SCHOOL OF ECONOMICS AND MANAGEMENT

Bachelor Programme in Economy and Society

# Economic Complexity and Income Inequalities Across Countries and Regions

A quantitative study comparing national and regional income inequalities in association with economic complexity across OECD countries and Spanish provinces

by

Johanna Marie Fredrich

jo1628fr-s@student.lu.se

Abstract

This thesis provides a comprehensive empirical study of the association between economic complexity and income inequality from a multilevel macro perspective. It explores the relationship between economic complexity and income inequality across the countries and regions of the OECD, with a particular focus on regional income inequality in Spain. This relationship is assessed using regression analysis throughout three parts focusing on different income inequality measurements as the dependent variables and the national and regional Economic Complexity Index as the explanatory variables. Part I focuses on the impact of economic complexity on national income inequality. The study finds that economic complexity negatively correlates with national income inequality across OECD countries. Part II looks at interregional inequalities, evaluating the impact economic complexity has on the level of regional inequalities across the OECD countries. The study cannot find a significant relationship between the variables. However, Part III, which focuses more specifically on the association between regional inequalities and economic complexity across Spanish regions, shows that regions with higher Economic Complexity Index levels also have higher income levels. Therefore, divergence over time because of differing economic complexity levels cannot be ruled out. Additionally, Part III assesses the effects of economic complexity on within-regional inequalities across the Spanish provinces. It finds that economic complexity is correlated with lower income inequality levels across the Spanish provinces, especially in predominantly urban and urban-rural mixed regions. The study also shows that the Economic Complexity Index better predicts income equality than GDP.

EOSK12

Bachelor's Thesis (15 credits ECTS)

June 2023

Supervisor: Finn Hedefalk

Examinator: Erik Bengtsson

Word count: 12 539

Source Cover page images: OEC (2023)

Table of Contents

**List of Figures** ........ 3

**List of Tables** ........ 4

**List of Abbreviations** ........ 4

**1. Introduction** ........ 5

1.1 Aim and Research Question ........ 6

1.2 Structure of the Thesis ........ 7

**2. Background** ........ 7

2.1 Theoretical Framework ........ 7

2.1.1 Economic Complexity Theory ........ 8

2.1.2 The ECI ........ 12

2.1.3 The Regional ECI ........ 14

2.2 Literature Review ........ 15

2.2.1 Causes and Types of Income Inequality ........ 15

2.2.2 Economic Complexity and Income Inequality ........ 17

2.2.3 Economic Complexity and Regional Income Inequality ........ 19

**3. Data** ........ 21

3.1 Source Data ........ 21

3.2 Sample ........ 22

3.2 Variables ........ 23

3.2.1 Variables Part I and II ........ 24

3.2.2 Variables Part III ........ 29

**4. Method** ........ 31

4.1 Part I: Economic Complexity and Inequality Across Individuals on the National Level ... 31

4.2 Part II: Economic Complexity and Inequality Across Regions at the National Level ........ 32

4.3 Part III: Economic Complexity and Regional Inequality in Spain .... 33
4.4 Econometric Robustness Tests .... 34
**5. Analysis .... 35**
5.1 Descriptive results .... 35
5.2 Regression Analysis Results .... 37
5.2.1 Part I: Economic Complexity and Inequality Across Individuals on the National Level .... 37
5.2.2 Part II: Economic Complexity and Inequality Across Regions at the National Level .... 39
5.2.3 Part III: Economic Complexity and Regional Inequality in Spain .... 41
**6. Discussion .... 46**
6.1 Limitations of the Study .... 49
**7. Conclusion and Policy Implications .... 50**
**8. References .... 53**
**9. Appendix .... 61**
9.1 Multicollinearity Test Results .... 61
9.2 Breusch-Pagan Test for Heteroskedasticity Results .... 62

# List of Figures

Figure 1: Spain's Product Space with specialization RCA ≥ 1 (2021) .... 10
Figure 2: Spain's Relatedness Space (2021), source: adapted from OEC (2023) .... 12
Figure 3: Methodological structure of the study .... 21
Figure 4: The Lorenz curve, source: adapted from Stiglitz et al. (2018) .... 24
Figure 5: ECI for the countries of the world (left) and the provinces of Spain (right), source OEC (2023) .... 26
Figure 6: Results linear regression analysis with trend line, dependent variable: Gini coefficient, independent variable: ECI 2019 .... 38

Figure 7: Results linear regression analysis, dependent variable: Theil Entropy Index, independent variable: ECI 2019 ........ 40
Figure 8: Results linear regression analysis with trend line, dependent variable: GDP per capita, independent variable: ECI 2020 ........ 42
Figure 9: Distribution of ECI 2020 values by geographic typology ........ 45
Figure 10: Results linear regression for Spanish provinces including all observations (left) and excluding outlier (right), dependent variable: Gini coefficient, independent variable: ECI ........ 45

## List of Tables

Table 1: Selected sample of OECD countries in alphabetical order ........ 22
Table 2: Sample of Spanish provinces in alphabetical order ........ 23
Table 3: Descriptive statistics for the variables of Part I and Part II ........ 36
Table 4: Descriptive statistics for Part III ........ 36
Table 5: Four-step regression analysis, dependent variable: Gini coefficient ........ 38
Table 6: Four-step regression analysis, dependent variable: Theil Entropy Index ........ 40
Table 7: Two-step regression analysis, dependent variable: GDP per capita ........ 43
Table 8: Four-step regression analysis, dependent variable. Gini coefficient ........ 46

## List of Abbreviations

**ECI** – Economic Complexity Index

**Gini** – Gini Coefficient

**GDP** – Gross Domestic Product

**GNI** – Gross National Income

**Geotypology** – Geographic Typology

# 1. Introduction

The question of why some countries or regions are rich while others are relatively poor might be one of the oldest debates in the social sciences. Adam Smith (1776) famously traced the prospering of nations back to the division of labour, arguing that countries grow when knowledge and skills are divided productively in the economy. Over time, scholars have come forward with many theories to explain the inequalities between individuals and locations. Kuznets (1955) related inequality levels between countries to aggregated income levels, a theory that has been heavily disputed ever since. It has been proven that high levels of income inequality can harm economic development and societal peace (Piketty, 2014; Stiglitz, 2016). Therefore, it is vital to constantly search for the underlying structural reasons for income inequality and keep it low.

Less discussed among scholars over time but also highly important is interregional income inequality, the disparity of incomes across regions, and intraregional income inequality, which is the inequality between individuals within a region. Sometimes income differences across regions can be higher than across countries, and these inequalities harm a country's economic performance and social peace. As Fish (1984) points out, intraregional inequality often differs from the national income distribution; therefore, even if income inequality converges nationally, intraregional dynamics can behave differently. Consequently, it is essential to look at national, interregional, and intraregional income inequality to cover the whole spectrum of macroeconomic dynamics of income inequality.

A recent concept that quantifies and predicts the prospering of nations and regions is *Economic Complexity*. Hidalgo & Hausmann (2009) developed the *Economic Complexity Index (ECI)*, which ranks the productive capabilities of economies. With some abstraction, it is essentially a modern reinterpretation of Smith's division of labour theorem. It measures the amount and diversity of collective knowledge a society holds and how it productively combines it to create more complex and interconnected production networks in the economy (Hausmann et al., 2013).

While high economic complexity seems to impact economic growth and development positively (Hidalgo & Hausmann, 2009; Le, Niem & Kim, 2022), its effects on income inequality have only been empirically researched in recent years, with the evidence remaining scarce and ambiguous.

Most studies focus on the relationship between economic complexity and income inequality across countries, while few studies have examined the effects on regional inequalities. The main contribution of this thesis is to explore the association of economic complexity with both national and regional income inequalities, allowing for a holistic study that compares the dynamics on different levels.

## 1.1 Aim and Research Question

This study explores the relationship between income inequality and economic complexity on multiple levels, from the national to the regional and from across to within countries. The study is geographically brought to examine the effects of economic complexity on a macro-level and to draw general conclusions for further, more specific research. However, there is a specific focus on two main variables, economic complexity as the independent variable and a measure of inequality as the dependent variable. Thus, while the geographical entities are vast, the focus on the variables is narrow.

The main contribution of this thesis is the analysis of the association of economic complexity and regional inequality on these multiple levels. Empirical research on the effect of economic complexity on inequality across and within regions is scarce. Most studies focus on a specific country or region or many different explanatory variables. This study, therefore, contributes with its narrow focus on economic complexity and income inequality while considering a large sample of OECD countries and a more detailed analysis of Spanish regions. This allows for a comprehensive picture of the general effects of economic complexity on national and regional economic inequality. It should also be noted that most research on the topic is recent, primarily published in the last two years, which makes this study a contribution to ongoing research in the field. The research question is the following:

*Research Question:*
*What is the association between economic complexity and income inequalities across countries and regions of the OECD?*

*Sub-research questions:*

*What is the association between economic complexity and interregional and intraregional inequality in Spain?*

To answer these questions, the thesis will use linear regression analysis looking at countries and regions of the OECD to explore the association between the national Economic Complexity Index (ECI) and inequalities across countries and regions. Additionally, the sub-research question is answered by looking at the regional ECI and regional income inequalities across the Provinces of Spain.

## 1.2 Structure of the Thesis

The thesis's structure is as follows: Section 1 explains the aim and research question of the thesis. Section 2 presents the theoretical concept for the thesis, economic complexity theory and the Economic Complexity Index (ECI). These are explained conceptually from a historical perspective, in relation to related concepts as well as mathematically. Then, the literature on the relationship between economic complexity and income inequality is reviewed in general and specifically with a perspective on regional inequality. In section 3 the Data of the thesis is explained, including a thorough description of the data and variables used. Subsequently, section 4 goes through the method of each Part separately. Section 5 analysis the results individually for each part, followed by section 6 which brings the results together and compares them to one another and the existing literature. Lastly, Section 7 summarizes the findings and gives policy suggestions based on the results.

# 2. Background

## 2.1 Theoretical Framework

The main theoretical framework for this thesis is Economic Complexity Theory, mainly developed by Hidalgo & Hausmann (2009). This section gives an overview of the emergence of the concept and provides a brief technical description of how it can be understood and applied. Furthermore, the Economic Complexity Index (ECI) is explained intuitively and mathematically, as it is the main explanatory variable in this study. Therefore, a thorough understanding of the general concept and how this Index measures economic complexity is essential.

### 2.1.1 Economic Complexity Theory

Economic complexity is a relatively recent concept that has only developed over the past decade. New methods and fine-grained data became available, enabling the analysis and quantification of complex economic structures (Hidalgo, 2021). It is a framework often used for regional and national development discourses based on international trade data and the industrial structure of a region or country. According to Hausman et al. (2013), economic complexity measures the diversity and connectivity of knowledge expressed in a country's productive output and reflects the structures that combine and hold knowledge.

Historically, the concept has evolved from a general methodological shift in the field of economics, or what Beinhocker (2006, p. 12) calls the "complexity economics revolution". According to Beinhocker (2006, p. 80), a difference between traditional economics and complexity economics is that conventional models usually assume the existence of an economy and analyse how it evolves in terms of growth or wealth distribution. The question of why certain economic sectors evolve in the first place is rarely debated. Here, theories of the complex economy take an inherently different approach in asking how the economy came to be in a location (Beinhocker, 2006, p. 80). Beinhocker (2006) argues that the discipline of economics is changing fundamentally with the introduction of complexity economics. He uses it as an umbrella term for many new economic streams that use new mathematical and statistical approaches to mostly traditional economic problems, often borrowed from the natural sciences (Beinhocker, 2006, p. 96). He identifies "five big ideas" that he sees as part of the 'revolution' economics. These ideas, which are inherently different from traditional economics, are:

- Dynamics: dynamic, open and nonlinear system
- Agents: individual modelling that uses incomplete information to create ideas which are adapted and improved over time
- Networks: interactions and networks between individual agents are modelled, which can change over time

- Emergence: Micro and macroeconomics are not strictly distinguished
- Evolution: Endogenous factors can create novelty and growth through an evolutionary process of selection, differentiation and amplification

Economic complexity theory has established itself as one of the most popular frameworks which uses many of the ideas described by Beinhocker. According to Hidalgo (2021) it is essentially based on three developments in the research. Firstly, the comeback of industrial policy, secondly, the increased use of AI and thirdly, the maturing of endogenous growth theory. As presented by Romer (1990), endogenous growth theory argues that economic growth is mainly fostered by technological change, driven by endogenous forces such as investments in human capital and innovation.

While economic complexity is based on the classical economic assumption of wealth accumulation and economic development through the division of labour and human capital accumulation, the methodology behind the concept differs from traditional economics. Hidalgo (2021) states that rather than aggregating the output (as in, i.e. GDP) under the assumption of existing input of capital and labour, economic complexity applies techniques from machine learning similar to matrix factorisation to trade data, industry employment, and patents. Using dimensional reduction techniques on fine-grained data allows for quantifying abstract production factors and how these translate into economic output, measured in exports. This methodology provides a comprehensive understanding of the geography of economic output and can predict the future development and diversification potentials of locations (Hidalgo, 2021).

Economic complexity measures an economy's productive structure, which reflects the amount and type of knowledge prevalent in the economy (Utkovski et al., 2018, p. 2). The export structure of a location is measured with data on the network linking countries and products according to their revealed comparative advantage (RCA) of the products (Utkovski et al., 2018, p. 2). The RCA was developed by Balassa (1964) and is based on the country's or region's relative share of a product in all world exports. Mathematically the RCA is expressed like this:

$$RCA_{cp} = \frac{X_{cp}}{\sum_c X_{cp}} / \frac{\sum_p X_{cp}}{\sum_{c,p} X_{cp}}$$

where $X_{cp}$ are the exports of the product $p$ by country $c$. So the $RCA_{cp}$ is the revealed comparative advantage that the country $c$ has in the product $p$ (Hausmann et al. 2013). The RCA is then used to create a matrix where 1 represents that a country has a comparative advantage larger than one in the product that it exports and 0 otherwise. The matrix shows the connection between the country and the products it makes, and it is used to measure economic complexity and the product space of a country (Hausmann et al., 2013).

$$M_{cp} = \begin{cases} 1\, if\ RCA \ \geq 1; \\ 0\, otherwise \end{cases}$$

Figure 1 shows an example of the product space for Spain for the year 2021. The different colours of the nodes depict the different product categories, and the size of the nodes represents the relative export value of the good. All coloured nodes represent products with an RCA above 1, while the grey nodes indicate that the products have no comparative advantage (RCA < 1).

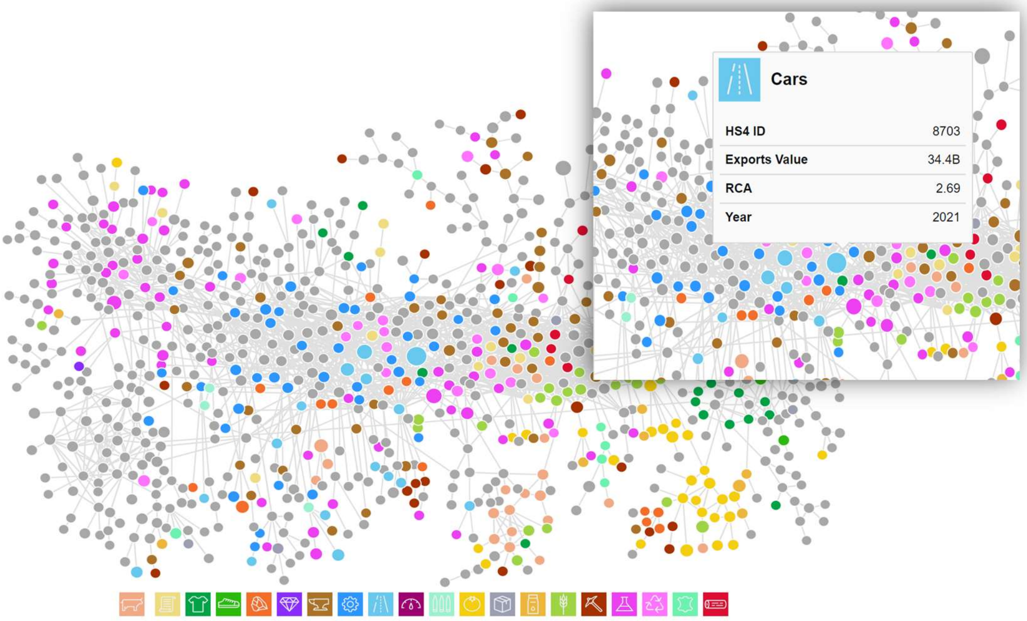


*Figure 1: Spain's Product Space with specialization RCA ≥ 1 (2021), source: adapted from OEC (2023)*

Another important concept for the theory of economic complexity is relatedness. It is calculated through a matrix which measures to what degree an activity is linked to a specific location, showing essentially how compatible an activity and a location are (Hidalgo, 2021, pp. 6-7). Relatedness, thus, measures the activities and knowledge in a specific location. It can predict whether a location is likely to enter a new activity, depending on the existence of similar activities with related knowledge exist in the location (Hidalgo, 2021, pp.6-7). Unlike traditional measurements of competitiveness, economic complexity looks at the diversity and number of products in a country's export basket as well as the capabilities necessary to produce the good (Sciarra et al., 2020).

Mathematically relatedness can be calculated in two different ways, both essentially define relatedness as a fraction of related activities that exist in a location (Hidalgo, 2021, p.7).

$$\omega_{cp} = \frac{\Sigma_{p'} M_{cp'} \phi_{pp'}}{\Sigma_{p'} \phi_{pp'}} \quad (1)$$

$$\omega_{cp} = \frac{\Sigma_{c'} M_{c'p} \phi_{cc'}}{\Sigma_{c'} \phi_{cc'}} \quad (2)$$

Formula (1) describes the relatedness $\omega$ between the location $c$ and the activity p, where $M_{cp'}$ is a matrix that indicates the existence of activity $p$ in a location $c$. $\phi_{pp'}$ is a measure of proximity between the activities $p$ and $p$', also known as the "metrics of proximity" (Hidalgo, 2021, p.7). Formula (2) is very similar, only that the proximity metrics measures the distance between locations $c$ and $c$' instead of the products (Hidalgo, 2021, p.7).

Figure 2 shows the relatedness space for Spain. The nodes represent the same products as in Figure 1, only that the colours depict the degree of relatedness in this illustration. Closely related products tend to be co-exported, and the higher the relatedness value, the more knowledge exists in the economy for this production. Thus the higher the probability that, in this case, Spain will export these products in the future. The main contribution of economic complexity to existing concepts is that it does not only analyse the economic structure of the past and present. Hausmann et. al. (2013) claim that relatedness provides a reliable prediction of economic growth and development trajectories for regions and countries.

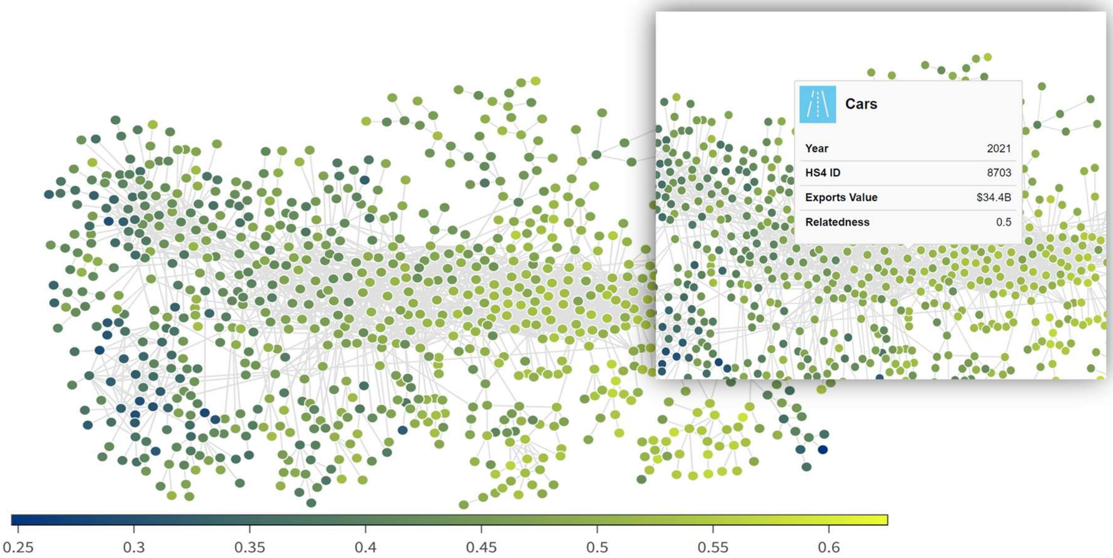


*Figure 2: Spain's Relatedness Space (2021), source: adapted from OEC (2023)*

Economic complexity differs from other economic development measures, as it is non-income-based. This means that it does not only measure the economy's current output (such as GDP) but also a location's potential for growth and development based on the existing structure of the economy. More traditional methods often do not capture these structures (Utkovski et al., 2018).

### 2.1.2 The ECI

Finally, the heart of Economic Complexity Theory is the Economic Complexity Index (ECI) which is used as the primary independent variable in this thesis. Hausman et al. (2013) created the ECI to quantify economic complexity in a country or region relative to other locations. The calculations of the ECI are based on international trade data on exports of goods which is used because of the detailed information on linkages it provides between countries and products.

According to Hidalgo (2021), the mathematical equation of the ECI is derived from two central assumptions of the economic complexity theory:

$$\text{i. } K_c = f(K_p, M_{cp}),$$

where, a locations complexity $K_c$ is a function $(f)$ of the complexity of the activities $K_p$ and the activities presented in the location $M_{cp}$, and

$$\text{ii. } K_p = g(K_c, M_{cp}),$$

where, an activity's complexity $K_p$ is the function $(g)$ of the complexity of the location $K_c$ and the places in which the activity exists $M_{cp}$.

The ECI is calculated by reducing these two assumptions to a linear equation form:

$$K_c = \overline{M_{cc'}} K_{c'} \quad (1)$$

$$K_p = \overline{M_{pp'}} K_{p'} \quad (2)$$

The equation is self-consistent and a relative measure, as a location's or activity's complexity can change depending on the entries or exits of other activities or locations. $\overline{M_{cc}}$ and $\overline{M_{pp'}}$ are eigenvectors as $K_c = K_c = 1$ and $K_p = K_p =1$, which means that $\overline{M_{cc'}} = \overline{M_{pp'}} = 1$. Since this is the case, the measurement of economic complexity is essentially the variation in the system (Hausmann et al., 2013, p.24).

The complexity of a location $K_c$ is the average complexity of the activities present in this location, and the complexity of an activity $K_p$ is the average complexity of the location where the activity exists. This looks mathematically as shown in formulas (3) and (4):

$$K_c = \frac{1}{M_c} \Sigma_p M_{cp} K_p \quad (3)$$

$$K_p = \frac{1}{M_p} \Sigma_c M_{cp} K_c \quad (4)$$

where, $K_c$ is the complexity of location $c$ (i.e. a country) and $K_p$ is the complexity of an activity $p$ (product or industry). As mentioned earlier $M_{cp}$ is the matrix representing the connection between location ($c$) and activity ( $p$), which is 1 if the RCA ≥ 1, and 0 otherwise.

With some algebraic transformation and by using a Z-transformation to normalise the results for the relative metric, the equation for the ECI is the following:

$$ECI = \frac{K_c - \overline{K_c}}{\sigma(K_c)} \quad (5)$$

where $\underline{K_c}$ denotes the average of $K_c$ and $\sigma(K_c)$ stands for the standard deviation of $K_c$. The standard deviation of $K_c$ is the measure of economic complexity. (The detailed calculations of the ECI can be revised in the paper by Hausmann et al. (2023)).

### 2.1.3 The Regional ECI

The regional ECI, which is the economic complexity index for a subnational entity such as provinces or municipalities, is calculated in a similar manner but with different observation units and is defined as:

$$ECI_c = \frac{1}{M_p} \Sigma_p M_{cp} PCI_p \quad (6)$$

Whereby $M_{cp}$ is defined by using a version of the RCA that is modified to a regional version:

$$R_c p = \frac{X_{cp}^{local} / X_c^{local}}{X_p^{world} / X^{world}}$$

where, the share of an activity locally ($X_{cp}^{local}/X_{c}^{local}$) is compared to the share of an activity in the world ($X_{p}^{world}/X^{world}$).

There are several limitations to the ECI. Hausmann et al. (2013, p.22) acknowledge that a weakness is that the ECI is only based on export data and not production, so goods mainly used in the domestic market and not exported are not included. However, based on the Ricardian trade model, it can be assumed that if a country does not export a good, it is likely to be relatively less efficient in producing the good than other countries (Ricardo, 1817). Another limitation of trade data is that countries may export goods they did not produce but only re-export. Hausmann et al. (2013, p.22) try to circumvent this problem by requiring countries to export a certain share of associated products to the product in question. It should also be mentioned that it is a limitation that only tradable goods are included and not services. Still, to measure whether the productive structure of an economy is associated with income inequalities, economic complexity theory and the ECI are the best possible tools available and are therefore used in this study. Especially for comparing national and regional inequalities in relation to economic complexity, the ECI is a reliable measurement as national and regional data is available.

## 2.2 Literature Review

The literature review will first briefly discuss previous research on national, interregional and intraregional income inequality. Subsequently, there is a thorough review of the research on the association of economic complexity and income inequality, followed by a specific focus on regional income inequality.

### 2.2.1 Causes and Types of Income Inequality

This thesis is concerned with three different types of income inequality; national income inequality across individuals in a country, interregional income inequality as measured as a relative level of regional disparities for a country, and intraregional inequality as measured as the income distribution across individuals in a region. Most literature can be found on national income inequality, while the other types are less commonly discussed.

Scholars have long been concerned with finding the sources of income inequality across countries. Kuznets (1955) famously claimed that income inequality was closely connected to the aggregate income levels of a country, establishing the renowned Kuznets curve. He argued that the relationship between income levels and income inequality follows an inverted U-shaped curve, indicating that when countries become richer, they will first become increasingly unequal and then progressively more equal. This observation has been heavily disputed over the decades, and many scholars, such as Piketty (2014), argue that income inequalities have been on the rise in most high-income countries since the mid-20th century. Scholars such as Acemoglu & Robinson (2012) have developed the idea that the level of inequality in a society depends more on the type of growth and the institutions and economic structures present than the state of development. Most scholars today agree that income inequality depends on more factors than only GDP.

A high level of income inequality in a country has been proven to affect economic growth negatively (Stiglitz, 2016, p.134). Scholars such as Stiglitz (2016), Piketty (2014) and Arestis & Gonzalez-Martinez, 2016) argue that inequalities between individuals have grown substantially in Western countries, exhibiting a negative impact on economic performance and social stability. Arestis & Gonzalez-Martinez (2016) find that the OECD countries show especially high increases in income polarization. This makes the OECD an especially interesting sample to study for this thesis.

Regional inequalities, especially interregional disparities, are more challenging to measure, and their importance is often neglected. Scholars such as Barro and Sala-i-Martin (1991) have argued that the ease of migration across regional borders within countries will automatically balance out inequalities across regions in highly developed countries. One of the most influencing hypotheses on the dynamics of interregional inequalities was developed by Williamson (1965), who argued that regional disparities follow the same inverted U-shaped curve as described by Kuznets (1955). He found that when a country becomes richer, its regions first diverge, followed by convergence in the later stages of development. This theory has been disputed, and Artelaris & Petrakos (2014) state that there has been highly contradictory evidence on the trajectory and causes of regional disparities over the last decades. Gbohoui, Lam & Lledo (2019) argue that disparities across OECD regions are still high and often more severe than differences in incomes across countries. They find that rather than only the income level, more multifaced causes such as the concentration of human and physical capital, migration, differences in economic development and population variations contribute to interregional inequality.

It is assumed that intraregional income inequalities are similar to national inequalities, as they are measured similarly only on a smaller scale (Artelaris & Petrakos, 2014). However, as early as 1984, Fish (1984) pointed out that intraregional inequalities often differ from the national income distribution and even if income inequality convergence nationally, intraregional dynamics can be different. The same relationship can be observed for interregional and intraregional convergence dynamics. According to Artelaris & Petrakos (2014), policies that focus on interregional convergence can often have the opposite effect on intraregional inequalities, implying a seemingly inevitable trade-off. As all types of inequalities have different causes and dynamics, this study will take a holistic approach to the topic and examine and compare the effects on all three types of income inequalities.

### 2.2.2 Economic Complexity and Income Inequality

Since economic complexity is a relatively new concept, empirical research on its relationship with income inequality is relatively scarce in the literature. However, long before the concept of economic complexity was defined, scholars have been concerned with the effect of a country's productive structure on its development and income distribution. Rosenstein-Rodan (1943), Prebisch (1950), and Singer (1950) argued that the industrial structure of an economy is directly connected to its ability to generate income and wealth. During their time, the most significant structural shift was the transformation of the agricultural to the industrialised society. They argued that the emergence of more complex economic structures and production is the most critical step for a country towards development and wealth.

Innovation is fundamental for transforming the productive structures of an economy and creating more complex economies (Balland et al., 2022, pp.8-9). Schumpeter (1942) argues that the key drivers of economic growth are structural transformation and innovation. His ideas on innovation are based on the theory of creative destruction, arguing that constant innovation is necessary for the capitalistic system to work efficiently. He claimed that innovation is driven by the diffusion of knowledge through social networks. This idea is closely connected to Hidalgo and Hausmann's (2009) theory of the productive network structures that define highly complex economies. Scholars have pointed out that innovation and increased complexity of the economy positively impact economic development (i.e. Hidalgo and Hausmann, 2009; Schumpeter, 1942). However, some scholars have pointed out the possible adverse effects of innovation and complex economies on equality (Pinheiro et al. 2022)

Most studies on the relationship between economic complexity and inequality are from the last few years, as the concept has only recently been established. Hartman et al. (2017) first explored the relationship between economic complexity and income inequality. They used the ECI to empirically measure the effect of economic complexity on income distributions across a large sample of countries. The result suggest that a higher level of economic complexity is associated with lower levels of income inequality. Hartman et al. (2017) state that "the mix of products that a country exports are a significant predictor of income inequality in both cross-sectional and panel regressions" (p. 81). Most scholars across the literature agree that there is a positive relationship between economic complexity and income equality (i.e. Hartman et al., 2017; Lee & Wang, 2021; Hartmann & Pinheiro, 2022). However, Hidalgo (2021) stresses that a positive relationship between the variables might only hold across comparable regions or countries.

Chu & Hoang (2020) find for a large sample of 88 countries that the relationship between economic complexity and income inequality follows a U-shaped distribution. Their empirical findings show that, especially for countries at low-income levels, an increase in economic complexity often also increases income inequalities. The fact that most scholars tested the relationship only for high-income countries could explain why Chu & Hoang (2020) found different results for their sample, which includes low-income countries. Overall, the results seem to differ depending on the sample of countries used for the analysis. Therefore, this study will research the relationship between economic complexity and national income inequality to explore the dynamics between the variables for the OECD countries, which are used as a sample.

### 2.2.3 Economic Complexity and Regional Income Inequality

The literature on economic complexity and regional income inequality is even more scarce, and there have only been a handful of empirical studies on the topic, all from the last few years. From a theoretical perspective, Hartmann & Pinheiro (2022) state that it is likely that a high ECI can create increased interregional inequalities, that is, higher disparities between regions. They argue that knowledge-intensive production often concentrates spatially; thus, they cluster in specific regions of a country due to agglomeration effects. Porter (2003) argues that differences in regional performance are significantly influenced by the clusters present in the region. He finds that wage differences in a region's traded industries shape regional performance and income disparities. He also contends that a higher share of employment in the tradable industry sector in a region contributes significantly to better regional performance.

Economists and geographers have long argued that agglomeration effects and economies of scale resulting from complex production can lead to interregional inequality where some regions thrive while others fall behind. Boschma & Lambooy (1999) mainly trace these spatial concentrations back to historical path dependencies, which is the idea that economic structures are advert to change and persist over time. Krugman (1991), on the other hand, argues from a neoclassical trade-theory perspective, claiming that industry can emerge anywhere where production agglomerates. He claims that this typically creates a core and periphery structure within countries, leading to regional differences in output between regions as production is clustered in the core.

Hartman & Pinheiro (2022) point out that as a country's economy develops, governments focus more on increasing human capital in all regions. This can lead to migration movements to the centres of complex economic activity, similar to the core in Krugman's model. Thus, rural or less developed regions are increasingly falling behind. These findings are similar to those observed by innovation economists. Many scholars have argued that innovation increases regional economic inequality and thus inhibits inclusive growth across and within regions (George, McGahan & Prabhu, 2012; Lee, 2018; Piketty, 2014; Pinheiro et al., 2021). Empirical research on the relationship between economic complexity and inequalities across regions is scarce. Pinheiro et al. (2021) found that high-income regions tend to diversify into high-complex activities, while low-income regions usually adopt less-complex activities, indicating that diversification can potentially increase polarization and aggravate income differences across regions.

Concerning intraregional income inequality, many scholars have found that higher economic complexity correlates with improved equality within a region. Török et al.'s (2022) study on Romania shows that regions with higher economic complexity tend to be more equal. Additionally, he finds empirical evidence that past economic complexity indices for regions in Romania were reliable predictors of regional growth in the future. Chávez et al. (2017) find a similar relationship for Mexican states, with more economically complex states showing higher income equality levels. Marco et al. (2022) also find a slightly positive relationship between economic complexity and income equality in Spanish regions. In a similar study for Brazil, Bandeira Morais et al. (2021) found that the relationship between economic complexity and income inequalities in the regions follows an inverted U-shape, with regions first becoming less equal when they develop at low-income levels and then becoming increasingly equal as income-levels continue to rise.

Hidalgo (2022, p.6) states that economic complexity as a framework was rapidly adopted by policymakers worldwide, such as in the smart specialisation strategy in Europe, new manufacturing strategies in the US or policies for China's special economic zones. Thus, evaluating the impact of economic complexity on society in terms of economic inequality becomes essential, as evidenced by several scholars shows that there seems to be a critical correlation between economic complexity and income inequality between individuals and regions (Hidalgo, 2021; Marco et al., 2022; Bandeira Morais et al., 2021; Gómez-Zaldívar, 2022; Török et al., 2022; Pinheiro et al., 2021).

### 3.2.4 The Study's Contribution to the Literature

The literature on economic complexity and national income inequality shows ambiguous results depending on the sample and time frame of the study. This thesis will explore the relationship between the two variables for the OECD countries for 2019. There is a gap in the literature regarding empirical evidence on economic complexity and interregional inequalities, which is why the thesis explores this relationship in detail. The literature on economic complexity and intraregional inequalities finds mostly negative correlations, but not exclusively. While there is also such a study on Spain, it looks at several variables, while this analysis is more focused and explores both intra- and interregional dynamics in Spain. This holistic analysis that compares across-country inequalities with regional inequalities within a country is unique in the literature. Unlike other studies, this country study is not the focus of the analysis but serves as an example to explore local effects.

# 3. Data

This thesis explores the relationship between economic complexity and regional inequality by conducting a quantitative study examining the relationship between economic complexity and income inequality. The empirical analysis is divided into three parts, as shown in Figure 5. The analysis progresses from vast geographic entities (countries) to smaller ones (regions).

**The relationship between economic complexity and** ...

| Sample: OECD countries (National level) | |
|---|---|
| **… Inequality across individuals at the national level (Part I)** | **… Inequality across regions at the national level (Part II)** |
| Sample: Spanish Provinces (Regional level) | |
| **… Inequality across and within regions (Part III)** | |

*Figure 3: Methodological structure of the study*

## 3.1 Source Data

The quantitative analysis of the relationship between economic complexity and regional inequality includes several regression models. The models use secondary sources, which are mainly datasets provided by the OEC (2023), the OECD (OECD Data, 2023), the Global Data Lab (2023) and the Spanish Statistical Office (INE, 2022, 2023, 2023a-f) using data for the year 2019 and 2020. The datasets were modified, and the units and scales were adjusted to be comparable and fitting for the analysis.

## 3.2 Sample

There are two samples used throughout the study. The first sample consists of 34 of the current member states of the OECD (Table 1). There are 38 OECD members, of which 34 have complete data availability for the analysis. The OECD member states Estonia, Iceland, Luxembourg, and Latvia were excluded because data on these countries was unavailable for the main independent variable. The sample selection of OECD members was based on two reasons. Firstly, the OECD provides extensive data resources for its member countries, including a national Gini coefficient, the dependent variable for most regressions. Additionally, data on economic complexity is available for all member countries of the OECD. Secondly, using only OECD countries creates the certainty of some degree of homogeneity and comparability between the countries. According to the OECD, the minimum requirement to become a member state is the adherence of the country with two essential requirements: "(i) democratic societies committed to rule of law and protection of human rights; and (ii) open, transparent and free-market economies." (OECD TUAC, 2023, p.1). It must be noted that it is not evident that all current member countries fulfil these requirements, however, this serves as an approximation that these countries are more comparable than a random selection of countries.

| | | |
|---|---|---|
| 1. Australia | 2. Austria | 3. Belgium |
| 4. Canada | 5. Chile | 6. Chile |
| 7. Colombia | 8. Costa Rica | 9. Czechia |
| 10. Denmark | 11. Finland | 12. France |
| 13. Germany | 14. Greece | 15. Hungary |
| 16. Ireland | 17. Israel | 18. Italy |
| 19. Japan | 20. Lithuania | 21. Mexico |
| 22. Netherlands | 23. New Zealand | 24. Norway |
| 25. Poland | 26. Portugal | 27. Slovakia |
| 28. Slovenia | 29. South Korea | 30. Spain |
| 31. Sweden | 32. Turkey | 33. United Kingdom |
| 34. United States | | |

*Table 1: Selected sample of OECD countries in alphabetical order*

The second sample for Part III consists of Spanish provinces, as seen in Table 2. The sample includes 46 out of 50 provinces in Spain. The provinces Avila, Las Palmas, Tenerife, and Zamora had to be excluded because of missing data. Additionally, in the second regression in this section, Navarra is excluded because of missing data on the Gini coefficient for the region.

| | | |
|---|---|---|
| 1. Álava | 2. Albacete | 3. Alicante |
| 4. Almería | 5. Asturias | 6. Badajoz |
| 7. Baleares | 8. Barcelona | 9. Burgos |
| 10. Cáceres | 11. Cádiz | 12. Córdoba |
| 13. Cantabria (Santander) | 14. Castellón | 15. Ciudad Real |
| 16. Cuenca | 17. Girona | 18. Granada |
| 19. Guadalajara | 20. Gipuzkoa | 21. Huelva |
| 22. Huesca | 23. Jaén | 24. La Coruña |
| 25. La Rioja | 26. León | 27. Lleida |
| 28. Lugo | 29. Málaga | 30. Murcia |
| 31. Orense | 32. Palencia | 33. Pontevedra |
| 34. Salamanca | 35. Segovia | 36. Sevilla |
| 37. Soria | 38. Tarragona | 39. Teruel |
| 40. Toledo | 41. Valencia | 42. Valladolid |
| 43. Vizcaya | 44. Zaragoza | 45. Madrid |
| 46. Navarra* | | |

**not included in all regression Models*

*Table 2: Sample of Spanish provinces in alphabetical order*

## 3.2 Variables

The variables for the analysis are constructed using data that provide international and regional data on the indicators used. Part I and II use the same explanatory and control variables as both parts use indicators on the country level. Part III uses different variables as these are on the regional level. Therefore, the variables for Part I and II are presented together, followed by a presentation of the variables for Part III.

### 3.2.1 Variables Part I and II

The analysis is divided into three parts which will be explained in more detail in the subsequent section. The dependent variable for Part I is the national Gini coefficient, measuring each country's income inequality level. For Part II, the Theil entropy index is the dependent variable to measure each country's regional inequality level.

*Gini Coefficient*

The Gini coefficient is the dependent variable for the models in Part I. It is calculated by the OECD using the concept of the Lorenz curve. The area of the diagonal Lorenz curve (line of perfect equality) is subtracted from the actual Lorenz curve (Canberra Group, 2011, p. 75-78). The Gini coefficient is thus represented by the blue area, as seen in Figure 3. The calculation is based on disposable household income and compares the cumulative proportions of the population to the cumulative proportions of each individual's income. The Gini coefficient ranks between 0 and 1, where 0 means perfect equality, and 1 indicates perfect inequality.

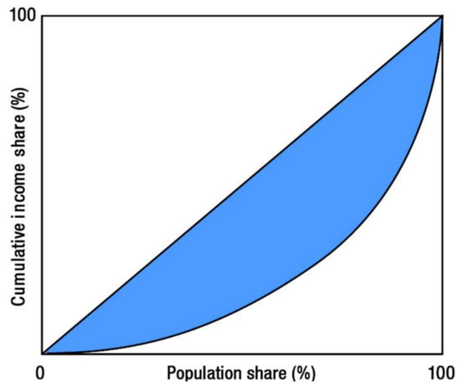


*Figure 4: The Lorenz curve, source: adapted from Stiglitz et al. (2018)*

Mathematically, the OECD calculates the Gini coefficient as follows:

$$G = \left(\frac{1}{2n^2\mu}\right)\sum_{i,j}^{n} \left|y_i - y_j\right|$$

where, $n$ stands for the number of people (population), $\mu$ represents the mean equivalised disposable household income of each individual, and $y_i$ and $y_j$ are the equivalised disposable household income of the $i$th and $j$th individuals in the population (Canberra Group 2011, p. 78).

The Gini coefficient is one of the most widely used indicators of inequality, and the OECD has good data availability of this indicator for its member states. All Gini coefficients are taken from the OECD database (OECD Data, 2023), except for the Gini for Colombia, which was not in the dataset and is therefore taken from a dataset provided by the World Bank (World Bank, 2023).

*Theil entropy index*

Part II uses the Theil entropy index for regional inequalities as its dependent variable. It is a measurement used by the OECD to measure regional disparities (OECD, 2016, pp. 176-177). In this case, the Theil entropy index is calculated with the Gross National Income (GNI) per capita for each region in each country which is provided by the OECD Data (2023a). Intuitively, the Theil entropy index compares the mean GNI per capita of each region to the mean GNI per capita of all regions combined and thus estimates the deviation of the region's income from the average national income. Mathematically the Theil entropy index can be expressed in the following way:

$$Theil = \frac{1}{N}\sum_{i=1}^{N} \frac{y_i}{\bar{y}} ln\left(\frac{y_i}{\bar{y}}\right)$$

where N is the number of regions within a country, $y_i$ is GNI per capita in the $i^{th}$ region, and $\bar{y}$ is the mean GDP per capita across all regions in the country. The calculation is done for every country individually. The Theil entropy index ranges from 0 to ∞, where a higher value indicates higher regional disparities and 0 represents an entirely equal distribution.

This study uses GNI rather than GDP or household income for two main reasons. Firstly, because of data availability for the analysed regions. Secondly, it includes the income earned by foreign residents. According to Bogin et al. (2017, p.2), GNI is a better estimation of the average personal income of a nation's residents, or in this case, region, than GDP.

*ECI*

The analysis uses two different independent variables. Parts I and II use the national Economic Complexity Index (ECI), which measures a country's economic complexity level. The data is provided by the Observatory of Economic Complexity (OEC 2023). Part III uses the regional Economic Complexity Index for Spanish provinces, also provided by the OEC (2023). The mathematical background of calculating the ECI is explained in section 2.1. Figure 4 shows the ECI graphically.

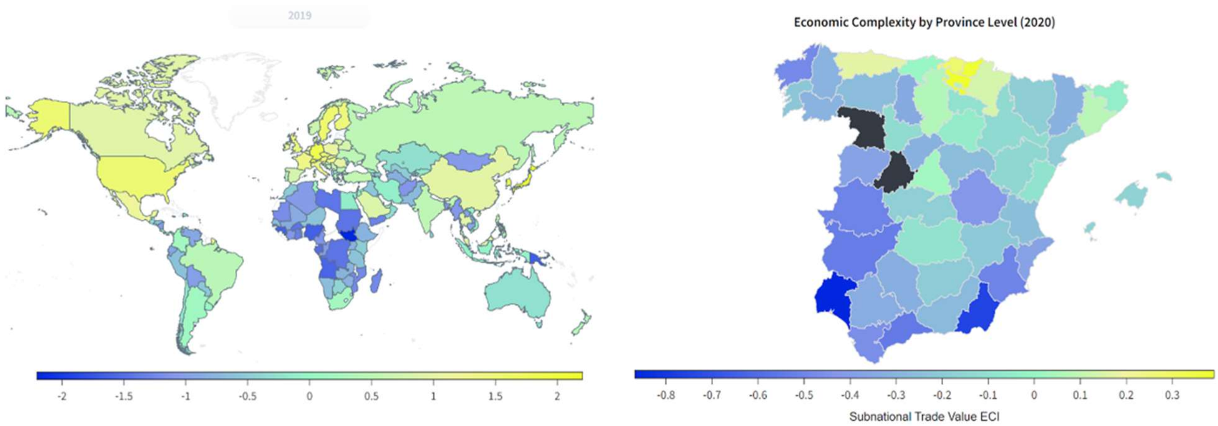


*Figure 5: ECI for the countries of the world (left) and the provinces of Spain (right), source OEC (2023)*

The data for all indicators is cross-sectional. For Part I and II, all data is from the year 2019. The latest available year with complete data available for all indicators was 2020, however, to avoid fluctuations and short-term discrepancies that result from the COVID-19 pandemic, the year 2019 was chosen. For Part III, data on the ECI was only accessible for the year 2020, therefore, the dependent variable measuring the regional Gini coefficient and the independent variable ECI is for the year 2020. Comparing the data to previous years, it became clear that fluctuations are not that high for the Gini, and especially the ECI shows minimal changes over the years, even including the pandemic years. For the controls, however, there are significant differences, especially for the one controlling tourism and unemployment, which is why the control variables are chosen for the year 2019. This limitation must be kept in mind, however, it does not have a significant influence on the general trends and relationships between the variables.

*Control Variables*

Part I and II use the same control variables. All control variables for these Parts are measured on the country level and are carefully selected to help understand the causes for inequality apart from economic complexity across individuals and regions. The data for all control variables is taken from a dataset from the Global Data Lab (2023), which provides human development and economic indicators indices. The control variables for Part I and II are divided into two categories:

1. *Human development indicators*
    - *Income (standard of living)*
    - *Health*
    - *Education*

2. *Economic indicators*
    - *GNI per capita*
    - *Economic growth rate (5-year average)*
    - *Unemployment rate*

*1. Human development control indicators*

Income is presented as an index that measures the overall standard of living in a country. As mentioned previously, Kuznets (1955) found that a higher income level is associated with lower inequality. Contrastingly, more recently, scholars such as Piketty and Saez (2014) find that since the second half of the 20th century, higher levels of income in an economy are related to higher income inequality. Therefore, there is a control for the standard of living through the income index to evaluate whether inequality is associated with the level of income rather than economic complexity. The other development indicators, health and education, control whether a higher overall level of health and education in a country can explain part of the differing inequalities in the selected OECD countries. Goldin & Katz (2009, p.4) show that, for the US, a higher average level of education has historically positively affected economic equality. The relationship between health and income inequality seems to be less clear. Epidemiologists and health economists such as Pickett & Wilkinson (2015) and Lorgelly & Lindley (2008) argue that the relationship has been debated heavily. While most scholars find a significant relationship between income inequality and health, the direction of causation is unclear. Additionally, studies find contradicting results on whether higher levels of inequality are positively or negatively associated with health.

*2. Economic control indicators*

The economic indicator GNI per capita is very similar to the income index; it measures the same thing but is presented as the natural logarithm of Gross National Income per capita in 2011 US$ PPP. Income is included in both groups of control variables because it serves both as an indicator of how well the economy is doing and indicates the standard of living and, thus, the level of human development.

Another economic control variable is the economic growth rate. The yearly growth rate is calculated by taking an average over five years, from 2015 to 2019, to avoid distortions of the general trend due to short-term fluctuations. As economic growth is closely related to income level, it is also closely related to the findings of Kuznets and Piketty mentioned above. While in Kuznets' time, economic growth seemed to affect income inequality positively, today, the opposite appears to be true, at least for high-income countries (Piketty & Saez, 2014). The analysis includes the growth rate to control for its effect on income inequality.

Lastly, the analysis controls for unemployment rates which are also calculated as a five-year average from 2015 to 2019 to avoid short-term variations. Sheng (2011) shows that over a time of 70 years, unemployment and income inequality have been positively correlated in the US. On the other hand, Burniaux et al. (2006) find that this relationship has been less consistent for the OECD countries. Using the same control variables for Part I and for Part II allows for a comparison of how these indicators affect inequality at the individual and regional levels.

### 3.2.2 Variables Part III

Part III looks firstly at regional ECI and regional economic performance and secondly at regional ECI and intraregional inequality for Spain. Part III uses GDP per capita and a regional Gini coefficient for the provinces of Spain as dependent variables. The main independent variables for the analysis are the ECI for the OECD countries and the regional ECI for the Spanish provinces, both provided by the OEC (2023). The regional Gini coefficient for the Spanish provinces and the regional GDP per capita are provided by the Spanish Instituto Nacional de Estadística (INE, 2023; INE, 2023a). The Gini coefficient is calculated similarly to the one in Part I, using disposable household income, whereby the net income of one household is divided by the household members (INE Atlas. 2023). The formula used is the same as the one by the OECD, only adjusted for the regional population instead of the national population.

The control variables in Part III differ from those in Part I and Part II because of the data availability for the Spanish provinces and because they are adapted to the Spanish economy. All control variables use data from different datasets from the Instituto National de Estadística. The control variables are the following:

- *GDP per capita*
- *Tourism*
- *Unemployment rate*
- *Industrial employment*
- *Service employment*
- *Population variations*
- *Geographic typology*

Rodríguez-Pose & Tselios (2009) define several important indicators for regional inequality in the EU. Among these are several of the chosen indicators for this analysis. They find, for instance, relationships between income inequality and income levels per capita, unemployment, industry, population and urbanisation.

As previously established, the overall economic performance and income levels could affect income inequality, which is why GDP per capita is included as a control variable. GDP per capita is used instead of other income and economic performance indicators because of data availability for the Spanish provinces. Provincial GDP per capita is calculated from data on GDP, and the data on population are provided by the INE (INE, 2023a; INE, 2023b).

Tourism is included as a control as it is a central part of the Spanish economy. In 2019 tourism contributed 12.6% to the total GDP of the country, with 12.7% of the population being employed in an activity related to tourism (INE, 2022, p.1). The indicator 'Tourism' is compiled using a dataset consisting of the aggregated number of tourists per province (INE, 2023c) for each year, divided by the population of the province (INE, 2023b). The unemployment rate is included as it is found to be a potential cause of differences in GDP (Rodríguez-Pose & Tselios, 2009). The data on unemployment is provided by the INE (INE, 2023d).

Industrial employment and service employment are two control variables that indicate the percentage of the population of a province that works in either the industrial or service sectors. Controlling industrial and service sector employment can show whether the type of sector employment affects income inequality within Spanish regions. The data is provided by the INE (INE, 2023e).

The population variances variable indicates whether a region is experiencing net positive or negative inflows and to what extent. It could also indicate the attractiveness of a region. Population variation is presented as a yearly percentage change of population calculated as a ten-year average from 2010 to 2019 with data from a dataset provided by INE (INE, 2023f).

Finally, the analysis uses dummy variables for the geographic typology of the region with data provided by Eurostat (Eurostat, 2018) that classifies the regions into three categories: PU (predominantly urban), IN (intermediate), and PR (predominantly rural). Creating a dummy variable for the geographic typology allows the analysis of whether the effects of economic complexity on income inequality are different in urban, rural, and mixed areas.

All control variables are chosen for the year 2019. While there were later years available, the year 2019 was selected to avoid the effects of the global pandemic, which had a short-term impact on many indicators.

## 4. Method

All parts of this study are based on a cross-sectional analysis, capturing the relationships between different indicators for the year 2019, and in the case of Part II, the years 2019 and 2020 are mixed. The theoretical model for the analysis is a multivariable linear regression model (MLR):

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_{control\, i} + \ldots + \varepsilon_i \qquad i = 1, \ldots, n$$

where, $y_i$ represents the dependent variable, $\alpha$ is the intercept, the coefficient $\beta_1$ estimates the effect of the independent variable $x_i$ and $\beta_2$ is the estimator of a control variable $x_{contorl\, i}$. Other control variables can be added to the model and $\varepsilon_i$ is the error term. All Parts use an adapted version of this model for the regressions. In the following sections, the methodology for each Part is presented separately.

### 4.1 Part I: Economic Complexity and Inequality Across Individuals on the National Level

Part I analyses the association of a country's ECI with income inequality between individuals across the sample of OECD countries and the year 2019. Most literature on economic complexity and inequality within countries suggests that economic complexity negatively correlates with national income inequality. A regression with the independent variable ECI and the dependent variable Gini shows that the relationship holds for the sample of the 34 OECD countries. The setup is the following:

$$Gini_i = \alpha + \beta_1 (ECI)_i + \beta_2 (control1)_i + \beta_3 (control2)_i + \ldots + \varepsilon_i \qquad i = 1, \ldots, 34 \qquad (1)$$

This setup uses the national Gini coefficient for each of the 34 OECD countries ($i$) as the dependent variable. The independent variable is the national ECI for each country ($i$). Control variables are added in a stepwise manner to the multiple regression analysis. In Model 1, the regression is unadjusted without any control variable, Model 2 includes development indicators as control variables and Model 3, economic indicators. Finally, Model 4 includes all control variables.

## 4.2 Part II: Economic Complexity and Inequality Across Regions at the National Level

This section analyses the relationship between economic complexity ad interregional inequality for the 34 OECD countries. The national ECI is again used as the independent variable, while the dependent variable is the Theil entropy index indicating the level of inequality between regions in a country. The regression estimates whether the level of economic complexity in a country, measured by the ECI, correlates with higher levels of inequality between regions, measured by the Theil entropy index.

The same countries are included in the second section as in the first, except for Israel, as it lacks regional data in the dataset. Therefore, the sample consists of 33 countries. The setup is the following:

$$Theil_i = \alpha + \beta_1(ECI)_i + \beta_2(control1)_i + \beta_3(control2)_i + \dots + \varepsilon_i \qquad _{i = 1, \dots, 33} \quad (2)$$

Similar to Part I, the regression is run in four steps. Model 1 is an unadjusted regression without any control variables, followed by Model 2 with development control variables, Model 3 with economic control variables and finally, Model 4 with all control variables added.

The test for the Breusch-Pagan test heteroskedasticity shows that when adding the control variables, the models in this section suffer from heteroskedasticity (see Appendix 9.2). Therefore, robust standard errors are used for the estimation of those models.

## 4.3 Part III: Economic Complexity and Regional Inequality in Spain

Part III explores the relationship between economic complexity and income inequality between and within the provinces of Spain. The section aims at explaining economic complexity and income inequality more specifically by zooming in on one country and analysing regional economic complexity and its relationship to regional GDP per capita and regional Gini coefficients.

*Economic complexity and economic inequality across Spanish provinces (interregional)*

The first subsection analyses the relationship between regional economic complexity levels and GDP per capita across Spanish provinces. While this analysis does not give any direct evidence of whether economic complexity is related to higher or lower inequalities across regions, a relationship does show whether economic complexity can potentially lead to certain regions forging ahead or whether the lack of economic complexity leads to regions falling behind as found in the literature. The setup for this regression is the following:

$$GDPpc_i = \alpha + \beta_1 (ECI)_{regional\,i} + \beta_2(control1)_i + \beta_3(control2)_i + \ldots + \varepsilon_i \qquad i = 1, \ldots, 46 \qquad (3)$$

The model is estimated in a two-step regression analysis; firstly, Model 1 is unadjusted and secondly, Model 2 with all the regional control variables.

*Economic complexity and income inequality within Spanish provinces (intraregional)*

The second part of the section deals with economic complexity and economic inequality within the provinces of Spain, measured by the Gini coefficient for each region.

$$Gini_{provinces\,i} = \alpha + \beta_1 (ECI)_{regional\,i} + \beta_2(control1)_i + \beta_3(control2)_i + \ldots + \varepsilon_i \qquad i = 1, \ldots, 45 \qquad (4)$$

In this analysis, a four-step regression Model is used again. Model 1 is unadjusted and only includes the independent variable regional ECI and the dependent variable Gini for the provinces. Model 2 is also an unadjusted regression excluding the province of Madrid, which had been identified as an outlier. Model 3 includes all control variables. Model 4 also includes all control variables, and the regression is run separately for the different geographic typologies of the provinces.

## 4.4 Econometric Robustness Tests

Every regression is tested for robustness to verify that the MLR models follow the Gauss-Markov assumptions so that the results are reliable and correctly specified. All models tested for misspecification with the test for multicollinearity. Additionally, the data is tested for heteroskedasticity with the Breusch-Pagan test. These two tests are considered the most important ones for cross-sectional data. However, it should be noted that more robustness tests could be undertaken to ensure full reliability. Regarding the scope of this thesis, it was considered satisfactory to only test for the most important violations.

*Multicollinearity*

All models in this thesis are tested for multicollinearity, with the test measuring the Variance Inflation Factor (VIF) (Appendix 9.1). In Stata, this is done by using the command 'vif' after the regression. If the VIF value is over ten, the model has a multicollinearity problem that should be corrected. According to Kutner et al. (2005, pp.278-279), it is common in social science that independent variables are inherently correlated and explain the same relationship. If this is the case, the regression results can be distorted, and coefficients show misleading results (Kutner et al., 2005, p.283). As many of the independent variables used as controls in these models explain economic or human development in some way, the problem of multicollinearity is taken into account for the regression, and the models are adjusted accordingly. The study consists of three sections with three different models that are explained separately.

*Heteroskedasticity*

Heteroskedasticity is a common problem in linear regression models using cross-sectional data, which describes the inconsistency of error variance across observations. According to Breusch &

Pagan (1979, p.1287), in the absence of homoskedasticity, ordinary least squares (OLS) become inefficient, and a bias occurs in the estimated standard errors, leading to invalid linear regression model estimations. Heteroskedasticity violates the Gauss-Markov assumptions, which ensure that the OLS estimator has the lowest variances, thus, is the most efficient estimator of all estimators. One possibility to test for heteroskedasticity is the Breusch-Pagan test, which can be carried out in Stata using the command 'hettest'. The test's null hypothesis is that the data is homoscedastic. If the null hypothesis is rejected at the five per cent level, the model is likely to suffer from heteroskedasticity, and robust standard errors should be used to obtain more efficient results of the inference to obtain the most accurate result. In Stata, the model can be estimated using robust standard errors by adding the command 'vce(robust)' behind the regression. All models in the analysis are tested for heteroskedasticity (see Appendix 9.2) using the Breusch-Pagan test, and the models that suffer from heteroskedasticity are then modified to be estimated with robust standard errors.

# 5. Analysis

## 5.1 Descriptive results

Tables 3 and 4 give an overview of the main descriptive results for all sections, including the dependent variable (ECI), the independent variables (Gini and GDP), and all the control variables.

| Variable | Obs. | Mean | Std. dev. | Min | Max |
|---|---|---|---|---|---|
| **ECI** | 34 | 1.093 | 0.607 | -0.256 | 2.205 |
| **Gini coefficient** | 34 | 0.325 | 0.068 | 0.222 | 0.513 |
| **Health** | 34 | 0.938 | 0.039 | 0.834 | 0.991 |
| **Education** | 34 | 0.868 | 0.065 | 0.691 | 0.942 |
| **Income** | 34 | 0.907 | 0.057 | 0.749 | 0.986 |
| **GNI per capita (log)** | 34 | 10.612 | 0.376 | 9.564 | 11.130 |
| **Economic growth rate** | 34 | 2.697 | 1.596 | 0.792 | 9.870 |
| **Unemployment rate** | 34 | 7.121 | 3.946 | 2.818 | 21.665 |
| **Theil entropy index** | 33 | 0.035658 | 0.031725 | 0.00276 | 0.1474 |

*All observations for the year 2019*

*Table 3: Descriptive statistics for the variables of Part I and Part II*

| Variable | Obs. | Mean | Std. dev. | Min | Max |
|---|---|---|---|---|---|
| **ECI** | 45 | -0.219 | 0.258 | -0.867 | 0.390 |
| **Gini** | 45 | 31.111 | 1.615 | 28.600 | 35.900 |
| **GDP per capita** | 45 | 24918.470 | 5073.467 | 17997.000 | 36624.000 |
| **Tourism** | 45 | 2.157 | 2.964 | 0.199 | 12.811 |
| **Unemployment rate** | 45 | 13.846 | 4.994 | 7.230 | 24.700 |
| **Industry employment** | 45 | 14.484 | 5.130 | 5.300 | 27.500 |
| **Service employment** | 45 | 66.040 | 5.287 | 58.500 | 80.600 |
| **population variance** | 45 | -0.172 | 0.450 | -1.033 | 0.581 |
| **geographic typology** | 45 | 1.156 | 0.903 | 0.000 | 2.000 |

*ECI and Gini coefficient for the year 2020, all other variables for the year 2019*

*Table 4: Descriptive statistics for Part III*

## 5.2 Regression Analysis Results

The results of the regression analysis are presented in the three sections separately.

### 5.2.1 Part I: Economic Complexity and Inequality Across Individuals on the National Level

Table 5 presents the regression results for the four-step regression model for Part I. The results for Model 1 show that when regressing ECI on the Gini coefficient without any control variables, there is a clear negative relationship between the ECI and the Gini coefficient. The p-value shows that the relationship is statistically significant at the one per cent level, and the adjusted R-square shows that around 19 per cent of the differences in Gini coefficients can be explained by the ECI. All models in this thesis use the adjusted R-square instead of the regular R-squared, as the former takes into account the number of variables in the model and gives thus a more accurate estimation of the actual part of the differences in the response variable that the variables can explain. Figure 6 graphically shows the relationship between the two variables.

In Model 2, development indicators are added as control variables to the regression. In this model, none of the variables shows a statistically significant relationship with the Gini coefficient alone. However, the adjusted R-squared value shows that 40 per cent of the differences in Gini values can be explained by these variables together.

Model 3 includes economic indicators as controls. This model shows that the Gini coefficient of a country is highly correlated with the GNI per capita of the country (significant at the five per cent level), while all other variables are not statistically significantly related. The adjusted R-squared is, however, still at 30 per cent.

Model 4 includes all indicators except for the GNI per capita, which was excluded from the model since the test for multicollinearity shows that income and GNI per capita are highly correlated, as demonstrated by the VIF results in the test for multicollinearity (see Appendix 9.1). Therefore, the regression with development and economic control variables runs without the GNI per capita. In Model 4, the ECI shows a statistically significant relationship at the 10 per cent level, as does education. The adjusted R-squared shows that 41 per cent of the variation in the Gini can be with the ECI and the control variables. Overall, the regression analysis suggests that economic complexity is negatively associated with inequality across the studied OECD countries.

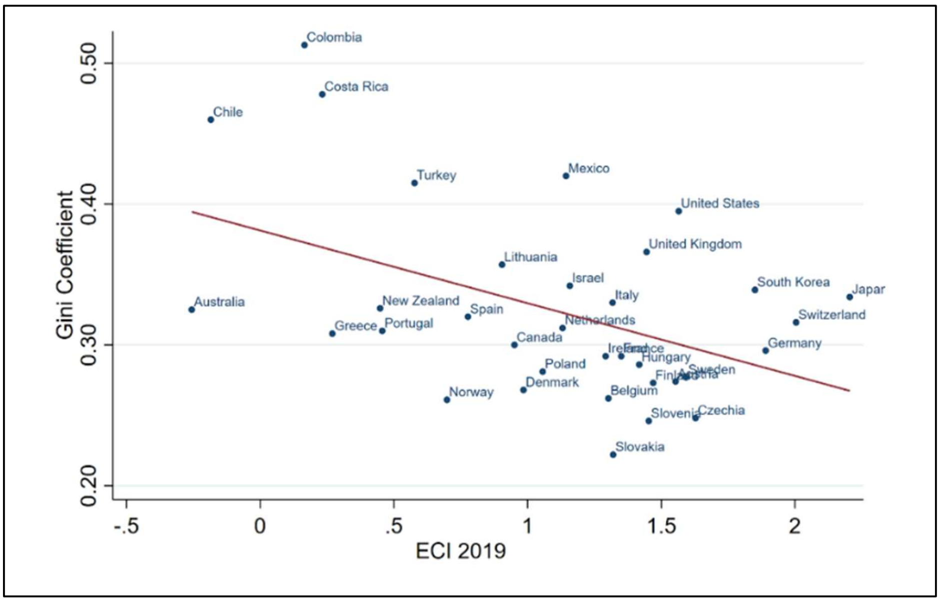


*Figure 6: Results linear regression analysis with trend line, dependent variable: Gini coefficient, independent variable: ECI 2019*

| Explanatory Variable | Model 1 unadjusted | Model 2 development control var. | Model 3 economic control var. | Model 4 all control var. |
|---|---|---|---|---|
| **ECI** | -0.052*** (0.018) | -0.024 (0.018) | -0.029 (0.019) | -0.035* (0.019) |
| **Health** | | 0.106 (0.319) | | 0.214 (0.340) |
| **Education** | | -0.359 (0.247) | | -0.440* (0.249) |
| **Income** | | -0.342 (0.357) | | -0.346 (0.360) |
| **GNI per capita** | | | -0.101** (0.029) | |
| **Economic growth rate** | | | -0.002 (0.006) | -0.002 (0.006) |
| **Unemployment rate** | | | -0.003 (0.003) | -0.004 (0.003) |
| R-squared | 0.213 | 0.471 | 0.46 | 0.52 |
| Adj. R-squared | 0.189 | 0.399 | 0.385 | 0.413 |

*$***p < 0.01$ $**p < 0.05$ $*p < 0.1$*
*Standard errors in parentheses,*
*34 observations*

*Table 5: Four-step regression analysis, dependent variable: Gini coefficient*

### 5.2.2 Part II: Economic Complexity and Inequality Across Regions at the National Level

The regression analysis for this Part uses the Theil entropy index as its dependent variable for the relationship between economic complexity and interregional inequality. The regression analysis is done in four steps, progressing from unadjusted to including development controls, economic controls, and finally, including all control variables. Similarly to the previous section, the multicollinearity test shows that GNI per capita and income are strongly correlated (see Appendix 9.1), which is why GNI per capita was excluded from the final regression with all control variables.

Figure 7 shows that the observations are quite randomly scattered, and there is seemingly no linear relationship between economic complexity and interregional inequality across the OECD countries. Interestingly, Slovakia, which had a very low Gini coefficient in section one, sticks out with the highest Theil entropy index. The data shows that while the income differences between individuals in the county are relatively small compared to other countries, the differences between the regions are comparably high. A similar pattern can be observed in Czechia. Conversely, Colombia has the highest Gini coefficient in the sample, so the highest income differences among individuals. However, Figure 5 shows that the differences between regions are rather average compared to the other countries in the sample.

While there is no clear trend to be detected in the graph, it is noticeable that some countries with the highest and the countries with the lowest ECIs also have the lowest levels of interregional economic inequality. However, those observations should be treated cautiously, and further research must be conducted to confirm this trend.

Table 6 shows that none of the four regression models exhibits statistically significant results at the one, five, or ten per cent level. Thus, we fail to reject the null hypothesis of no correlation between ECI and the Theil entropy index. This confirms the observation from the graphical depiction that the distribution of the Theil entropy indices has no linear relationship with the ECI or any other control variable. Therefore, it can be concluded that this analysis could not find any association between economic complexity and the level of interregional inequality across the OECD countries.

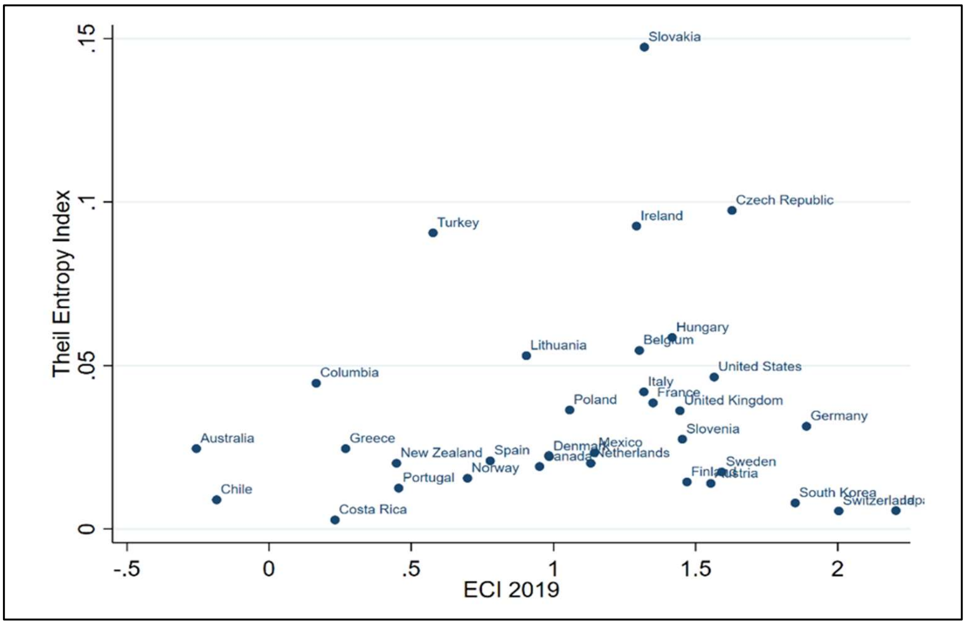


*Figure 7: Results linear regression analysis, dependent variable: Theil Entropy Index, independent variable: ECI 2019*

| Explanatory Variable | Model 1 unadjusted | Model 2 development control var. | Model 3 economic control var. | Model 4 all control var. |
|---|---|---|---|---|
| **ECI** | 0.005 (0.009) | 0.005 (0.009)¹ | 0.009 0.011 | 0.013 (-0.012)¹ |
| **Health** | | 0.022 (-0.159)¹ | | 0.168 (-0.195)¹ |
| **Education** | | 0.175 (-0.199)¹ | | -0.184 (-0.205)¹ |
| **Income** | | -0.122 (-0.231)¹ | | -0.286 (-0.283)¹ |
| **GNI per capita** | | | -0.003 0.018 | |
| **Economic growth rate** | | | 0.002 0.004 | 0.003 (0.003)¹ |
| **Unemployment rate** | | | -0.002 0.002 | -0.003* (0.001)¹ |
| Adj. R-squared | -0.023 | 0.0592 | 0.088 | 0.168 |

****p < 0.01 **p < 0.05 *p < 0.1*
*Standard errors in parentheses*
*¹ robust standard errors*
*33 observations*

*Table 6: Four-step regression analysis, dependent variable: Theil Entropy Index*

### 5.2.3 Part III: Economic Complexity and Regional Inequality in Spain

Part III looks at the relationship between economic complexity and regional inequalities by zooming in on the Spanish Provinces. This one-country analysis is both interregional and intraregional.

*Economic Complexity and Economic Inequality Across Spanish Provinces (interregional)*

While Part II found no relationship between economic complexity and interregional inequality, a closer look at one country shows that economic complexity seems related to a region's economic performance. Figure 8 shows that ECI and GDP per capita are indeed strongly correlated. This is supported by the regression analysis in Table 7, which estimates that the ECI of a region has a statistically significant relationship with the GDP of a region at the five per cent level. This holds for the unadjusted regression and the regression with control variables. The adjusted R-squared indicates that 51 per cent of the differences in GDP per capita can be explained by the ECI for the unadjusted regression and over 80 per cent for the adjusted one. This shows that the ECI and the control variables explain the differences in GDP per capita to a large extent.

Besides the ECI, tourism and population variance are all positively related to GDP per capita in Spanish provinces, while unemployment is negatively correlated. All of these are statistically significant at the five per cent level. Notably, industry employment is not correlated with GDP per capita. At the same time, the ECI, which is largely based on industrial sectors, is strongly correlated, and can explain large parts of the differences in the economic performance of the regions. An explanation might be that while most activities included in the economic complexity matrixes are inherently industrial (Hidalgo & Hausmann 2009), it does not mean that all industrial activities are complex.

These findings are relevant to evaluating the question of how economic complexity relates to regional inequality. In Part II, no correlation was found, however, the limitations of the method and the use of cross-sectional data should not be underestimated. Part III finds that economic complexity is highly correlated with a region's economic performance, which does not prove but might suggest that increasing economic complexity in some regions could have diverging effects across the regions in Spain. Especially taking into account the findings by scholars that economic complexity is positively correlated with economic growth – i.e. Hidalgo & Hausmann, 2009; for country performance, Chávez et al. (2017) for Mexican regions or Stojkoski & Kocarev (2017), for Southeastern and Central Europe. Thus, following these findings and endogenous growth theory, it is likely that the regions with high levels of economic complexity will continue to grow more in the long run, while the regions with lower levels of economic complexity might stagnate. Considering that the regions with higher complexity levels already show higher GDP per capita levels, divergence cannot be excluded as a possible scenario for the future of the Spanish provinces. However, it should be noted that due to the cross-sectional nature of this study, further research needs to be conducted on the patterns of convergence or divergence across Spanish provinces to determine the actual effects of economic complexity on interregional inequality.

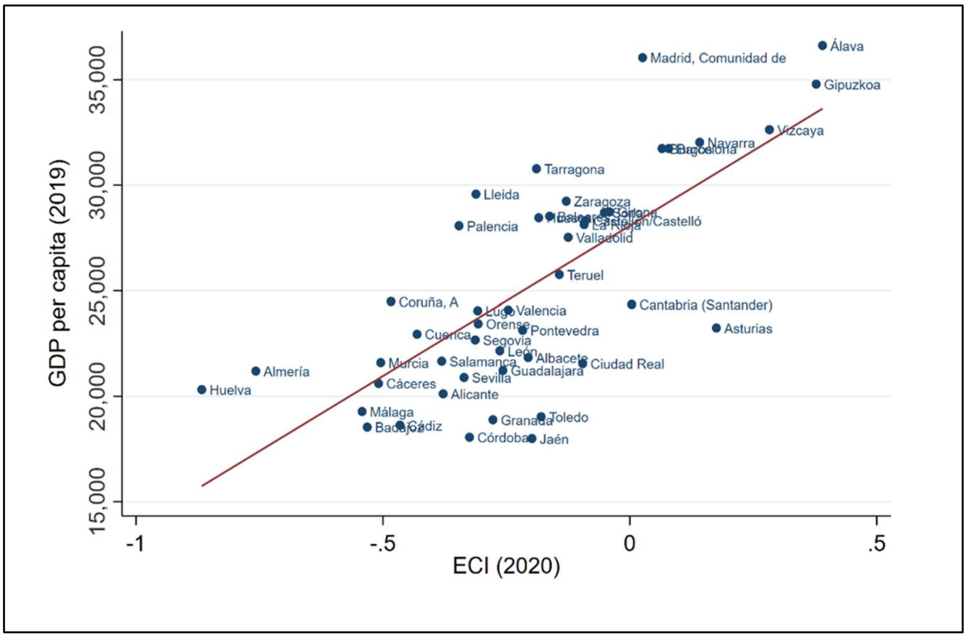

*Figure 8: Results linear regression analysis with trend line, dependent variable: GDP per capita, independent variable: ECI 2020*

| Explanatory Variable | Model 1<br>unadjusted | Model 2<br>with control var. |
|---|---|---|
| **ECI** | 14230.08**<br>(2040.846) | 5161.41**<br>(2211.11) |
| **Tourism** | | 407.7**<br>(168.11) |
| **Unemployment rate** | | -606.46**<br>(116.66) |
| **Service employment** | | -169.93<br>(118.60) |
| **Industry employment** | | 96.06<br>(141.33) |
| **Population variance** | | 2569.55**<br>(895.54) |
| Adj. R-squared | 0.514 | 0.812 |

****p < 0.01  **p < 0.05  *p < 0.1*
*Standard errors in parentheses, 46 observations*

*Table 7: Two-step regression analysis, dependent variable: GDP per capita*

*Economic Complexity and Income Inequality within Spanish Provinces (intraregional)*

This subsection explores economic complexity and inequality between individuals within Spanish provinces (intraregional). The regression analysis determines whether regions that are more complex, as measured by the regional ECI, show lower or higher levels of income inequality, as measured by the regional Gini coefficient. The four-step regression model shown in Table 8 gives conflicting results that can provide some direction of a possible correlation. The unadjusted model shows no statistically significant correlation between economic complexity and the Gini coefficient for the provinces.

However, as Figure 9 demonstrates, there is one significant outlier in the sample, which is the capital Madrid. Glaeser et al. (2009, p.626) argue that big cities typically have a more diverse population regarding skill levels and often show higher inequalities than rural areas. With 2.8 million inhabitants, Madrid is by far the biggest city in Spain. Excluding Madrid from the sample gives a higher correlation, which is close but not sufficiently statistically significant on the ten per cent level (Model 2). When including control variables in the regression (Model 3), the ECI becomes statistically significant on the ten per cent level, and thus, the variables seem to be somewhat negatively correlated. However, it also becomes evident that other control variables explain the differences in Gini more thoroughly, such as tourism and unemployment. Both variables show a statistically significant positive correlation at the five per cent level. Thus, a higher Gini coefficient is positively associated with higher unemployment levels and tourism.

Model 4 in Table 8 includes the dummy variable geographical typology, which controls whether a region is predominantly urban, rural, or intermediate. Figure 9 shows the distribution of the ECI by geographic typology. It shows that rural areas tend to have the lowest ECI values on average, while urban areas have, on average higher ECI values. Intermediate regions have the widest spread in ECI values from very high to very low. Notably, the correlation between economic complexity and intraregional inequality is the highest for intermediate provinces. In these regions, the ECI is the only variable that shows a statistically significant relationship with the Gini coefficient, and the adjusted R-value is at 45 per cent.

Hence, economic complexity is the most important factor and seems to explain a high percentage of the differences in inequalities in these intermediate regions, with a higher ECI indicating lower inequalities in these provinces. In urban regions, economic complexity is also negatively correlated with the Gini coefficient, however, only on the ten per cent level, and many of the control variables also show some level of correlation that can be considered statistically significant. Interestingly, GDP per capita, which had a very strong correlation with the ECI, shows a positive correlation with the Gini coefficient, while the ECI shows a negative correlation. For the rural provinces, economic complexity does not seem to affect inequalities, so there might be other, in this study, unexplored factors that drive inequalities in these regions.

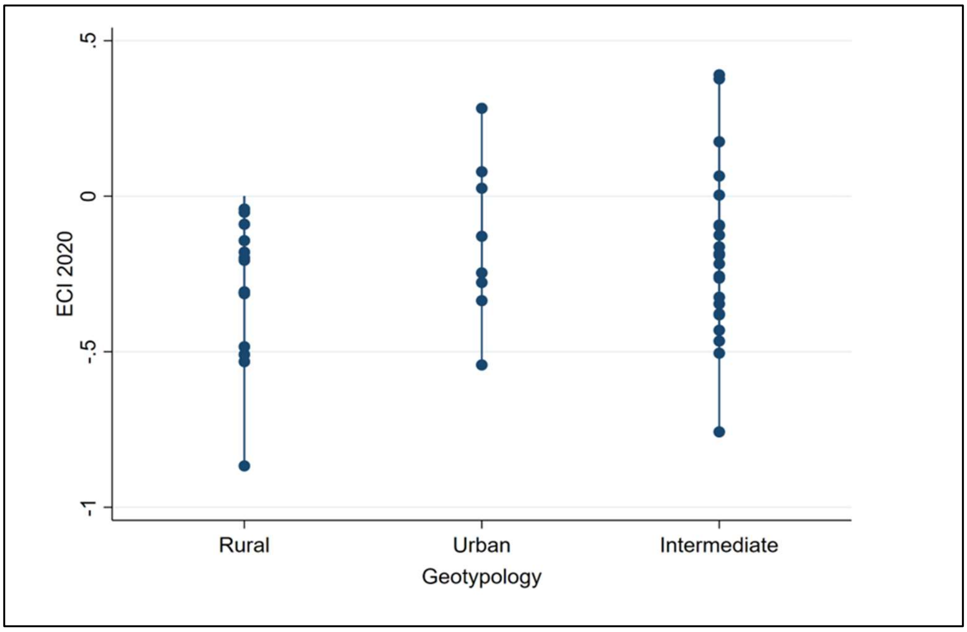


*Figure 9: Distribution of ECI 2020 values by geographic typology*

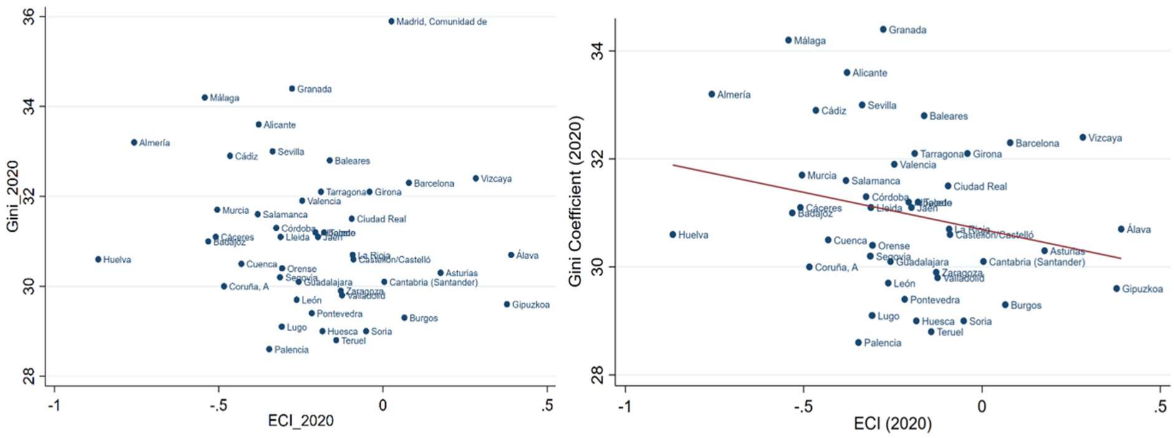


*Figure 10: Results linear regression for Spanish provinces including all observations (left) and excluding outlier (right), dependent variable: Gini coefficient, independent variable: ECI*

| Explanatory Variable | Model 1 unadjusted | Model 2 excluding outlier | Model 3 with control var. | Model 4 controlled for geographic typology Urban | Rural | Intermediate |
|---|---|---|---|---|---|---|
| **ECI** | -0.940<br>0.942 | -1.376<br>0.843 | -1.541*<br>(0.777) | -3.772*<br>(0.502) | 2.131<br>(1.529) | -2.328**<br>(1.006) |
| **GDP per capita** | | | 0.000<br>(0.000) | 0.001*<br>(0.000) | 0.000<br>(0.000) | -0.000<br>(0.000) |
| **Tourism** | | | 0.174**<br>(0.084) | -0.146<br>(0.029) | 0.425<br>(0.237) | 0.232<br>(0.155) |
| **Unemployment rate** | | | 0.156**<br>(0.070) | 1.055*<br>(0.119) | 0.148<br>(0.098) | 0.146<br>(0.121) |
| **Service employment** | | | 0.1*<br>(0.050) | 1.125*<br>(0.118) | 0.086<br>(0.081) | 0.120<br>(0.084) |
| **Industry employment** | | | 0.020<br>(0.058) | 0.87*<br>(0.121) | -0.018<br>(0.109) | 0.116<br>(0.087) |
| **Population variance** | | | 0.535<br>(0.476) | 2.64<br>(9.595) | 1.162<br>(0.63) | 1.142<br>(0.753) |
| Adj. R-squared | -0.0001 | 0.037 | 0.463 | 0.267 | 0.3826 | 0.454 |

****p < 0.01 **p < 0.05 *p < 0.1*
*Standard errors in parentheses, 46 observations*
*note: model 3 and 4 also exclude the outlier*

*Table 8: Four-step regression analysis, dependent variable. Gini coefficient*

## 6. Discussion

The different parts of the regression analysis give insights into the varying relationships between economic complexity and income inequality, depending on whether the focus lies upon inequalities between individuals nationally, between regions or between individuals within regions. Part I on economic complexity and national income inequality across the OECD countries shows a clear relationship between economic complexity and lower income inequality for the cross-sectional analysis. This holds even when controlling for development and economic indicators such as income level, education or unemployment. This is in line with most findings in the literature, such as the research by Hartmann et al. (2017), Lee & Wang (2012), and Hartmann & Pinheiro (2022). While this study shows a clear relationship, this does not necessarily imply causation. As Hartmann et al. (2017) state, the association of economic complexity with income inequality is likely to be much more complex. A highly complex product mix in a location requires a high level of knowledge in the economy (Hidalgo & Hausmann, 2009), which means that a high percentage of the population must be well-educated. It can therefore be assumed that economic complexity is most likely associated with a certain set of institutions fostering education and inclusive growth, as argued by Acemoglu & Robinson (2012). Following their argument, structures and institutions that foster inclusiveness and economic and social mobility will allow for the efficient growth and use of knowledge in society. The results in Part I show that GNI per capita is also correlated with income inequality, even more so than the ECI when only controlling for economic indicators. However, when including both development and economic indicators, the ECI is the only variable that correlates with income inequality, and income is no longer correlated. Thus, it can be argued that, unlike Kuznets' (1955) claims, income levels are not the only indicator of inequality, and economic complexity seems to be correlated with the level of inequality, at least across the sample of OECD countries. As mentioned before, these countries are relatively homogenous, with similar economic and political structures, which makes them comparable but also inhibits this study from drawing general conclusions for all countries.

No significant relationship could be found for Part II for between economic complexity and the level of interregional inequality across the OECD countries. The literature presents strong theoretical arguments that higher economic complexity could be correlated to higher levels of between-regional inequalities due to clustering and agglomeration effects of highly complex industries (Hartmann & Pinheiro, 2022; Krugman, 2019; Porter, 2003). Empirically, this effect could not be found in this study, which suggests that economic complexity is not correlated with interregional income inequality for the sample of the OECD countries. However, it might be the case that the cross-sectional analysis does not show such dynamics or that the Theil entropy index is not the optimal measurement to detect these trends. The lack of literature on the topic indicates that scholars have not found a correlation either or that measuring interregional inequalities brings many limitations and difficulties, making it difficult to make conclusions. It might, however, also simply be the case that there is no correlation and that a country's economic complexity level does not correlate with the level of interregional inequality.

However, the findings of Part III indicate that such general conclusions of no correlation should be treated with caution. The analysis for the Spanish provinces shows that higher economic complexity is strongly associated with higher income levels in a region. This suggests the existence of industrial clusters that generate higher income in certain regions. Therefore, it could be the case that economic complexity creates regional disparities within a country, as could be argued for Spain from these findings, albeit that these disparities cannot be shown in an across-country study. Stojkoski & Kocarev (2017) argue that comparing countries with regard to their level of regional inequality is extremely difficult as many country-specific factors cannot be accounted for in the analysis.

Thus, while it can be concluded that economic complexity correlates with higher income levels on a regional level, it remains unclear whether it actually amplifies regional disparities. However, following Hidalgo & Hausmann's (2009) theory that the ECI is a reliable indicator of economic growth, divergence cannot be ruled out as a possible implication of increasing economic complexity in the future if other regions do not catch up with increasing their regional complexity. Gbohoui et al. (2019) find that during the last two decades, growth across OECD regions has been much higher in regions with higher GDP levels. While this is a highly speculative assumption, the relationship between GDP and economic complexity indicates there is a possibility of divergence between regions if the levels of economic complexity differ widely between them.

The relationship between economic complexity and intraregional income inequality, as analysed in Part III, is slightly more apparent. Economic complexity seems to be related positively to income equality for the Spanish provinces, although only weakly, when controlling for socioeconomic factors such as unemployment, GDP, or population variances. This is in line with the findings of Marco et al. (2022), Bandeira Morais et al. (2021), Chávez et al. (2017), Török et al. (2022), who have all recently tested the association between economic complexity and intraregional inequality for different countries. All found that higher economic complexity negatively correlates with intraregional income inequality. Notably, in this study, economic complexity shows a correlation with intraregional inequality, while GDP seems not to be associated on a statistically significant level. This shows once more that the economy's structure seems more relevant to the inequality levels in regions rather than the income level.

Notably, when estimating models separately by geographic typologies, it becomes evident that, especially for intermediate regions (neither predominantly urban nor predominantly rural), the ECI is strongly negatively associated with income inequality. Urban regions also show a similar correlation, while inequalities in rural regions do not seem to be affected by the level of economic complexity. The reasons for this might be related to the fact that the ECI is generally higher for urban regions than for rural regions, with intermediate regions lying primarily in between. Further research must be conducted to evaluate the reasons for these relationships in detail.

In general, it can be concluded that inequalities between individuals, as measured with the Gini coefficient in this study, are negatively associated with economic complexity on both the country and regional level. Thus, higher economic complexity levels seem to have a positive association with income equality, presenting a win-win situation, as higher economic complexity levels are arguably favourable for economic development and income equality. For the association between economic complexity and interregional income inequality, the results are unclear and need to be further investigated with more complex methods and ideally time-series or panel data to see an effect over time.

## 6.1 Limitations of the Study

There are several limitations to this study. It must be noted that the thesis uses cross-sectional data, thus, it looks at the association between economic complexity and income inequality at a specific point in time, which limits the reliability of the results for general conclusions. It should also be noted that this thesis aims to explain the relationship between economic complexity and income inequality mainly in a statistical sense, therefore, the reasons why economic complexity is associated with more or less inequality us only discussed briefly. The main contribution of this thesis is to explore the relationships between the variables across different geographical scopes and entities, while the underlying causes are left to further research.

Additionally, there are limitations to the indicators used in the analysis. Firstly, economic complexity is a relatively new concept, and its degree of measurement is limited for different countries. Depending on how accurate and transparent the countries provide data about their production and exports, it might be biased. The availability of trade data and the classification of complex products can also be inaccurate. Secondly, measuring national and intraregional inequality with the Gini coefficient is widely used and the best method available considering the scope of the research. However, while it gives a good idea of how unequal income is distributed in a country, the index does not provide any insights into how income is distributed across the deciles. Thus, similar Gini values can come from very different distribution patterns.

Similarly, measuring interregional inequality with the Theil entropy index as in Part II has several limitations. While it is an officially used measurement by, for instance, the OECD, it has many shortcomings, especially when comparing regional disparities between countries. The index does not use populated or size-weighted values, which results in some differences due to the sizes of regions of the different countries. Stojkoski & Kocarev (2017) argue that valuing each country as equal when comparing them bears problems with it as each country has individual characteristics and structures that affect the prevalence and effects of economic complexity.

Lastly, in Part III, the dependent and independent variables use data from 2020, while the control variables are from 2019. This is because data for the main variables were only available for 2020, however, it was decided to use the control variables with data from 2019 to circumvent the fluctuations in data due to the effects of the global pandemic. After careful consideration, it was decided to use these different years for the variables, as the main variables show almost no fluctuation due to the pandemic, while the control variables did. Therefore, an analysis using these different years was judged to be most representative of the general relationship between them.

# 7. Conclusion and Policy Implications

This study identifies the association between economic complexity and income inequality across countries and regions of the OECD. The results show that economic complexity negatively correlates with income inequality on the individual level, thus between individuals nationally across the OECD countries and between individuals within Spanish regions. While it should be stressed that this study can only conclude that there is a statistical association and not a causation, the results have important policy implications. As the productive structure shows a positive association with equality and income levels, policymakers should aim at increasing and diversifying the product mix of economies in accordance with the institutions and structures that allow knowledge to evolve and spread.

The results regarding the association between economic complexity and income inequality across regions are less clear. For economic complexity and interregional inequality across the OECD countries, there is no statistical relationship found. However, economic complexity and income levels are shown to be highly correlated across Spanish Provinces, which implies that there might be a dynamic of divergence that could not be explored by the method or data used in this study. Policymakers should, therefore, bear in mind that the productive structure can be strongly related to income levels, as shown for the Spanish provinces. The focus should thus lie on the diversification and increase of complexity for regions that perform relatively poorly to prevent these regions from falling behind.

For intraregional inequality and economic complexity in Spanish provinces, the results suggest that the ECI has different associations with inequality depending on whether a region is predominantly rural, urban, or mixed. The differing effect of economic complexity on inequality in urban and rural areas is important for policymakers to consider. Further research should evaluate whether different industrial policies should be implemented depending on the geographic typology of a region.

This study serves as an overview of the statistical relationships between economic complexity and inequalities across different macro-levels – from national, over interregional, to intraregional. It shows that the recent concept of economic complexity might be an additional variable that can help solve the long-disputed economic riddle of what determines the levels of inequality across nations and regions. However, the results also clearly show that the relationship between economic complexity and regional disparities is difficult to measure and must be further explored.

What can be concluded from this study is that economic complexity is associated with income inequality across individuals and that it seems to be a better indicator of inequality patterns across the OECD than aggregated income measures such as GDP or GNI. Therefore, rather than only focusing on GDP growth, policies should aim to improve a location's productive structure through investments in human capital and industrial policy. Future research should focus not only on the statistical relationship but also on the qualitative aspects to fully understand how the economy's productive structure can foster economic development while increasing equality in society. As policymakers rapidly adopt the economic complexity framework for development and growth policies, exploring and understanding the association between economic complexity and income inequality is vital to create equal and prospering economies and societies.

# 8. References

Acemoglu, D. & Robinson, J. (2012). Why Nations Fail: The Origins of Power, Prosperity and Poverty, New York: Crown

Arestis, P. & Gonzalez-Martinez, A. R. (2016). House Prices and Current Account Imbalances in OECD Countries, *International Journal of Finance & Economics*, vol. 21, no. 1, pp.58–74, Available Online: https://doi.org/10.1002/ijfe.1523 [Accessed 15 May 2023]

Artelaris & Petrakos (2014) Intraregional Spatial Inequalities and Regional Income Level in the European Union: Beyond the Inverted-U Hypothesis, *International Regional Science Review*, vol. 39, no. 3, Available online: https://doi.org/10.1177/0160017614532652 [Accessed 15 May 2023]

Balassa, B. (1964). The Purchasing-Power Parity Doctrine: A Reappraisal, *Journal of Political Economy*, vol. 72, no. 6, pp.584–596, Available Online: https://www.journals.uchicago.edu/doi/abs/10.1086/258965?journalCode=jpe [Accessed 13 May 2023]

Balland, P.-A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N. & Rigby, D. (2022). The New Paradigm of Economic Complexity, *Research Policy*, vol. 51, no. 3, p.104450, Available Online: https://doi.org/10.1016/j.respol.2021.104450 [Accessed 13 May 2023]

Bandeira Morais, M., Swart, J. & Jordaan, J. A. (2021). Economic Complexity and Inequality: Does Regional Productive Structure Affect Income Inequality in Brazilian States?, *Sustainability*, vol. 13, no. 2, p.1006, Available Online: https://doi.org/10.3390/su13021006 [Accessed 13 May 2023]

Barro, R. J. & Sala-i-Martin, X. (1991). Convergence across States and Regions, *Brookings Papers on Economic Activity*, vol. 22, no. 1, pp.107–182, Available online: https://www.brookings.edu/wp-content/uploads/1991/01/1991a_bpea_barro_salaimartin_blanchard_hall.pdf [Accessed 13 May 2023]

Beinhocker, E. D. (2006). The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics, Boston Massachusetts: Harvard Business Press

Bogin, B., Scheffler, C. & Hermanussen, M. (2017). Global Effects of Income and Income Inequality on Adult Height and Sexual Dimorphism in Height: BOGIN et Al., *American Journal* *of Human Biology*, vol. 29, Available online: https://doi.org/10.1002/ajhb.22980 [Accessed 13 May 2023]

Boschma, R. A. & Lambooy, J. G. (1999). Evolutionary Economics and Economic Geography, *Journal of Evolutionary Economics*, vol. 9, no. 4, pp.411–429, Available online: https://doi.org/10.1007/s001910050089 [Accessed 15 May 2023]

Breusch, T. S. & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation, *Econometrica*, vol. 47, no. 5, pp.1287–1294, Available online: https://doi.org/10.2307/1911963 [Accessed 15 May 2023]

Burniaux, J.-M., Padrini, F. & Brandt, N. (2006). Labour Market Performance, Income Inequality and Poverty in OECD Countries, Paris: OECD, Available Online: https://www.oecd-ilibrary.org/economics/labour-market-performance-income-inequality-and-poverty-in-oecd-countries_882154447387 [Accessed 15 May 2023]

Canberra Group (2011). Handbook on Household Income Statistics, Second Edition, Available Online: https://unece.org/fileadmin/DAM/stats/groups/cgh/Canbera_Handbook_2011_WEB.pdf [Accessed 14 May 2023]

Chávez, J., Mosqueda, M. & Gómez-Zaldívar, M. (2017). Economic Complexity and Regional Growth Performance: Evidence from the Mexican Economy, *Review of Regional Studies*, vol. 47, pp.201–219, Available online: https://doi.org/10.52324/001c.8023 [Accessed 15 May 2023]

Chu, L. K. & Hoang, D. P. (2020). How Does Economic Complexity Influence Income Inequality? New Evidence from International Data, *Economic Analysis and Policy*, vol. 68, pp.44–57, Available online: https://doi.org/10.1016/j.eap.2020.08.004 [Accessed 15 May 2023]

Eurostat (2018). Archive: Urban-rural typology, *Eurostat*, Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Urban-rural_typology#Database [Accessed 8 May 2023]

Fisch, O. (1984). Regional Income Inequality and Economic Development, *Regional Science and Urban Economics*, vol. 14, no. 1, pp.89–111, Available online: https://doi.org/10.1016/0166-0462(84)90045-0 [Accessed 8 May 2023]

Gbohoui, W., Lam, W. & Lledo, V. (2019). The Great Divide: Regional Inequality and Fiscal Policy, *IMF Working Papers*, vol. 19, Available online: https://doi.org/10.5089/9781498311625.001 [Accessed 12 May 2023]

George, G., McGahan, A. M. & Prabhu, J. (2012). Innovation for Inclusive Growth: Towards a Theoretical Framework and a Research Agenda, *Journal of Management Studies*, vol. 49, no. 4, pp.661–683, Available online: https://doi.org/10.1111/j.1467-6486.2012.01048.x [Accessed 10 May 2023]

Glaeser, E. L., Resseger, M. & Tobio, K. (2009). Inequality in Cities*, *Journal of Regional Science*, vol. 49, no. 4, pp.617–646, Available online: https://doi.org/10.1111/j.1467-9787.2009.00627.x [Accessed 10 May 2023]

Global Data Lab (2023) Subnational HDI (v7.0), Available online: https://globaldatalab.org/shdi/table/ [Accessed 15 May 2023]

Goldin, C. & Katz, L. (2009). The Future of Inequality: The Other Reason Education Matters So Much, Scholarly Articles, Harvard University Department of Economics, Available Online: https://econpapers.repec.org/paper/hrvfaseco/4341691.htm [Accessed 15 May 2023]

Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M. & Hidalgo, C. A. (2017). Linking Economic Complexity, Institutions, and Income Inequality, *World Development*, vol. 93, pp.75–93, Available online: https://doi.org/10.1016/j.worlddev.2016.12.020 [Accessed 15 May 2023]

Hartmann, D. & Pinheiro, F. L. (2022). Economic Complexity and Inequality at the National and Regional Level, *Handbook of Complexity Economics*, Available Online: http://arxiv.org/abs/2206.00818 [Accessed 14 May 2023]

Hausmann, R., Hidalgo, C., A., Coscia, S., B., M., Simoes, A., Yıldırım, M., A. (2013) The Atlas of Economic Complexity: Mapping Paths to Prosperity, Updated edition., Cambridge, MA: The MIT Press, Available online: https://growthlab.cid.harvard.edu/files/growthlab/files/atlas_2013_part1.pdf [Accessed 15 May 2023]

Hidalgo, C. A. (2021). Economic Complexity Theory and Applications, *Nature Reviews Physics*, vol. 3, no. 2, pp.92–113, Available online: https://doi.org/10.1038/s42254-020-00275-1 [Accessed 15 May 2023]

Hidalgo, C. A. (2022). The Policy Implications of Economic Complexity, Available online: https://arxiv.org/abs/2205.02164 [Accessed 15 May 2023]

Hidalgo, C. A. & Hausmann, R. (2009). The Building Blocks of Economic Complexity, *Proceedings of the National Academy of Sciences*, vol. 106, no. 26, pp.10570–10575, Available online: https://doi.org/10.1073/pnas.0900943106 [Accessed 15 May 2023]

INE (2023). Gini Index and Income Distribution P80/P20, Instituto Nacional de Estadística, Available online: https://www.ine.es/jaxiT3/Tabla.htm?t=37686&L=1 [Accessed 5 May 2023]

INE (2022). Aportación del turismo al PIB de la economía española: por PIB y sus componentes, valor absoluto/porcentaje/índice y period, Instituto Nacional de Estadística, Available online: https://www.ine.es/jaxi/Datos.htm?path=/t35/p011/rev19/serie/l0/&file=03001.px [Accessed 5 May 2023]

INE (2023a). P.I.B. a precios de mercado y valor añadido bruto a precios básicos por ramas de actividad: Precios corrientes por provincias y period, Instituto Nacional de Estadística, Available online: https://www.ine.es/jaxi/Datos.htm?path=/t35/p010/rev19/l1/&file=02001.px [Accessed 5 May 2023]

INE (2023b). Demografía y población, Instituto Nacional de Estadística, Available online: https://www.ine.es/dynInfo/Infografia/Territoriales/capitulo.html#!tabla [Accessed 5 May 2023]

INE (2023c). Viajeros y pernoctaciones por comunidades autónomas y provincias, Instituto Nacional de Estadística, Available online: https://www.ine.es/jaxiT3/Datos.htm?t=2074 [Accessed 5 May 2023]

INE (2023d). Tasas de actividad, paro y empleo por provincia y sexo, Instituto Nacional de Estadística, Available online: https://www.ine.es/jaxiT3/Datos.htm?t=3996 [Accessed 5 May 2023]

INE (2023e). Distribución porcentual de los activos por sector económico y provincial Instituto Nacional de Estadística, Available online: https://www.ine.es/jaxiT3/Datos.htm?t=3994 [Accessed 5 May 2023]

INE (2023f). Variación anual en la población por provincias, Instituto Nacional de Estadística, Available online: https://www.ine.es/jaxiT3/Datos.htm?t=2918 [Accessed 5 May 2023]

INE Atlas (2023). Household Income Distribution Atlas, Instituto Nacional de Estadística, Available online: https://www.ine.es/dynt3/metadatos/en/RespuestaDatos.htm?oe=30325 [Accessed 5 May 2023]

Krugman, P. (1991). Increasing Returns and Economic Geography, *Journal of Political Economy*, vol. 99, no. 3, Available online: https://pr.princeton.edu/pictures/g-k/krugman/krugman-increasing_returns_1991.pdf, [Accessed 15 March 2023]

Kutner, M. H. (ed.). (2005). Applied Linear Statistical Models, 5th ed., Boston: McGraw-Hill Irwin, Available online: https://users.stat.ufl.edu/~winner/sta4211/ALSM_5Ed_Kutner.pdf [Accessed 15 March 2023]

Kuznets, S. (1955). Economic Growth and Income Inequality, *The American Economic Review*, vol. 45, no. 1, pp.1–28, Available online: https://www.jstor.org/stable/1811581 [Accessed 15 March 2023]

Le, T. T. M., Niem, L. D. & Kim, T. (2022). Economic Complexity and Economic Development in ASEAN Countries, *International Economic Journal*, vol. 36, no. 4, pp.556–568, Available online: https://doi.org/10.1080/10168737.2022.2142643 [Accessed 15 March 2023]

Lee, C.-C. & Wang, E.-Z. (2021). Economic Complexity and Income Inequality: Does Country Risk Matter?, *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, vol. 154, no. 1, pp.35–60, Available online: https://ideas.repec.org/a/spr/soinre/v154y2021i1d10.1007_s11205-020-02543-0.html [Accessed 15 March 2023]

Lee, N., Morris, K. & Kemeny, T. (2018). Immobility and the Brexit Vote, *Cambridge Journal of Regions, Economy and Society*, vol. 11, no. 1, pp.143–163, Available online: https://eprints.lse.ac.uk/86367/1/Lee_Immobility%20and%20Brexit_2018.pdf [Accessed 15 March 2023]

Lorgelly, P. K. & Lindley, J. (2008). What Is the Relationship between Income Inequality and Health? Evidence from the BHPS, *Health Economics*, vol. 17, no. 2, pp.249–265, Available online: https://doi.org/10.1002/hec.1254 [Accessed 15 March 2023]

Marco, R., Llano, C. & Pérez-Balsalobre, S. (2022). Economic Complexity, Environmental Quality and Income Equality: A New Trilemma for Regions?, *Applied Geography*, vol. 139, p.102646, Available online: https://doi.org/10.1016/j.apgeog.2022.102646 [Accessed 15 March 2023]

OEC (2023) The Observatory of Economic Complexity, Available Online: https://oec.world/ [Accessed 14 May 2023]

OECD (2016), OECD Regions at a Glance 2016, OECD Publishing, Paris. Available online: http://dx.doi.org/10.1787/reg_glance-2016-en [Accessed 15 March 2023]

OECD Data (2023). Income inequality, Available online: https://data.oecd.org/inequality/income-inequality.htm [Accessed 15 March 2023]

OECD Data (2023a). OECD Regional Statistics, Available online: https://www.oecd-ilibrary.org/urban-rural-and-regional-development/data/oecd-regional-statistics/regional-economy_6b288ab8-en?parentId=http%3A%2F%2Finstance.metastore.ingenta.com%2Fcontent%2Fcollection%2Fregion-data-en [Accessed 15 March 2023]

OECD TUAC (2023) OECD Membership and the Values of the Organisation, Available online: https://tuac.org/news/oecd-membership-and-the-values-of-the-organisation/ [Accessed 15 March 2023]

Pickett, K. E. & Wilkinson, R. G. (2015). Income Inequality and Health: A Causal Review, *Social Science & Medicine*, vol. 128, pp.316–326, Available online: https://doi.org/10.1016/j.socscimed.2014.12.031 [Accessed 15 March 2023]

Piketty, T. (2014). Capital in the Twenty-First Century, Cambridge Massachusetts: Harvard University Press

Piketty, T. & Saez, E. (2014). Inequality in the Long Run, *Science*, vol. 344, no. 6186, pp.838–843 Available online: https://www.science.org/doi/10.1126/science.1251936 [Accessed 15 March 2023]

Pinheiro, F. L., Balland, P.-A., Boschma, R. & Hartmann, D. (2022). The Dark Side of the Geography of Innovation: Relatedness, Complexity and Regional Inequality in Europe, *Regional Studies*, vol. 0, no. 0, pp.1–16 Available online: https://doi.org/10.1080/00343404.2022.2106362 [Accessed 15 March 2023]

Porter, M. (2003). The Economic Performance of Regions, *Regional Studies*, vol. 37, no. 6–7, pp.549–578 Available online: https://doi.org/10.1080/0034340032000108688 [Accessed 15 March 2023]

Prebisch, R. (1950). The Economic Development of Latin America and Its Principal Problems, [e-journal], Available Online: https://repositorio.cepal.org/handle/11362/29973 [Accessed 14 May 2023]

Ricardo, D. (1817). On the Principles of Political Economy and Taxation, Kitchener, Ontario: Batoche Books, Available online: https://socialsciences.mcmaster.ca/econ/ugcm/3ll3/ricardo/Principles.pdf [Accessed 15 March 2023]

Rodríguez-Pose, A. & Tselios, V. (2009). Education and Income Inequality in the Regions of the European Union*, *Journal of Regional Science*, vol. 49, no. 3, pp.411–437 Available online: https://doi.org/10.1111/j.1467-9787.2008.00602.x [Accessed 15 March 2023]

Romer, P. M. (1990). Endogenous Technological Change, *Journal of Political Economy*, vol. 98, no. 5, pp.S71–S102 Available online: https://www.jstor.org/stable/2937632 [Accessed 15 March 2023]

Rosenstein-Rodan, P. N. (1943). Problems of Industrialisation of Eastern and South-Eastern Europe, *The Economic Journal*, vol. 53, no. 210/211, pp.202–211 Available online: https://doi.org/10.2307/2226317 [Accessed 15 March 2023]

Schumpeter, J. A. (1942). Capitalism, Socialism, and Democracy, SSRN Scholarly Paper, 1496200, Available Online: https://papers.ssrn.com/abstract=1496200 [Accessed 14 May 2023]

Sciarra, C., Chiarotti, G., Ridolfi, L. & Laio, F. (2020). Reconciling Contrasting Views on Economic Complexity, 1, *Nature Communications*, vol. 11, no. 1, p.3352 Available online: https://www.nature.com/articles/s41467-020-16992-1 [Accessed 15 March 2023]

Sheng, Y. (2011). Unemployment and Income Inequality: A Puzzling Finding from the US in 1941-2010, *SSRN Scholarly Paper*, 2020744, Available Online: https://papers.ssrn.com/abstract=2020744 [Accessed 15 May 2023]

Singer, H. W. (1950). The Distribution of Gains between Investing and Borrowing Countries, *The American Economic Review*, vol. 40, no. 2, pp.473–485 Available online: https://www.jstor.org/stable/1818065 [Accessed 15 March 2023]

Smith, A, (1776). The Wealth of Nations . Oxford, England,
Stiglitz, J. E. (2016). Inequality and Economic Growth, pp.134–155, *The Political Quarterly*, Available online: https://academiccommons.columbia.edu/doi/10.7916/d8-gjpw-1v31 [Accessed 15 March 2023]

Stiglitz, J. E., Fitoussi, J., Durand, M. (2018) Chapter 3. The need to follow up on the Stiglitz-Sen-Fitoussi Commission, *Beyond GDP, OECD*, Available online: https://www.oecd-ilibrary.org/sites/9789264307292-6-en/index.html?itemId=/content/component/9789264307292-6-en [Accessed 15 March 2023]

Stojkoski, V. & Kocarev, L. (2017). The Relationship Between Growth and Economic Complexity: Evidence from Southeastern and Central Europe, *MPRA Paper,* Available Online: https://mpra.ub.uni-muenchen.de/87432/ [Accessed 15 May 2023]

Török, I., Benedek, J. & Gómez-Zaldívar, M. (2022). Quantifying Subnational Economic Complexity: Evidence from Romania, 17, *Sustainability*, vol. 14, no. 17, p.10586 Available online: https://www.mdpi.com/2071-1050/14/17/10586 [Accessed 15 March 2023]

Utkovski, Z., Pradier, M. F., Stojkoski, V., Perez-Cruz, F. & Kocarev, L. (2018). Economic Complexity Unfolded: Interpretable Model for the Productive Structure of Economies, *PLOS ONE*, vol. 13, no. 8, p.e0200822 Available online: https://doi.org/10.1371/journal.pone.0200822 [Accessed 15 March 2023]

Williamson, J. G., (1965). Regional Inequality and the Process of National Development: A Description of the Patterns, *Economic Development and Cultural Change*, vol. 13, no. 4, Available online: https://www.jstor.org/stable/1152097 [Accessed May 3 2023]

World Bank (2023) GINI Index for Colombia [SIPOVGINICOL], retrieved from FRED, Federal Reserve Bank of St. Louis, Available online: https://fred.stlouisfed.org/series/SIPOVGINICOL [Accessed May 3 2023]

# 9. Appendix

## 9.1 Multicollinearity Test Results

| Variable | VIF for Table 5 | VIF for Table 6 | VIF for Table 7 | VIF for Table 8 |
|---|---|---|---|---|
| Income | 18399.06* | 22446.06* | | |
| GNI per capita | 18380.79* | 22346.13* | | |
| Education | 3.20 | 3.37 | | |
| ECI | 2.22 | 2.05 | 2.63 | 1.55 |
| Unemployment rate | 1.78 | 1.85 | 2.68 | 4.61 |
| Economic growth rate | 1.41 | 1.71 | | |
| Industry Employment | 1.11 | 1.07 | 4.43 | 3.21 |
| Service Employment | | | 3.06 | 2.69 |
| Tourism | | | 1.93 | 2.24 |
| Population variation | | | 1.3 | 1.69 |
| GDP per capita | | | | 5.41 |

The test results for multicollinearity indicates that the regression in the Tables 5 and 6 have a multicollinearity problem when all control variables are included. This is determined by the VIF value of Income and GNI per capita that is extremely high and far above the critical value 10.

* VIF > 10

## 9.2 Breusch-Pagan Test for Heteroskedasticity Results

The results for the Breusch-Pagan test for heteroskedasticity show that Model 2 and Model 4 of the regression model in Table 5 suffer from heteroskedasticity, that is, the P value (Prob > chi2) is below the critical value 0.05. Therefore robust standard error are used in these regression models. Otherwise, non of the models shows problems with heteroskedasticity, thus normal standard errors are used in the regression analysis of all other models.

Breusch-Pagan Test for Heteroskedasticity for Table 5

| Model | Model 1 unadjusted | Model 2 development control var. | Model 3 economic control var. | Model 4 all control var. |
|---|---|---|---|---|
| Prob > chi2 | 0.225 | 0.011 | 0.148 | 0.001 |

*Null hypothesis: Constant variance (homoskedasticity)*

Breusch-Pagan Test for Heteroskedasticity for Table 6

| Model | Model 1 unadjusted | Model 2 development control var. | Model 3 economic control var. | Model 4 all control var. |
|---|---|---|---|---|
| Prob > chi2 | 0.075 | 0.778 | 0.735 | 0.984 |

*Null hypothesis: Constant variance (homoskedasticity)*

Breusch-Pagan Test for Heteroskedasticity for Table 7

| Model | Model 1 unadjusted | Model 2 with control var. |
|---|---|---|
| Prob > chi2 | 0.648 | 0.814 |

*Null hypothesis: Constant variance (homoskedasticity)*

Breusch-Pagan Test for Heteroskedasticity for Table 8

| Model | Model 1 | Model 2 | Model 3 | Model 4 controlled for geographic typology | | |
|---|---|---|---|---|---|---|
| | unadjusted | excluding outlier | with control var. | Urban | Rural | Intermediate |
| Prob > chi2 | 0.705 | 0.705 | 0.196 | 0.675 | 0.415 | 0.081 |

*Null hypothesis: Constant variance (homoskedasticity)*