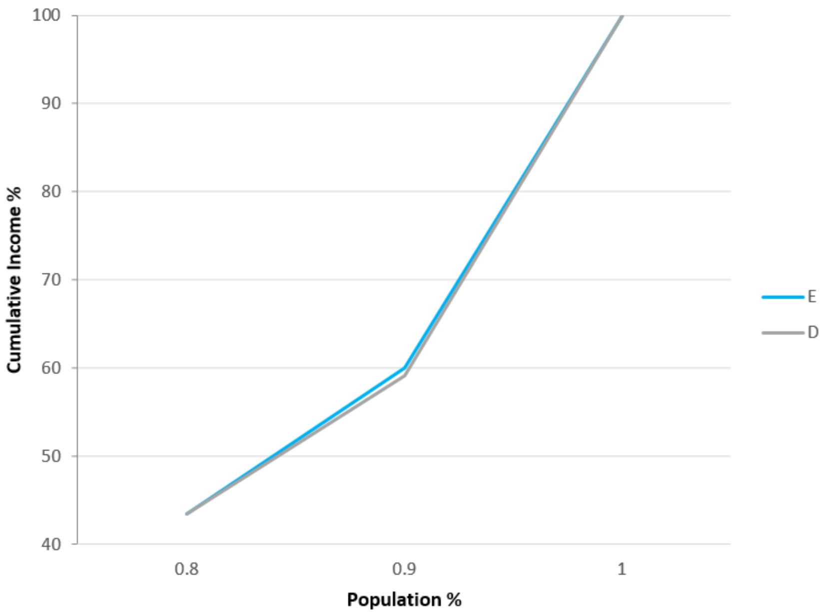

Figure 5.26: Graphical representation of part of the accumulated functions from D and E distributions.

- Distribution A2, has the same kind of variation as A, but with a lower percentage in the 1^{st} decile, making the GPlus coefficient sensitive to the change as it can be seen in Table 5.14. From Brazil 2017 and distribution A to distribution A2, the GPlus increases from 0.560 to 0.562, showing a bit more of inequality on the lower tail and fitting with the Gini Index change. This is also reflected in the GMinus, which decreases given the overall behaviour of the upper tail when compared with the full distribution.

5.9 Mathematical Sensitivities

Let's now assume there is a perturbation in the income earned by 1 individual in the population. Assumptions:

- The individual where the income perturbation, ξ, occurs, will occupy the same $x - axis$ position after and before the perturbation, i.e., the individual will keep his own order (x_0), in the cumulative proportion of population ranked by income.
- In order to keep also the cumulative proportion of income, the income increment should be given in a way that the income for that individual doesn't pass the income earned by the individual in the following position.
- According to the above, a perturbed Lorenz curve, $\tilde{F}(x)$ is defined as per 5.2 and an example is presented in Figure 5.27:

$$\tilde{F}(x) = \begin{cases} \frac{F(x)}{1+\delta}, & \text{if } 0 < x < x_0 \\ \frac{F(x)+\delta}{1+\delta}, & \text{if } x_0 < x < 1 \end{cases} \tag{5.2}$$

where, δ is the impact of the income increment ξ to one individual in population size p, in order to normalize the perturbed distribution and keeping the main characteristics of a Lorenz curve.

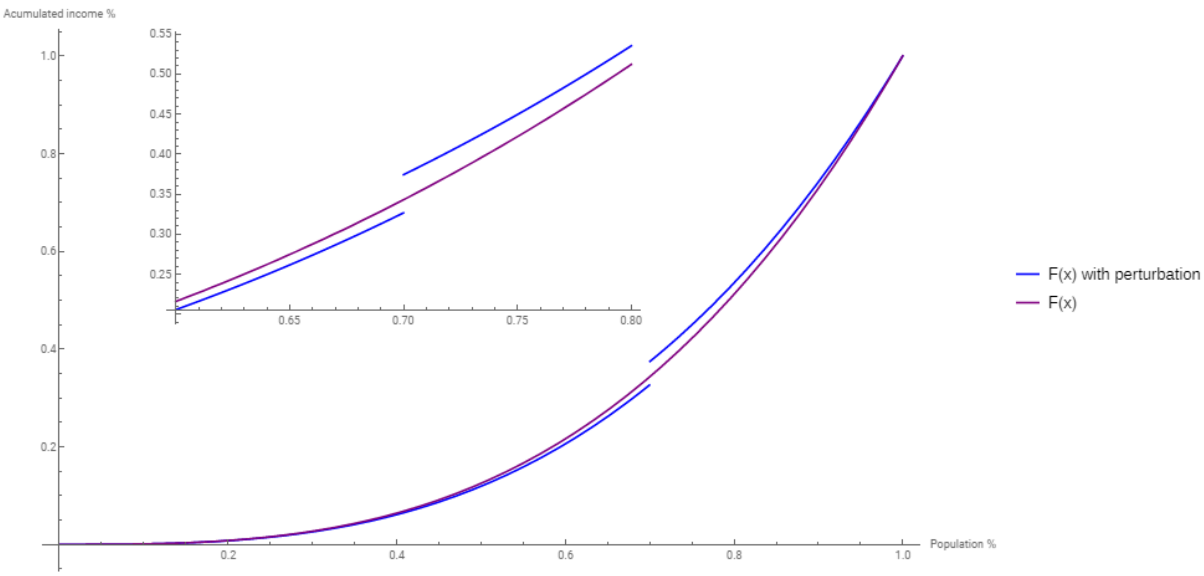

Figure 5.27: Example of a Lorenz curve after ($\tilde{F}(x)$) and before ($F(x)$) a small income increment (perturbation) for a person on position $x_0 = 0.6$.

Let's first analyse what happens to the Gini Index with this perturbation. It is already known that the Gini Index is given in function of the Lorenz curve $F(x)$ as per Equation 5.3:

$$G = 1 - 2\int_0^1 F(x)\mathrm{d}x <=> \int_0^1 F(x)\mathrm{d}x = \frac{1-G}{2}. \tag{5.3}$$

The perturbed Gini Index will then be given by Equation 5.4:

$$\tilde{G} = 1 - 2\int_0^1 \tilde{F}(x)\mathrm{d}x, \tag{5.4}$$

replacing 5.2 in 5.4:

$$\begin{aligned} \tilde{G} &= 1 - 2\int_0^{x_0} \frac{F(x)}{1+\delta}\mathrm{d}x - 2\int_{x_0}^1 \frac{F(x)+\delta}{1+\delta}\mathrm{d}x \\ &= 1 - 2\int_0^{x_0} \frac{F(x)}{1+\delta}\mathrm{d}x - 2\int_{x_0}^1 \frac{F(x)}{1+\delta}\mathrm{d}x - 2\int_{x_0}^1 \frac{\delta}{1+\delta}\mathrm{d}x, \end{aligned} \tag{5.5}$$

expanding and simplifying 5.5:

$$\tilde{G} = 1 - \left[2\int_0^1 F(x)\mathrm{d}x + 2\int_{x_0}^1 \delta \mathrm{d}x\right](\frac{1}{1+\delta}), \tag{5.6}$$

replacing 5.3 in 5.6, and solving, the following is obtained:

$$\begin{aligned} \tilde{G} &= \left[1 + \delta - 2\int_0^1 F(x)\mathrm{d}x - 2\delta + 2\delta x_0\right](\frac{1}{1+\delta}) \\ &= [G - \delta + 2\delta x_0](\frac{1}{1+\delta}). \end{aligned} \tag{5.7}$$

Let's now explore what happens to $\Delta G = \tilde{G} - G$, for different x_0 using Equation 5.7. When x_0 is near 0, i.e., when the small income increment is given to the poorest person in the population, ΔG will be negative, as $\tilde{G}$ is lower than G and inequality decreases. When x_0 is near 1, the difference is positive, i.e., $\tilde{G}$ is higher than G and inequality increases. An interesting point to investigate is the critical point where the income increment changes the inequality trend, $\Delta G = 0$:

$$\begin{aligned} &\Delta G = 0 \\ &\Leftrightarrow \left[G - \frac{\delta}{p} + 2\frac{\delta}{p}x_0\right](\frac{1}{1+\frac{\delta}{p}}) - G = 0 \\ &\Leftrightarrow \left[G - \frac{\delta}{p} + 2\frac{\delta}{p}x_0 - G - G\frac{\delta}{p}\right](\frac{1}{1+\frac{\delta}{p}}) = 0 \\ &\Leftrightarrow -\frac{\delta}{p} + 2\frac{\delta}{p}x_0 - G\frac{\delta}{p} = 0 \\ &\Leftrightarrow x_0 = \frac{1+G}{2}. \end{aligned} \tag{5.8}$$

Further to this, for example, when the original Gini index, G, is 0.50, this implies that as long as the recipient of income increment (x_0) is exactly in the position equivalent to the 75% of the distribution ($\frac{1+0.5}{2}$), inequality will not change, and $\tilde{G} = 0.5$. On the other hand, if the small increment goes to a household in the upper 25%, inequality will increase. This result is in line with that obtained by Hoffman (2001) who considered the effect of an "infinitesimal increment" which does not change the income order [71].

A graphical representation of ΔG for several Lorenz curves type x^a is represented in Figure 5.28 and, as expected, the higher the Gini Index (larger a), the higher the critical point.

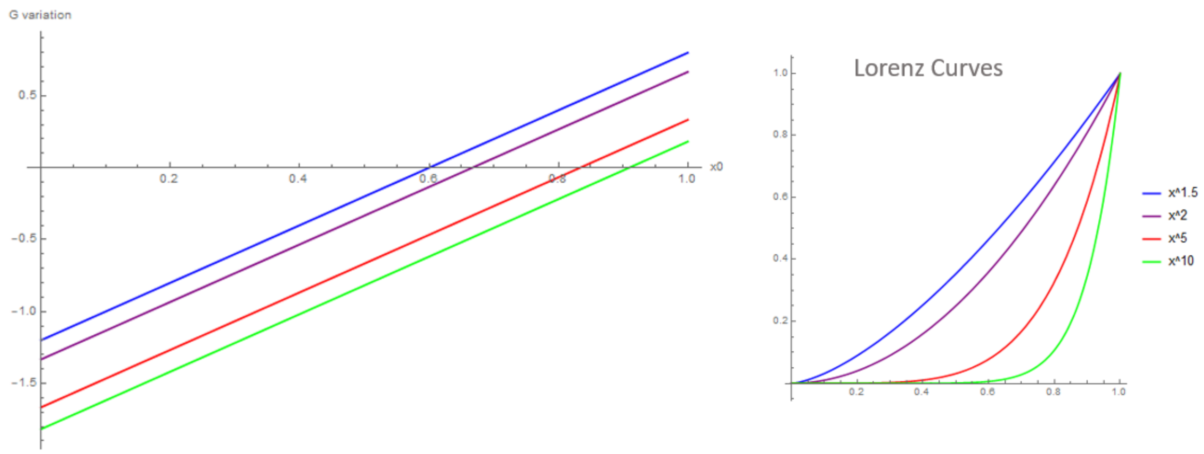

Figure 5.28: Δ G for several Lorenz curves (x^a).

Let's now do the same for GMinus and GPlus coefficients, using the same perturbed curve (Equation 5.2). In order to simplify the calculations, we assume the following identities:

$$\begin{aligned}
\ln(a+\delta) &\simeq \ln a + \frac{\delta}{a}, \\
\frac{1}{a+\delta} &\simeq \frac{1}{a}\left(1-\frac{\delta}{a}\right), \\
A &= \int_0^1 F(x)dx, \\
S &= \int_0^1 F(x)\ln F(x)dx, \\
C &= \int_0^1 (1-F(x))\ln(1-F(x))dx, \\
D &= \int_0^1 1-F(x)dx, \\
G_- &= -1-\frac{2S}{A}, \\
G_+ &= 1+\frac{2C}{D}, \\
\frac{2S}{A} &= (-1-G_-), \\
\frac{2C}{D} &= (-1+G_+).
\end{aligned}$$

The perturbed GMinus will be given as per Equation 5.9:

$$\tilde{G}_- = -1 - 2\frac{\int_0^{x_0}\frac{F(x)}{1+\delta}\ln\frac{F(x)}{1+\delta}\mathrm{d}x + \int_{x_0}^1 \frac{F(x)+\delta}{1+\delta}\ln\frac{F(x)+\delta}{1+\delta}\mathrm{d}x}{\int_0^{x_0}\frac{F(x)}{1+\delta}\mathrm{d}x + \int_{x_0}^1\frac{F(x)+\delta}{1+\delta}\mathrm{d}x}. \tag{5.9}$$

Simplifying, changing the notation ($F(x)=f$) and omitting the $\mathrm{d}x$ (considering the integration is always with respect to x):

$$
\begin{aligned}
\tilde{G}_- &= -1-2\frac{\int_0^{x_0} F(x)\ln\frac{F(x)}{1+\delta}\mathrm{d}x + \int_{x_0}^1 (F(x)+\delta)\ln\frac{F(x)+\delta}{1+\delta}\mathrm{d}x}{\int_0^{x_0} F(x)\mathrm{d}x + \int_{x_0}^1 (F(x)+\delta)\mathrm{d}x} \\
&= -1-2\frac{\int_0^{x_0}(f\ln f - f\ln(1+\delta)) + \int_{x_0}^1 f\ln\frac{f+\delta}{1+\delta} + \delta\int_{x_0}^1 \ln\frac{f+\delta}{1+\delta}}{\int_0^{x_0} f + \int_{x_0}^1 f + \int_{x_0}^1 \delta},
\end{aligned}
\tag{5.10}
$$

developing using the identity $\ln(a+\delta) \simeq \ln a + \frac{\delta}{a}$:

$$
\begin{aligned}
\tilde{G}_- &= -1-2\frac{\int_0^{x_0}(f\ln f - f\delta) + \int_{x_0}^1 f\ln(f+\delta) - \int_{x_0}^1 f\ln(1+\delta) + \delta\int_{x_0}^1 \ln(f+\delta) - \delta\int_{x_0}^1 \ln(1+\delta)}{\int_0^1 f + (1-x)\delta} \\
&= -1-2\frac{\int_0^{x_0}(f\ln f - f\delta) + \int_{x_0}^1 f\ln f + \delta\int_{x_0}^1 \frac{f}{f} - \delta\int_{x_0}^1 f + \delta\int_{x_0}^1 \ln f + \delta^2\int_{x_0}^1 \frac{1}{f} - \delta^2(1-x)}{A+(1-x)\delta},
\end{aligned}
\tag{5.11}
$$

using $\frac{1}{a+\delta} \simeq \frac{1}{a}\left(1-\frac{\delta}{a}\right)$ and $\delta^2 = 0$:

$$
\begin{aligned}
\tilde{G}_- &= -1-2\frac{\int_0^{x_0}(f\ln f - f\delta) + \int_{x_0}^1 f\ln f + \delta\int_{x_0}^1 1 - \delta\int_{x_0}^1 f + \int_{x_0}^1 \delta\ln f}{A\left(1+\frac{(1-x)}{A}\delta\right)} \\
&= -1-2\frac{\int_0^1 f\ln f - \delta\int_0^1 f + \delta\left((1-x)+\int_{x_0}^1 \ln f\right)}{A}\left(1-\frac{(1-x)}{A}\delta\right),
\end{aligned}
\tag{5.12}
$$

referring now the non perturbed identities to obtain the final GMinus perturbed expression:

$$
\begin{aligned}
\tilde{G}_- &= -1-2\frac{S+\left(1-x-A+\int_{x_0}^1 \ln f\right)\delta}{A}\left(1-\frac{(1-x)}{A}\delta\right) \\
&= -1-2\frac{S+\left(1-x-A+\int_{x_0}^1 \ln f\right)\delta}{A} + \frac{2S}{A}\left(\frac{(1-x)}{A}\delta\right) \\
&= -1-2\frac{S}{A} + 2\frac{\left(-1+x+A-\int_{x_0}^1 \ln f\right)\delta}{A} + (-1-G_-)\left(\frac{(1-x)}{A}\delta\right) \\
&= -1+1+G_- + \frac{\delta}{A}\left(-2+2x+2A-2\int_{x_0}^1 \ln f - 1 - x + (x-1)G_-\right) \\
&= G_- + \frac{\delta}{A}\left(-3+3x+2A-2\int_{x_0}^1 \ln f + (x-1)G_-\right).
\end{aligned}
\tag{5.13}
$$

It's now possible to compute $\Delta G_- = \tilde{G}_- - G_-$, and find the critical points ($\Delta G_- = 0$) for several curves. As already observed, GMinus is mostly sensitive to high earners, but it is also sensitive to changes in the poorest side of population. The sensitivities results for GMinus are represented in Figure 5.29 for several Lorenz curves. For curve type x^a the higher the inequality, the higher are the critical points of GMinus. If a small income increment is given to a person in the second quintile of a distribution with the shape of x^2, the perturbed GMinus will decrease, meaning less inequality on the richest side.

However, if the increase is given to a person in the top 10, the inequality will increase.

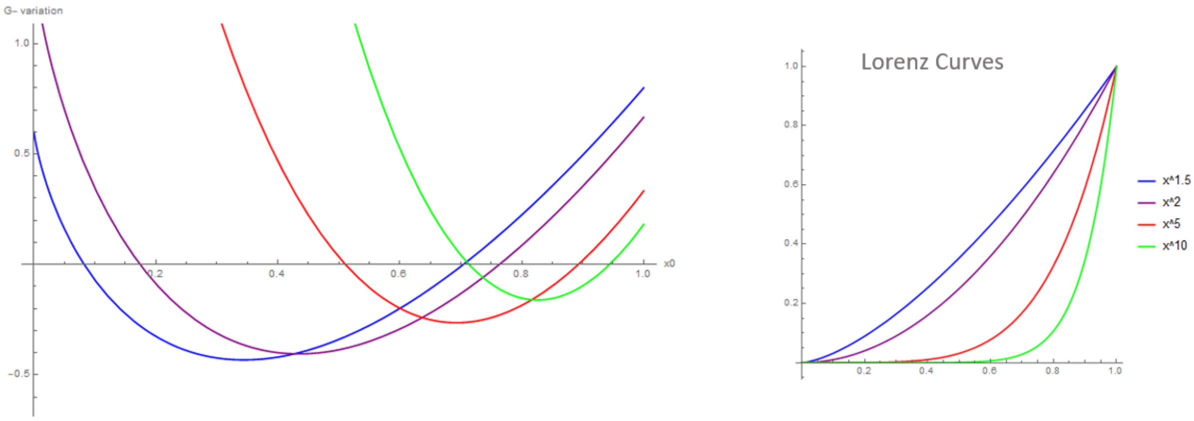

Figure 5.29: $\Delta\ G_-$ for several Lorenz curves (x^a).

Doing the same for perturbed GPlus in Equation 5.14:

$$\tilde{G}_+ = 1 + 2\frac{\int_0^{x_0} 1 - \frac{F(x)}{1+\delta}\ln(1-\frac{F(x)}{1+\delta})\mathrm{d}x + \int_{x_0}^1 1 - \frac{F(x)+\delta}{1+\delta}\ln(1-\frac{F(x)+\delta}{1+\delta})\mathrm{d}x}{\int_0^{x_0} 1 - \frac{F(x)}{1+\delta}\mathrm{d}x + \int_{x_0}^1 (1-\frac{F(x)+\delta}{1+\delta})\mathrm{d}x}, \tag{5.14}$$

simplifying and changing the notation to $F(x) = f$ and omitting the $\mathrm{d}x$ (considering the integration is always with respect to x):

$$\begin{aligned}
\tilde{G}_+ &= 1 + 2\frac{\int_0^{x_0} \frac{(1-F(x)+\delta}{1+\delta}\ln\frac{1-F(x)+\delta)}{1+\delta} + \int_{x_0}^1 \frac{(1-F(x)-\delta+\delta)}{1+\delta}\ln\frac{1-F(x)-\delta+\delta}{1+\delta}}{\int_0^{x_0} \frac{(1-F(x)+\delta)}{1+\delta} + \int_{x_0}^1 \frac{1-F(x)+\delta-\delta}{1+\delta})} \\
&= 1 + 2\frac{\int_0^{x_0}(1-f+\delta)\ln\frac{1-f+\delta}{1+\delta} + \int_{x_0}^1 (1-f)\ln\frac{1-f}{1+\delta}}{\int_0^{x_0}(1-f+\delta) + \int_{x_0}^1 1-f} \\
&= 1 + 2\frac{AUX1 + AUX2}{AUX3},
\end{aligned} \tag{5.15}$$

where AUX1 is given by Equation 5.16, AUX2 is given by Equation 5.17 and AUX3 is given by Equation 5.18, simplified using the identities described in GMinus calculations.

$$\begin{aligned}
AUX1 &= \int_0^{x_0}(1-f)\ln(1-f+\delta) - \int_0^{x_0}(1-f)\ln(1+\delta) + \int_0^{x_0}(\delta)\ln(1-f+\delta) - \int_0^{x_0}\delta\ln(1+\delta) \\
&= \int_0^{x_0}(1-f)\left[\ln(1-f) + \frac{\delta}{1-f}\right] - \int_0^{x_0}(1-f)\delta + \int_0^{x_0}\delta\left[\ln(1-f) + \frac{\delta}{1-f}\right] - \int_0^{x_0}\delta^2 \\
&= \int_0^{x_0}(1-f)\ln(1-f) + \int_0^{x_0}\delta - \int_0^{x_0}(1-f)\delta + \int_0^{x_0}\delta\ln(1-f) + 0 + 0,
\end{aligned} \tag{5.16}$$

$$\begin{aligned}
AUX2 &= \int_{x_0}^1 (1-f)\ln\frac{1-f}{1+\delta} \\
&= \int_{x_0}^1 (1-f)\ln(1-f) - \int_{x_0}^1 (1-f)\delta,
\end{aligned} \tag{5.17}$$

$$AUX3 = \int_0^{x_0} (1-f) + \int_0^{x_0} (\delta) + \int_{x_0}^{1} 1-f. \tag{5.18}$$

Getting Equation 5.15 and replacing the results from Equations 5.16, 5.17 and 5.18 we obtain:

$$\begin{aligned}
\tilde{G}_+ &= 1 + 2\frac{\int_0^1 (1-f)\ln(1-f) - \delta\int_0^1 (1-f) + (\delta x_0) + \int_0^{x_0} \delta \ln(1-f)}{\int_0^{x_0}(1-f) + \int_0^{x_0}(\delta) + \int_{x_0}^1 1-f} \\
&= 1 + 2\left[C - \delta D + \delta x_0 + \int_0^{x_0} \delta \ln(1-f)\right]\left[\frac{1}{D}(1-\frac{\delta x_0}{D})\right] \\
&= 1 + \frac{2}{D}\left[C - \delta D + \delta x_0 + \int_0^{x_0} \delta \ln(1-f)\right]\left[1 - \frac{\delta x_0}{D}\right] \\
&= 1 + \frac{2C}{D} + 2\delta\left[-\frac{Cx_0}{D^2} - 1 + \frac{x_0}{D} + \frac{1}{D}\int_0^{x_0} \ln(1-f)\right] \\
&= G_+ + 2\delta\left[-\frac{Cx_0}{D^2} - 1 + \frac{x_0}{D} + \frac{1}{D}\int_0^{x_0} \ln(1-f)\right] \\
&= G_+ + \frac{2\delta}{D}\left[-\frac{x_0 G_+}{2} + \frac{3x_0}{2} - D + \int_0^{x_0} \ln(1-f)\right].
\end{aligned} \tag{5.19}$$

It's now possible to compute $\Delta G_+ = \tilde{G}_+ - G_+$, and find the critical points ($\Delta G_+ = 0$) for several curves.

As already observed in previous sections, GPlus is mostly sensitive to low earners, as can be also observed in Figure 5.30 for several Lorenz curves. For a curve type x^a the higher the inequality, the higher the critical points of GPlus. If a small income increment is given to a person in the second quintile of a distribution with the shape of x^2, the perturbed GPlus will decrease, meaning less inequality in poorest side. However, if the increase is given to a person in the top 10, the inequality will increase.

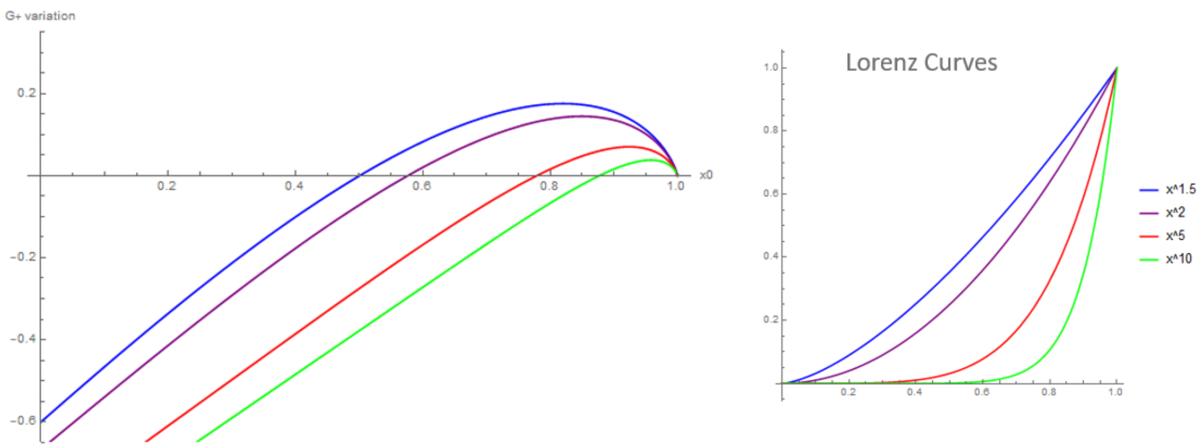

Figure 5.30: ΔG_+ for several Lorenz curves (x^a).

It's interesting to study these new coefficients' sensitivities for curves with the same Gini Index, as per the distributions represented in Table 5.2. Recalling the new coefficients' results for the 3 distributions represented in Figure 5.31, all with Gini Index = 0.2: Curve B is the one with an higher GPlus (0.303) and curve C with the lower (0.204); and Curve C is the one with a higher GMinus (0.383) and B the one with the lower (0.134). Given this, we expect to have a larger range of possibilities to increment the income and reduce bottom inequality in curve C.

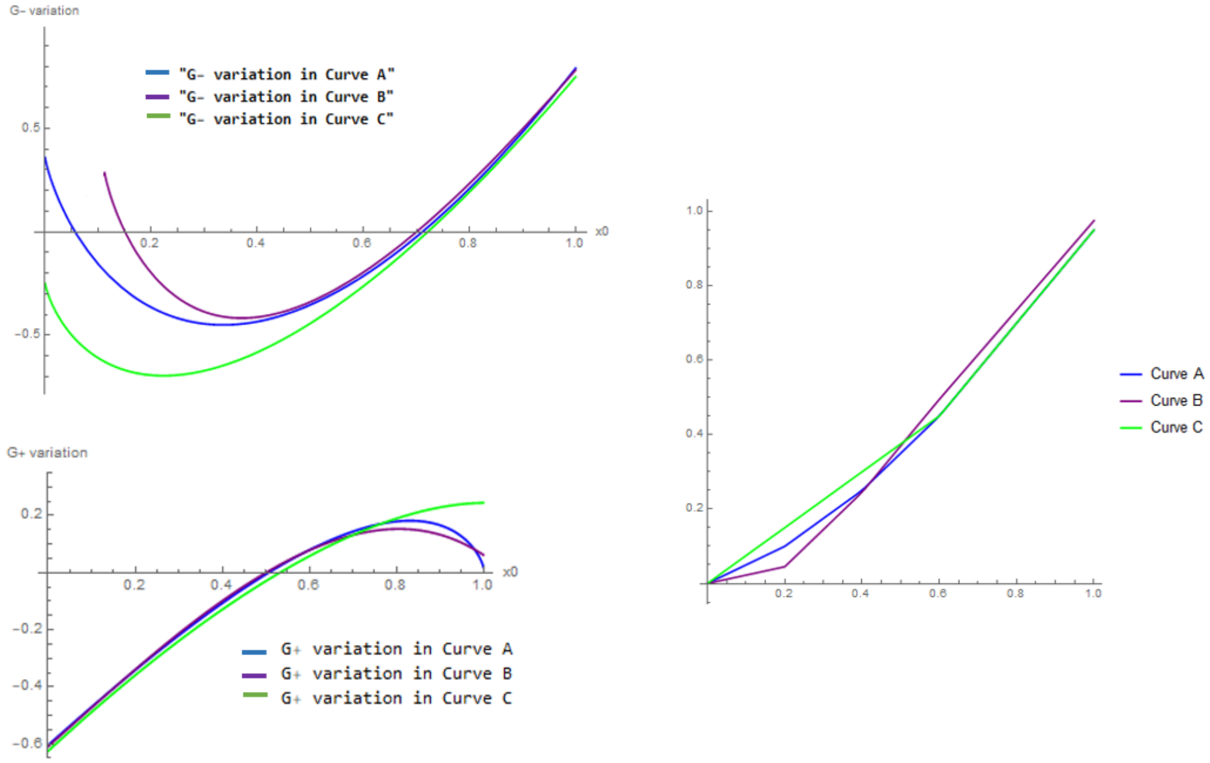

Figure 5.31: ΔG_{-} and ΔG_{+} for distributions with the same Gini Index, represented in Table 5.2.

5.10 Re-Scale

When comparing distributions with different Gini Indexes it is not easy to take meaningful conclusions from the new factors without looking at the Gini Index itself. In order to reduce this problem and improve data analysis it is proposed in this section, to normalize the results, using the Gini Index as base. In this way, it is possible to have an accordance between the distributions results, allowing a better individual comparison.

For example, the Gini Index in 2018 in Portugal is 0.3200 and in Spain is 0.3311. It is possible to say that the Gini Index variation from Portugal to Spain is 3.5%. If we do this using the raw GPlus and GMinus results, the difference will not have the same meaning, and to avoid that, there is the need of normalization.

In Tables 5.15 and 5.16 the normalized results for both coefficients are presented for several distributions analysed in the previous sections, using the new factors results re-scaled based on the Gini Index, GTop and GBottom (G_T and G_B), given in Equations 5.20 and 5.21. GTop is the re-scaled result for G- and GBottom for G+.

$$G_T = \frac{G-}{GiniIndex} \tag{5.20}$$

$$G_B = \frac{G+}{GiniIndex} \tag{5.21}$$

Table 5.15: Re-scaled new coefficients for some European Countries in 2018.

	European Countries in 2018						
	Slovenia	Belgium	Finland	Netherlands	Sweden	Denmark	Poland
GBottom	1.20	1.21	1.16	1.17	1.20	1.16	1.18
GTop	1.42	1.40	1.52	1.47	1.40	1.50	1.46
	Germany	Portugal	Greece	Spain	Italy	Lithuania	Bulgaria
GBottom	1.16	1.16	1.17	1.19	1.17	1.15	1.12
GTop	1.48	1.53	1.47	1.42	1.43	1.53	1.65

Table 5.16: Re-scaled new coefficients for OECD (average) between 1999 and 2018.

	OECD Average								
	1999	2000	2001	2004	2005	2006	2007	2008	2009
GBottom	1.168	1.171	1.190	1.172	1.175	1.172	1.174	1.171	1.174
GTop	1.545	1.536	1.487	1.534	1.521	1.530	1.526	1.535	1.525
	2010	2011	2012	2013	2014	2015	2016	2017	2018
GBottom	1.175	1.178	1.180	1.179	1.178	1.178	1.179	1.180	1.179
GTop	1.521	1.512	1.508	1.522	1.512	1.512	1.510	1.506	1.507

The results in Table 5.15 allow a comparison of the variation in GPlus and GMinus coefficients between the other European countries and Portugal. The variation percentage is presented in Figure 5.32.

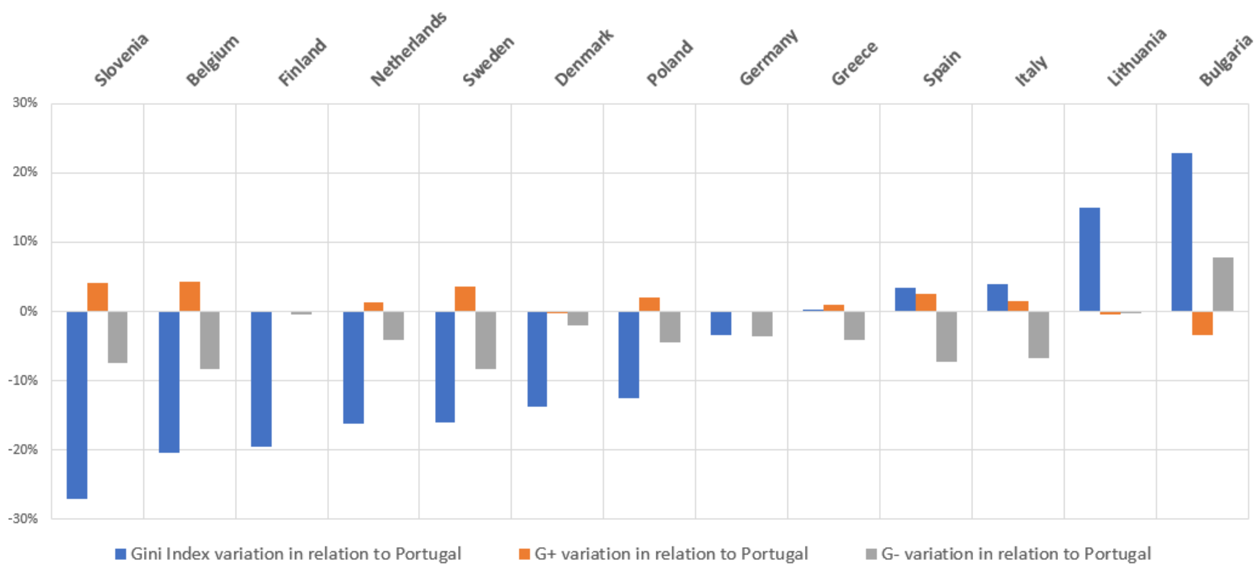

Figure 5.32: Variation percentage of Gini Index, GPlus and GMinus between European countries and Portugal in 2018, using the re-scaled factors.

The increase of 3.5% in the Gini Index from Portugal to Spain can now be seen as a balance between an increase in GPlus and a decrease in GMinus. It is also possible to verify that in all the countries except

Bulgaria, there was a decrease in GMinus when compared with the Portuguese GMinus. It means that Portugal is one of the countries with a higher inequality on the richest side of the population.

These re-scaled results also allow a better comparison over the years. In Figure 5.33 it is possible to observe the growth of inequality coefficients in each year, obtained from the re-scaled factors.

From 2004 to 2005 in OECD the Gini Index decreased, which can be associated with the decrease of GMinus. In each entry, when the Gini Index decreases, the GMinus also decrease, but, the GMinus can decreases, even if the Gini Index increases as between 2009 and 2010.

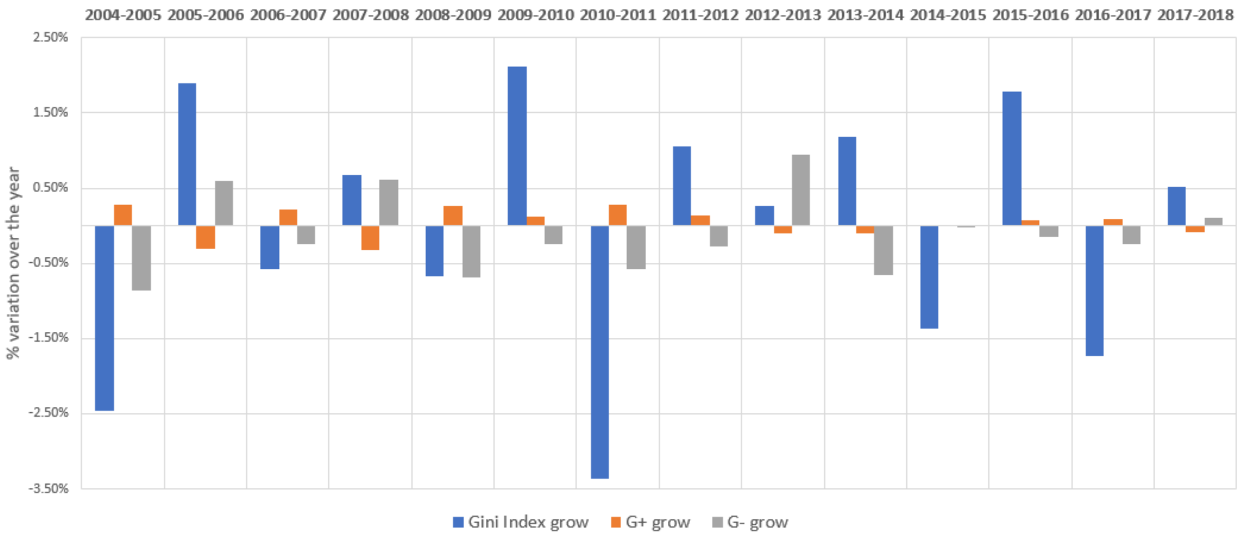

Figure 5.33: Gini Index, GPlus and GMinus grow in each year in OECD, using the re-scaled factors.

Chapter 6

Conclusions and Future Work

This work has succeeded in the creation and analysis of the new complementary coefficients for income inequality measure, based on the entropy concept.

The main aspect to note is that the new coefficients allow for a better comparison of income distributions, mainly in distributions with the same Gini Index and specifically in the distributions' tails.

It was proved that GMinus index is more sensitive to changes in higher income (top quintiles/deciles) population and GPlus in bottom quintiles/deciles. Also, they allow a better understanding of the increase or decrease of the middle class population.

Combining these with the already known Gini Index gives a significant gain of information when analysing this type of subject.

Calculating GPlus and GMinus for several real world case scenarios demonstrates the importance of these complementary coefficients to the inequality measures as they revealed promising results.

It is also important to note these coefficients can be applied to several other topics (not only income), as for example resource consumption, and can easily be interpreted, contributing to the efficiency in research and reporting.

It was observed that GPlus and GMinus are not symmetric and we believe that, although these behave quite well in the analysis performed, there is a better coefficient that can be derived from one of these, in order to have a symmetrical pair. This is of significant importance when analysing populations where both poor and richer tails have high inequality. GMinus, as proved, is more sensitive to upper tail but it is also sensitive to bottom variations, so a new coefficient that is entirely symmetrical to GPlus seems to be a better solution to this kind of tasks.

In summary, we presented new coefficients that can be used to present clear results about the distributions' tails without the use of graphical support, which was not easy to do using other inequality measures.

6.1 Future Work

Given the fact that the developed coefficients are new, there are a lot of interesting knowledge areas that can be explored. As relevant future work, here are the following suggestions:

- Complete the theoretical (mathematical) analysis of each coefficient, studying in more detail the monotony, bounds and bottom and top sensitivities.
- Explore the MLF factor and its implications and relate it with the variations of GPlus and GMinus for real cases examples, given the proper context and check for example what happen to the coefficient in cases of transfers between the population.
- Explore the new possible symmetrical coefficient. GMinus is sensitive to upper and bottom tails variations so a new coefficient that is entirely symmetrical to GPlus can had value when solving the described tasks, as it can give us only the ingormation needed to the upper tail, without having the effect of bottom tail variations.
- Authors as Demetrius [12] and Fernandez [72] developed many descriptions of the life table entropy, that can be easily compared with GPlus and GMinus and for that reason, a further look into the evolutionary entropy area with a comparison to the new coefficients will be interesting.
- Apply the coefficient to more real cases, exploring for example gender and ethnic groups.
- Apply the new coefficients in different studies such as resource consumption, effects of corruption...
- In the Human Sciences area it will be interesting to compare the coefficients with historical and political facts occurred in order to draw some conclusions about the changes in income percentage for bottom and top quintiles given by GPlus and GMinus coefficients.

Bibliography

[1] *World Social Report 2020 : inequality in a rapidly changing world / Department of Economic and Social Affairs*. United Nations, New York, 2020. ISBN 9789211303926.

[2] F. Cingano. Trends in income inequality and its impact on economic growth. (163), 2014. doi: https://doi.org/https://doi.org/10.1787/5jxrjncwxv6j-en. URL https://www.oecd-ilibrary.org/content/paper/5jxrjncwxv6j-en.

[3] T. Goda. Global trends in relative and absolute income inequality. *Ecos de Economia*, 20:46 – 69, 06 2016. ISSN 1657-4206. URL http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-42062016000100003&nrm=iso.

[4] A. Hyde, T. Vachon, and M. Wallace. Financialization, income inequality, and redistribution in 18 affluent democracies, 1981–2011. *Social Currents*, 5(2):193–211, Apr. 2017. doi: 10.1177/2329496517704874. URL https://doi.org/10.1177/2329496517704874.

[5] D. Fine. *Inequality: A persisting challenge and its implications*. McKinsey Global Institute, 2019. URL https://www.mckinsey.com/industries/public-and-social-sector/our-insights/inequality-a-persisting-challenge-and-its-implications#.

[6] A. Cole. *Income Data Is a Poor Measure of Inequality*. Tax Foundantion - Special report, 2014.

[7] F. Abergel. *Econophysics and data driven modelling of market dynamics*. Springer, Cham, 2015. ISBN 978-3-319-08473-2.

[8] T. Lux. Applications of statistical physics in finance and economics. Economics Working Papers 2007-05, Christian-Albrechts-University of Kiel, Department of Economics, 2007. URL https://ideas.repec.org/p/zbw/cauewp/5533.html.

[9] B. Bressan. *From physics to daily life : applications in biology, medicine, and healthcare*. Wiley Blackwell, Weinheim an der Bergstrasse, Germany, 2014. ISBN 978-3-527-33261-8.

[10] K. Sharma and A. Chakraborti. Physicists' approach to studying socio-economic inequalities: Can humans be modelled as atoms?, 2016.

[11] O. Penrose. Foundations of statistical mechanics. *International Series of Monographs in Natural Philosophy*, 22, 1970.

[12] L. Demetrius. Adaptive value, entropy and survivorship curves. *Nature*, 275:213–214, 1978.

[13] The world bank - world development indicators. `http://data.worldbank.org/data-catalog/world-development-indicators`. Accessed: 2020-04-01.

[14] Boltzman. Lectures on gas theory. *Berkeley: University of California Press*, 1964.

[15] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3): 379–423, July 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x. URL `https://doi.org/10.1002/j.1538-7305.1948.tb01338.x`.

[16] I. Muller. *A history of thermodynamics : the doctrine of energy and entropy*. Springer, Berlin London, 2007. ISBN 3-540-46226-0.

[17] R. Clausius. Ueber die bewegende kraft der wärme und die gesetze, welche sich daraus für die wärmelehre selbst ableiten lassen. *Annalen der Physik und Chemie*, 155(3):368–397, 1850. doi: 10.1002/andp.18501550306.

[18] S. Mishra and B. M. Ayyub. Shannon entropy for quantifying uncertainty and risk in economic disparity. *Risk Analysis*, 39(10):2160–2181, Apr. 2019. doi: 10.1111/risa.13313.

[19] D. Pele, E. Lazar, and A. Dufour. Information entropy and measures of market risk. *Entropy*, 19(5): 226, May 2017. doi: 10.3390/e19050226.

[20] Y. Tao, X. Wu, T. Zhou, W. Yan, Y. Huang, H. Yu, B. Mondal, and V. M. Yakovenko. Exponential structure of income inequality: evidence from 67 countries. *Journal of Economic Interaction and Coordination*, 14(2):345–376, Dec. 2017. doi: 10.1007/s11403-017-0211-6. URL `https://doi.org/10.1007/s11403-017-0211-6`.

[21] P. Churski, K. Kopczewska, A. Ochojski, and A. Polko. *Measuring regional specialisation. New Approach.* 09 2017. ISBN 978-3-319-51504-5. doi: 10.1007/978-3-319-51505-2.

[22] L. Vozna. The notion of entropy in economic analysis: the classical examples and new perspectives. 3:1–16, 06 2016. doi: 10.1515/jheec-2016-0001.

[23] N. Georgesçu-Roegen. *The entropy law and the economic process.* A Harvard paperback: Economics. Harvard University Press, 1971. ISBN 9780674257818. URL `https://books.google.pt/books?id=aoe7AAAAIAAJ`.

[24] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, Mar. 1951. doi: 10.1214/aoms/1177729694. URL `https://doi.org/10.1214/aoms/1177729694`.

[25] F. Balibrea. On clausius, boltzmann and shannon notions of entropy. *Journal of Modern Physics*, 07 (02):219–227, 2016. doi: 10.4236/jmp.2016.72022. URL `https://doi.org/10.4236/jmp.2016.72022`.

[26] L. Boltzmann. *Wiener Berichte*. Number 75. in WA II, 1877.

[27] L. Demetrius. Demogrpahic parameters and natural selection. *Proc. Natl. Acad. Sci. USA*, 71: 4654–4647, 1974.

[28] Collins dictionary. *HarperCollins Publishers*. https://www.collinsdictionary.com/pt/dictionary/english/inequality.

[29] Utip - university of texas - inequality project. `https://utip.lbj.utexas.edu/default.html`. Accessed: 2020-04-01.

[30] C. V. J. Income inequality. URL `https://www.britannica.com/topic/income-inequality`.

[31] World inequality database. `https://wid.world`. Accessed: 2020-04-01.

[32] L. Agnello, S. K. Mallick, and R. M. Sousa. Financial reforms and income inequality. *Economics Letters*, 116(3):583–587, 2012. ISSN 0165-1765. doi: https://doi.org/10.1016/j.econlet.2012.06.005. URL `https://www.sciencedirect.com/science/article/pii/S0165176512003278`.

[33] B. Nolan. *Changing inequalities and societal impacts in rich countries : thirty countries' experiences*. Oxford University Press, Oxford, United Kingdom, 2014. ISBN 9780199687428.

[34] A. B. Atkinson. Bringing income distribution in from the cold. *The Economic Journal*, 107(441): 297–321, Mar. 1997. doi: 10.1111/j.0013-0133.1997.159.x. URL `https://doi.org/10.1111/j.0013-0133.1997.159.x`.

[35] M. Roser and E. Ortiz-Ospina. Income inequality. *Our World in Data*, 2013. https://ourworldindata.org/income-inequality.

[36] G. A. Cohen. The pareto argument for inequality. *Social Philosophy and Policy*, 12(1):160–185, 1995. doi: 10.1017/S026505250000460X.

[37] E. Santarelli. *Entrepreneurship, Growth, and Innovation : the Dynamics of Firms and Industries*. Springer Science+Business Media, Inc, Chapter - Gibrat Law, 2006. ISBN 978-0-387-28868-0.

[38] A. Chakraborti, Y. Fujiwara, A. Ghosh, J. ichi Inoue, and S. Sinha. Physicists' approaches to a few economic problems. In *New Economic Windows*, pages 237–286. Springer International Publishing, 2015. doi: 10.1007/978-3-319-08473-2_11. URL `https://doi.org/10.1007/978-3-319-08473-2_11`.

[39] I. Trapeznikova. Measuring income inequality. *IZA World of Labor*, 2019. doi: 10.15185/izawol.462. URL `https://doi.org/10.15185/izawol.462`.

[40] L. G. Bellù and P. Liberati. Charting income inequality: The lorenz curve. 09 2005.

[41] D. Thon. An axiomatization of the gini coefficient. *Mathematical Social Sciences*, 2(2):131–143, 1982. ISSN 0165-4896. doi: https://doi.org/10.1016/0165-4896(82)90062-2. URL `https://www.sciencedirect.com/science/article/pii/0165489682900622`.

[42] R. Aaberge. Axiomatic characterization of the gini coefficient and lorenz curve orderings. *Journal of Economic Theory*, 101(1):115–132, 2001. ISSN 0022-0531. doi: https://doi.org/10.1006/jeth.2000.2749. URL https://www.sciencedirect.com/science/article/pii/S0022053100927492.

[43] T. Gajdos and J. A. Weymark. Multidimensional generalized gini indices. *Economic Theory*, 26 (3):471–496, Oct. 2005. doi: 10.1007/s00199-004-0529-x. URL https://doi.org/10.1007/s00199-004-0529-x.

[44] J. "Sanchez-Perez, L. Plata-Perez, and F. Sanchez-Sanchez. "an elementary characterization of the gini index". *"Munich Personal RePEc Archive"*, "2012".

[45] B. Jancewicz. Income inequalitie: axioms of income inequality and people perceptions. (25):21–42, June 2016. doi: 10.7206/dec.1733-0092.69. URL https://doi.org/10.7206/dec.1733-0092.69.

[46] *Income inequality metrics*. Wikipedia.

[47] A. Cobham, L. Schlogl, and A. Sumner. Inequality and the Tails: The Palma Proposition and Ratio Revisited. Technical report, United Nations, Department of Economics and Social Affairs, Sep 2015.

[48] S. P.-D. Rita Neves Costa. Not all inequality measures were created equal. *IZA World of Labor*, 31, Dec 2019.

[49] J. Fellman. Income inequality measures. *Theoretical Economics Letters*, 08(03):557–574, 2018. doi: 10.4236/tel.2018.83039. URL https://doi.org/10.4236/tel.2018.83039.

[50] J. Johnston. H. Theil. Economics and Information Theory. *The Economic Journal*, 79(315):601–602, 09 1969. ISSN 0013-0133. doi: 10.2307/2230396. URL https://doi.org/10.2307/2230396.

[51] M. O. Lorenz. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9(70):209–219, 1905. ISSN 15225437. URL http://www.jstor.org/stable/2276207.

[52] R. Aaberge. Axiomatic characterization of the Gini coefficient and Lorenz. ICER Working Papers 06-2000, ICER - International Centre for Economic Research, May 2000.

[53] A. Sen. *On Economic Inequality*. Oxford University Press, 1997.

[54] L. Ceriani and P. Verme. The origins of the gini index: extracts from variabilità e mutabilità (1912) by corrado gini. *The Journal of Economic Inequality*, 10(3):421–443, June 2011. doi: 10.1007/s10888-011-9188-x. URL https://doi.org/10.1007/s10888-011-9188-x.

[55] H. Dalton. The measurement of the inequality of incomes. *The Economic Journal*, 30(119):348, Sept. 1920. doi: 10.2307/2223525. URL https://doi.org/10.2307/2223525.

[56] G. Kelly. Gini coefficients, social network analysis, and markov chains: quantitative methods for analyzing the distribution of benefits in natural resource-dependent communities. 2012.

[57] S. Tuters. Wilkinson, r., & pickett, k. (2009).the spirit level: Why more equal societies almost always do better. london: Allen lane. *Leadership and Policy in Schools*, 11(1):129–134, Jan. 2012. doi: 10.1080/15700763.2011.577928. URL https://doi.org/10.1080/15700763.2011.577928.

[58] N. Auger, G. Zang, and M. Daniel. Community-level income inequality and mortality in québec, canada. *Public Health*, 123(6):438–443, June 2009. doi: 10.1016/j.puhe.2009.04.012. URL https://doi.org/10.1016/j.puhe.2009.04.012.

[59] R. Chakraborty. Stability and outcomes of common property institutions in forestry: Evidence from the terai region of nepal. *Ecological Economics*, 36:341–353, 02 2001. doi: 10.1016/S0921-8009(00)00237-8.

[60] R. M. Fum and R. Hodler. Natural resources and income inequality: The role of ethnic divisions. *Economics Letters*, 107(3):360–363, 2010. URL https://EconPapers.repec.org/RePEc:eee:ecolet:v:107:y:2010:i:3:p:360-363.

[61] V. Pérez-Cirera and J. Lovett. Power distribution, the external environment and common property forest governance: A local user groups model. *Ecological Economics*, 59:341–352, 09 2006. doi: 10.1016/j.ecolecon.2005.11.002.

[62] S. Gupta, H. Davoodi, and R. Alonso-Terme. Does corruption affect income inequality and poverty? *Economics of Governance*, 3:23–45, 02 2002. doi: 10.1007/s101010100039.

[63] X. jun Wang, J. yun Zhang, S. Shahid, A. ElMahdi, R. min He, X. gong Wang, and M. Ali. Gini coefficient to assess equity in domestic water supply in the yellow river. *Mitigation and Adaptation Strategies for Global Change*, 17(1):65–75, June 2011. doi: 10.1007/s11027-011-9309-7. URL https://doi.org/10.1007/s11027-011-9309-7.

[64] L. G. Bellù. Inequality analysis-the gini index. http://www.fao.org/3/am352e/am352e.pdf. Accessed: 2020-05-02.

[65] Unu wider - world income inequality database - wiid. https://www.wider.unu.edu/database/wiid. Accessed: 2020-05-01.

[66] HSRC. Income inequality and limitations of the gini index: the case of south africa. 12(5):9, Nov. 2014. URL http://www.hsrc.ac.za/en/review/hsrc-review-november-2014/limitations-of-gini-index.

[67] Oecd dataset. https://data.oecd.org/, . Accessed: 2020-04-01.

[68] J. Conway. *The Book of Numbers*. Springer New York, New York, NY, 1996. ISBN 978-1-4612-4072-3.

[69] W. R. Inc. Mathematica, Version 12.2. URL https://www.wolfram.com/mathematica. Champaign, IL, 2020.

[70] Organisation for economic co-operation and development. https://www.oecd.org/about/members-and-partners, . Accessed: 2020-10-21.

[71] R. Hoffman. Effect of the rise of a person's income on inequality. *Brazilian Review of Econometrics*, (21):237–262, 2001.

[72] O. E. Fernandez and H. Beltrán-Sánchez. The entropy of the life table: A reappraisal. *Theoretical Population Biology*, 104:26–45, Sept. 2015. doi: 10.1016/j.tpb.2015.07.001. URL https://doi.org/10.1016/j.tpb.2015.07.001.

Appendix A

Datasets

The several datasets used in this work are detailed below.

A.1 UNU-WIDER World Income Inequality Database - WIID

The **UNU-WIDER World Income Inequality Database**— widely known by its acronym **WIID** [65]— provides information on income inequality for 189 developed, developing, and transition countries (including historical entities) in an organized and user-friendly manner. The version used in this work is from May 2020.

From this database the following were used: the Gini coefficients, income deciles and quintiles, country names and years. We've only considered the countries/years with the same parameters of information, ie, Income Net or gross, measured by household and with an equivalent scale. A screenshot of this database can be found in Figure A.1.

B	E	F	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	AB	AD	AG
country	**year**	**gir**	**q1**	**q2**	**q3**	**q4**	**q5**	**d1**	**d2**	**d3**	**d4**	**d5**	**d6**	**d7**	**d8**	**d9**	**d10**	**resource**	**scale**	**reference_**
Slovenia	1997	22.9	9.95	15.18	18.94	23.03	32.89	3.96	5.99	7.09	8.09	9	9.94	10.85	12.18	13.74	19.15	Income (net)	Equivalized	Person
Norway	2011	22.9	10.1	15.5	18.7	22.5	33.2	3.8	6.3	7.4	8.2	8.9	9.7	10.7	11.8	13.5	19.6	Income (net)	Equivalized	Person
Sweden	1992	22.9	9.77	15.42	19.01	23.03	32.76	3.69	6.08	7.25	8.17	9.09	9.92	10.93	12.1	13.81	18.95	Income (net)	Equivalized	Person
Czechia	1994	23	11.34	14.72	18.12	22.48	33.34	5.05	6.29	7.02	7.7	8.61	9.51	10.61	11.87	13.94	19.4	Income (net)	No adjustment	Household
Czechia	1996	23	10.99	14.63	17.84	22.62	33.92	4.9	6.08	6.9	7.73	8.43	9.41	10.52	12.1	14.18	19.74	Income (net)	No adjustment	Household
Taiwan	2010	23	10.95	14.87	18.01	22.03	34.15	4.8	6.15	7.05	7.82	8.57	9.44	10.38	11.65	13.64	20.51	Consumption	Per capita	Person
Denmark	1995	23	10.67	14.95	18.23	22.51	33.65	4.48	6.19	7.07	7.88	8.68	9.55	10.58	11.93	13.82	19.83	Income (net)	Per capita	Person
Austria	1987	23	9.73	15.45	18.39	21.94	34.5	4.38	5.35	7.57	7.88	8.78	9.61	10.92	11.02	15.16	19.34	Income (net)	Per capita	Person
Denmark	2004	23	10.12	14.94	18.98	23.06	32.88	4.25	5.87	6.93	8.01	9.02	9.96	10.94	12.12	13.72	19.16	Income (net)	Equivalized	Person
Slovenia	2007	23	9.79	15.23	18.94	23.2	32.83	3.92	5.87	7.12	8.11	9.01	9.93	10.97	12.23	14.03	18.8	Income (net)	Equivalized	Person
Sweden	2004	23	9.9	15.3	18.8	23	32.9	3.8	6.1	7.2	8.1	9	9.9	10.9	12.2	13.9	19	Income (net)	Equivalized	Person
Czechia	1992	23	10.66	14.86	18.36	22.47	33.66	4.55	6.11	7.01	7.85	8.72	9.64	10.61	11.86	13.64	20.02	Income (gross)	Equivalized	Person

Figure A.1: Part of WIID Database.

A.2 The World Inequality Database - WID

Presented for the first time in January 2011 by T. Atkinson and T. Piketty, the World Inequality Database offers one of the most extensive available databases on the historical evolution of the world distribution of income and wealth, both within countries and between countries [31].

Appendix B

OECD Data details

In order to obtain the new coefficients results for OECD overall, the quintile data from WIID database for each country and year was used to calculate the average OECD quintile data in line with Figure B.1, where Y represents the use of each data entry in the data used.

From the 37 countries from OCDE it was only possible to find income percentage by quintile/decile with the same scale for 34 countries. For that reason, the following countries were not considered in this work: Czech Republic, Korea and New Zealand. Also, 2002 and 2003 have less than 20 countries available and for that reason the new coefficients were not calculated for these years.

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Australia	N	N	Y	N	Y	Y	N	N	N	Y	N	Y	N	N	N	Y	N	N	N	N
Austria	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Belgium	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Canada	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Chile	Y	Y	N	N	Y	N	N	Y	N	N	Y	N	Y	N	Y	N	Y	N	Y	N
Colombia	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Denmark	N	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Estonia	N	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Finland	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
France	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Germany	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Greece	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Hungary	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Iceland	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N
Ireland	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Israel	N	N	Y	N	N	N	Y	N	Y	N	N	Y	N	Y	N	Y	N	Y	N	N
Italy	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Japan	N	N	N	N	N	Y	N	N	N	Y	Y	N	N	N	N	N	N	N	N	N
Latvia	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Lithuania	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Luxembourg	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Mexico	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Netherlands	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Norway	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Poland	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Portugal	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Slovakia	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	Y	N	N	N	N
Slovenia	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Spain	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sweden	N	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Switzerland	N	Y	N	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Turkey	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
United Kingdom	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
United States	N	Y	N	N	N	Y	N	N	Y	N	N	Y	N	N	Y	N	N	Y	N	N
Grand Total	**22**	**27**	**21**	**12**	**17**	**27**	**26**	**27**	**30**	**30**	**29**	**31**	**28**	**29**	**29**	**31**	**28**	**30**	**27**	**21**

Figure B.1: Summary of Country/year in WIID database used for OECD calculations.

Alphabetical Index