UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M. FANNO"

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

"Economic Complexity, institutions, and income inequality: evidence from Italian regions "

RELATORE:

CH.MO PROF. Roberto Antonietti

LAUREANDO: David Fanton

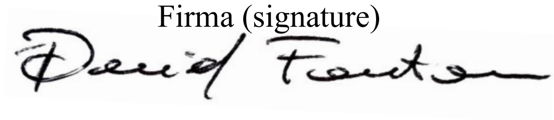
MATRICOLA N. 2022907

ANNO ACCADEMICO 2022 – 2023

Dichiaro di aver preso visione del "Regolamento antiplagio" approvato dal Consiglio del Dipartimento di Scienze Economiche e Aziendali e, consapevole delle conseguenze derivanti da dichiarazioni mendaci, dichiaro che il presente lavoro non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Dichiaro inoltre che tutte le fonti utilizzate per la realizzazione del presente lavoro, inclusi i materiali digitali, sono state correttamente citate nel corpo del testo e nella sezione 'Riferimenti bibliografici'.

I hereby declare that I have read and understood the "Anti-plagiarism rules and regulations" approved by the Council of the Department of Economics and Management and I am aware of the consequences of making false statements. I declare that this piece of work has not been previously submitted – either fully or partially – for fulfilling the requirements of an academic degree, whether in Italy or abroad. Furthermore, I declare that the references used for this work – including the digital materials – have been appropriately cited and acknowledged in the text and in the section 'References'.

Firma (signature)



INDEX:

INTRODUCTION

The aim of this thesis is to investigate the relationship between income inequality and economic complexity, mediated by institutions, looking at the relationship between the different regions of Italy, carrying out a within country analysis.

In the last twenty years there has been a growing interest in the study of economic complexity: the production of scientific articles that contains "economic complexity" in the title, abstract or keywords grew from 952 in 2002 to more than 16.000 in 2021. Several aspects have been analysed and studied, such as the relationship between economic complexity and GDP, economic growth, sustainability and inequality (Hidalgo, 2021).

However, there is an open debate on the relationship between economic complexity and income inequality. Indeed, if we look at the relationship between countries, there is a strong consensus in stating that higher levels of economic complexity are associated with lower levels of income inequality. On the other hand, a within country analysis does not show the same consensus between economists. This work aims to present a within country analysis at a regional level for Italy. This specific regional level empirical analysis, to our knowledge, has never been carried out before in Italy. However, there are some reference papers that have carried out a similar analysis in other countries such as Brazil, US and Mexico.

The work is organized as follows. In **chapter 1** we start explaining the origin of economic complexity from an historical point of view. Then we move to explain the general concept and the theoretical explanation of why this measure is important and why it has received an increasing attention from researchers. Thus, we explain how the Economic Complexity Index (ECI) is calculated following mainly the Hidalgo and Hausmann (2009) notation and steps. Consequently, we move to look at the main drivers and factors that affects economic complexity, quickly presenting the main applications that can be found in the literature. Therefore, we explain the relationship that links Economic complexity and institutions, looking at growth theory models and to institutional theory. Finally, we look at the relationship that links Economic complexity and GDP, both from a theoretical and an empirical perspective.

In **chapter 2**, we will briefly analyse the concept and data of income inequality. First, we present different measures of wage dispersion. Then we focus the analysis on a global level looking at data and drivers of income inequality. The literature divides them in two main categories, those that depend on the market conditions and those that depend on labour institutions. Finally, we look at data and drivers in Italy, highlighting the regional differences and the reasons why we observe them.

In **chapter 3**, we will analyse the relationship between income inequality and economic complexity starting from the Kuznets (1955) curve that links inequality to economic growth. Indeed, many authors studied the relationship between GDP growth and income inequality finding different results. In fact, we will present studies stating that this relationship is negative, others that is positive, others that is inverted U-shaped and others that do not find any kind of relationship. There will be presented also the different theoretical explanations of these results. Then, we will describe which are the findings of different studies in analysing the relationship between income inequality and economic complexity. This review will start looking at the within countries link, that shows a consensus in establish that higher economic complexity leads to lower inequality. However, examining the within country relationship, the results are contradictory: some research found positive relationships, other negative. In both the within country and between country will be presented the related theoretical explanations that the authors pointed out.

Finally, in **chapter 4** we present our dataset, explaining the sources of the data and illustrating some core information of the variables used in our models and their distributions. Moreover, we present some graphical representation of the data highlining some key characteristic of them, above all the North/South gap. Then, we present our estimation strategy describing the reasons why we chose a Fixed-Effect model with Driscoll and Kraay's covariance matrix estimator and an Instrumental Variable strategy. We estimated our model by first considering all the regions and then dividing them into geographical areas (North-West, North-East, Centre, South). Finally, we present our results: we register that, by considering all the regions, higher values of economic complexity lead to higher levels of income inequality. However, North-Eastern regions follows the opposite path, so higher economic complexity leads to lower inequality. We explain this result by looking at two main indicators that differentiate the N-E from other areas: institutional quality and technological entropy.

CHAPTER 1 - ECONOMIC COMPLEXITY

1.1 ORIGIN

The origin of "economic complexity" comes from the 21st century and it is based on the discussion of complex systems in natural science, such as mathematics and physics. Indeed, the link between complexity and economics must be sought in the "Santa Fe Perspective", a group of scientists who were working on the Economics Program at the Santa Fe Institute for the Study of Complex Systems (Fontana, 2010). This group aimed to apply complex systems tools to economics.

Therefore, to understand what economic complexity is, we should start looking at the definition of a complex system. One of the first definitions made in the perimeter of the Santa Fe Perspective, was formulated by Cowan and Feldman in 1986, which describes *Complex adaptive systems* as:

> "[…] systems comprising large numbers of elements the properties of which are modifiable as a result of environmental interaction [...] Complex adaptive systems process information and can modify their internal organization in response to such information. In general, complex adaptive systems are highly nonlinear and are organized on many spatial and temporal scales."
>
> (Cowan and Feldman, 1986 as cited in Fontana, 2010)

Certainly, Santa Fe's agenda had the ambition to change the neoclassical economic paradigm in favour of describing economic phenomena and their dynamic processes through the mathematics of stochastic processed computer simulation. In other words, there was a strong interest in using more sophisticated tools to apply in economics, taking knowledge from different fields of science.

The neoclassical theory uses mathematical tools to solve linear and homogeneous maximization problems, taking into account axioms, proofs, theorems, and so on. Complexity economics does not need strong assumptions, because complexity methods do not require knowing specifically the nature or composition of every single factor at work. Economic complexity can be used to estimate the effects of different inputs, without knowing their nature or their dimensions. If traditional economics needs to assume the nature of these inputs, complexity economics can be used to learn factors directly from the data (Hidalgo, 2021).

Using the methods of complexity economics, Hidalgo and Hausmann (2009) have built a new view of economic growth and development theory, exploiting trade data. One of the first explanations of the development of nations comes from Adam Smith (1776) who elaborates on the idea that the division of labour and the division of knowledge can increase efficiency and thus productivity through specialization. The limit identified by Smith is the dimension of the market itself. However, in an increasingly globalized world where the input and output markets are larger and larger, we should not see a relevant difference in GDP per capita, because countries can exploit a world-scale specialization. More recently, the trade literature focuses mainly on the Heckscher-Ohlin model: it suggests that countries should specialize in those activities that intensively use the resources where it has a relative advantage. However, not even this model finds an explanation for why we see a divergence in growth (Ourens, 2012). According to Hidalgo and Hausmann (2009) and Acemoglu and Robinson (2012), the reasons why we see these differences should be imputed to the existence of not tradable capabilities that are institutions, regulations, and property rights, which are key determinants in the development of countries. If the latter authors identify the reasons for these divergences mainly in the institutional framework, differentiating into "inclusive" and "extractive" economic institutions, the former authors do not specifically identify the reasons, speaking only of "nontradable capabilities".

1.2 THEORY

Nevertheless, the economic complexity approach allows to retrieve indirect measures of these "nontradable capabilities" and this is exactly the novelty of this approach. To better explain the underlying idea, it is useful to use the same analogy used by the authors: the Lego analogy. Think of a country or region that produces different outputs (e.g. cars, computers…). To produce a single output each country needs specific capabilities or resources or combinations of them. In our analogy each product is represented by a Lego model, a country or a region is symbolised by a bucket of Lego pieces while each capability is represented by a single piece of Lego. The more Lego pieces you have in the bucket, the more complex the Lego model you can build. This means that the more capabilities a country has, the more complex the product it can produce. Therefore, what the economic complexity method does, is to retrieve the diversity and exclusivity of Lego pieces by looking at Lego models, thus looking only at the final output that a country produces and trade (Hidalgo and Hausmann, 2009; Antonietti, 2022). It is important to note that this process is possible if we consider this data as a bipartite network that connects outputs to countries and considering that this bipartite network is the result of a tripartite network that links countries, capabilities, and products.

FIGURE 1: bipartite and tripartite network

Source: Hidalgo and Hausmann, 2009

With the availability of data, we can easily uncover the connections between products and countries. However, the connection between countries/outputs with capabilities is not directly observable. Nevertheless, a country can only produce a product only if it has the necessary capabilities to create it. Hence, the presence of a link in the bipartite network (which we can observe) also signals the presence of the required capabilities.

This mechanism makes it possible to calculate the complexity of a country, a region, or a province at different times, without saying anything about the procedure of collecting or developing these capabilities (Hidalgo and Hausmann, 2009). Using this outcome-based method, we can say that if two goods are related, and thus more likely to be produced in combination, it is because they need similar capabilities, thus similar institutions, knowledge, technologies... On the opposite, if two goods are very different it is more difficult that these two goods are produced in combination (Hidalgo et al., 2007). This relatedness can be measured, and the Economic Complexity Index (ECI) can be calculated.

1.3 MEASUREMENT

The first calculation of the ECI was made using trade data. However, industry data, patent data, occupational data, investment data, and other data can be used to calculate the ECI (Antonietti, 2022). Using trade data, the ECI is computed using the method of reflections. Each step in the following section follows Hidalgo and Hausmann (2009) and Hidalgo's (2021) notation proposed in their research, if it is not indicated differently.

The first step is to consider exports with a revealed comparative advantage (RCA), thus the only products considered are those that are competitive on the world market. In other words, we will consider only products in which the country considered holds a “strong” position. To do so, we need to compute the Balassa index of specialization RCA_{cp} (Balassa, 1986):

$$RCA_{cp} = \frac{x_{cp}}{\sum_p x_{cp}} \Bigg/ \frac{\sum_c x_{cp}}{\sum_{cp} x_{cp}} \tag{1}$$

Where x_{cp} is the export value of country c and product p. The Balassa index can return values greater than or less than one: if $RCA_{cp} < 1$, then country c is not specialised in the production of the product p. On the other side, if $RCA_{cp} > 1$, then country c is specialised in the production of the product p. Instead of the Balassa index, other indices can be used. Often, it is useful to normalize RCA_{cp} with population data, to remove noises that come from fluctuations in commodities prices, currency exchange rates, or seasonal employment.

The next step is to define a binary specialisation matrix M:

$$M_{cp} = \begin{cases} 1 \ \ if \ \ R_{cp} \geq R^* \\ 0 \ \ if \ \ R_{cp} < R^* \end{cases} \tag{2}$$

Where $R^* = 1$. Then, $M_{cp} = 1$ if country c has a revealed comparative advantage in the product p, $M_{cp} = 0$ otherwise. Thanks to this matrix, it is possible to derive the ubiquity, that is the number of countries that have a $RCA_{cp} > 1$, and the diversity, that is the number of products in which a country has a $RCA_{cp} > 1$. Stated differently, ubiquity is a measure of the sophistication of a product; indeed, if a product is exported from a few countries, it might mean that the capabilities needed to produce it are rare and therefore more complex and sophisticated. On the other hand, if a country has high levels of diversity, it means that it has many capabilities to produce different products. Formally:

$$M_c = \textstyle\sum_p M_{cp} = \text{diversity} \tag{3}$$

$$M_p = \textstyle\sum_c M_{cp} = \text{ubiquity} \tag{4}$$

However, we cannot use these indicators alone. For example, consider country A which is the sole exporter of a specific product, such as a mining good that is only available in A due to geographical reasons. Thus, country A will have a high level of ubiquity even if the product under consideration is very simple and not sophisticated. Now consider a country B that is the only exporter of another good, which is so sophisticated and complex to produce that it needs specific skills and capabilities that are only available in B due to investment, institutional or regulatory reasons. We cannot say that countries A and B have a similar complexity by basing our analysis only on ubiquity, and the complexity index must take these problems into account. The same story could be made for diversity: considering two countries, a large country and a small country, it is likely that the larger country will have greater diversity due only to its size. Indeed, it is mechanical that a larger country exports more different products and that it will have more products that have an RCA>1. The solution is to consider both ubiquity and diversity. To do so, the ECI is obtained by an iterative method of reflections, which means finding the eigenvalue of the following matrix:

$$\widetilde{M}_{cc'} \equiv \sum_p \frac{M_{cp}M_{c'p}}{M_c M_p} = \frac{1}{M_c}\sum_c \frac{M_{cp}M_{c'p}}{M_p} \quad (5)$$

The ECI is the second largest eigenvector K_c of the matrix M. To retrieve the final ECI, we have to make a standardisation as follows:

$$ECI_c = \frac{K_c - \overline{K}}{std(K)} \quad (6)$$

Where $\overline{K}$ is the average value of K_c and $std(K)$ is the standard deviation of K. Calculating the ECI in this way, it means that countries that have an ECI>0 are locations with a complexity larger than the average location in the dataset considered.

There are some interesting properties of the ECI to consider (Hidalgo, 2021):

- The complexity of a country does not increase until an added activity in the country exceeds the average complexity, this means that low-sophisticated goods do not increase the overall complexity.
- Countries with similar complexity values have similar specialisation models as well.
- The ECI strongly correlates with traditional measures of technology sophistication, such as R&D data or patent data. However, unlike these traditional measurements, the ECI does not need to know anything about which activities are more or less sophisticated, because the information comes directly from the data.
- The ECI does not correlate with measurements of population size, diversification, or concentration, which means that ECI does not depend on the dimension of the country considered.

Following the methodology explained before, the Atlas of Economic Complexity has built a ranking based on the ECI, of all the countries whose data are available. Here are the top five and the last five countries on the list:

TABLE 1

RANK	COUNTRY	ECI	RANK	COUNTRY	ECI
1	Japan	2.27	**129**	Nigeria	-1.73
2	Switzerland	2.14	**130**	Gabon	-1.83
3	Germany	1.96	**131**	Guinea	-1.91
4	South Korea	1.95	**132**	Liberia	-2.24
5	Singapore	1.87	**133**	Angola	-2.51

Source: Atlas of Economic Complexity on 2020 data

To make the concept clearer, it is useful to look at what these countries export, so for which products they have a RCA. For simplicity, we will look only at Japan and Angola, respectively the first and the last of the list. However, the underlying reasoning can be easily applied to the other countries. Japan exports a wide variety of products: about 20% of its exports come from the services sector, 20% from the machinery sector, 17% from the vehicle sector, 15% from the electronics sector, and 14% from the chemical sector. The remaining part of exports is divided between the metals, stones, agricultural, mineral, and textile sectors. A completely different story for Angola: more than 93% of the exports are represented by the minerals sector (of which oil accounts for more than 85% of exports) and 6% are represented by the stones sector (data comes from Atlas of Economic Complexity). Therefore, it is clear that countries with higher ECI have more products with a RCA, so they are more specialised in producing products that need many different capabilities, skills, and knowledge. In addition, they not only export more sophisticated products but also export a greater variety of products, even very different from each other.

1.4 DRIVERS AND APPLICATIONS

There have been few attempts to identify drivers and factors that affect economic complexity. These can be grouped into five main factors:

- **Institutions**: seen as the presence and quality of institutions. The theory of this aspect could be ambivalent. Indeed, good or "inclusive" institutions could lead to a higher level of economic complexity because they create the correct political, academic, and entrepreneurial environment to increase capabilities and skills; bad or "extractive" institutions may not develop those features that are necessary to create a complex economy. On the other hand, it could be the opposite, so a complex economy modifies institutions through the creation of unions or the increase of learning opportunities and the creation of more conscious workers/students/citizens. The empirical analysis found that there is a positive relationship between the availability and the quality of institutions and economic complexity (Sweet and Eterovic-Maggio, 2015; Lapatinas et al., 2019b; Antonietti and Burlina, 2022; Vu, 2022).

- **Spatial agglomeration**: seen as the possibility of creating an economy of scale in large urban areas. Indeed, complex activities require a high level of specialisation and coordination. These two tasks are easier if the people and knowledge that need to be coordinated are present in the same area. In addition, spillovers and the diffusion of knowledge are easier in large cities, due to geographical proximity, the presence of universities, and the presence of different activities, industries, and organizations. In this respect, empirical analysis has found that in metropolitan areas of the US tend to be concentrated those activities that are knowledge-intensive suggesting that spatial agglomerations (i.e. metropolitan areas, big cities…) increase the economic complexity of an area (Balland and Rigby, 2017; Balland et al.,2020)
- **Technology**: seen as the possibility of being connected via internet. If there are more people using internet it means there are more people sharing ideas and content. In addition, a high percentage of people using internet implies that the population has a good knowledge of basic technology. In this regard, empirical work has found a positive and strong relationship between internet use and economic complexity (Lapatinas, 2019).
- **Foreign direct investments**: FDIs from developed countries, that is, countries that are also technologically more advanced, can spread knowledge and capabilities in the least developed host country. This happens because MNEs, which go to the least developed country, have more advanced inputs that give to the host country, directly increasing the complexity. Indirectly, the MNEs can generate spillovers in the host economy, creating new capabilities. Here too, there are several empirical confirmations that FDIs increase economic complexity (Javorcik et al., 2018; Khan et al., 2020; Antonietti and Franco, 2021)
- **Entropy**: this concept comes from physics, and it is considered the tendency of things to break down and of particles to wiggle and rearrange themselves into new building blocks. Then, the higher level of entropy the higher the number of feasible options and combinations. In economic terms, a higher level of entropy means higher possible combinations of capabilities that translate into more complex products, increasing the

level of economic complexity of the location. Again, some empirical work has found that the availability of skills, technologies and capabilities (which in this case are represented by the level of entropy) is correlated with economic complexity. (Bahar et al., 2020; Antonietti and Burlina, 2022).

There have been several applications of economic complexity that have studied the impact on **economic growth** (Hidalgo and Hausmann, 2009; Hausmann et al., 2009; Stojkoski et al., 2016; Tacchella et al., 2018; Poncet and de Waldemar, 2013; Domini, 2019; Chávez et al., 2017), **sustainability** (Neagu and Teodoru, 2019; Can and Gozgor, 2017; Lapatinas et al., 2019a; Romero and Gramkow, 2021; Swart and Brinkmann, 2020; Dong et al., 2020; Dordmond et al., 2020), **human development and health** (Ferraz et al., 2018; Lapatinas, 2016; Vu, 2020), **income inequality** (Hartmann et al., 2017; Lee and Vu, 2019; Chu and Hoang, 2020; Marco et al., 2022; Bandeira Morais et al., 2021).

As regards the first element, the link between economic complexity and economic growth is well established: GDP per capita, long-term economic growth, and other economic indices are strongly positively correlated with economic complexity. These results are robust even after controlling for many varied factors and have been replicated for different countries, different data, and both nationally and sub-nationally. Regarding sustainability, there are some studies linking greenhouse gas emissions and green jobs to economic complexity. Indeed, a U shape relationship has been found between environmental outputs and economic complexity: green indices first deteriorate with complexity and after reaching a certain level, improve. However, this relationship turns out to be constantly negative considering the air quality. Human development and health indicators too are positively correlated to economic complexity. The relationship between economic complexity and inequality will be deeply analysed in chapter 3.

1.5 ECONOMIC COMPLEXITY AND THE ROLE OF INSTITUTIONS

It is important to emphasise the role of the institutions in the discussion on economic complexity. Indeed, as pointed out before, there is a prominent body of literature related to growth theory models that highlight the importance of good institutions for fostering economic growth and prosperity: the so-called "institutional theory". One of the first contributions come from North (1990), that defined institutions as:

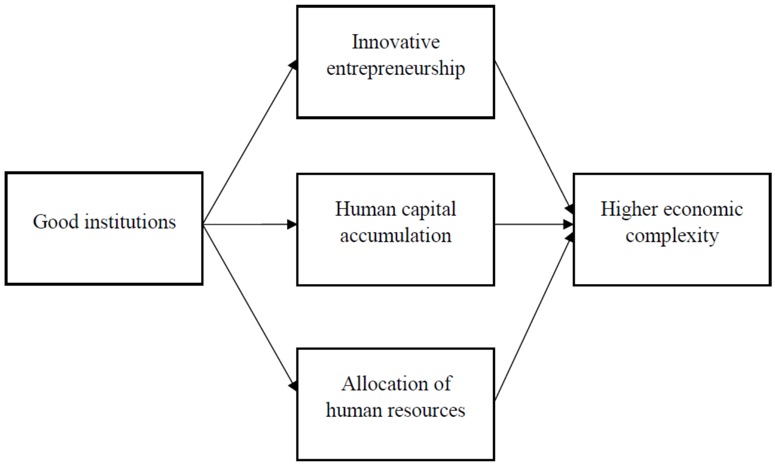
> "[…] the rules of the game in a society, or more formally, are the humanly devised constraints that shape human interaction."
>
> (North, 1990)

In his seminal contribution, North (1990) identifies both formal and informal institutions: the firsts are constituted by laws, constitutions, and regulations enforced by the authority at the different levels of power from central governments and local administrations. The latter are made up of conventions, norms of behaviour, and codes of conduct that are generally more difficult to change because the origins of these informal institutions have historical roots (North, 1990). Therefore, good institutions are those that allow for improving the economic situation of a country, increasing GDP per capita, reducing poverty and inequality. Strong property rights, competitive markets, political stability, an efficient legal system, and low corruptions are all examples of good inclusive institutions.

In this respect, many studies have empirically established that one of the main determinants to consider when looking at economic performances, is the type and quality of institutions (Vu, 2022). For example, Easterly and Levine (2003) have tested whether economic development depends on geographical endowments like the presence of disease, environmental conditions, or living in a tropical location. Indeed, we see that many countries with similar geographical endowments do not have good economic conditions. Therefore, one might think that geographical reasons are the causes. However, what they have found is that geographical endowments affect development only through institutions, without any direct effect, concluding that institutions are the true determinants of development. Rodrik et al. (2004), using an instrumental variable approach, estimated the contributions of institutions, geography, and trade in determining the level of income of different countries; the result was that institutions are the real key variable that determine the income level and geography has an indirect effect through institutions. Acemoglu and Johnson (2005), using an instrumental variable approach, found that "property right institutions", defined as protecting citizens against expropriation by the government or other dominant social classes, affect long-run economic growth, financial development, and investments. In Acemoglu et al. (2005), they describe in detail the theoretical background, several empirical estimations, and one quasi-experiment (the Korean case) to prove that differences in economic institutions are the fundamental cause of differences in economic development.

However, is this also true in the discussion of economic complexity? The main idea is that better institutions increase the economic complexity index, through higher incentives for innovative entrepreneurship, better human capital, and a more efficient allocation of resources towards more productive activities. Vu (2022) tried to test if the quality of institutions has a positive effect on economic complexity. Indeed, countries with better institutions should tend to accumulate more of the aforementioned capabilities reflected in the ability to produce and export more sophisticated products. On the other hand, countries with poor institutions should be more likely to have a less sophisticated productive structure. As said in previous sections, the types of products a country produces and exports, determine the level of economic complexity. In this line of research, Hausmann and Rodrik (2003) and Hausmann et al. (2007) emphasised the importance of the role of cost uncertainties: they theorize that economic development and the creation of new capabilities, are auto-discovery processes in which a country learns how to improve its economic structure and its productive activities. This process is partly about discovering the cost structures of the economy, so cost uncertainties play a key role in the whole process where less uncertainty allows for a better estimate of the investment returns. The most important players in this process are the pioneer investors because, thanks to their engagement in discovering new investment opportunities, they allow us to discover the cost structures. However, this activity creates positive externalities since other investors could enter the market by exploiting the pioneering discoveries of investors. Therefore, it is at this point that institutions play a key role. Indeed, well-functioning institutions, strong property rights, a good judicial system, laws, regulations, and so on, are key drivers in this auto-discovery process and a country with these characteristics can internalize these externalities. This internalization allows pioneers to protect their investments by encouraging their investment activities (Vu, 2022). Moreover, Vu (2022) stressed the connection between institutions and human capital formation: good institutions provide also a good education system and more generally better human capital (among others see Acemoglu et al., 2014; Dias et Tebaldi, 2012). Indeed, higher levels of human capital determine a higher level of economic performance and economic complexity, ceteris paribus. This happens because a more qualified and educated workforce can learn and recombine new capabilities faster and more efficiently. Vu (2022) summarizes these ideas in the following scheme:

FIGURE 2: institutions and economic complexity.



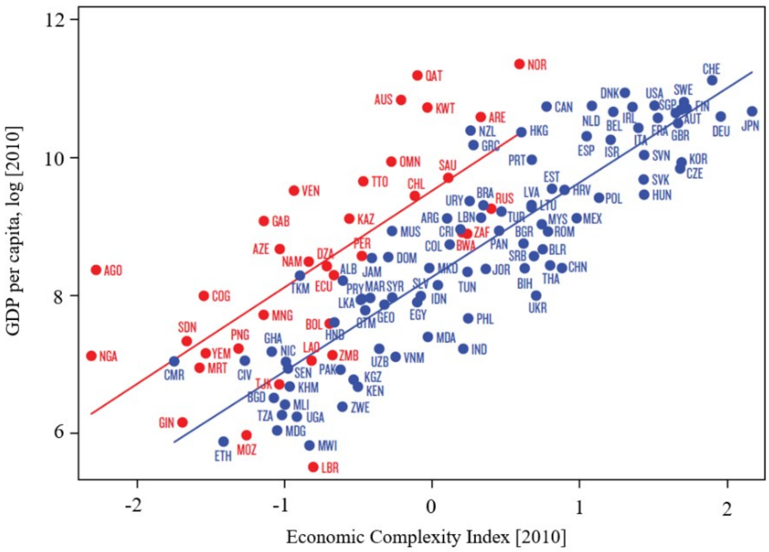
Source: Vu, 2022

These intuitions were empirically tested with different estimation strategies: OLS estimates confirmed the fundamental role of institutions in determining the economic complexity of a country. Moreover, IV estimates suggest a causal interpretation; IV estimates were conducted using two different instruments first alone and then combined (exposure to UV-R, the settlers' mortality rate) (Vu, 2022).

1.6 ECONOMIC COMPLEXITY AND GDP

As mentioned before, the connection between economic complexity and growth indicators such as GDP per capita or economic growth is strong and well documented. Indeed, the ECI is a good predictor of future economic growth and the correlation between the future level of income and the ECI is strong and positive even after controlling for many other factors (Hidalgo, 2021). Indeed, it has been found that ECI predicts long-term economic growth even in some tens of years. Economic complexity represents the number of capabilities of a country or a region. In other words, it reflects the amount of knowledge. Putting it in this way, it is more comprehensible that the amount of knowledge in a country predicts the level of future income. However, looking at the data, a slightly different story emerges: some countries with low levels of complexity have a medium or high level of income per capita. This does not invalidate the general idea that ECI is a predictor of income per capita; indeed, giving a closer look at the data, we can see that countries that have a low level of complexity but high income per capita are those countries that rely their economy on natural resources, that are the simplest products but that can generate a lot of income. The following graph shows the correlation between ECI in 2010 and GDP per capita in the same year for 128 countries:

FIGURE 3: ECI and GDP per capita

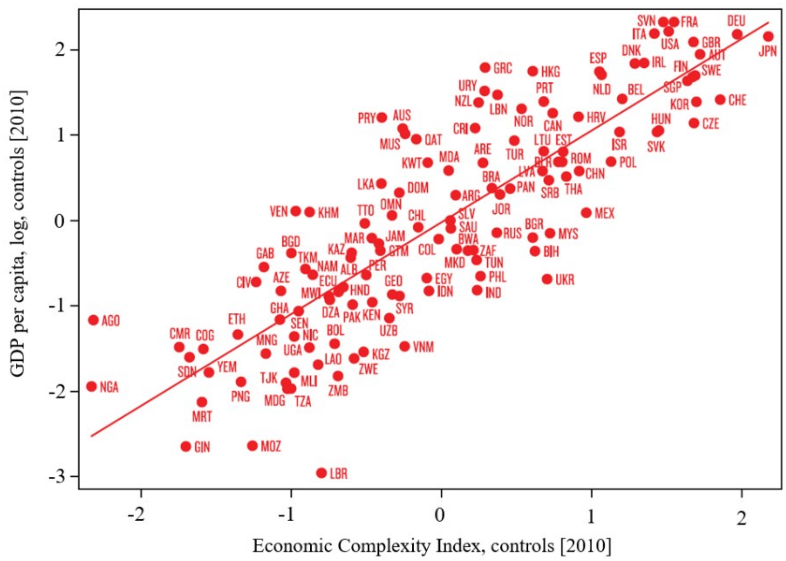


Source: Hausmann et al, 2014

In red are countries where exports of natural resources are above 10% of their total GDP, while in blue countries with limited exports of natural resources. It is easy to see that many countries are relatively rich even if they are not complex at all, look for example the case of Angola (AGO) which is the last in the economic complexity ranking, but it has an income per capita bigger than many other countries.

In the following figure, Hausmann et al. (2014) have taken into account this problem of natural resources, which has little to do with knowledge and know-how, but it is more a geographical or geological fact than an economic one.

FIGURE 4: ECI and GDP per capita after controlling for natural resources.



Source: Hausmann et al, 2014

In figure 4, the relationship, after controlling for natural resources, is quite clear: for higher levels of ECI, we see higher levels of GDP per capita. Nevertheless, this correlation is not perfect, we can indeed see some countries are above and below the red line. Considering again Angola. It has a level of GDP per capita higher than it should be considering its level of economic complexity. On the other side, if we look at India, we should expect a higher level of GDP per capita, given its level of EC.
This does not contradict the theory presented above. Indeed, in Angola's case, we can read the data as follows: the country is too rich compared to the knowledge it has, and it cannot sustain that level of income in the future so we can expect that the GDP per capita will decrease. On the contrary, India has the knowledge to become richer and, according to the view presented, India will become richer. It must be noted that the driver for more or less GDP per capita is economic complexity (Hausmann et al, 2014).

All these considerations were also proved more formally by Hausmann et al. (2014). The regression considers the country's initial level of EC, the initial level of income, and the growth rate of natural resources export to demonstrate that the EC predicts future growth. The analysis suggests that countries move their level of income following their level of know-how; in case this is not observed, as, for Angola and India, the level of income will be correct over time through higher or lower growth rates. Obviously, for countries, it is possible to acquire new capabilities to boost their production of complex products and then increase their GDP. However, this depends on how easy it is to acquire these new capabilities, and this is linked to how many capabilities close to the new capacities needed to produce the new products, the country possesses. The results of these analyses show that the Economic Complexity Index can explain more than 50% of the variance of the future ten years' growth of more than 100 countries; this is a much higher percentage than any other variable used in the past growth literature.

Felipe et al. (2012), using and adapting the methodology described in section 1.3, derived the Economic Complexity Index both for countries (124 countries) and products (5107 products). In their paper, they validate what was discussed previously. They find that the more complex products are the machinery, chemical, and metal products, while the simplest are the raw materials, wood, textile, and agricultural products. Moreover, complex products are made in highly connected cores, while simple products are made in the peripheries of the product space. In addition, they classified countries by their level of income and the level of complexity of the exported products:

TABLE 2

5 MOST COMPLEX PRODUCTS	TOP FIVE EXPORTERS	AVERAGE GDP PER CAPITA (2005 PPP$)
(1) Cyclic hydrocarbons	Netherlands, USA, Japan, Germany, UK	~ 33.400
(2) Metalworking machine	USA, Japan, Netherlands, Malaysia, UK	~ 29.400
(3) Particle accelerators	USA, UK, Japan, France, Netherlands	~ 32.600
(4) Methacrylic acid	Germany, USA, Japan, Belgium, UK	~ 33.000
(5) Carbide tooltips	Sweden, Germany, Israel, Japan, USA	~ 31.400

Source: author's elaboration from Felipe et al. (2012) data

TABLE 3

5 LEAST COMPLEX PRODUCTS	TOP FIVE EXPORTERS	AVERAGE GDP PER CAPITA (2005 PPP$)
(5107) Sawlogs and veneer logs	Gabon, Malaysia, Congo, Cameroon, Equatorial Guinea	~ 10.200
(5106) Cashew nuts	Côte d'Ivoire, United Republic of Tanzania, Guinea-Bissau, Indonesia, Benin	~ 1.400
(5105) Manioc	Thailand, Vietnam, Costa Rica, Indonesia, Germany	~ 10.400
(5104) Natural rubber	Indonesia, Thailand, Malaysia, Vietnam, Côte d'Ivoire	~ 4.700
(5103) Cocoa beans	Côte d'Ivoire, Ghana, Indonesia, Nigeria, Cameroon	~ 1.800

Source: author's elaboration from Felipe et al. (2012) data

It follows immediately that countries exporting the most complex products are also those with a higher per capita income compared to countries exporting less complex products. Then, richer countries export complex products, and poorer countries export less complex products.

Taking into account all these results, we can reasonably say that the role of a country's productive structure is prominent in determining growth paths. The specific characteristics of the exported products strictly depend on the aforementioned capabilities that a country has. Economic growth and development can be reached through the creation and improvement of these capabilities, indirectly measured by the Economic Complexity Index.

CHAPTER 2 – INCOME INEQUALITY

2.1 THEORY AND MEASUREMENTS

There are many inequalities that can be considered like gender, wealth, political or life inequality. In this work, we will focus on income inequality, that is defined as how the households' income is distributed within a population, the less equal the distribution, the higher income inequality is.

Different measures have been proposed, weighing different aspects in different ways. However, they are all different faces of the same medal. The most common indexes are, according to the US Census Bureau, the Gini index and the shares of aggregate household income received by each quintile. Here a brief explanation of some common indexes produced by different statistical departments all over the world (US Census Bureau):

- **Atkinson index**: it takes the name from the British economist that developed the measure. It is widely used to determine which end of the distribution contributed most to the observed inequality. This measure can also have a normative interpretation, given that can be imposed different coefficients to weight incomes by choosing different levels of inequality aversion. If the inequality aversion is equal to 1, the Atkinson index become more sensitive to the lower part of the income distribution, while if it is equal to 0, the index become more sensitive to changes in the upper part of income distribution.
- **Equivalence Adjustment of Income**: this measure attempts to address the problem of the number of people in a household. Indeed, other measures treat the same income for a single person household in the same manner of a four people household. In fact, an equivalent-adjusted income of a household made by a single person earnings $ 20.000 is roughly double of the equivalent-adjusted income of a household made by four people earnings the same amount of money.
- **Mean Log Deviation**: it is a summary of the difference between the shares of income and the share of population, after having transformed these two measures with the natural logarithm. It takes a value equal to 0 when all households have the same income, while taking higher positive values when all income is concentrated in a few people.
- **Theil index**: it measures by how much the population is far from the "ideal" equal society in which the income is the same for everyone. It is a measure related to the

entropy because it gives as result the difference between the maximum possible entropy of the data and the observed entropy.

- **Gini index**: it takes the name from the Italian statistician Corrado Gini. Since it combines the detailed share data into a single statistic that sums up the income distribution throughout the whole income distribution, it is defined as a summary measure. If the index is equal to 0 then it indicates perfect equality, where the income is equal for everyone. On the other hand, if it is equal to 1 then it indicates perfect inequality, where only one person receives all the income. This index is based on the difference between the Lorenz curve and the perfectly equal income curve. In fact, the Lorenz curve represents the actual income distribution of a country, while the other one represents the situation in which everyone have the same income.
- **Income quantiles share ratio** (S80/S20 ratio): it is calculated as the ratio between total income of the population in the top quintile (20% of the population with the highest income) and the income of the population in the bottom quintile (20% of the population with the lowest income). As this ratio increases, inequality increases because it means that the total income of the wealthiest increases or the total income of the poorest decreases, leading to more income inequality.

2.2 DRIVERS OF INCOME INEQUALITY

The phenomenon of income inequality appears in most of developed countries during the last decades of the 20th century, and several research tried to study the reasons and the drivers of this trend. As we will see later, the link between economic development and wage inequality was deeply studied since Kutznets (1955), without reaching a strong consensus between economists on the real nexus. However, there have been many other attempts that tried to find the origins of income inequality.

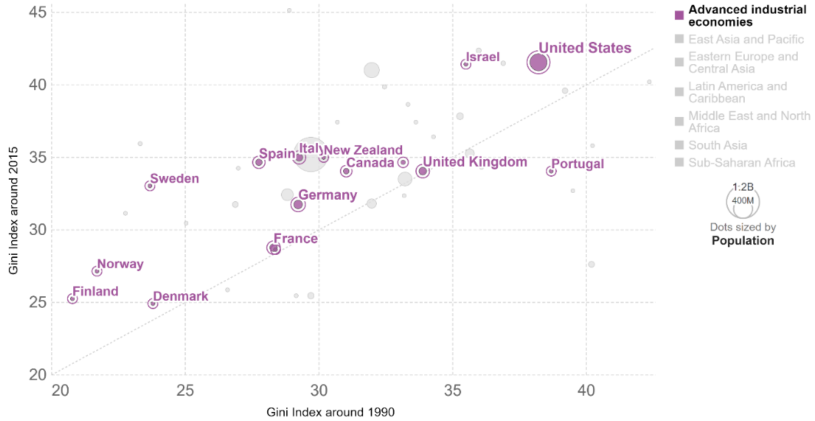
Many studies (one among many: Caminada and Wand, 2011) have examined domestic policies regarding welfare systems as crucial tools to influence income inequality levels in a country; a poorly designed tax and benefit system can lead to high inequality while an efficient one can low the inequality. Indeed, through benefit systems and social transfers, inequality can be reduced. However, there are others kind of causes that can be categorized in two main areas: the first is that which belongs to “market forces”, the second belongs to labour institutions. In the following, “market forces” drivers are explained.

Katz and Murphy (1992) argued that the demand and supply of workers' skills were the reasons for the rise in inequality in the US: building a simple model of supply and demand, they linked the increase in inequality in US to the increase in the demand of high-skilled and educated workers. In addition to this, the wage premium for educated people was increasing, leading to a larger inequality due to the increase in top salaries. In other words, they pointed out that if we observe wage dispersion is due either to the high demand for workers belonging to the right side of the income distribution, which leads to an increase in higher wages, or to the high supply of workers belonging to the left side of the income distribution, which leads to a decrease in lower wages. In fact, if we increase the income of the highest earners or reduce the income of the lowest earners, inequality increases.

Acemoglu and Autor (2011) found different other reasons that pushes income inequality. First, the latest innovations in the production system cause a more polarized structure of the workforce, due to the elimination of simpler jobs replaced by automation systems, leaving low-skilled workers without a job. This phenomenon is particularly strong since this group of workers is also the more fragile and they face greater difficulties in learning other skills to apply for other jobs. In this context, highly skilled workers not only earn more, but it is also easier for them to adapt or learn new skills to apply to other jobs. Second, a drop in the real wage of low skilled workers. Third, evident changes in the earnings structure in different part of the income distribution related to the first point. Fourth, what they called the "convessification" of skill and schooling returns, which means that the returns of skills and education do not follow a growing path, but a convex path, that is, to reach a point in which education pay more than non-education the level of it must be above a certain threshold, emphasizing the polarization.

In the following figure 5, is represented the Gini index in 1990 versus the Gini index measured in 2015 of some advanced industrial economies.

FIGURE 5: Gini index in 1990 vs 2015 of advanced industrial economies



Source: Our World in Data

In the horizontal axis the Gini index in 1990, while in the vertical axis the Gini index in 2015. The grey dotted line cutting the graph diagonally represents the line where a point should be if the Gini index in 1990 and 2015 were the same. More simply, France and the UK are in the diagonal line, so in the two years that we are considering the Gini index is the same. Instead, if a country is above the diagonal line, it means that the Gini index in 2015 is larger than in 1990. What the graph tells is that all the advanced industrial economies considered, with the exception of Portugal, in 2015 have a higher income inequality compared to 1990, while France and UK did not change the level of wage dispersion.

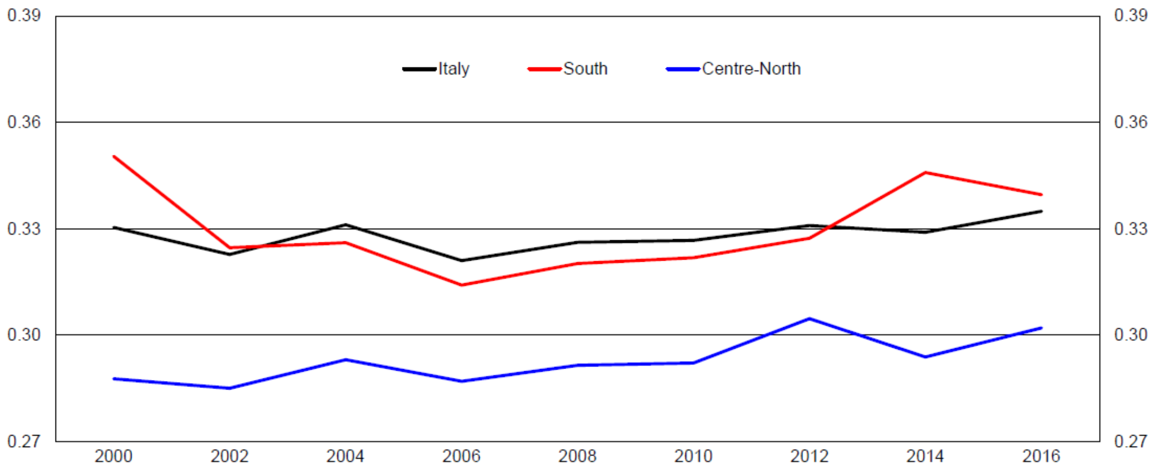
Looking at the figure 5, a problem comes to light. Indeed, if you look at countries that share similar (even if not equal) market conditions, such as trade openness, education and production technologies, as in the case of EU countries or the UK and the US, the "market forces" motivations do not always accurately predict the substantial heterogeneity in income inequality trends (Devicienti et al. 2019), even if the redistributive fiscal policies and the effects of different welfare systems are taken into account. Certainly, the small differences in the market conditions have a role in these differences, however it is evident that there are other forces to consider (Devicienti et al. 2019). Many studies shift the attention from the cited drivers (i.e. "market forces" drivers) to labour market institutions drivers.
Koeniger et al. (2007) studied data of eleven OECD countries, focusing on the market institutions such as unemployment insurance, unions, firing regulations, minimum wages and the impact on income inequality. They found that changes in these kinds of institutions explain as much as is explained by technology measures. Indeed, employment protection, higher union density and higher minimum wage reduce income inequality. Before them, also Blau and Kahn (1996) found "strong evidence for the importance of labour market institutions in explaining international differences in the levels of wage inequality", in particular centralized systems of collective bargaining.
Simón (2010), studying wage dispersion in EU, found that it varies markedly. Also looking at figure 5, it can be noticed that the violet dots of EU countries are not near to each other, meaning that they have different Gini indexes. However, the explanation of these differences, according to Simón (2010), must be searched in differences in inter-firm wage differentials. Citing directly from his paper: "this finding suggests that, in general, international differences in inter-firm wage differentials are mostly driven by differences in returns to firm characteristics and not by heterogeneity in firm populations. The importance of differences in returns to workplace characteristics is also consistent with the hypothesis that there is more spread to inter-firm wage differentials in countries with decentralized wage-setting systems." (Simón, 2010). Many other researchers have studied the impact of labour market institutions

on income inequality. Summarizing, Di Nardo et al. (1996) identified the reasons in the declining in minimum wages and union strength; Piketty and Saez (2003) in the changes in social norms; Barth et al. (2016), Faggio et al. (2010) and Card et al. (2013) suggested that a considerably part of the increase in income inequality (respectively in the US, UK and Germany) comes from between-firms rather than within-firms. This means that in the same company the distribution of income is not as unequal as it is between one firm and another, due to different wage policies put in place by the firms. On this point, Card et al. (2013) suggested that the changes in the collective bargaining institutions, is the main cause of this specific aspect (Devicienti et al. 2019).

2.3 INCOME INEQUALITY IN ITALY: DATA AND DRIVERS

In the analysis of income inequality in Italy, it is important to keep in mind the different paths that the indicators undertake in the different Italian regions. To better explain this sentence, look at figure 6:

FIGURE 6: Gini index in Italy from 2000 to 2016.



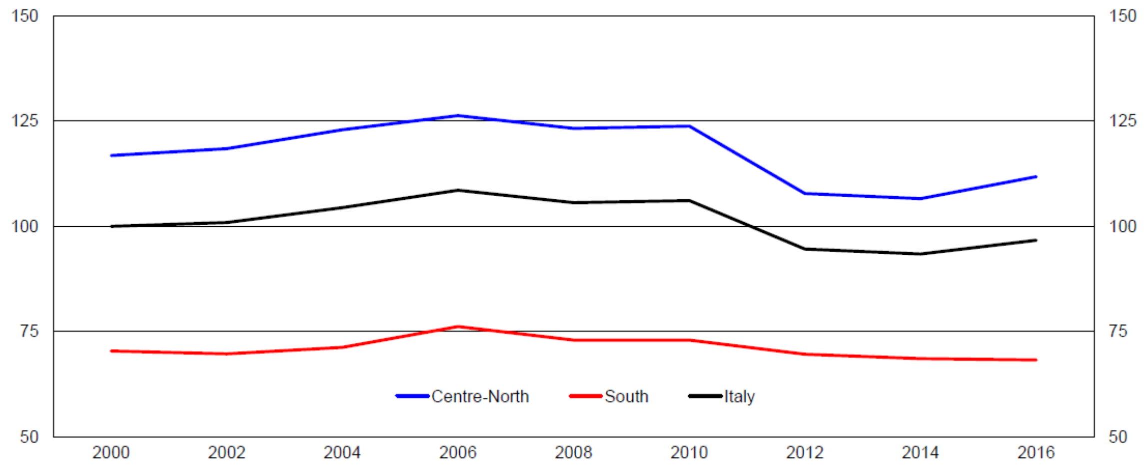
Source: Ciani and Torrini (2019).

By dividing Italy in two macrozones, Centre-Nord and South, it immediately comes to light the substantial difference between the two areas. Indeed, to look at data and to understand the phenomenon, it is important to focus on the North-South divide that characterize income inequality. Comparing other countries that have similar regional differences, such as Germany and Spain, the impact of these disparities is bigger in Italy (Ciani and Torrini, 2019). However, during the period considered in figure 6, the overall inequality in Italy is stable and does not show big variations. If we look only the South, the inequality decreases significantly in the 2000-2006 period, an then it starts to rise during the economic crisis and the recovery.

In the Centre-North, however, inequality remains stable from 2000 to 2006 and then increases. Other measures of income inequality, like the Mean Log Deviation, shows similar results (Ciani and Torrini, 2019).

In addition, looking at household income, southern regions have a lower income compared to northern regions:

FIGURE 7: Average income in Italy from 2000 to 2016 (Italy 2000=100).



Source: Ciani and Torrini (2019).

It can be noted that the rate of income growth from 2000 to 2006 is more or less the same between the two areas, showing respectively an increase of 8.2% for the southern regions and a growth of 8.1% in the northern regions. More than 75% of total growth is due to the increase in labour income, which means that the most important factor in the description of the change comes from it. Then, after 2006, the average income decreases for both areas, -10.4% for the South and -11.5% for the North, affecting the northern regions more; similarly, the most important factor was represented by labour income. Moreover, the average number of employed households, decreased more in the south then in the north respectively of -8.6% and of -7.4%. However, the labour equivalent income[1] per worker decreased more in the Centre-North (-10.1%) compared to the South (-6.7%) (Ciani and Torrini, 2019).

A detailed analysis of the data shows that the decrease in inequality in the South between 2000 and 2006 is driven by the increase of the lower part of the income distribution (10th percentile), while higher incomes have been stable. On the other hand, the most vulnerable are

[1] Equivalent income is the ratio between total household income and the number of equivalent adults. The latter is calculated using the OECD-modified equivalence scale which assigns a value of 1 to the head of household, of 0.5 to each additional member over the age of 14 and of 0.3 to each child below that age (Ciani and Torrini, 2019)

those with low incomes, so during the economic crisis the increase in inequality is driven by the decrease in lower wages, lower government transfers and higher unemployment. Similarly, the reason for the increase in inequality in the North is due to the decrease in lower wages, but the contraction was much smaller compared to the South.

As briefly mentioned before, welfare systems are crucial tools to face inequality. In European countries, we see several types of systems and diverse results in applying different approaches. In the case of Italy, Di Caro (2018) analysed the personal income-tax system in its redistributive capacity, showing that few tax instruments are relevant to the redistribution, and that the effects of these instruments are different if they are applied in a region or in another one. Moreover, the reason of these disparities between southern regions and the others, can be found in the difference in the education level of the population (higher education has the effect of reducing inequality), lower fraction of adults with labour income in the South and different values of work intensity, that is lower in the South, but it shows similar values if you consider the top decile of the income distribution. Even if there are relevant demographic differences, this last factor does not play a crucial role in explaining the gap (Ciani and Torrini, 2019).

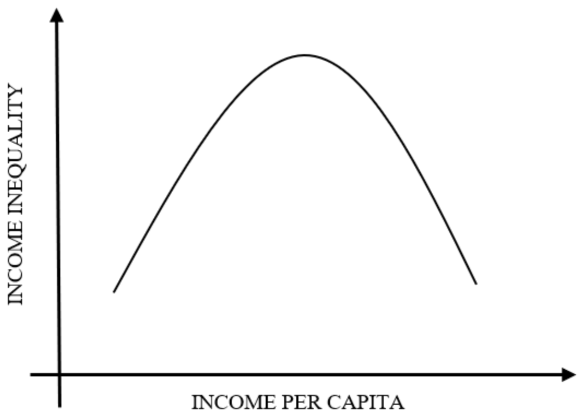
Looking at labour market institutions drivers, Boeri et al. (2021) analysed the importance of wage bargaining: they looked at Germany and Italy, two countries that show similar geographical differences in firm productivity but different model of wage setting. In the first country, the system allows for local bargaining, while Italy sets wages based on nationwide contracts. Data show that, in Italy, there is no connection between the level of nominal wages and local productivity and that differences in nominal wages are limited. In contrast, Germany experiences larger geographic wage gaps and a link between local wages and local productivity. Then, the model applied to Italy leads to an increase in unemployment in those provinces where the productivity is low because employers cannot adjust wages according to the productivity level. Higher unemployment directly increases income inequality.

CHAPTER 3 – INCOME INEQUALITY AND ECONOMIC COMPLEXITY

3.1 INCOME INEQUALITY AND ECONOMIC GROWTH

There have been several attempts to find a relationship between economic growth and inequality. One of these first efforts dates back to the middle of the last century, from Kuznets (1955). In his research, he tried to study the relationship and the reasons that link income inequality and GDP per capita. Although he pointed out in his concluding remarks that his study was "5 per cent empirical information and 95 per cent speculation" (Kuznets, 1955) the work is extremely important in laying the foundations for future studies, and that is what happened. Indeed, data and statistical measures were very poor in 1955, and this justifies the "5 per cent empirical information"; however, the findings were extremely useful for future studies. He found that the relationship between these two variables has an inverted U-shape: increasing per capita income increases income inequality first and, after reaching a given point, decreases as shown in the following figure:

FIGURE 8: Kuznets curve



Source: author's elaboration.

He thought that, in the early stages of the economic development, when income per capita increased, income inequality also increased. When the economy began to become more industrialised and modern structures began to penetrate the entire socio-economic life, then increasing in income per capita led to lower income inequality (De Dominicis et al., 2008). In fact, economic development and industrialization begin to spread in urban agglomeration and then begin to spread throughout the region or country. This was particularly evident in the early decades of the 19th century: large cities became increasingly wealthy, while the countryside and rural areas did not benefit from industrialization. Therefore, the disparity between cities and rural areas has increased, also leading to greater income inequality. When development begins to spread even in rural areas, the income gap between the city and the countryside is reduced, reducing income inequality while there is an increase in GDP per capita.

After Kuznets, there have been several efforts to explain this relationship, even if the main energies were not focusing on GDP per capita but in economic growth.
One of the explanations between who have argued that if we see an increasing in economic growth, we will also see an increase in income inequality, relay on the importance of the saving rate. Actually, in developed countries, the saving rate of the wealthier social classes is higher than the one of the poor. If poor people must use most of their income in consumption, the rich can (in relative terms) save more. Therefore, more redistribution from rich to poor means that the saving rate of the whole economy will reduces, and with it the growth rate.
Another explanation that supports this positive relationship comes from the incentive compatibility theory: more redistribution might lead to a lower incentive for rich to work and to produce income, lowering the economic growth. Finally, if there are large investments that are difficult to divide between different capitals and that are classified as sunk costs, the concentration of wealth is necessary for the creation of economic activities (De Dominicis et al., 2008).

On the opposite side, researchers who tried to explain the negative relationship looked at four different theories, called by De Dominicis et al. (2008), the endogenous fiscal policy theory, the socio-political instability theory, the borrowing and investment in education theory and, finally, the joint education/fertility theory.
The **endogenous fiscal policy** approach highlights the importance of the distortionary mechanisms put in place by governments in case of high income inequality. Certainly, in these cases, governments will try to reduce inequality by introducing redistributive measures via (mainly) taxation. This will hit capital investment and will also tighten economic growth. In other words, more inequality leads to stronger fiscal distortions that hits economic growth.
The **socio-political instability** framework relies on the idea that a strong concentration of wealth in the hands of few people, could lead to an increase in illegal activities and to social protests. In this socio-economic environment, uncertainty increases discouraging investments and, in turn, economic growth.
The **borrowing and investment in education** theory suggests that higher levels of inequality limit the human capital acquisition. Undeniably, in presence of borrowing constraints, there are fewer people that can access education and so less people can increase their human

capital. As highly demonstrated in many researches (look for instance: Lucas, 1988; Barro and Lee, 199; Hanushek and Kimko, 2000; Hanushek and Wossmann, 2012) human capital significantly affects economic growth.

Finally, the **joint education/fertility** theory tries to link fertility, schooling and growth. Higher fertility rates mean fewer resources a family can invest in education, so less human capital and then less growth. De la Croix and Doepke (2003) developed a theoretical framework in which fertility and education decisions are mutually dependent. In their study they found that, on average, poor parents decide to have more children and to invest less in education compared to rich parents. Subsequently, higher inequality leads to more "poor parents" that invest less in education. This choice is detrimental to overall human capital and will eventually reduce economic growth.

Therefore, looking at theoretical explanations do not help to find a definitive answer. Nevertheless, empirical confirmations are even more difficult to find since they are contradictory. Indeed, there have been found positive relationships, negative relationships, inverted U-shaped, not unique or inconclusive relationships between income inequality and economic growth (Shin, 2012). Here some past literature and the relative results (the list is not complete):

TABLE 4

RELATIONSHIP	AUTHORS
negative	Murphy et al. (1989), Perotti (1993), Alesina and Rodrik (1994), Persson and Tabellini (1994), Perotti (1996), Alesina and Perotti (1996), Acemoglu (1997), Helpman (2004), Tachibanaki (2005), Sukiassyan (2007)
positive	Okun (1975), Bourguignon (1990), Benabou (1996), Li and Zou (1998), Aghion and Howitt (1998), Forbes (2000)
inverted U	Kuznets (1955), Chen (2003)
not unique or inconclusive	Amos (1988), Barro (2000), Banerjee and Duflo (2003), Weil (2005), Shin et al. (2009)

Source: author's elaboration from Shin's (2012) data.

The reasons for these different results should be search in the type of dataset and in the empirical estimation used (Sukiassyan, 2007). However, there are several studies that, even if they rely on the same dataset, they found different results. To be more precise, many researches from the 1990s and the early years of the new millennial, used the same dataset coming from Deininger and Squire (1996) that contains comprehensive cross-country inequality data. Nevertheless, the sample used were different, in order to exploiting different qualities of the data selected. It is the case of Persson and Tabellini (1994), Alesina and Perotti (1996), Perotti (1996), Li and Zou (1998), Barro (2000), Forbes (2000), Banerjee and Duflo (2003). In these cases, therefore, the reasons for the contrasting results must be searched in the different estimation strategies and empirical specifications. In the following table are summarized the different estimation strategies used and the sign of the relationship between economic growth and income inequality:

TABLE 5

AUTHORS	ECONOMETRIC MODEL	FINDING
Persson and Tabellini (1994)	OLS	Negative relationship
Alesina and Perotti (1996)	2SLS, 3SLS	Negative relationship
Perotti (1996)	OLS	Negative relationship
Li and Zou (1998)	Panel fixed effects	Positive relationship
Barro (2000)	3SLS	Inconclusive
Forbes (2000)	Panel fixed effects	Positive relationship
Banerjee and Duflo (2003)	Fixed Random effects	Inconclusive

Source: author’s elaboration from Sukiassyan (2007) data.

It is clear that the connection is difficult to establish and to study. Many scholars have theorised the connection between the production structure and income per capita. However, as saw before, using GDP per capita or economic growth was not conclusive. Using measures such as the number of workers employed in a certain sector, measures of diversity and aggregation, measures of diversification, do not really take into account the sophistication of the production structure (Hartmann et al., 2017). Still, new empirical strategies and increasing attention to economic complexity have put new light on this nexus. In addition, the results on the relationship between income inequality and institutions have added an important aspect in the development of the analysis considered (Acemoglu and Robinson, 2012).

3.2 INCOME INEQUALITY AND ECI BETWEEN COUNTRY

From a theoretical perspective, the production structure of a country, captured by the ECI, could have both positive or negative effects on income inequality. The production structure, indeed, could affect positively or negatively the occupational choice of workers, binding their available options on education, human capital development, their bargaining power or the possibility of having strong and effective unions. If technological development, industrialization and an increasing level of complexity provide new jobs and new learning opportunities, then workers can benefit in terms of income per capita, reducing also inequality. It is the case in Germany, for example. On the contrary, if industrialization leads to the reduction of the unions power, to increasing global competition that forces workers to accept low-wage jobs, then income inequality increases: it is the case of the US (Hartmann et al., 2017). Likewise, while simple products are made thanks to resources abundance and cheap labour, complex industrial products or complex services tends to rely on large tacit knowledge, not measured by the classical indicators of human capital or year of schooling. Higher levels of tacit knowledge give the possibility to workers to increase the efficacy of unions and the effectiveness of their negotiation. This tacit knowledge is captured by the ECI, then for higher levels of complexity the inequality should decrease.
Besides, the economic structure is likely to affect also political power: countries that have based their economy on natural resources and on not diverse products, tend to be ruled by the leading economic social classes, which protect their "business" and finally fostering income inequality (Hartmann et al., 2017).
On the opposite side, it could happen that increasing economic complexity will also increase the income inequality. In a more complex productive structure with highly diversified products and tasks, the capabilities required and the type of jobs available will increase. Therefore, a high level of specialisation is required and workers who can learn new tasks and can learn them quickly will have an advantage that turns into higher wages. On the contrary, workers that cannot follow the increasing complexity, will be left behind. In simpler words, a more complex economic structure will lead to higher income inequality (Hodgson, 2003).

There have been several studies that tried to empirically test this relationship. In the following pages there are some of them summarised:

- **Hartmann et al.** (2017) analysed more than 70 countries from 1962 to 2012, merging different datasets. The ECI was calculated as explained in Chapter 1. The Gini index was the measure of income inequality. Other controls are GDP per capita, years of schooling, population, and institutional factors such as corruption control, political stability, government effectiveness, regulatory quality, voice and accountability. In the bivariate analysis both economic complexity and GDP per capita show a negative relationship with income inequality. However, what is interesting is the difference between the fit of the two models. The relationship with GDP per capita shown an R^2 of 0.36, while the one with ECI an R^2 of 0.58, and this difference is statistically significant. Moreover, the negative relationship between ECI and income inequality is stable across all the considered years. In addition, looking at other productive structure indicators, it has not been found any case in which "income per capita or measures of productive structures are significantly preferred as predictor variable for income inequality in comparison to ECI" (Hartmann et al., 2017). In the cross-sectional pooled regression analysis, they regressed income inequality against economic complexity, GDP per capita and its square, average years of schooling, population, and the institutional factors aforementioned. They found that ECI is a negative predictor of income inequality, that ECI is the most significance variable in the regression and the variable is the most predictive of income inequality. Fixed-effect panel analysis also found a strong and negative relationship between the two variables of interest, even if the observed temporal variation of income inequality and economic complexity is low: one s.d. increase in ECI translates in 0.03 reduction of Gini index. Overall, despite the econometric strategy, the result is that economic complexity reduces inequality, and this is robust to many controls.

- **Fawaz and Rahnama-Moghadamm** (2019), considered more than 120 countries from 1964 to 2013. They used a spatial autoregressive model (SAR), to analyse if income inequality in a country depends on the income inequality and economic complexity of economically related countries. The variable considered are the Gini index, the ECI, the GDP per capita and its square, real domestic and real foreign investment, one variable that captures the relationship between income inequality in the observed country and income inequality among its economically proximate countries, and finally one variable that captures the relationship between income inequality and the economic complexity of the observed economy's trading partners. The main finding of our interest is that trading with more economically complex countries is correlated with a reduction in income inequality.
- **Lee and Vu** (2019) in their models used the Gini index, ECI, GDP and its square, years of schooling, population, trade openness, and FDI. In the OLS model they analysed more than 90 countries from 1980 to 2014. First, they pointed out the possible bias of reverse causality and the need of using an IV to overcome it; finding a

consistent instrument that affects income inequality only through the effect on economic complexity, seems a very difficult task. This is why they first chosen to look at a simple OLS. The result is that economic complexity affects negatively income inequality, and the result is strong and statistically significant, while the other inequality determinants considered (FDI, trade openness, education, institutions) are statistically not significant, meaning that ECI captures most of the variability. To address the possible endogeneity of OLS, they also estimate a dynamic panel data model, looking at 113 countries from 1965 to 2014. As explanatory variable, they also included a lagged income inequality and a lagged economic complexity. The findings here are not consistent with those of OLS models: an increase in economic complexity is associated with a higher level of income inequality. This was confirmed both in the short run and in the long run.

- **Chu and Hoang** (2020) analysed more than 80 countries from 2012 to 2017. They used the ECI and the Gini index. The difference here is that they tried to test the relationship also considering government spending and trade openness. The idea behind is that both of these two variables reduce the positive impact of economic complexity. Indeed, public expenditure has, among other things, the aim of reducing income inequality, while for trade openness has been tested that in developing countries it could reduce the wage gap between skilled and unskilled workers. In addition, they considered in their analysis education, GDP per capita and institutional variables. Another difference is that they deal with the possible problem of reverse causality between ECI and inequality by using an IV approach; the instrument chosen is research and development expenditure (as a percentage of GDP) in a dynamic model that consider the one-period lag Gini index. GMM and 2SLS are the estimation strategies used. They found that economic complexity positively affects income inequality, so higher levels of economic complexity mean higher inequality both in GMM and in 2SLS. However, by calculating the marginal effect of the ECI on income inequality at different levels of other explanatory variables, they found that above a certain threshold of education, government spending and trade openness, more economic complexity helps to reduce income inequality. In conclusion, these two authors found that the relationship is not linear, and other factors must be taken into account in studying this nexus.
- **Sepehrdoust et al.** (2021) focused on selected developing countries from 2000 to 2019. Gini index, ECI, scientific productivity and risk types (economic risk, financial risk and political risk) were used as variables. The methodology was a panel-VAR

model. Thanks to the chosen methodology, the authors analysed the impulse response function of the Gini index to an exogenous shock of the Gini index itself and of the other considered variables. The result was that, in developing countries, an increase in economic complexity leads to an increase in income inequality. However, they pointed out the strong possibility that the outcome is result of the selected sample: indeed, complexity helps to reduce inequality above a certain threshold, as pointed out by other researchers. Though, looking at developing countries only, the level of complexity remains low and insufficient to reduce income inequality.

- **Lee and Wang** (2021) looked at 43 countries from 1991 to 2016. In their fixed effect model, they consider the Gini index and, as independent variables, the ECI, GDP, trade openness, FDI, population, government expenditure. With this approach, they found that economic complexity reduces income inequality. Moreover, by dividing the sample between developed and developing countries, they found that for developed countries more complexity means less inequality while for developing countries more complexity means more inequality, consistently with the findings of Sepehrdoust et al. (2021). To address the problem of possible heterogeneity and to investigate whether changes in this relationship are due to country risk, the authors has developed a finite mixture model (FMM) adding as independent variables the country risk, political risk, economic risk and financial risk. The model suggests dividing countries between in two categories, high risk countries and low risk countries. The results are similar to those found in the previous model, with the only difference that for high-risk countries, an increase in economic complexity has no statistical effects on inequality, while for low-risk countries, the effect is strong and statistically significant (more complexity leads to less inequality). Furthermore, they tried to deal with the endogeneity problem due to the possibility of reverse causality or omitted variable bias. Using an IV-FMM and lagged ECI as instrument variables, they found consistent results with the previous analyses.
- **Pham et al.** (2023) using two-stage GMM estimation for a sample of 99 countries from 2002 to 2016, found that both economic development and the shadow economy have nonlinear effects on income inequality: the relationship between economic complexity and income inequality has a U-shape, while the impact of the informal economy on income inequality follows an inverted U-shaped pattern.
- **Amarante et al**. (2023) using a Fixed Effect estimation of panel data for the period from 1995 to 2018 for approximately 190 countries, found a non-linear relationship between the Economic Complexity and Income inequality. They found an inverted U-

shape relationship between these two variables. This means that economic complexity reduces income inequality after certain thresholds, which is the case of high-income economies.

In the following table the summarized results exposed in the previous pages:

TABLE 6

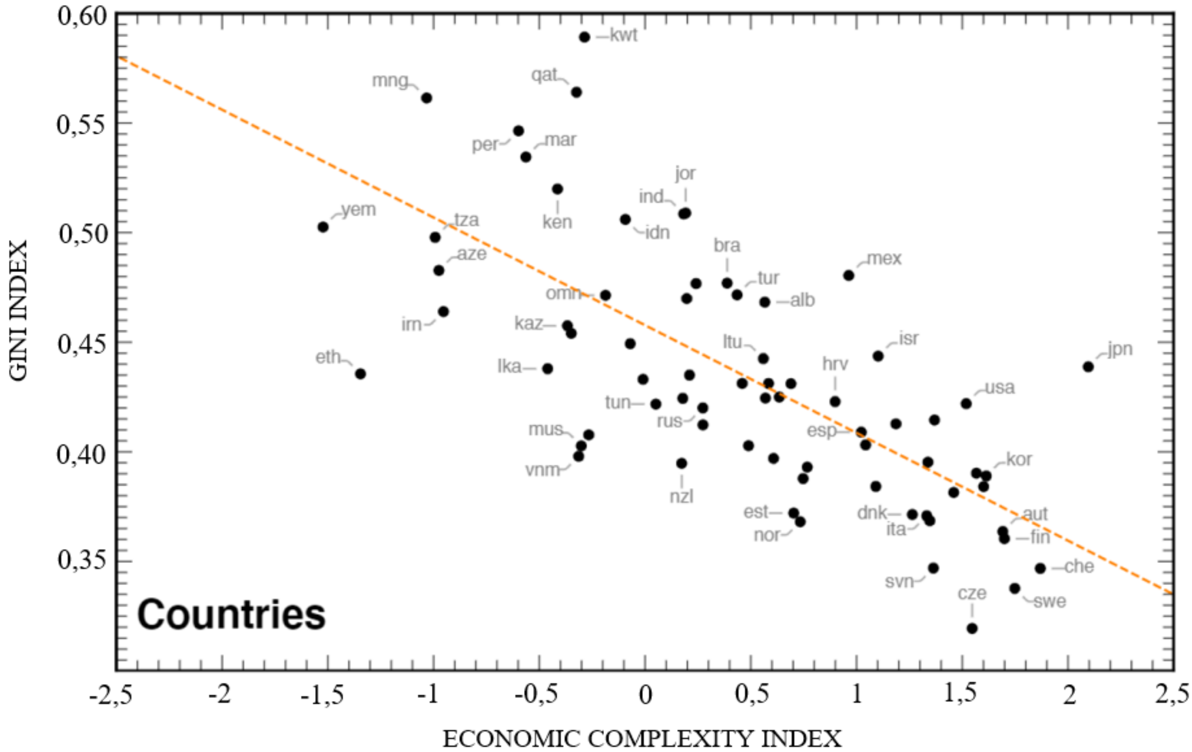
AUTHORS	ESTIMATION METHOD	COUNTRIES	FINDING
Hartmann et al. (2017)	Bivariate, Cross-sectional pooled regression, FE	70	Negative relationship
Fawaz and Rahnama-Moghadamm (2019)	SAR	120	Negative relationship
Lee and Vu (2019)	OLS, Dynamic panel data	90/113	Contradictory
Chu and Hoang (2020)	IV-GMM, IV-2SLS	80	Inverted U-shaped
Sepehrdoust et al. (2021)	Panel-VAR	Developing countries	Inverted U-shaped
Lee and Wang (2021)	FE, FMM, IV-FMM	43	Inverted U-shaped
Pham et al. (2023)	GMM	99	Inverted U-shaped
Amarante et al. (2023)	FE	190	Inverted U-shaped

Source: authors' elaboration

We can conclude that there is a consensus on the effect of economic complexity on income inequality between countries, even if there are some differences if the negative relationship is valid only above a certain threshold or if it is always valid. In general, the inverted U-shaped relationship predominates.

In the following figure, is represented a bivariate analysis that links income inequality (Gini index) and economic complexity in different countries. Even if the methodology is simple, it is useful to have a graphical representation of the phenomenon that clearly show the correlation between the two variables of interest:

FIGURE 9: Economic Complexity and inequality



Source: Hartmann and Pinheiro (2022).

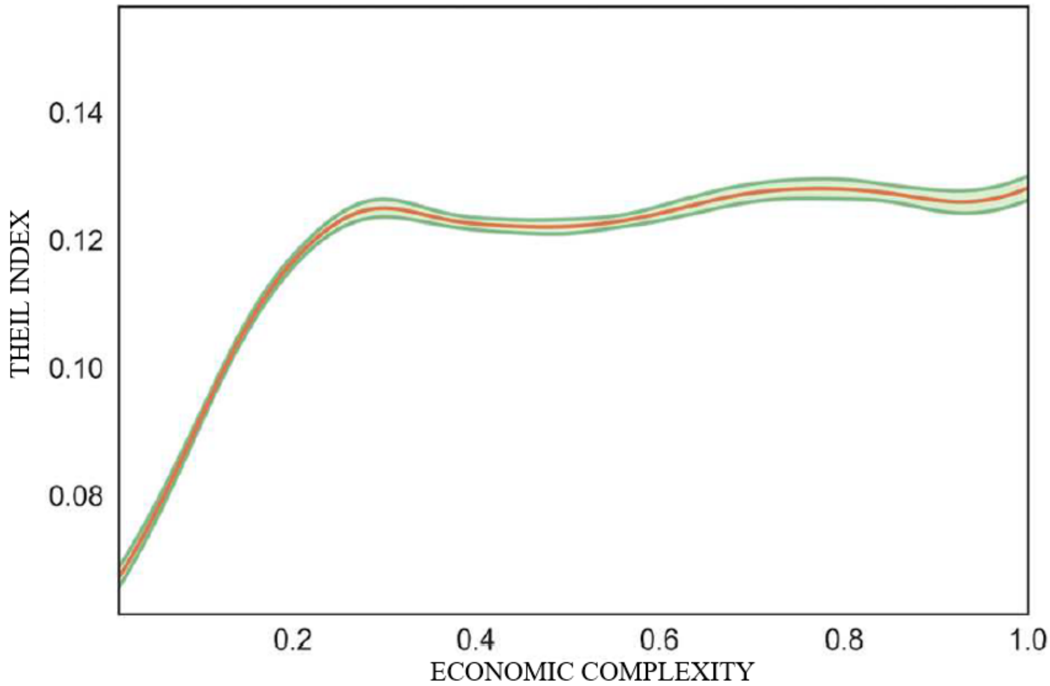
3.3 INCOME INEQUALITY AND ECI WITHIN COUNTRY

If the between country analysis shows a consensus in the direction of the relationship between income inequality and economic complexity, the within country analysis is less clear especially looking at empirical results, that are scarce. Instead of starting with various theoretical explanations, it is more useful to look at some research that empirically analysed the within country relationship:

- **Sbardella et al.** (2017), look at the relationship between wage inequality and economic complexity at a county-level (more than 3100 counties) in the US from 1990 to 2014. The economic complexity of the different counties has been calculated using a slightly different approach compared to the one described before, however the idea behind is the same and the differences can be neglected. In fact, they used Fitness as a measure of Economic Complexity. For inequality they used the Theil index. They used a non-parametric estimation. The US is a developed country with the highest level of total GDP and one of the highest levels of GDP per capita and it is also a country with high income inequality. Given the US is a high developed country, one should expect that, if the relationship is scale-invariant, the path follows a negative trend, so high complexity leads to less inequality. Moreover, the institutional factors have not effect in a cross-sectional counties analysis. This happens because the institutional framework is homogenous since it is determined more at a federal or state level than at a county level.

The result of the non-parametric analysis is shown in the following figure:

FIGURE 10: Economic Complexity and inequality



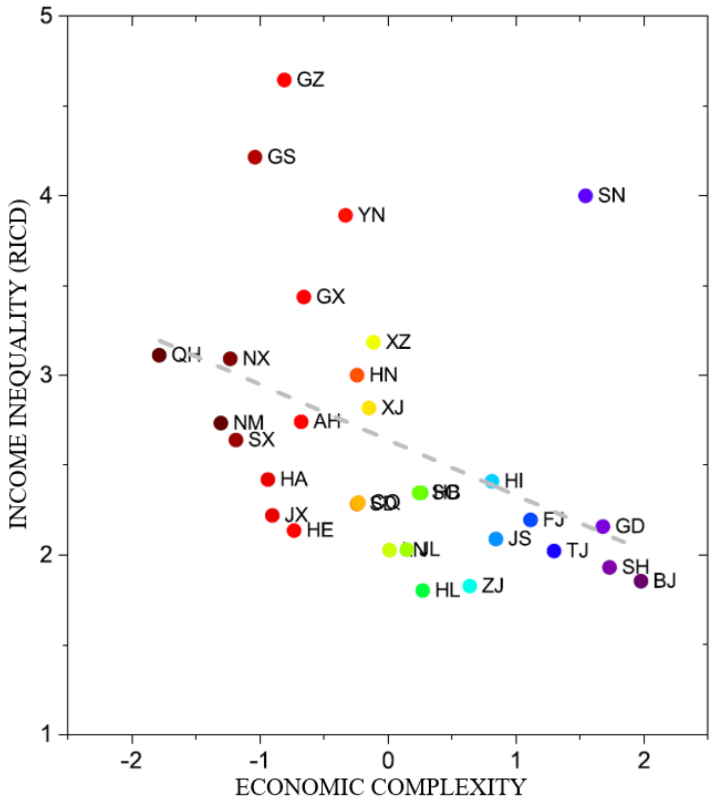
Source: Sbardella et al. (2017).

The relationship is positive, so for higher level of economic complexity, inequality increases. Therefore, also in the US case, the nexus goes in the opposite direction compared to the one found in the between country analysis. This means that the key determinants change at different scales.

- **Gao and Zhou** (2018) quantified the economic complexity of China's provinces analysing 25 years' firm export data from 1990 to 2015 for 31 provinces (then all the chines' provinces/regions except the two special administrative regions). One of the main limitations in this study is that firms' data they use contains only a small fraction of the total numbers of Chinese firms. This approach could lead to potential biases that could invalidate the overall analysis. However, given the small body of literature present in this specific field, it is useful to look also at these attempts.
  After having calculated the regional economic complexity index (ECI), they look at the relationship between ECI and income inequality. This last measure is calculated as the income inequality between urban and rural areas in China measured with the

relative income differences (RICD): this indicator is the ration between RICU to RICR (where RICU is the relative income in urban area and RICR in rural area). Their efforts were mainly focused in calculating the economic complexity index, then they briefly look at the correlations between the RICU/RICR and economic complexity, and finally at the correlation between the RICD, that is their measure of income inequality, and economic complexity. The following figure 11 represent this last correlation:

FIGURE 11: Economic Complexity and inequality



Source: Gao and Zhou (2018).

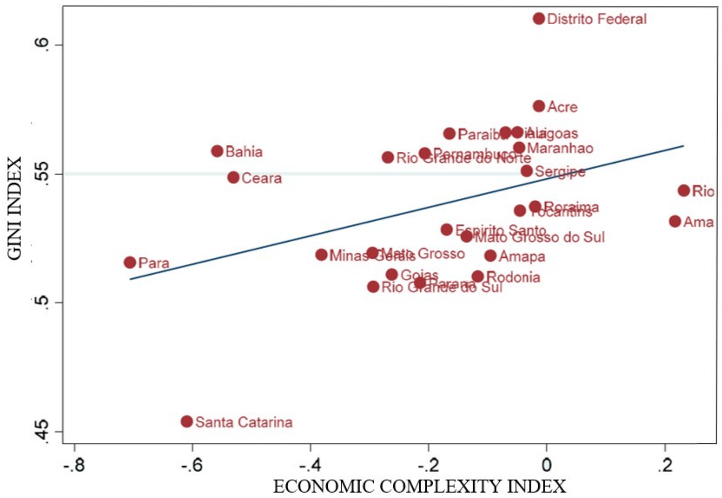
What the graph shows is that provinces with higher economic diversity and relative income have less income inequality. This result is robust after controlling for population, urbanization, schooling, innovation and trade.

- **Zu et al.** (2020) shift the analysis in the China's prefecture-level regions (126 units) in 2013. The fact that they limit the analysis only for this year is due to data availability and might rise some concerns on the generalization of these results; indeed, this is a strong limitation of the research. They studied how regional export product/destination structures have shaped income inequality, then not looking only at the complexity of the production structure but also looking at the destination of the exports. They used the ECI and the Gini index to look at the variables of interest. They found that a higher level of economic complexity and a higher percentage of products exported to more complex countries/regions mean a lower level of inequality only for urbanised units, while for rural area income inequality increases.

- **Bandeira et al.** (2021) analysed the income inequality-economic complexity nexus at regional level within Brazil, exploring the impact of economic complexity at a regional dimension. Indeed, in Brazil it is likely that many determinants of economic complexity vary across regions, think for example on the high concentration of industries in urban agglomerations, that are differently distributed in Brazil. The methodology used was similar to the one used by Hartmann et al. (2017). They estimate the models using pooled OLS and Random Effects estimations (to deal with low variability in the data considered). That is applied to 27 federative units with annual data from 2022 to 2014. The variables considered were the Gini index, Theil index, ECI and its square, GDP and its square, years of schooling, population, and many others controls connected to income inequality in Brazil (share of white workers, of self-employed workers, of agricultural workers, active workers in urban area). Adding the square of ECI and GDP, allow to look at any non-linear effect of these two variables, as theorised by the Kuznets curve. In fact, the increase in economic complexity could firstly have a beneficial effect for high skilled workers that can immediately adapt they capabilities to new productive structures; this is reflected in an increase in inequality. However, when the economic complexity reaches a certain point, then also low skilled labour can benefit from higher complexity, reducing income inequality.

  In a first very simple bivariate analysis, they look at the correlation between ECI and the Gini coefficient of the 27 federative units considered. They immediately found that São Paulo is a significant outlier, with a ECI of ≈ 5, compared to the other 26 units that go from a value of -0.7 to 0.2. This suggests looking also at these 26 units separately, because more similar from an economic complexity point of view. The result of the bivariate analysis (figure 11) shows a positive correlation between the two variables. If you also consider São Paulo, the correlation reverses, becoming slightly negative, meaning that for high income regions, the relationship reverses.

FIGURE 12: Economic Complexity and inequality



Source: Morais et al. (2021).

With a deeper analysis, the OLS and the RE estimations found substantially the same result: the relationship has an inverted U-shape, supporting the idea described before. They also divided the analysis looking at regions with different level of development, finding the inverted U-shape materialised only in high developed regions, suggesting that a certain level of development needs to be reached before economic complexity starts to have an impact on inequality. Moreover, they deal with the fact that the ECI is based on exports to other countries; this feature can be problematic because it does not consider the fact that a region could export complex products to another region. Then, this complexity is not capture by the ECI. To test if other characteristics of regional productive structures are important, they add in the regression the level of regional industry and occupation diversity.

- **Marco et al.** (2022) focused their research in analysing the trilemma between income inequality, environmental degradation and economic complexity at a sub-national level in Spain, more precisely at a province level (50 provinces) from 2002 to 2016. Therefore, they build up the trilemma indexes, normalized between 0 and 1. For income inequality they used the Gini index, for economic complexity the ECI, and for environmental data they used CO2 emissions. They estimate a linear and non-linear functional form to investigate the presence of this trilemma; then, the aim is to look if by increasing one of the considered variables, the other two reduces. If this happens, it means that it is not possible to increase economic complexity without reducing environmental quality or income equality. Looking at the graphical representation (figure 13 and 14) of the phenomenon, it becomes clear the different trade-offs between Economic complexity (CI), income equality (IE), environmental quality (EQ), and economic growth (EG).

FIGURE 13: pre crisis (2002-08) vs post crisis (2009-16)

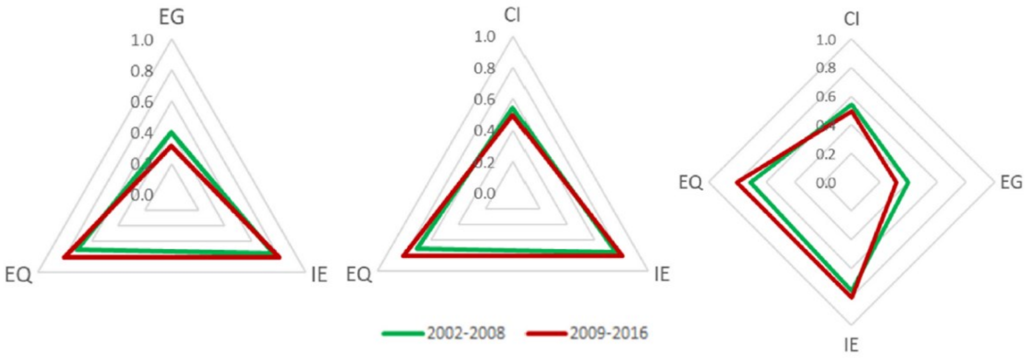
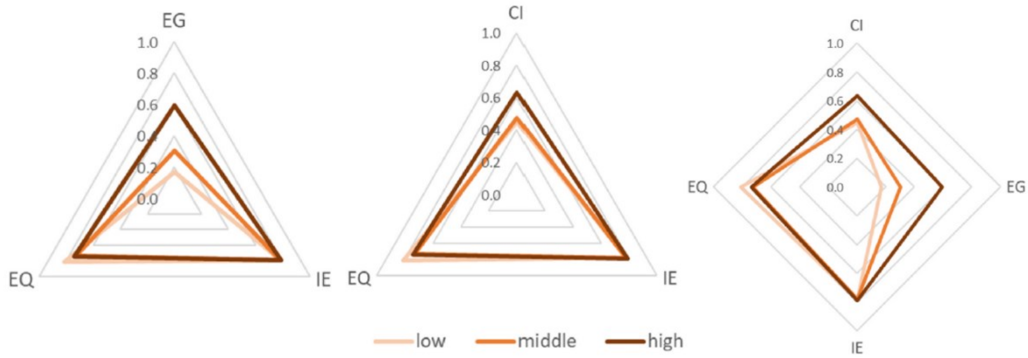


FIGURE 14: by income groups (2002-16)



Source: Marco et al. (2022).

Focusing on the income equality vs economic complexity and economic growth trade off, figure 13 shows that between the pre-crisis period and the post-crisis period, a decrease in economic growth and a decrease in economic complexity leads to a higher income equality. This has been showed both for the trilemma and for the quadrilemma. On the other hand, figure 14 looks at different income groups of provinces: here it seems that changes in economic growth or complexity do not affect income equality. However, for environmental quality, the trilemma and the quadrilemma persist. Concluding, they found robust evidence that a trilemma exists not only between the economic growth and the other two variables considered, but also considering the economic complexity index.

- **Gómez-Zaldívar et al**. (2022) have analysed the relationship in Mexican states (31 federal states) from 2004 to 2019. The two main variables of interest are the Gini coefficient for income inequality, and the ECI for economic complexity. The methodology chosen is a Random Effect model, considering as control variable GDP per capita, population, education, corruption control, government effectiveness, voice and accountability, and rule of law. The result is strong and significant: states with more complex economy shows lower levels of inequality. This result is robust for many different checks, suggesting that increasing specialization has led to the creation of more job opportunities, and the resulting economic growth has helped to increase the labour force (by decreasing the non-active population and unemployment) which in turn reduces inequality. The Mexican case is interesting from several points of view: first, in the years considered, some Mexican states have faced an important economic transformation resulted from improved trade opportunities making it a good case study. Second, during the considered period the inequality increased at a country level however the paths are different by considering different states. Third, they found, similarly to other studies, that economic complexity has a spatial relationship with inequality. This means that inequality does not depend exclusively to the economic complexity of the considered state but also to the neighbours' level of economic complexity.

In the following table the summarized results on the relationship between economic complexity and income inequality:

TABLE 7

AUTHORS	ESTIMATION METHOD	COUNTRY	FINDING
Sbardella et al. (2017)	Non-parametric	USA	Positive relationship
Gao and Zhou (2018)	Bivariate/multivariate	China	Negative relationship
Zu et al. (2020)	OLS	China	Positive relationship for non-urbanised area
Bandeira et al. (2021)	Cross-sectional pooled regression, RE	Brazil	U-shaped
Marco et al. (2022)	Linear and non-linear	Spain	Positive relationship
Gómez-Zaldívar et al. (2022)	RE	Mexico	Negative relationship

Source: author's elaboration.

Summing up, at regional level, there is not a total consensus: different approaches found different results. Therefore, it must be explored which are the theoretical explanation of why we observe positive or negative relationship between our variable of interests. Indeed, the reasons why higher complexity leads to lower inequality can be reconducted to the ones that explain also the between country analysis, with a regional point of view instead of national. For the positive relationship the authors pointed out the following arguments:

First, large urban agglomerations show, on average, higher ECI compared to rural areas, for the reasons that have been discussed before (more circulation of different ideas, economy of scale, high specialization and coordination…). However, there are many researches that show what higher urbanization means: as pointed out by Bettencourt and West (2010), cities are the "world's centres of creativity, power and wealth". Indeed, as a city increases its dimension, also economic activity, productivity, GDP, number of patents, economic complexity and many other indicators increase. On the other hand, there are some drawbacks such as the increase in crime, pollution, and income inequality. This latter point has been deeply analysed by Heinrich Mora et al. (2021). They found that by increasing the population size of a city, income inequality (measured with the Gini index) also increases. Giving a closer look to the data, the beneficial effects of agglomeration go mainly to the wealthier decile of the population. We see that if the city size increases, the aggregate GDP increases in every decile. In the least decile the relationship is perfectly proportional, then an increment in the city size indicator of 1 means an increase in the GDP of 1, while for the top decile the increase in more than proportional, leading to a higher inequality. Then, according to Heinrich Mora et al (2021), more urbanization leads to higher inequality both within the city but also between cities and rural area. If we think that for larger agglomeration, we will have a larger ECI, then it could be that for higher levels of economic complexity also inequality will increase, despite the negative association found in comparing countries.

Second, in a within country analysis, the institutional factors may not be so important as in a between countries analysis (Marco et al., 2022). Certainly, looking at different regions or different province in the same country, institutions may not vary a lot, given that many formal and informal rules are common to all the country. Indeed, the environment analysed is more "homogeneous" from this point of view. According to what exposed in previous pages, higher levels of tacit knowledge (measured with ECI) give workers the opportunity to increase the efficacy of their actions even in changing institutions towards more inclusive ones. However, workers from one region or one city may not be able to have sufficient power to change laws or regulations and so institutions. Hence, few large areas with high levels of economic complexity may not have the ability to influence the whole country institutions. On the other hand, low-skilled workers in large urban agglomeration are relatively easy to substitute, reducing their contractual power in asking higher salaries, institutional changes and redistributive measures.

Third, taking back Heinrich Mora et al (2021), also Marco et al. (2022) pointed out that most complex areas are those with higher productivity and wealth that, in turn, could increase inequality. Moreover, within the same country, the labour mobility is easier; this could lead to a migration of low-skilled labour to wealthier areas to search for better working conditions and higher salaries. This phenomenon could hit poorer workers because a large supply of low-skilled labour, put downward pressure on salaries, increasing inequality. In addition to this, manufacturing industries may decide to exit the urbanized area, and with them also the middle-income workers, looking for better logistics or lower rental costs, leaving more voice and power to workers and owners of high-skilled services.

Fourth, regions with high economic complexity are those that are also more economic developed. Sbardella et al. (2017) showed that “industrial systems develop in a nested fashion” meaning that the production structure evolves accordingly to what already is produced in the area. For example, it is difficult to see a region or province devoted primarily to agricultural production that begins to produce complex products such as software or services. Changes require strong transformation that goes form high-skilled workforce to logistic to legislation. The type of productive system is closely linked to income inequality: greater complexity means more differentiated production, then the possible types of jobs increase. More type of jobs means higher inter-sectoral wage gap. Moreover, the more complex the economy, the more renumerations and the variability of salaries increase, resulting in increased inequality.

CHAPTER 4 - EMPIRICAL ANALYSIS: THE RELATIONSHIP BETWEEN ECI AND INCOME INEQUALITY IN ITALIAN REGIONS

4.1 DATASET AND DATA

The dataset used in the empirical analysis covers the years from 2004 to 2016 of 19 Italian regions and two autonomous provinces (Provincia Autonoma di Trento and Provincia Autonoma di Bolzano). It contains data that come from different sources: Italian National Institute of Statistics (ISTAT), European Statistical Office (EUROSTAT), Nifo and Vecchione (2014) that created the Institutional Quality Index (IQI), and Territorial Statistical Atlas of Infrastructure (ASTI). Data from Nifo and Vecchione (2014) and ASTI were available at provincial level; to find the corresponding regional data, provincial data were used to calculate the regional arithmetic mean. Data from ISTAT and EUROSTAT were available only at a regional level. This and the high regional heterogeneity explain the choice of an analysis between regions and not between provinces.

4.2 DEPENDENT VARIABLE

The two variables of interests in our estimations are the level of income inequality and the level of economic complexity. For income inequality, the variables used in the estimations are:

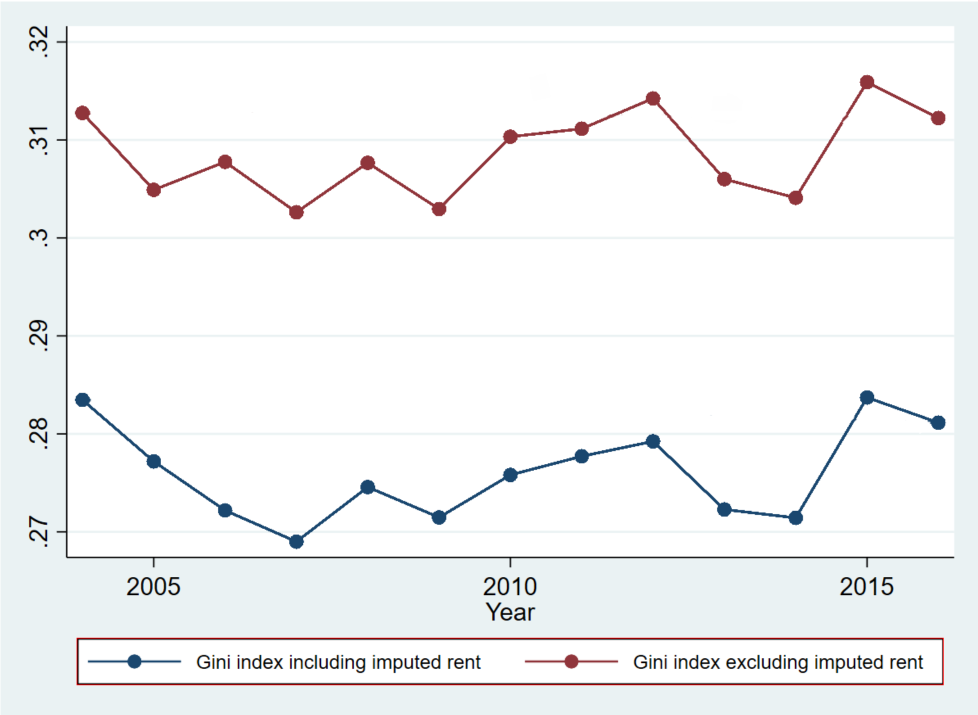
- **Gini index including imputed rent**. Source: ISTAT.
- **Gini index excluding imputed rent**. Source: ISTAT.
- **Income quintiles share ratio S80/S20**. Source: EUROSTAT.

It must be pointed out that, as explained in the methodological note that ISTAT provides, the value of the imputed rent is estimated by the owner on the basis of the price he believes should be paid to live in rent in his house. From this estimate, any interest paid is deducted on the home loan. The part of the cost of the loan intended to repay the borrowed capital is not subtracted because it corresponds to a debt reduction, i.e. an increase in the household's assets. By analogy, tenants who pay a subsidised rent, will report the difference between the market price and the rent paid. The rent is considered net of the ordinary maintenance costs, which correspond to the depreciation of the real estate stock.

Differentiating between Gini index including imputed rent and excluding imputed rent is important. Since the measures of inequality are based on the net income of individuals and families, it does not include items relating to the possession or rental of the dwelling. Therefore, in analysing distributive indexes little attention is paid to the link between economic well-being and characteristics of the house, although it represents most of the wealth for the majority of the population in Italy. A failure to consider the benefits and/or charges related to the house could leads to distorted interpretations. Differentiating between including and excluding imputed rent helps to deal with this problem. Basically, including imputed rents means considering this housing problem by attributing a monetary value to the fact that the individual/family lives in a home property. However, it must also kept in mind that not always a higher value of the dwelling, and therefore of the imputed rent, indicates an increase in the availability of potential consumption. Just think about the common case of retired people that are houseowners but receive a small pension (Farina and Franzini, 2015).

Looking at data, inclusion of the in-kind housing advantage in the concept of economic well-being reduces the absolute value of the Gini index:

FIGURE 15: estimated mean Gini index including and excluding imputed rent

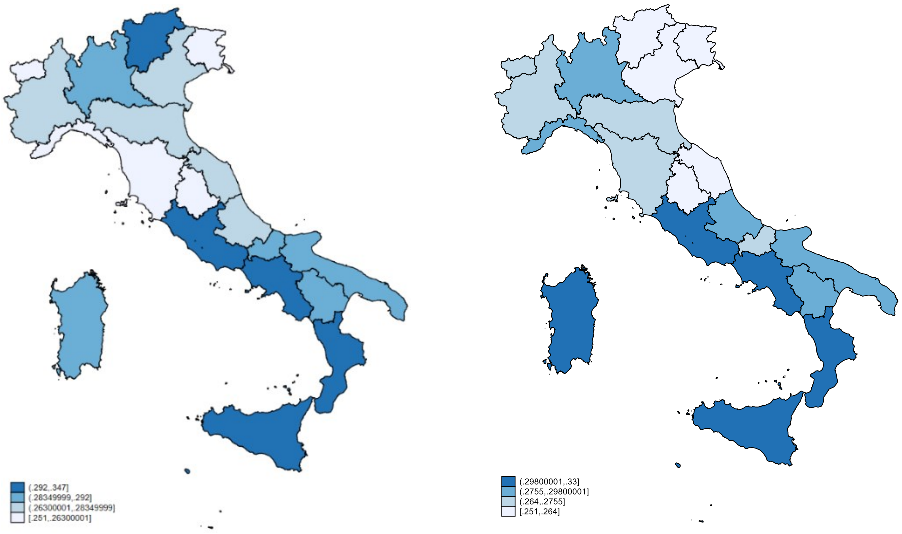


Source: authors' elaboration.

As can be seen in figure 15, the Gini index including imputed rent is lower than the one excluding imputed rent. Nevertheless, the dynamic of the two variables is similar, validating the fact that the two lines describe the same phenomenon. However, figure 15 represents the average of the Gini index of the 21 units (19 regions plus 2 autonomous provinces) in the dataset and it does not consider possible regional differences across different regions. Figure 16 and figure 17 show clearly that the Gini index including imputed rent, or the one excluding imputed rent describe the same phenomenon in relative terms, even if the absolute values are different. Indeed, comparing figure 16 a) with figure 17 a) and figure 16 b) with figure 17 b), and by comparing regions, we see that it is not so influent the fact that we include the imputed rent or do not do so: in fact, where the Gini index is high/low, it will be high/low independently of the inclusion of the rent.

Figure 16: Gini index including imputed rent

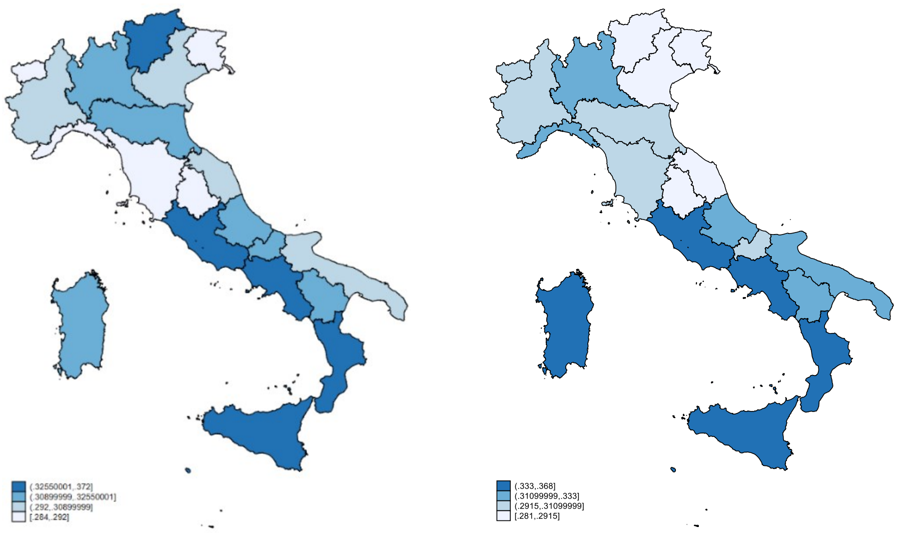
a) 2004 b) 2016



Source: authors’ elaboration.

Figure 17: Gini index excluding imputed rent

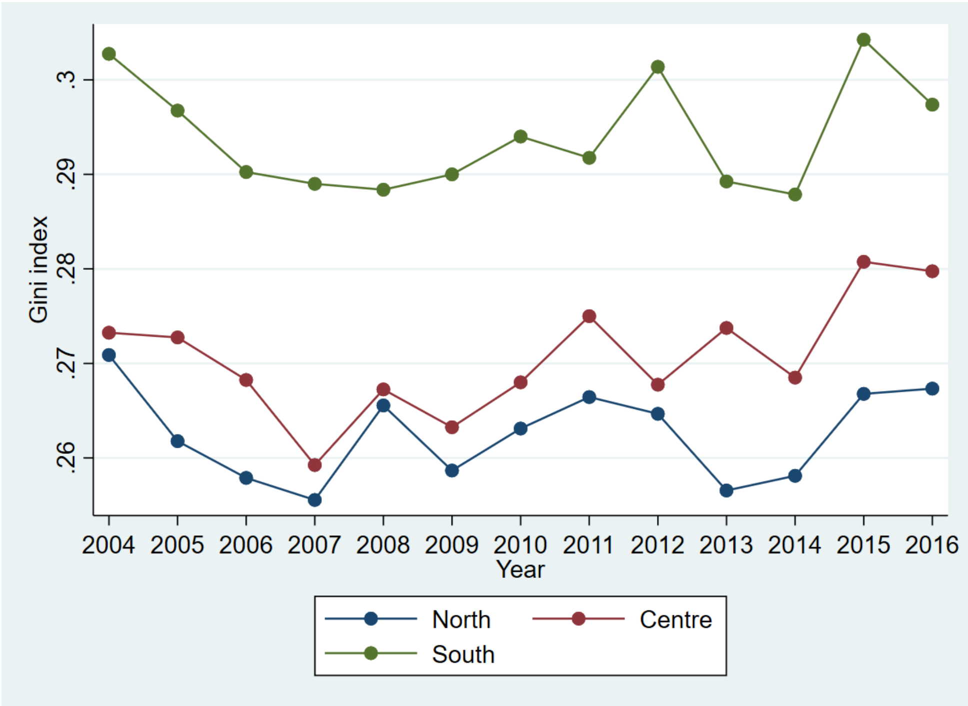
a) 2004 b) 2016



Source: authors’ elaboration.

Since we have established that the Gini index including imputed rent and excluding imputed rent broadly follow the same path, we can focus our attention in only one index. In analysing these data, we can see the same story described in chapter 2: a substantial gap between the northern regions and the southern regions. In the following figure 18, the gap is even clearer:

FIGURE 18: estimated means of Gini index (included imputed rent) by area

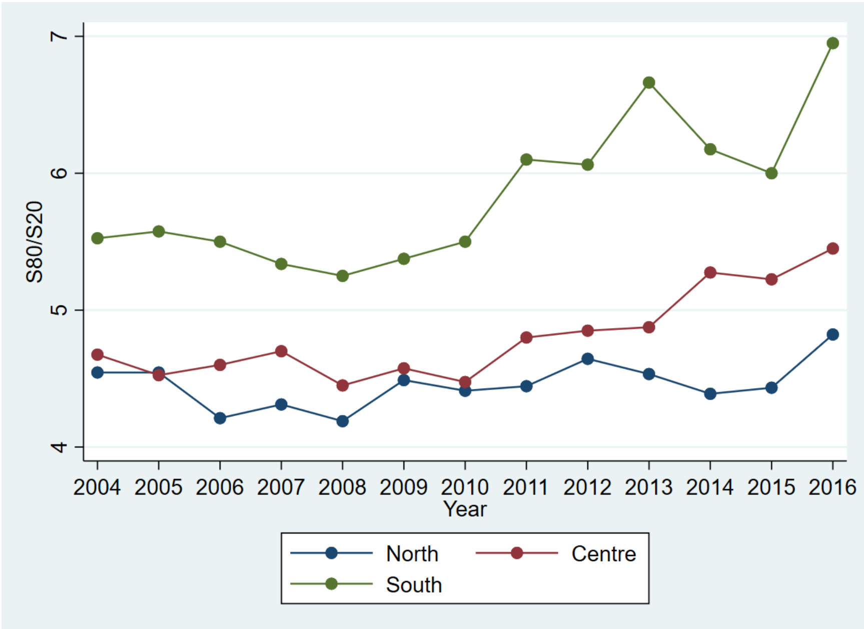


Source: authors' elaboration.

Following the ISTAT classification, in the north were included the following regions: Piemonte, Valle d'Aosta, Emilia-Romagna, Veneto, Friuli Venezia Giulia, Lombardia, Liguria and Trentino-Alto Adige. For centre: Toscana, Umbria, Marche and Lazio. For south: Abruzzo, Basilicata, Molise, Calabria, Puglia, Campania, Sicilia and Sardegna. Note that until 2012, centre regions and northern regions were sharing quite similar values, while from 2013 to 2016 also the gap of these two areas became relevant.

As mentioned, the other variable that describe income inequality is the income quintiles share ratio S80/S20. As explained in chapter 2, it is calculated as the ratio between total income of the population in the top quintile (20% of the population with the highest income) and the income of the population in the bottom quintile (20% of the population with the lowest income). For higher values, higher inequality. Even if this index is different compared to the Gini index, it tells the same story, that is in southern regions the income inequality is higher compared to centre/northern regions, and that form 2013 also the gap between northern regions and centre regions became relevant:

FIGURE 19: estimated means of S80/S20 by area



Source: authors' elaboration.

FIGURE 20: Income quintiles share ratio S80/S20

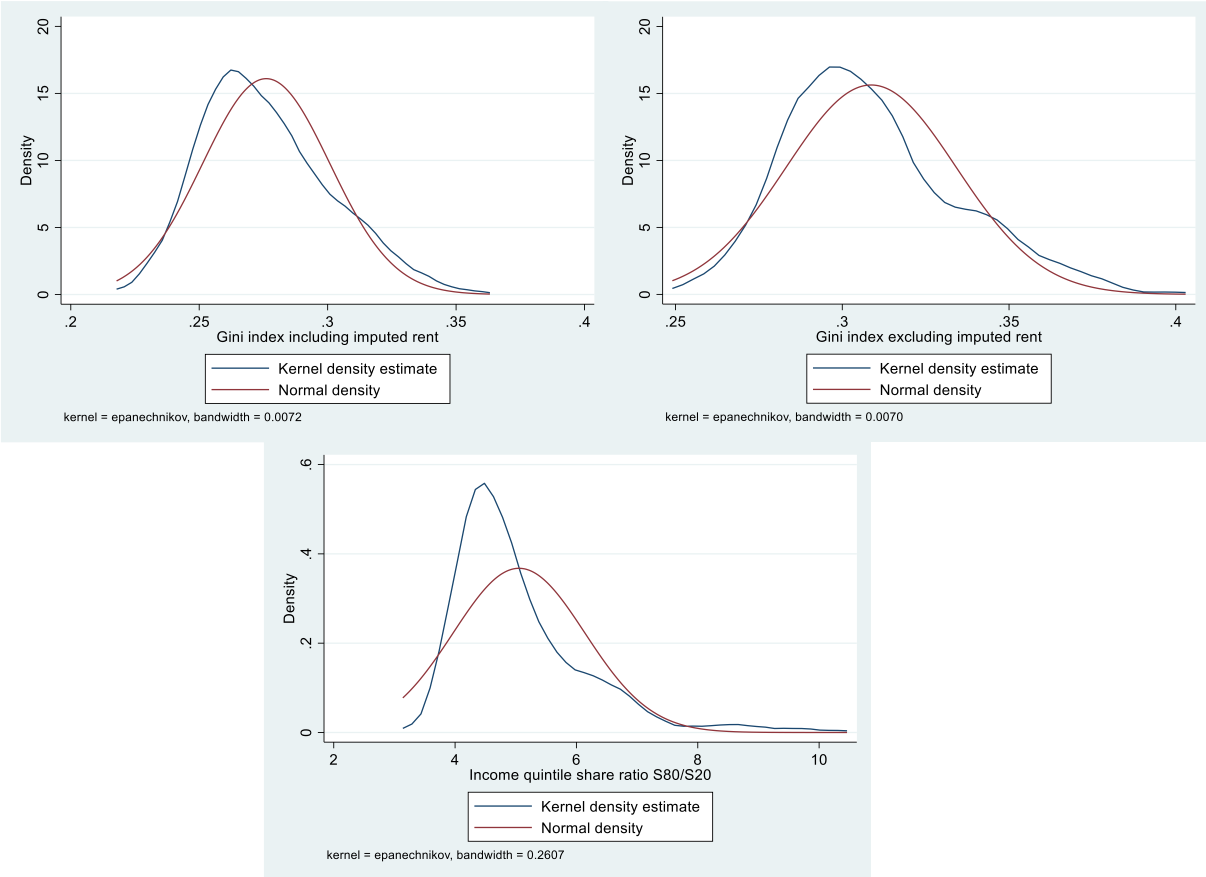
a) 2004

b) 2016

Source: authors' elaboration.

In our dataset, the indexes used to explain income inequality assume the following distribution:

FIGURE 21: kernel density estimates of income inequality data (all years)



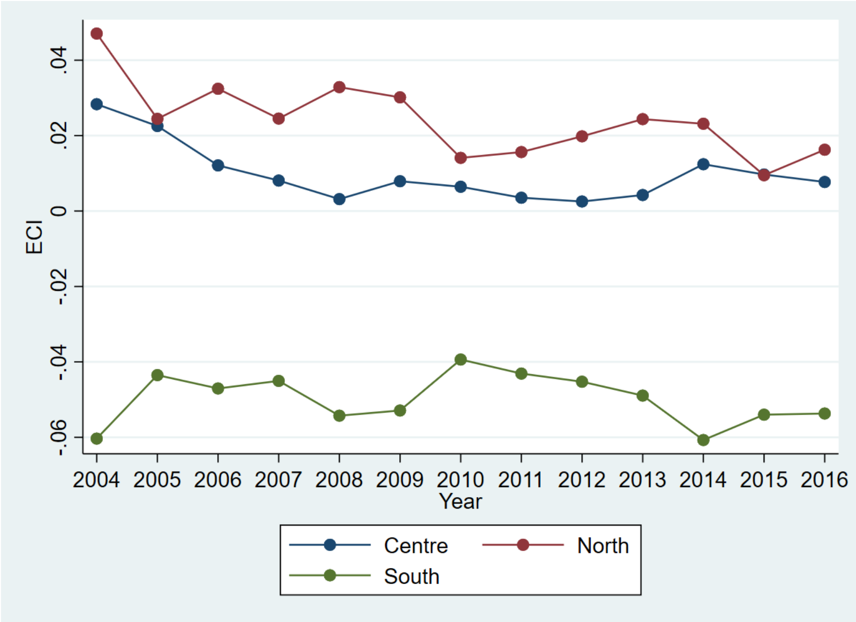
Source: authors' elaboration.

The Gini index (both including and excluding imputed rent) shows a more normal distribution, while the Income quintile share ratio, is less normal and seems asymmetric. It is necessary to keep in mind that these graphs show all the data from year 2004 to 2016, see annex 1 for the distributions divided by year. Looking at the same graphs selecting only data from the same year, the Gini indexes show a quite normal distribution, while the Income quintile share ratio shows a less normal distribution.

4.3 FOCAL REGRESSOR

The other variable of interest is the **economic complexity index** (available only from 2004 to 2016). The calculation method for retrieve this index is explained in chapter 1. The indicator was calculated by Antonietti and Burlina (2022) using ISTAT data and following the methodology developed by Hidalgo and Hausmann (2007, 2009). Here the path of this indicator over years, divided by regions.

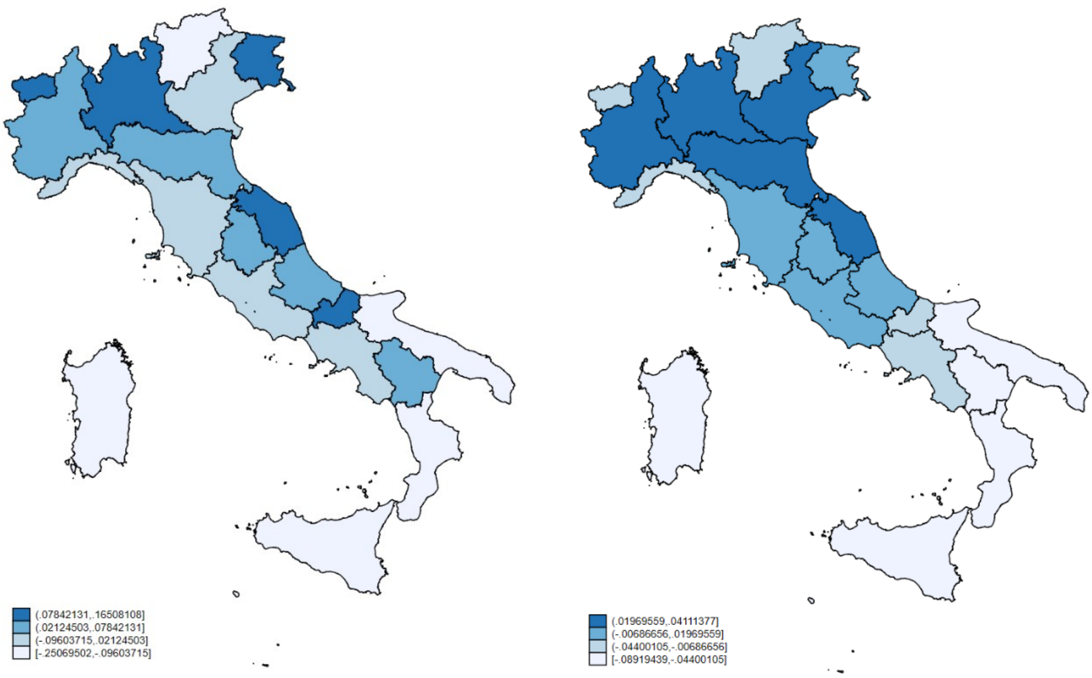
FIGURE 22: estimated means of ECI by area



Source: authors' elaboration.

FIGURE 23: Economic Complexity Index

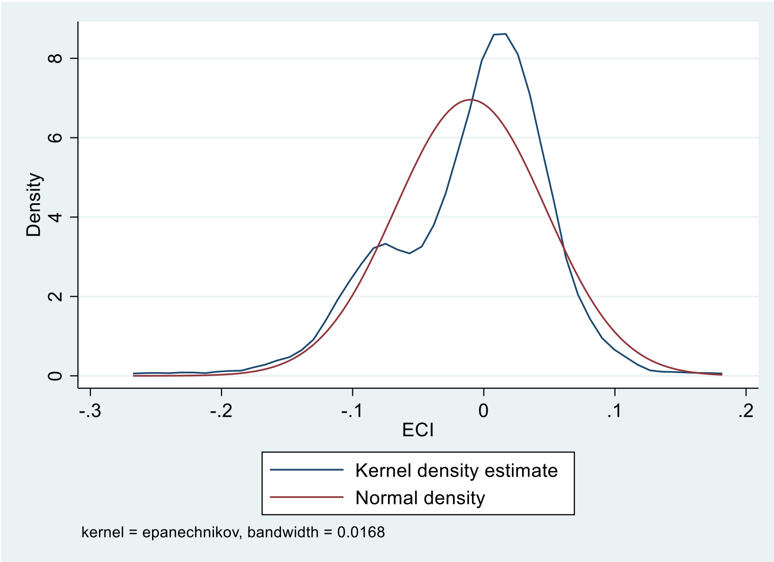
a) 2004 b) 2016



Source: authors' elaboration.

The kernel density estimation for the ECI is represented in figure 23 (see annex 1 for estimations of single years):

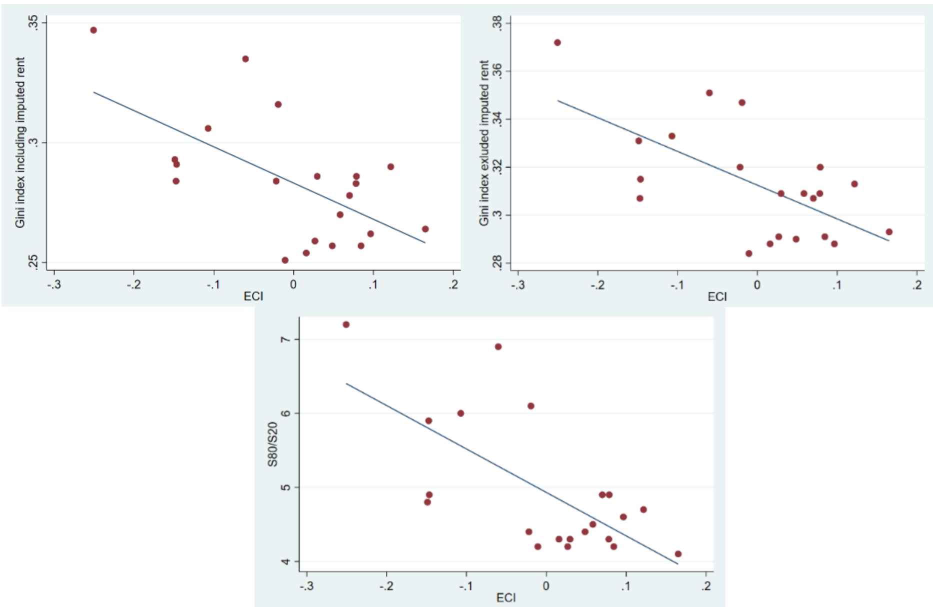
FIGURE 24: kernel density estimates of ECI (all years)



Source: authors' elaboration.

Data shows that in northern regions the ECI is higher than the ECI in southern regions. Moreover, by comparing the graphs of Gini index and the ECI, we see that southern regions have the highest levels of income inequality and the lowest levels of economic complexity, while for norther regions it seems the reverse. Indeed, the correlations between the indexes that represent income inequality and the ECI, are all negative (see annex 2 for more graphs):

FIGURE 25: Correlation between income inequality and ECI in 2004



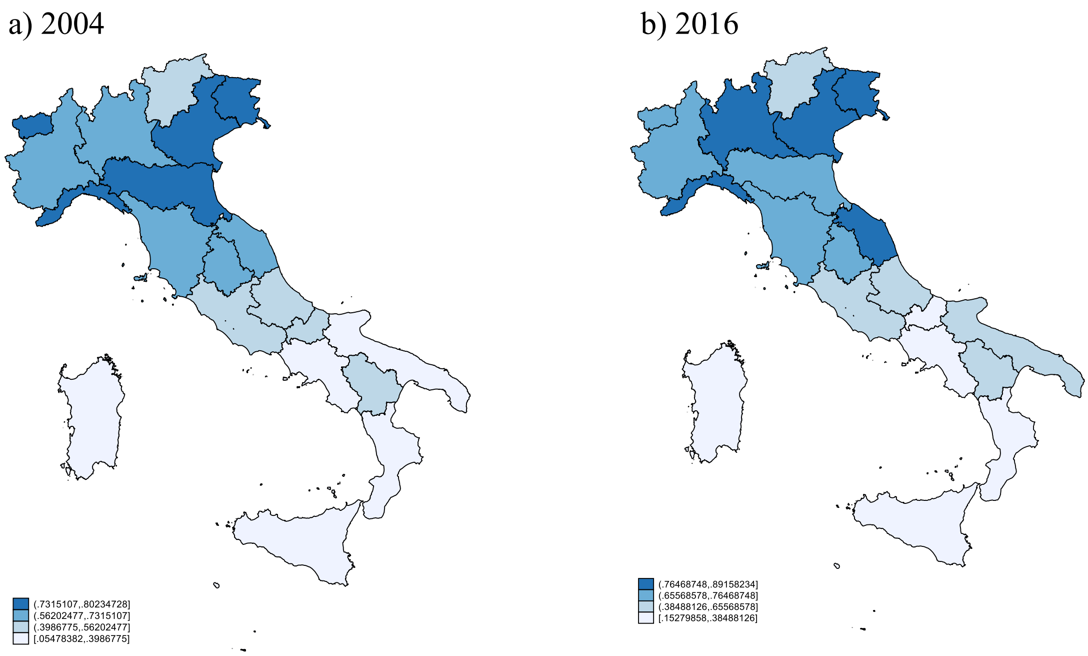
Source: authors' elaboration.

4.4 OTHER VARIABLES

The other variables used in the models will be presented are institutional quality indexes, GDP per capita (in 2015 prices), population, share of graduate on the overall population population density and entropy skills.

Institutional quality index: it contains five different indexes concern 5 major pillars of institutional quality. These are descripted by Nifo and Vecchione (2014) in IQI website: (1) voice and accountability capturing the citizens degree of participation in public elections, civic and social associations, the number of social cooperatives, the INVALSI test and the cultural liveliness measured in terms of books published; (2) government effectiveness measuring the endowment of social and economic structures in Italian provinces and the administrative capability of provincial and regional governments in terms of health policies, waste management and environment; (3) regulatory quality concerning the degree of openness of the economy, the rate of firms mortality, indicators of business environment and business density; (4) rule of law summarizing data on crime against persons or property, magistrate productivity, trial times, tax evasion and shadow economy; (5) corruption collecting data on crimes against the Public Administration, the number of local administrations overruled by the federal authorities and the Golden-Picci Index. These indexes are summarized by the authors in the IQI index.

FIGURE 26: IQI index



Source: authors’ elaboration.

Not surprising, as mentioned in previous chapters, it seems that a relationship exists between the ECI and the IQI. Even if the direction of the causal relationship has not proved by past research, it can be noticed that regions with higher ECI have also a higher IQI, meaning that regions with a high economic complexity have a higher institutional quality index and vice versa.

Gross Domestic Product (GDP) per capita, in 2015 prices, that comes from ISTAT dataset. This variable also shows the North-South gap recorded for the IQI, with northern regions that shows higher levels of GDP per capita, while southern regions lower levels.
Population, that comes from ISTAT dataset and it is the total population registered at 1st of January.
Share of graduates, that comes from ISTAT dataset and it is the share of graduates on the overall population.
Population density, that comes from ASTI dataset and it is the population for every 1 km^2.
Entropy skills, calculated in Antonietti and Burlina (2022). It represents the entropy (or the variety) of skills at regional level; referring to Hidalgo and Hausmann's words, it is a measure of how many types of legos there are in an economy.

TABLE 8: summaries statistics of the data

Variable	Obs.	Mean	Std. Dev.	Min	Max
Gini index included imputed rent	336	.2760952	.0247769	.225	.356
Gini index excluded imputed rent	336	.308663	.0255202	.256	.396
Income share ratio	336	5.053846	1.085022	3.4	10.2
ECI	336	-.010144	.0573476	-.250695	.1650811
IQI	336	.5863444	.2453343	.0547838	1
GDP per capita (2015 prices)	336	28441.94	7768.012	16035.75	43732.97
Population	336	2826758	2410139	121692	9958447
Share of graduate	336	9.381028	1.917764	5.432587	15.91089
Population density	336	175.9051	108.3741	37.65819	426.2866
Entropy skill	189	7.386708	.3250479	6.18226	8.481618

Source: authors' elaboration.

These data are summarized over the years from 2004 to 2016.

4.5 METHODOLOGY

These simple correlations shown above, indeed, could be spurious correlations: applying a cross sectional method, we expect to confirm these results. By applying this kind of methodology, we are trying to find the effect of the ECI on the inequality. However, as shown in the previous sections, in regions with higher ECI we see a lower level of inequality, and by adding some controls, we are not dealing with the possible biases that are in place.

Referring to Angrist and Pischke (2009), we are not in an ideal experiment in which we are randomly assigning the treatment. In this case the treatment is also affected by the peculiar characteristics of the regions. We should search, then, for more robust estimates.

Before running the regressions that will follow, every variable is transformed in natural logarithm. This practice is important both for statistical and for economic proposes: the residuals can have a skewed distribution and, by using this transformation, we should obtain more symmetrically distributed residuals. Similarly, the idea behind the transformation is to remove heteroscedasticity of the data. Moreover, using this kind of change, the interpretation log-log is as following: if we change x by one percent, we'd expect y to change by β1 percent.

The dependent variables that we will look at are different indices of income inequality, all in natural logarithm:

- Gini index included imputed rent, abbreviated to *Gini rents;*
- Gini index excluded imputed rent, abbreviated to *Gini no rents*;
- Income share ratio, abbreviated to *Income 80/20*.

The explanatory variables are all in natural logarithm or are standardized to mean zero:

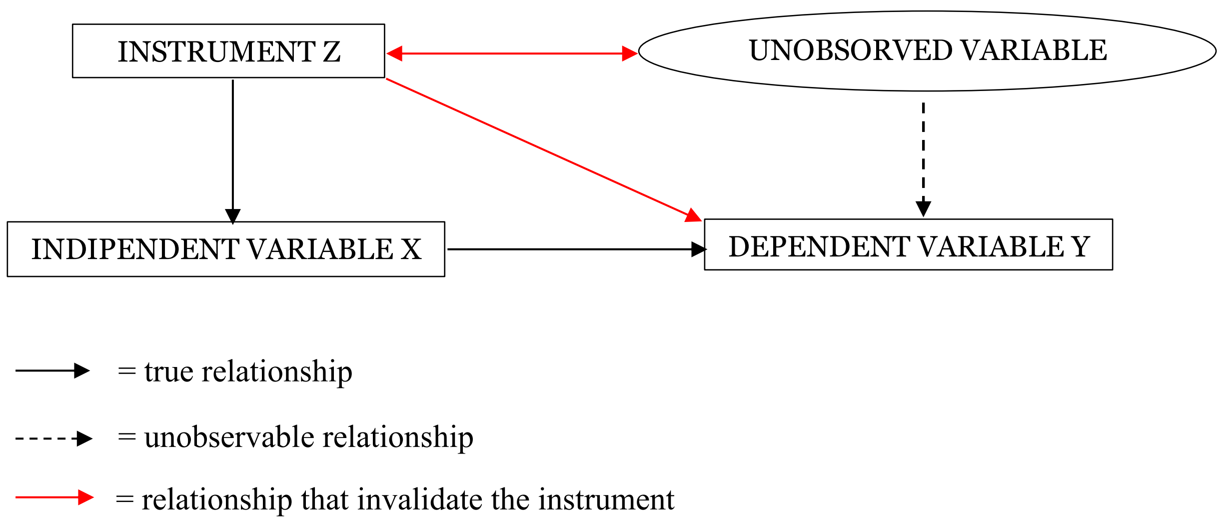
- Economic complexity index (*ECI*);
- Population (*pop*) and population density for every 1 km^2 (*popdens*);
- Share of graduates on the overall population (*sharegrdt*) to have a measure of human capital;
- An institutional quality index (*iqi*);
- GDP per capita (*gdppc*).

As said before, using OLS estimation strategies have two main problems: the first one is that it is very likely that we have a selection bias or an omitted variable bias. Formally, the omitted variable bias can be expressed by:

$$\widehat{\beta_i} = \beta_i + \gamma \frac{Covariance(X_i;z_i)}{Variance(X_i)} \tag{7}$$

Where $\widehat{\beta_i}$ is the estimated parameter, β_i is the real coefficient, X_i is the explanatory variable and z_i is the omitted variable. Indeed, if there are endogenous regressors in a model, it will cause OLS estimator to fail, as one of the assumption OLS is that there is no correlation between an explanatory variable and the error term. However, in our data, regions with higher GDP per capita, higher institutional quality, greater economic complexity are the northern regions which have always been regions with lower income disparities. This means that the estimations calculated with OLS have not a causal interpretation or, in other words, that higher economic complexity does not causes lower income inequality.

To solve this problem, we can change estimation strategy by using a Fixed Effect model or a Random Effect model: suppose that there are time-invariant regional characteristics correlated with the ECI which are unobservable or omitted from the data. Some intrinsic regional characteristics that are not measurable, such as the "informal" rules made by habits, social or cultural norms (the z_i), could affect the level of economic complexity through, for example, the entrepreneurs' propensity of innovation or diversification. The omission of these variables leads to a bias that could be larger enough to reverse the results. Another solution is to use an Instrumental Variable (IV) approach: this solution needs a strong and valid instrument. The challenge is to find one or more instruments that influence the inequality only through the Economic complexity. The instrument (Z) should affect the dependent variable (Y) only through the independent variable (X). Any other relationship invalidates the instrument:



In our case, the X is the ECI, the Y is the income inequality, and the instrument is the skill entropy/variety.

The skill entropy or, more precisely, the entropy in the labour force's skills, is calculated by Antonietti and Burlina (2022) as explained in the following section.
They considered two datasets: the INAPP-ISTAT survey on occupations ("Indagine Campionaria sulle Professioni"- ICP), and the ISTAT's Labour Force Survey (LFS). The ICP dataset contains information on the tasks, skills, attitudes to work, and working conditions for approximately 800 job titles, as classified by the International Standard Classification of Occupations (ISCO), obtained through a sample survey of 16,000 workers. The LFS dataset contains information on many labour market outcomes and indicators (e.g. work experience, employment status, main job characteristics, job seeking…). Then, they proceeded to calculate the indicator of entropy in the labour force's skills. From the ICP they calculate the average level of skills in the available job categories in Italy (a total of eight categories). From the LFS they took the distribution of occupations, and they calculate the regions' employment shares by type of occupation. These shares are used as weights to calculate the average level of each skill in each region. Finally, they compute the skill entropy as follows:

$$Skill\ entropy_r = \sum_{s=1}^{S} N_s log_2 \left(\frac{1}{N_s}\right) \tag{8}$$

where N_s is the average level of each skill *s* in the region *r*.

Indeed, a greater variety of skills/competences should foster a greater recombinant potential of knowledge, increasing the likelihood of developing exclusive and sophisticated products and finally increasing the degree of complexity of the system (Antonietti and Burlina, 2022). The skill variety is a valid instrument, since the effect on income inequality passes only through the economic complexity; this can be checked also looking at the correlations between income inequality and our instrument:

TABLE 9: correlations between income inequality and skill entropy

	Gini index included imputed rent	**Gini index excluded imputed rent**
Skill entropy	-0.0529	-0.0377

Source: authors' elaboration.

These correlations are not statistically significant, suggesting that between the two variables there is not a direct link. In fact, the distribution of these skills or capabilities does not in itself influence income distribution, and no literature has been found to support this link. However, the distribution of these skills directly affects the production of products and their degree of complexity. Then we can conclude that the skill variety effect on inequality is not direct, and it passes from a greater degree of complexity of goods and services of the region. In the estimations below (see table 16) we see that the instrument is also a strong instrument, meaning that the effect of skill variety on economic complexity is significant.

The second problem we encounter is the time dimension. Even if we were estimating a model without selection bias, it is not plausible that the effect of economic complexity on inequality is immediate. Indeed, what is more likely is that the effect of economic complexity on income inequality is postponed of some years. The income adjustment mechanism described in previous chapters due to higher or lower economic complexity, needs some time to happen: for example, the production structure could affect positively or negatively the occupational choice of workers, changing their bargaining power or the possibility of having strong and effective unions. In addition, higher level of economic complexity could provide new jobs or new learning opportunities, that could affect the wage distribution. Changes in production structures or the creation of new jobs and learning opportunities need some time to influence the wage structure of workers. To solve the second problem, we consider that the economic complexity could affects the income inequality with some years of delay, i.e. we put lagged variables and, for the IV model, that the skill entropy (our instrument) affects economic complexity after one year.

By combining these two solutions, we can run a Fixed Effect model (or a Random Effect model) and an IV model with lagged independent variables. The only choice that remains now is the choice between RE of FE. If effects are fixed, then the pooled OLS and RE estimators are inconsistent, while the FE estimator is not. In the FE model, to addressing consistency, we need some assumptions. Consider the following model:

$$y_{it} = \beta_0 + \beta_1 X_{it} + u_i + v_{it} \tag{9}$$

In this model the error term has two components: the time-invariant component u_i and the time-varying component v_{it}; u_i can be eliminated through differencing and v_{it} is an idiosyncratic error, meaning that $E(v_{it} \mid u_i, x_{it}) = 0$. RE models and pooled OLS models, need also the assumption that the time invariant component of the error term u_i, is distributed independently of X_{it}, that means $E(v_{it} \mid x_{it}) = 0$. In summary, FE models need less assumptions compared to RE models.

To make the most appropriate choice between FE or RE, we can run the Hausmann test (1978). The Durbin-Wu-Hausmann test, is a statistical test used to detect endogenous variables in the specified model. In this context, it can also help to choose between FE model and RE model. The null hypothesis is that the preferred model is RE model; the alternative hypothesis is that the preferred model is FE. Basically, the null hypothesis is that there is no correlation between the time-invariant error (u_i) and the variables in the model. If we reject the null, it means that there is correlation and then the assumption of $\mathrm{E}(\mathrm{v}_{it} \mid \mathrm{x}_{it}) = 0$ is violated. However, the Hausmann test needs homoskedasticity of the residuals, otherwise the resulting test would have an asymptotic size smaller or larger than the nominal size of the test. Testing for heteroskedasticity of the models estimated both with FE and RE, of the restricted and the unrestricted model, we found evidence of heteroskedasticity. Given the heteroskedasticity, the models will be estimated with robust standard errors. Therefore, we cannot perform the Hausmann test. An alternative to check which model is preferred between FE or RE is proposed by Wooldridge (2002). This test of fixed vs. random effects can also be seen as a test of overidentifying restrictions; in presence of homoscedasticity this test statistic is asymptotically equivalent to the usual Hausman fixed-vs-random effects test, and the null hypothesis is equal to the one of Hausman test. The result is:

TABLE 10: Test of overidentifying restrictions, fixed vs random effects.

H_0 = the preferred model is random effects

	Chi-sq (4)	P-value
Sargan-Hansen statistic	103.955	0.0000

Source: authors' elaboration.

Suggesting that the null hypothesis must be rejected. Our estimations, then, will use the Fixed effect model. By using a FE model, we are controlling for all time-invariant differences between the regions, so the estimated coefficients cannot be biased because of omitted time-invariant characteristics.

A third problem we encountered, is that in largescale micro-econometric panel datasets is common that cross-sectional and temporal dependencies can arise. Ignoring the possible correlation of error terms over time and between subjects, can lead to biased statistical inference. We established before that we are in presence of heteroskedasticity, and we can adjust this by estimating robust standard errors. However, it is reasonable to think that we are also in presence of autocorrelation (Hoechle, 2007). In the following table the statistical test for autocorrelation:

TABLE 11: Test for autocorrelation

H_0 = no first-order autocorrelation

	F (1, 20)	P-value
`xtserial`	22.96	0.0001

Source: authors' elaboration.

For all our estimations, we reject the null, so we are in presence of first order autocorrelation. Moreover, it is difficult to think that disturbances of our panel model are cross-sectionally independent. In fact, our regional data are likely to show spatial correlation between regions. For these reasons, and in order to reduce as much as possible the eventual bias that is still in place, we will use Driscoll and Kraay's covariance matrix estimator for the FE estimations (Hoechle, 2007).

We are going to estimate the coefficients of one restricted model (our baseline) and the non-restricted model (with different covariates). In addition, we checked if time fixed effects are needed when running a FE model. It is a joint test that has as null hypothesis to test if the dummies for all years are equal to 0. Here the result of the non-restricted model:

TABLE 12: Test if all year dummies are jointly not significant.

H_0 = year dummies are equal to 0

	F (12, 20)	P-value
`Testparm` i.year	7.08	0.0001

Source: authors' elaboration.

For all our estimations, we reject the null, so we will include year fixed effects.

4.6 EMPIRICAL ESTIMATIONS AND RESULTS

The baseline model is:

$$\text{Income inequality}_{it} = \beta_0 + \beta_1 ECI_{it} + \beta_2 \text{i.year} + u_i + \text{v}_{it} \qquad (10)$$

Where i is the region, and t is time.

Four different models are estimated, differing by the t of the ECI variable: the first model has t, the second $t-1$, the third $t-2$ and the last $t-3$.

TABLE 13

	(1) **FE no lags** Gini rents	(2) **FE 1 lag** Gini rents	(3) **FE 2 lags** Gini rents	(4) **FE 3 lags** Gini rents
ECI	-0.0483	-0.0369	0.173*	0.132
	(0.106)	(0.0920)	(0.0696)	(0.0652)
Year FE	✓	✓	✓	✓
Obs.	273	273	273	273
F-stat.	0.208	0.161	6.16***	4.09**
Within R^2	0.1612	0.1946	0.2249	0.2144

Driscoll and Kraay's covariance matrix, standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: authors' elaboration.

As shown by the results, our variable of interest is statistically significant at 0.05 level only when we consider $ECI_{i,t-2}$, so only when the Economic Complexity index is lagged of 2 periods. This means that our variable of interest influences income inequality 2 years after it changes. This confirms our story that incomes adjust after some years. Our model also shows that an increase in Economic Complexity determines an increase in income inequality in a within country analysis. This result is robust also considering other measures of income inequality, such as the Gini index excluded imputed rent and the Income share ratio. The results are shown in annex 3. The only difference we registered is that, when we put as dependent variable the Gini index excluded imputed rent, our variable of interest is significant also when it is lagged of 3 periods, suggesting that Economic Complexity has an effect after 2-3 years, while when we put as dependent variable the Income share ratio, the ECI is significant only when it is lagged of 3 years.

Since it is confirmed that the effect of the economic complexity on income inequality is lagged of 2-3 years, our non-restricted model will look at covariates lagged of different periods. Adding the controls, the non-restricted model is:

$$\text{Income inequality}_{it} = \beta_0 + \beta_1 ECI_{i,t-n} + \beta_2 pop_{i,t-n} + \beta_3 sharegrdt_{i,t-n} + \beta_4 iqi_{i,t-n} + \beta_5 \text{i.year} + u_i + \text{v}_{i,t} \quad \text{With } n = 1, 2, 3. \tag{11}$$

And the estimations:

TABLE 14

	(1) **FE 1 lag** Gini rents	(2) **FE 2 lag** Gini rents	(3) **FE 3 lag** Gini rents
ECI	-0.0399	0.198*	0.162**
	(0.0966)	(0.0773)	(0.0519)
pop	0.171	0.308	0.231
	(0.156)	(0.194)	(0.176)
sharegrad	-0.0281	0.0887	0.217*
	(0.0551)	(0.0607)	(0.0876)
iqi	0.00546	-0.00853	-0.00400
	(0.0173)	(0.0191)	(0.00750)
Year FE	✓	✓	✓
Observations	273	273	273
F-stat.	1.923	2.653*	28.56***
Within R^2	0.1985	0.2418	0.2552

Driscoll and Kraay's covariance matrix, standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: authors' elaboration.

These estimations confirm the results of the restricted model, meaning that higher values of economic complexity increase income inequality. Moreover, adding the controls, we see that the effect of economic complexity is significant for 2-3 lags. The general results are robust for all the three indexes of income inequality (see annex 4 for the estimations). However, our controls are not statistically significant, except for the share of graduates on the overall population (sharegrdt) only when we consider the Gini index as dependent variable in the model estimated with 3 lags.

In the following estimations we added other controls, to see if our results are robust to different covariates; here are only the results with the dependent variable the Gini index included imputed rents, to see all the estimations, see annex 5.

TABLE 15

	(1) **FE 1 lag** Gini rents	(2) **FE 2 lag** Gini rents	(3) **FE 3 lag** Gini rents	(4) **FE 1 lag** Gini rents	(5) **FE 2 lag** Gini rents	(6) **FE 3 lag** Gini rents
ECI	-0.0367 (0.0990)	0.206^{*} (0.0793)	0.170^{**} (0.0529)	-0.0328 (0.0906)	0.191^{*} (0.0697)	0.157^{*} (0.0568)
popden	0.177 (0.170)	0.356 (0.209)	0.326 (0.199)	✗	✗	✗
pop	✗	✗	✗	0.213 (0.136)	0.335 (0.164)	0.204 (0.154)
sharegrad	-0.0290 (0.0548)	0.0862 (0.0618)	0.213^{*} (0.0859)	-0.0130 (0.0497)	0.102 (0.0643)	0.207^{*} (0.0835)
iqi	0.00513 (0.0175)	-0.00929 (0.0190)	-0.00484 (0.00749)	✗	✗	✗
gdppc	✗	✗	✗	-0.124 (0.126)	-0.0893 (0.134)	0.0792 (0.125)
Year FE	✓	✓	✓	✓	✓	✓
Obs.	273	273	273	273	273	273
F-stat.	1.92	2.65^{*}	28.56^{***}	3.00^{*}	6.92^{***}	6.37^{**}
Within R^2	0.1985	0.2418	0.2552	0.2030	0.2435	0.2568

Driscoll and Kraay's covariance matrix, standard errors in parentheses.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Source: authors' elaboration.

All the results do not change when we included the additional control variables: the nature and significance of the estimated effects of economic complexity on income inequality are robust to all our controls. It is evident, however, that our controls do not add something relevant to our estimations. Except for the share of graduates on overall population (only in the models lagged of 3 years), all our covariates are not significant. This means that it is sufficient, for our purposes, to consider only the ECI and the fixed effects. Indeed, adding GDP per capita, we run into a problem of multicollinearity, since the ECI and the GDP per capita are highly correlated (more than 0.8): we highlighted in chapter 1 that ECI is a very good predictor of the economic growth and GDP per capita, insomuch that some authors used the two indicators as alternative for this reason. We have similar problems considering the IQI (more than 0.6 correlation). Several studies have shown that good institutions determine higher levels of economic complexity (see chapter 1), then an institutional quality measure will not add much, since it is already captured by the ECI.

This result is also confirmed by the IV estimation by considering the Gini indexes:

TABLE 16

	(1) **First Stage**	(2) **GMM** Gini rents	(2) **GMM** Gini no rents
Skill entropy$_{t-4}$	0.0218**** (0.00629)	✗	✗
ECI$_{t-3}$	✗	1.218* (0.737)	1.644* (0.865)
Year FE	✓	✓	✓
Region FE	✓	✓	✓
Observations	252	252	252
F stat.	66.23****	37.40****	31.80****
R^2	0.8602	0.7414	0.6710

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Source: authors' elaboration.

At the first stage we observe that our instrument is highly significant, the Kleibergen-Paap Wald F statistic is lager then 10, then it is a sufficiently strong instrument. At the second stage (estimated through a GMM estimation) we see that the Economic complexity is statistically significant at a 10% level, for both Gini index including and excluding imputed rents. As the previous results, we see that higher levels of ECI increase income inequality. Moreover, we see that the exogeneity test (H_0: the specified endogenous regressors can be treated as exogenous) accept the null hypothesis at 10% significance level. We can consider the ECI_{t-3} as exogenous, therefore using a FE model we obtain more efficient estimates.

Then, our results are suggesting that, analysing Italian regions, we find that increasing the economic complexity will lead to higher inequality. This is consistent with what Sbardella et al. (2017), Marco et al. (2022), and partially Zu et al. (2020) found in their research. The effect of an increment in the ECI, it is not immediate: the Gini index responds after 2-3 years, while the Income share ratio after 3 years. To explain these results, we should focus our attention to different factors. First of all, the effect of institutional factors could be relevant enough in the Italian case. As we showed in the above graphs, we have strong differences between regions. Another factor to consider is the labour mobility: indeed, between regions mobility is relatively easier compared to between countries. More economically complex regions could attract low skill workers, increasing inequality. Moreover, regions with a higher ECI are likely to have a more complex productive structure and a higher differentiation; these two aspects have an effect on inter-sectoral wage gap.

4.7 INTERPRETING THE RESULTS BY MACRO AREA

To better understand what are the aspects that are more influent in our case study, we estimated our models by differentiating by area: North-West, North-East, Centre, and South (with islands) focusing our attention on two main factors: the **institutional quality** and the **technology entropy.**

By estimating these models with the same FE methodology, we found the following results:

- For the South and Islands, the effect of the ECI is not significant, meaning that economic complexity does not affect the level of income inequality for southern regions.
- For the North-West and the Centre, the effect of the ECI is positive and statistically significant, for the North-West to 2-3 lags, for the Centre to 1-2-3 lags. This confirms our result found at national level.
- For the North-East, the effect of the ECI is negative and statistically significant to 1-2 lags. This means that in North-Eastern regions, higher economic complexity drives to lower income inequality.

Here the estimations:

TABLE 17

North-West

	(1) **FE 1 lag** Gini rents	(2) **FE 2 lags** Gini rents	(3) **FE 3 lags** Gini rents	(4) **FE 1 lag** Gini no rents	(5) **FE 2 lags** Gini no rents	(6) **FE 3 lags** Gini no rents	(7) **FE 1 lag** Income 80/20	(8) **FE 2 lags** Income 80/20	(9) **FE 3 lags** Income 80/20
ECI	0.252	0.411*	0.0698	0.228	0.470*	0.111	0.305	0.972	1.213**
	(0.273)	(0.179)	(0.180)	(0.299)	(0.168)	(0.187)	(0.622)	(0.684)	(0.337)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	65	65	65	65	65	65	65	65	65
F-stat.	0.85	5.29***	0.15	0.58	7.81***	0.35	0.24	2.02	12.92***
Within R^2	0.3268	0.3892	0.3420	0.2906	0.3627	0.3060	0.3289	0.4136	0.4133

North-East

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ECI	-1.025**	-0.415	0.419	-0.953*	-0.333	0.610	-2.058*	-1.483*	-1.093
	(0.317)	(0.215)	(0.793)	(0.363)	(0.209)	(0.731)	(0.757)	(0.616)	(0.584)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	65	65	65	65	65	65	65	65	65
F-stat.	10.49***	3.72*	0.28	6.88***	2.54	0.70	7.38***	5.79**	3.51*
Within R^2	0.3559	0.3244	0.3291	0.2480	0.2169	0.2512	0.4170	0.3824	0.3562

Centre

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ECI	0.430	0.464***	0.199	0.491*	0.503**	0.228	-0.0928	0.511	0.862*
	(0.247)	(0.104)	(0.206)	(0.188)	(0.127)	(0.209)	(0.296)	(0.480)	(0.317)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	52	52	52	52	52	52	52	52	52
F-stat.	3.04*	19.98***	0.93	6.83***	15.76***	1.19	0.10	1.13	7.38***
Within R^2	0.5900	0.5943	0.6011	0.4778	0.4794	0.4684	0.6361	0.6211	0.6159

South (with islands)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ECI	-0.0753	0.120	0.143	-0.0307	0.0926	0.146	0.117	0.0737	0.595
	(0.107)	(0.146)	(0.135)	(0.100)	(0.146)	(0.114)	(0.317)	(0.276)	(0.349)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	78	78	78	78	52	78	78	78	78
F-stat.	0.45	0.90	2.21	0.03	15.76***	0.55	0.05	3.95*	0.25
Within R^2	0.2217	0.2594	0.2896	0.2817	0.4794	0.3066	0.4210	0.4433	0.4045

Driscoll and Kraay's covariance matrix, standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

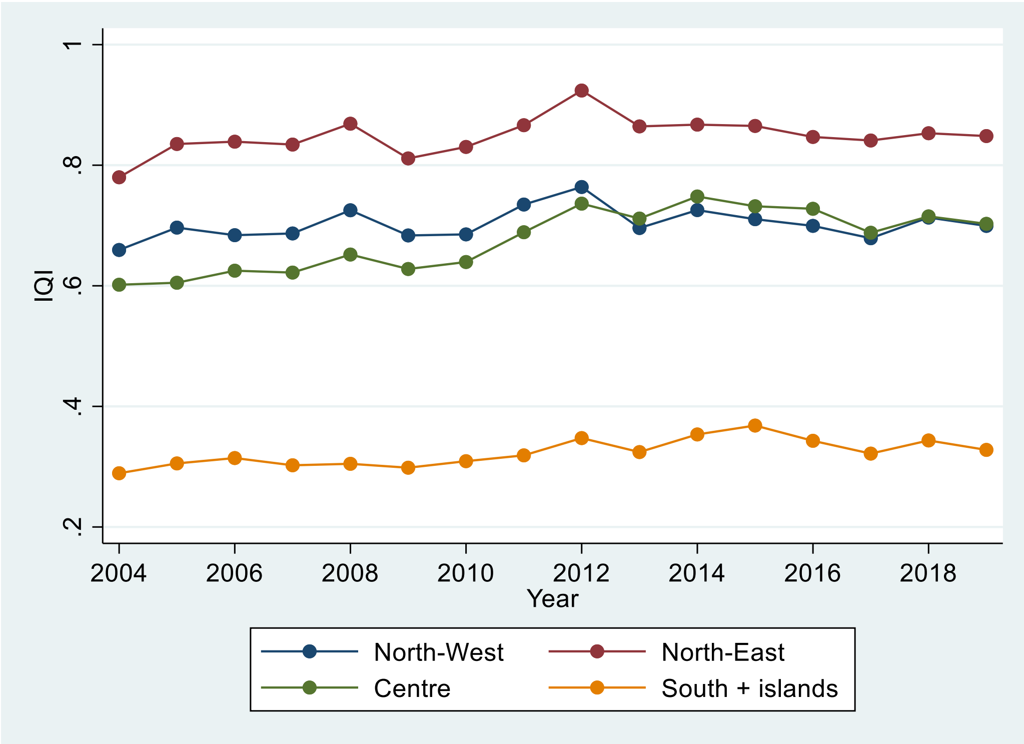
Source: authors' elaboration.

Therefore, we see that the North-west and the Centre confirm the result found by considering all the regions in the estimation, while the North-East shows an opposite result.

4.7.1 INSTITUTIONAL QUALITY

First, we see that in North-Eastern regions the value of the Institutional Quality Index (IQI) is higher compared to the other areas.

FIGURE 27: Estimated means of IQI by geographical area.



Source: authors' elaboration.

We also performed a t-test, to check if the IQI in the North-East is statistically different compared to other areas. The test we used assumes homogeneity of variance that, in our data, is respected (Seber, 1984).

TABLE 18: Test if IQI in North-East is different to other areas.

H_0 = means are equals

	P-value
Wilks' lambda	0.0000
Pillai's trace	0.0000
Lawley-Hotelling trace	0.0000
Roy's largest root	0.0000

Source: authors' elaboration

For all the tests, we reject the null, then the IQI for North-East is statistically different compared to other areas.

Higher values of IQI mean that the institutions in N-E regions work better compared to other regions. This factor is important in the wage settings mechanism, since better institutions allow for stronger unions and higher contractual power of the employees. Indeed, the specific aspects in which we found that N-E has better institutions are voice and accountability, government effectiveness, regulatory quality, and rule of law (see annex 6). For **voice and accountability**, it is reasonable to think that the degree of civic participation, and the cultural liveliness, means also higher participation to unions; **higher government effectiveness** means higher administrative capabilities of regions; better **regulatory quality** means more density of businesses and a more open economy; higher **rule of law** tells us that there is less tax evasion and a better judicial system.

4.7.2 TECHNOLOGY ENTROPY

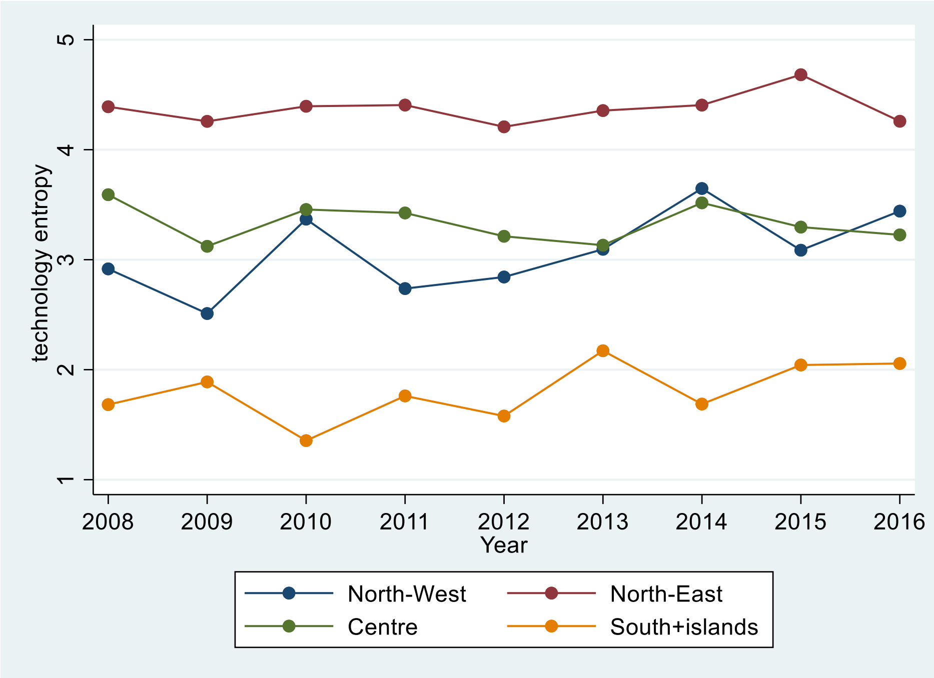
Second, we also register that the entropy (or variety) of technology is higher in North-East. This indicator, calculated by Antonietti and Burlina (2022), is proxied by the number of patents application submitted in a year. Data come from the OECD-REGPAT database. The authors used the International Patent Classification (IPC) to obtain the annual share of patents for every four-digit IPC code (P_g). Then the technology entropy measure is obtained as follows:

$$Technology\ entropy_r = \sum_{g=1}^{G} P_g log_2 \left(\frac{1}{P_g}\right) \tag{12}$$

where G is the total number of IPC codes g, and P_g is the proportion of patents for

each four-digit IPC code. In North-East, this indicator is statistically higher compared to other regions:

FIGURE 28: Estimated means of technology entropy by geographical area.



Source: authors' elaboration.

And the statistical test:

TABLE 19: Test if technology entropy in North-East is different to other areas.

H_0 **= means are equals**

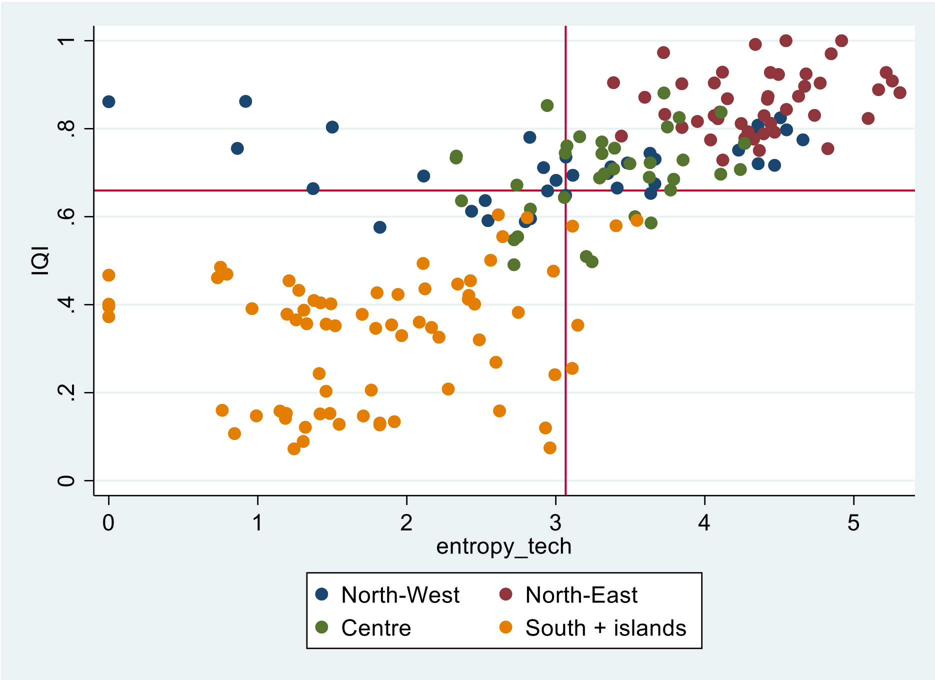
	P-value
Wilks' lambda	0.0000
Pillai's trace	0.0000
Lawley-Hotelling trace	0.0000
Roy's largest root	0.0000

Source: authors' elaboration.

4.7.3 COMBINING INSTITUTIONAL QUALITY AND TECHNOLOGY ENTROPY

Figure 27 and 28 show that, while the Central regions and the North-Western regions have similar values of IQI and technology entropy, the North-Eastern regions have significantly higher values for both of these 2 indicators. This is also evident in the following graph 28 (in red we plotted the median of the two variables):

FIGURE 29: technology entropy, IQI and the respective medians.



Source: authors' elaboration.

It is evident that North-Eastern regions are on the top right, above the median values of IQI and technology entropy. On the contrary, North-Western and Central regions have values nearer to the median, while Southern regions have values under the median.

Higher values of technological entropy, means larger possible combinations of tech capabilities that translate into more complex products, increasing the level of economic complexity. This last aspect, combined with the higher level of institutional quality could lead to a virtuous path of sectoral diversification and skill that implies higher ECI which, in turn, thanks to a high quality of institutions, leads to lower inequality.

Moreover, the value of ECI does not vary consistently between regions of N-W, N-E and Central regions. This implies that is not the "amount" of complexity that explain why we register an opposite result in N-E regions. However, we see that N-E regions have lower income inequality. We can conclude that the combination of the above four institutional aspects with high technological entropy leads to the reversal of the positive association between economic complexity and inequality. On the opposite side, in presence of median values of institutional quality and technological entropy, higher economic complexity leads to higher income inequality. Finally, if we are in an environment of low institutional quality and low technological entropy, the economic complexity has no effect on income inequality.

With the aim of exploring this results, other analyses have been conducted on the basis of two main ideas: the first is the link between economic complexity, income inequality and the shadow economy; the second is the weight of the manufacturing sector. Both were done with a focus on the dynamics in the North-East. Referring to the first point, Pham et al. (2023), performing a between country analysis, highlighted that income inequality is significantly and non-linearly linked with economic complexity and grey economy: between economic complexity and income inequality there exists a U-shaped relationship while between shadow economy and income inequality the relationship follows an inverted U-shaped. Given these findings and adding to our dataset the employment irregularity rate (the source was ISTAT), treating it as a proxy for the shadow economy, we tried to look if there exists such a relationship in analysing Italian regions. By adding this last covariate to the FE estimations, we found no significant relationships between shadow economy and income inequality when considering in the model the ECI. For the second point, the weight of the manufacturing industry, we considered two different indexes: the weight of value added of the manufacturing industry and the weight of labour units of the manufacturing industry (the source was ISTAT). Data shows that in North-East and North-East the weight of manufacturing is higher compared to other regions, with higher weigh in North-West compared to North-East. The idea was to test if the level of manufacturing influenced income inequality. Indeed, in the context of the NE, where we have a higher level of institutional quality and a higher level of technological distribution, it could be that the weight of manufacturing industry has a negative effect on income inequality. Indeed, this could have been explained by a higher unionization rate, the presence of national contracts, and perhaps a relatively greater homogeneity of jobs due to the greater weight of the manufacturing sector compared to the service sector, that ranges, for example, from restaurants to banks to transportation services; this type of activity has greater intra-sectoral wage dispersion, which is reflected in the overall income inequality of a region. However, running our estimates both considering all regions and by area, we found no strong evidence to support this explanation.

SUMMARY AND CONCLUSIONS

The concept of Economic complexity is taking more and more importance in the economic debate. Seminal contributions to this topic have been made and from these researches has been born a flourishing literature on economic growth, sustainability, human development and health, income inequality. The main factors that have been identified as main drivers of economic complexity are institutions, spatial agglomeration, technology, foreign direct investments, and entropy. Not surprisingly, the Economic complexity index is a very good predictor of economic growth and GDP per capita in all the countries: the ECI predicts long-term economic growth even in some tens of years.

Another important topic that always attracts the attentions of policy makers and economists is the income inequality, how it is determined, and which policies can reduce it. Since Kutznets (1955), the link between income inequality and economic growth was deeply studied, finding different results that pointed out different interpretation of the phenomenon. The raise of the attention for this new indicator, the Economic complexity index, has put new lights on this problem. In fact, the result of many researches have pointed out the negative or an inverted U-shaped relationship between economic complexity and income inequality in a between countries analysis: this means that higher complexity pushes income inequality down if we consider countries as the object of study, particularly if we consider high-developed countries. The investigation, then, turned towards a within countries analysis. In the analysis of different areas (regions, states…) within the same country, we do not find the same consensus. In effect, some studies pointed out that the effect of economic complexity on income inequality is positive, while others found that the effect is negative.

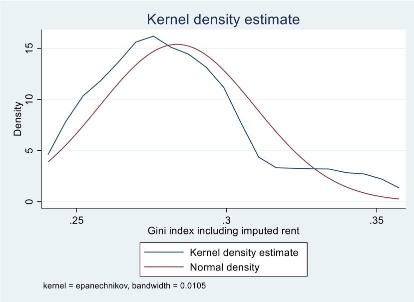
Our research fits directly into this debate: we tried to understand what the relationship is in analysing Italian regions. By running a Fixed Effect model and IV model with lagged independent variables, we found that higher level of economic complexity determines a higher level of income inequality with 2/3 years of delay. This result is robust to different controls. The reason why we chose to focus our attention on the Italian regional dimension, is because before, to our knowledge, this type of analysis was never done considering the Italian areas. The second reason is that many factors directly affecting the economic complexity of a region may differ greatly from region to region. Indeed, in the Italian case, it is well known the regional differences that exist, especially considering the North-South gap but also between other areas.

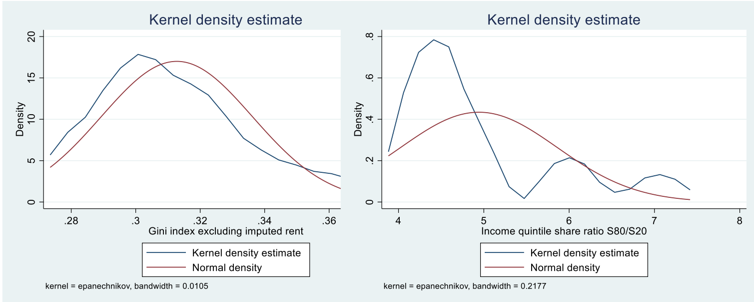
The reasons why we see an increase on income inequality when the economic complexity increases could be found in the characteristics of the workforce: indeed, more complexity may disproportionally benefit high skilled workers. In addition, greater complexity also means greater development and diversification which is reflected in more available options and thus a more differentiated income distribution. In fact, when complexity rises it is not automatic that it will influence workers skills or their learning opportunities; this influence can only be reached if the institutions are advanced enough to face with this phenomenon. If the institutions are not sufficiently developed, we may find ourselves in a situation where a worsening of disparities in the level of competence is expected due to increasing complexity, with an elite of highly skilled and talented employees on the one hand and a considerable number of unqualified and marginalized people on the other. A vicious circle could be established: along with the increased sophistication of skills, there will likely be a corresponding increase in the variety of products and tasks, as well as the talent and jobs available. There is expected to be a growing need for higher levels of capability in some areas as the economic system becomes more complicated. To handle the growing aspects of a more sophisticated system, new skills are created. Workers with advanced and transferable skills as well as learning ability and high levels of flexibility will become in high demand.

In the Italian case, the "amount" of education and learning opportunities could be considered quite similar in all Italian regions, even if there are still some North-South differentials. However, between North-East and North-West/Centre the educational offer is comparable. The difference is the quality of their institutions, that affects the quality of education. Indeed, in the case of N-E we are in a setting of high-quality institutions and high skill entropy, in particular technological entropy. These skills are formed by general and specific components, by tacit and codifiable factors. Looking at the German case, there emerges the importance of spreading these skills, which need on-the-job training from schools to factories, making them flexible and easier to transfer; in addition, it is critical that the system teach students and workers how to adapt what they already know and how to evolve these skills (Hodgson, 2003). The data show a low level of inequality and a high level of skill entropy in N-E, associated with good quality institutions leading to higher quality learning opportunities. This could be the mechanism that explains the negative relationship between ECI and inequality in that area. In other words, the ability of workers to better manage complexity at all income levels (especially at lower levels) has the effect of reducing inequality.

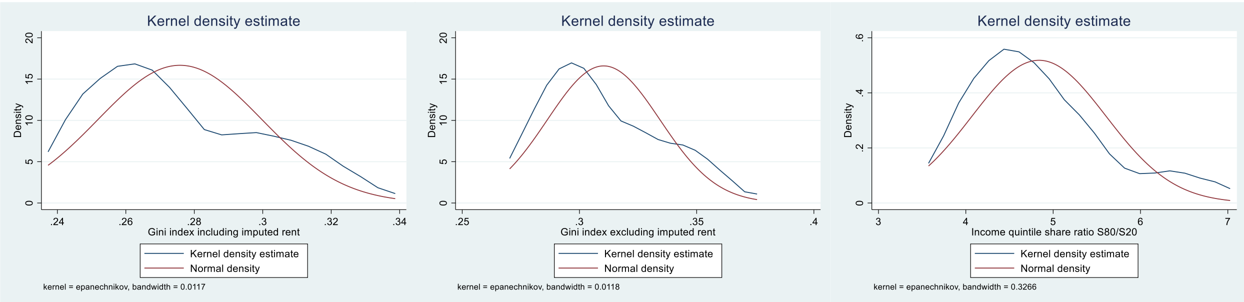
ANNEX 1: kernel density estimates of income inequality data

2004

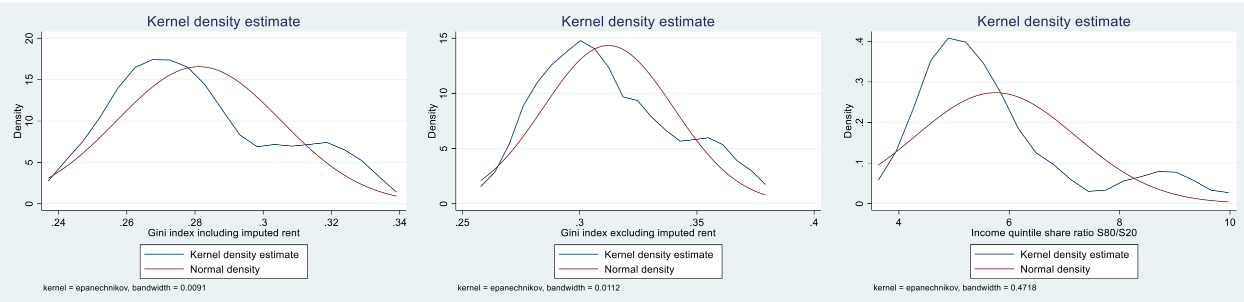




2010



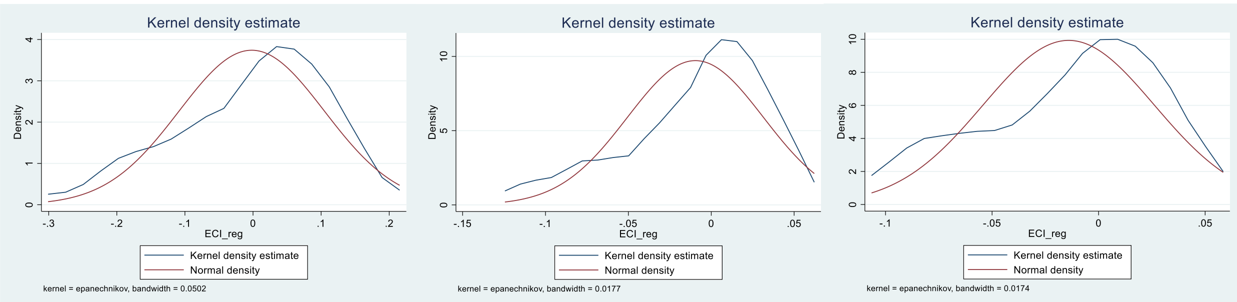
2016



Source: authors' elaboration.

Kernel density estimates of ECI data

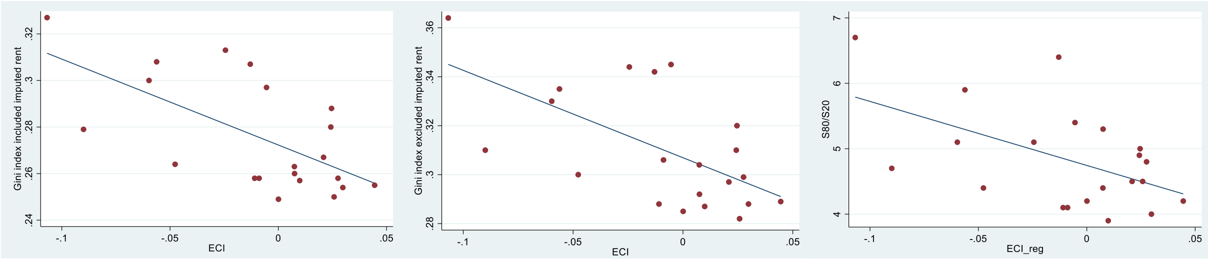
2004 2010 2016



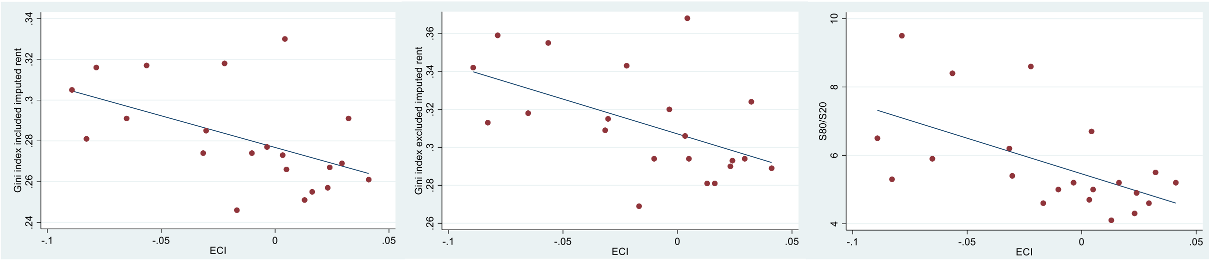
Source: authors' elaboration.

ANNEX 2: Correlation between income inequality and ECI in different years

2010



2016



Source: authors' elaboration.

ANNEX 3: Restricted model estimations

	(1) **FE no lags** Gini no rents	(2) **FE 1 lag** Gini no rents	(3) **FE 2 lags** Gini no rents	(4) **FE 3 lags** Gini no rents
ECI	-0.0445	0.00274	0.196^{*}	0.174^{**}
	(0.0941)	(0.0882)	(0.0701)	(0.0549)
Year FE	✓	✓	✓	✓
Obs.	273	273	273	273
F-stat.	0.224	0.00	7.839^{***}	10.11^{***}
Within R^2	0.1119	0.1568	0.1800	0.1780

Driscoll and Kraay's covariance matrix, standard errors in parentheses.
* p < 0.05, ** p < 0.01, *** p < 0.001

	(1) **FE no lags** Income 80/20	(2) **FE 1 lag** Income 80/20	(3) **FE 2 lags** Income 80/20	(4) **FE 3 lags** Income 80/20
ECI	0.0851	0.0575	0.236	0.717^{**}
	(0.268)	(0.264)	(0.268)	(0.193)
Year FE	✓	✓	✓	✓
Obs.	273	273	273	273
F-stat.	0.101	0.047	0.777	13.77^{***}
Year FE	0.2912	0.3076	0.3385	0.3375

Driscoll and Kraay's covariance matrix, standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: authors' elaboration.

ANNEX 4: Non-restricted model estimations

	(1) **FE 1 lag** Gini no rents	(2) **FE 2 lag** Gini no rents	(3) **FE 3 lag** Gini no rents	(1) **FE 1 lag** Income 80/20	(2) **FE 2 lag** Income 80/20	(3) **FE 3 lag** Income 80/20
ECI	-0.0143	0.214^{*}	0.198^{***}	-0.120	0.182	0.735^{**}
	(0.0982)	(0.0763)	(0.0418)	(0.246)	(0.291)	(0.223)
pop	-0.171	-0.0689	-0.108	-0.472	-0.330	-0.410
	(0.121)	(0.196)	(0.164)	(0.432)	(0.403)	(0.405)
sharegrad	-0.0323	0.0673	0.207^{*}	-0.320^{*}	-0.0267	0.164
	(0.0420)	(0.0753)	(0.0895)	(0.124)	(0.130)	(0.135)
iqi	0.00948	-0.0136	-0.00830	0.139^{***}	0.0441	-0.0149
	(0.0190)	(0.0216)	(0.00991)	(0.0201)	(0.0434)	(0.0448)
Year FE	✓	✓	✓	✓	✓	✓
Obs.	78	273	273	273	273	273
F-stat.	0.28	2.13	29.5^{***}	17.71^{***}	0.82	5.48^{**}
Within R^2	0.1620	0.1861	0.2137	0.3646	0.3435	0.3444

Driscoll and Kraay's covariance matrix, standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: athors' elaboration

ANNEX 5: Non-restricted model estimations with other controls

	(1) **FE 1 lag** Gini no rents	(2) **FE 2 lag** Gini no rents	(3) **FE 3 lag** Gini no rents	(4) **FE 1 lag** Gini no rents	(5) **FE 2 lag** Gini no rents	(6) **FE 3 lag** Gini no rents	(7) **FE 1 lag** Gini no rents	(8) **FE 2 lag** Gini no rents	(9) **FE 3 lag** Gini no rents
ECI	-0.0168	0.215^{*}	0.200^{***}	-0.00273	0.203^{*}	0.190^{**}	0.00158	0.195^{*}	0.175^{**}
	(0.0981)	(0.0778)	(0.0409)	(0.0888)	(0.0683)	(0.0465)	(0.0881)	(0.0681)	(0.0539)
popden	-0.162	-0.0227	-0.0274	✗	✗	✗	✗	✗	✗
	(0.133)	(0.208)	(0.178)						
pop	✗	✗	✗	-0.115	-0.0220	-0.115	✗	✗	✗
				(0.101)	(0.173)	(0.150)			
sharegrad	-0.0318	0.0665	0.205^{*}	-0.0125	0.0901	0.206^{*}	✗	✗	✗
	(0.0418)	(0.0764)	(0.0893)	(0.0380)	(0.0790)	(0.0851)			
iqi	0.00973	-0.0137	-0.00847	✗	✗	✗	✗	✗	✗
	(0.0192)	(0.0217)	(0.0102)						
gdppc	✗	✗	✗	-0.166	-0.155	0.0162	-0.183	-0.127	0.0717
				(0.116)	(0.120)	(0.103)	(0.123)	(0.117)	(0.121)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	273	273	273	273	273	273	273	273	273
F-stat.	0.52	2.20	34.86^{***}	0.53	4.21^{**}	8.22^{***}	0.15	4.73^{**}	12.62^{***}
Within R^2	0.1614	0.1856	0.2126	0.1702	0.1919	0.2132	0.1686	0.1855	0.1797

Driscoll and Kraay's covariance matrix, standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: authors' elaboration

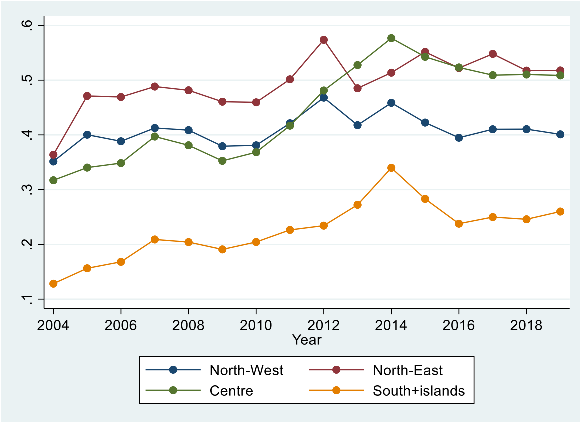
	(1) **FE 1 lag** Income 80/20	(2) **FE 2 lag** Income 80/20	(3) **FE 3 lag** Income 80/20	(4) **FE 1 lag** Income 80/20	(5) **FE 2 lag** Income 80/20	(6) **FE 3 lag** Income 80/20	(7) **FE 1 lag** Income 80/20	(8) **FE 2 lag** Income 80/20	(9) **FE 3 lag** Income 80/20
ECI	-0.127	0.179	0.735**	0.0249	0.231	0.726**	0.0520	0.233	0.715**
	(0.247)	(0.292)	(0.226)	(0.274)	(0.265)	(0.200)	(0.276)	(0.264)	(0.191)
popden	-0.457	-0.264	-0.266	✗	✗	✗	✗	✗	✗
	(0.453)	(0.428)	(0.461)						
pop	✗	✗	✗	-0.190	-0.159	-0.298	✗	✗	✗
				(0.370)	(0.379)	(0.388)			
sharegrad	-0.318*	-0.0269	0.163	-0.247*	0.0291	0.213	✗	✗	✗
	(0.123)	(0.131)	(0.138)	(0.106)	(0.109)	(0.146)			
iqi	0.140***	0.0444	-0.0148	✗	✗	✗	✗	✗	✗
	(0.0205)	(0.0439)	(0.0457)						
gdppc	✗	✗	✗	-0.772**	-0.495	-0.355	-0.874**	-0.503	-0.316
				(0.244)	(0.367)	(0.291)	(0.267)	(0.383)	(0.276)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	273	273	273	273	273	273	273	273	273
F-stat.	16.72***	0.90	4.73**	3.56*	1.19	8.71***	8.70	1.49	14.72***
Within R^2	0.3639	0.3427	0.3426	0.3666	0.3533	0.3507	0.3568	0.3528	0.3432

Driscoll and Kraay's covariance matrix, standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

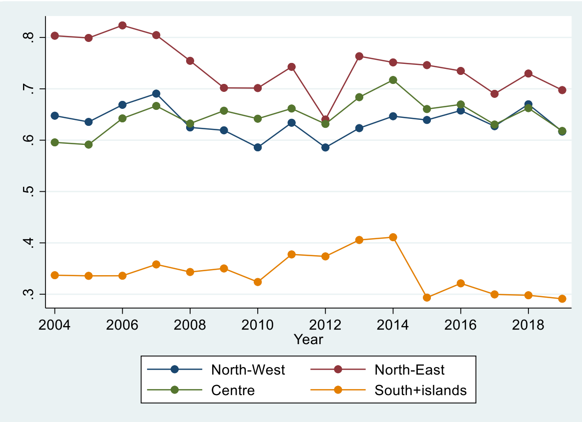
Source: authors' elaboration

ANNEX 6: Institutional factors that contribute to differentiate the N-E.

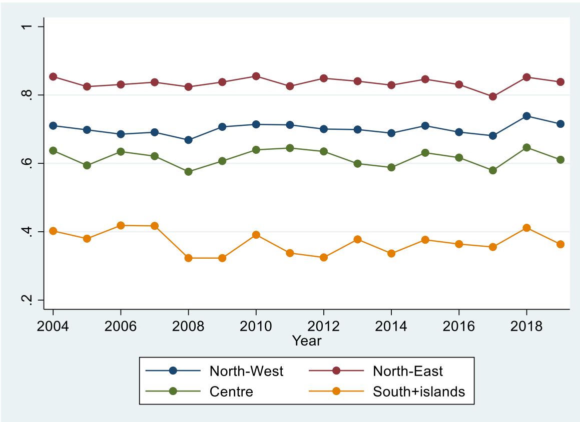
Government effectiveness



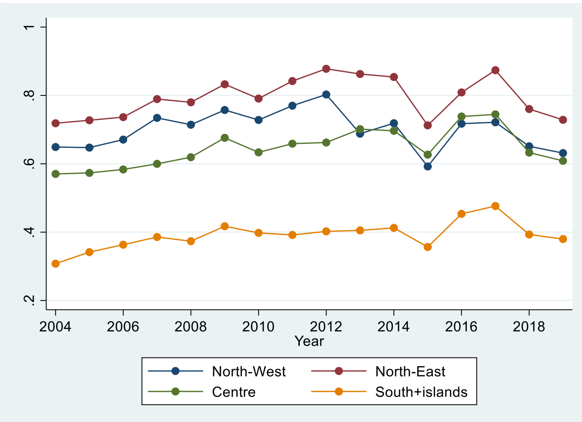
Regulatory quality



Rule of law



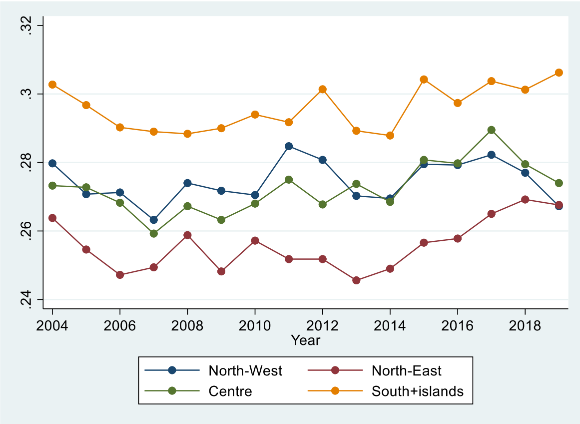
Voice and accountability



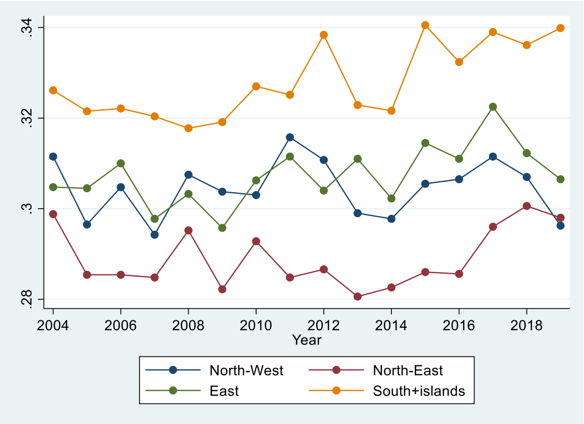
Source: authors' elaboration

ANNEX 7:

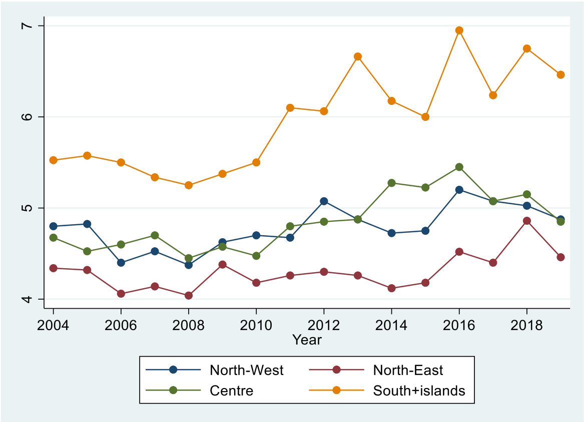
Gini index included imputed rents



Gini index excluded imputed rents



Income share ratio



Source: authors' elaboration

REFERENCES

1. Acemoglu, D. (1997). Matching, heterogeneity, and the evolution of income distribution. *Journal of Economic Growth 2* (1), 61–92.
2. Acemoglu, D., Autor, D. (2011). Skills, tasks and technologies: implications for employment and earnings. InD. Card andO. Ashenfelter (eds.), *Handbook of Labor Economics*, Vol. 4B. Amsterdam: Elsevier, pp. 1043–171.
3. Acemoglu, D., Gallego, F. A., Robinson, J. A. (2014). Institutions, human capital, and development. *Annu. Rev. Econ*., 6(1), 875-912.
4. Acemoglu, D., Johnson, S. (2005). Unbundling Institutions. *Journal of Political Economy, 113*(5), 949-995.
5. Acemoglu, D., Johnson, S., Robinson, J. A. (2005). Institutions as a Fundamental Cause of Long-Run Growth. In P. Aghion, S. N. Durlauf (Eds.), *Handbook of Economic Growth* (Vol. 1, pp. 385-472): Elsevier.
6. Acemoglu, D., Robinson, J. A. (2012). Why nations fail: The origins of power, prosperity, and poverty. *Currency*.
7. Aghion, P., Howitt, P. (1998). Endogenous growth theory. *MIT Press, Cambridge*.
8. Alesina, A., Perotti, R. (1996). Income distribution, political instability, and investment. *European Economic Review 40 (6),* 1203–1228.
9. Alesina, A., Rodrik, D. (1994). Distributive politics and economic growth. *Quarterly Journal of Economics 109 (2)*, 465–490.
10. Amarante, V., Lanzilotta, B., Torres, J. (2023). Inequality and productive structure. *United nations University*, WIDER Working Paper 2023/9
11. Amos, O.M. (1988). Unbalanced regional growth and regional income inequality in the latter stages of development. *Regional Science and Urban Economics 18 (4),* 549–566.
12. Angrist, J. D., Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. *Princeton university press.*
13. Antonietti, R., Burlina, C. (2022). Exploring the entropy-complexity nexus. Evidence from Italy. *Economia Politica – Journal of Analytical and Institutional Economics*, forthcoming.
14. Antonietti, R., Franco, C. (2021). From FDI to economic complexity: a panel Granger causality analysis. *Structural Change and Economic Dynamics 56*: 225-239.
15. ASTI (2022). https://asti.istat.it/asti/ [Accessed on 15 July 2022]
16. Atlas of Economic Complexity (2022). https://atlas.cid.harvard.edu/ [Accessed on 30 September 2022]
17. Bahar, D., Rapoport, H., Turati, R. (2020). Birthplace diversity and economic complexity: Cross-country evidence. *Research Policy 103991*, DOI: 10.1016/j.respol.2020.103991.
18. Balassa, B. (1986). Comparative advantage in manufactured goods: a reappraisal. *The Review of Economics and Statistics*, *68(2),* pp 315-19.

19. Balland, P. A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P. A., Rigby, D. L., Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour*, *4(3):* 248–254.
20. Balland, P.A., Rigby, D. (2017). The geography of complex knowledge. *Economic Geography, 93(1):* 1-23.
21. Bandeira Morais, M., Swart, J., Jordaan, J. A. (2021). Economic Complexity and Inequality: Does Regional Productive Structure Affect Income Inequality in Brazilian States? *Sustainability, 13(2),* 1006.
22. Banerjee, A., Duflo, E. (2003). Inequality and Growth: what can the data say? *Journal of Economic Growth 8 (3)*, 267–299.
23. Barro, R. J., Lee, J. W. (1993). International comparisons of educational attainment. Journal of monetary economics, 32(3), 363-394.
24. Barro, R.J. (2000). Inequality and growth in a panel of countries. *Journal of Economics Growth 5 (1),* 5–32.
25. Barth, E., Bryson, A., Davis, J. C., Freeman, R. (2016). It's where you work: increases in the dispersion of earnings across establishments and individuals in the United States. *Journal of Labor Economics*, 34: S67–S97.
26. Benabou, R. (1996). Inequality and growth. *In: Bernanke, B.S., Rotemberg, J. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge*.
27. Bettencourt, L., West, G. (2010). A unified theory of urban living. *Nature*, 467(7318), 912-913.
28. Blau, F. D., Kahn, L. M. (1996). International differences in male wage inequality: institutions versus market forces. *Journal of Political Economy*, 104(4), 791-837.
29. Boeri, T., Ichino, A., Moretti, E., and Posch, J. (2021). Wage equalization and regional misallocation: evidence from Italian and German provinces. *Journal of the European Economic Association*, 19(6), 3249-3292.
30. Bourguignon, F.J. (1990). Growth and inequality in the dual model of development: the role of demand factors. *Review of Economic Studies 57 (2)*, 215–228.
31. Can, M., Gozgor, G. (2017). The impact of economic complexity on carbon emissions: evidence from France. *Environ. Sci. Pollut. Res. 24*, 16364–16370.
32. Card, D. Heining, J., Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics*, 128: 967–1015.
33. Chávez, J. C., Mosqueda, M. T., Gómez- Zaldívar, M. (2017). Economic complexity and regional growth performance: evidence from the Mexican economy. *Rev. Reg. Stud. 47*, 201–219.
34. Chen, B.-L. (2003). An inverted-U relationship between inequality and long-run growth. *Economics Letters 78 (2)*, 205–212.
35. Chu, L. K., Hoang, D. P. (2020). How does economic complexity influence income inequality? New evidence from international data. *Economic Analysis and Policy*, 68, 44-57.

36. Chu, L. K., Hoang, D. P. (2020). How does economic complexity influence income inequality? New evidence from international data. *Econ. Anal. Policy 68*, 44–57.
37. Ciani, E., Torrini, R. (2019). The geography of Italian income inequality: recent trends and the role of employment. *Politica economica*, 35(2), 173-208.
38. De Dominicis, L., Florax, R. J., De Groot, H. L. (2008). A meta-analysis on the relationship between income inequality and economic growth. *Scottish Journal of Political Economy*, 55(5), 654-682.
39. De la Croix, D., Doepke, M. (2003). Inequality and growth: why differential fertility matters. *American Economic Review*, 3, 4, pp. 1091–113
40. Deininger, K., Squire, L. (1996). A new data set measuring income inequality. *World Bank Economic Review* 10(3), 565–591.
41. Devicienti, F., Fanfani, B., Maida, A. (2019). Collective bargaining and the evolution of wage inequality in Italy. *British Journal of Industrial Relations*, 57(2), 377-407.
42. Di Caro, P. (2018). Redistribution in real-world PIT: Evidence from Italian tax records *(No. wp2018-2).*
43. Di Nardo, J., Fortin, N., Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. *Econometrica*, 64 (5): 1001–44.
44. Dias, J., Tebaldi, E. (2012). Institutions, human capital, and growth: The institutional mechanism. *Structural Change and Economic Dynamics*, 23(3), 300-312.
45. Domini, G. (2019). Patterns of Specialisation and Economic Complexity Through the Lens of Universal Exhibitions. *1855–1900 LEM Papers Series 2019/20 Laboratory of Economics and Management*.
46. Dong, Z., Chen, W., Wang, S. (2020). Emission reduction target, complexity and industrial performance. J. Environ. Manag. 260, 110148.
47. Dordmond, G., de Oliveira, H. C., Silva, I. R., Swart, J. (2020). The complexity of green job creation: an analysis of green job development in Brazil. Environ. Dev. Sustain.
48. Easterly, W., Levine, R. (2003). Tropics, Germs, and Crops: How Endowments Influence Economic Development. *Journal of Monetary Economics*, 50(1), 3-39.
49. EUROSTAT (2022). https://ec.europa.eu/eurostat/data/database [Accessed on 17 July 2022]
50. Faggio, G., Salvanes, K. G., Reenen, J. V. (2010). The evolution of inequality in productivity and wages: panel data evidence. *Industrial and Corporate Change*, 19: 1919–51.
51. Farina, F., Franzini, M. (2015). La casa, il benessere e le disuguaglianze. *Egea*.
52. Fawaz, F., Rahnama-Moghadamm, M. (2019). Spatial dependence of global income inequality: The role of economic complexity. *The International Trade Journal*, 33(6), 542-554.
53. Felipe, J., Kumar, U., Abdon, A., Bacate, M. (2012). Product complexity and economic development. *Structural Change and Economic Dynamics* 23(1): 36–68.

54. Ferraz, D., Moralles, H. F., Campoli, J. S., de Oliveira, F. C. R., Rebelatto, D. A. (2018). Economic complexity and human development: DEA performance measurement in Asia and Latin America. Gestão Produção 25, 839–853.
55. Fontana, M. (2010). The Santa Fe Perspective on Economics: emerging patterns in the science of complexity. The Santa Fe Perspective on Economics, 1000-1030.
56. Forbes, K. (2000). A reassessment of relationship between inequality and growth. *American Economic Review 90 (4)*, 869–886.
57. Gao, J., Zhou, T. (2018). Quantifying China's regional economic complexity. *Physica A: Statistical Mechanics and its Applications*, 492, 1591-1603.
58. Gómez-Zaldívar, M., Osorio-Caballero, M. I., Saucedo-Acosta, E. J. (2022). Income inequality and economic complexity: evidence from Mexican states. *Regional Science Policy & Practice*.
59. Hanushek, E. A., Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. American economic review, 90(5), 1184-1208.
60. Hanushek, E. A., Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. Journal of economic growth, 17(4), 267-321.
61. Hartmann, D., Guevara, M. R., Jara- Figueroa, C., Aristarán, M., Hidalgo, C. A. (2017). Linking economic complexity, institutions, and income inequality. *World Dev. 93*, 75–93.
62. Hartmann, D., Pinheiro, F. L. (2022). Economic complexity and inequality at the national and regional level. arXiv preprint arXiv:2206.00818.
63. Hausman , J.A. (1978). Specification tests in econometrics. *Econometrica*, 46 (6) (1978), pp. 1251-1271.
64. Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A. (2014). The atlas of economic complexity: Mapping paths to prosperity. *Mit Press*.
65. Hausmann, R., Hwang, J., Rodrik, D. (2007). What You Export Matters. *Journal of Economic Growth, 12*(1), 1-25.
66. Hausmann, R., Rodrik, D. (2003). Economic Development as Self-Discovery. *Journal of Development Economics, 72*(2), 603-633.
67. Heinrich Mora, E., Heine, C., Jackson, J. J., West, G. B., Yang, V. C., Kempes, C. P. (2021). Scaling of urban income inequality in the USA. *Journal of the Royal Society Interface*, 18(181), 20210223.
68. Helpman, E. (2004). The Mystery of Economic Growth. *Belknap Harvard*.
69. Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics, 3(2),* 92-113.
70. Hidalgo, C. A., Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, *106(26)*, 10570-10575.
71. Hidalgo, C. A., Klinger, B., Barabási, A. L., Hausmann, R. (2007). The product space conditions the development of nations. *Science, 317(5837*), 482-487.

72. Hodgson, G. M. (2003). Capitalism, complexity, and inequality. *Journal of Economic Issues*, 37(2), 471-478.
73. Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The stata journal*, *7*(3), 281-312.
74. IQI (2022). https://sites.google.com/site/institutionalqualityindex/home?authuser=0 [Accessed on 17 July 2022]
75. ISTAT (2022). https://www.istat.it/en/ [Accessed on 15 July 2022]
76. Javorcik, B.S., Lo Turco, A., Maggioni, D. (2018). New and improved: does FDI boost production complexity in host countries? *Economic Journal 128(614)*: 2507–2537.
77. Katz, L. F., Murphy, K.M. (1992). Changes in relative wages, 1963–1987: supply and demand factors. *Quarterly Journal of Economics*, 107: 35–78.
78. Khan, H., Khan, U., Khan, M.A. (2020). The causal nexus between economic complexity and FDI: empirical evidence from time series analysis. *The Chinese Economy 53 (5)*: 374–394.
79. Koeniger, W., Leonardi, M., Nunziata, L. (2007). Labor market institutions and wage inequality. *ILR review*, 60(3), 340-356.
80. Kutznets, S. (1955). Economic Growth and Income Inequality. *The American Economic Review 45(1): 1-28.*
81. Lapatinas, A. (2016). Economic complexity and human development: a note. Econ. Bull. 36, 1441–1452.
82. Lapatinas, A. (2019). The effect of the Internet on economic sophistication: an empirical analysis. *Economics Letters*, 174: 35-38.
83. Lapatinas, A., Garas, A., Boleti, E., Kyriakou, A. (2019a). Economic complexity and environmental performance: evidence from a world sample. *MDPI Open Access J. 12*, 1–18.
84. Lapatinas, A., Kyriacou, A., Garas, A. (2019b). Taxation and economic sophistication: evidence from OECD countries. *PLOS ONE 14(3)*: e0213498
85. Lee, C. C., Wang, E. Z. (2021). Economic complexity and income inequality: Does country risk matter?. *Social Indicators Research*, 154(1), 35-60.
86. Lee, K.-K., Vu, T. V. (2019). Economic complexity, human capital and income inequality: a cross- country analysis. *Jpn. Econ. Rev. 71*, 695–718.
87. Li, H., Zou, H. (1998). Income inequality is not harmful for growth: theory and evidence. *Review of Development Economics 2 (3)*, 318–334.
88. Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, 22(1), 3-42.
89. Marco, R., Llano, C., Pérez-Balsalobre, S. (2022). Economic complexity, environmental quality and income equality: A new trilemma for regions? *Applied Geography, 139*, 102646.
90. Murphy, K.M., Shleifer, A., Vishny, R.W. (1989). Income distribution, market size, and industrialization. *Quarterly Journal of Economics 104 (3),* 537–564.

91. Neagu, O., Teodoru, M. C. (2019). The relationship between economic complexity, energy consumption structure and greenhouse gas emission: heterogeneous panel evidence from the EU countries. *Sustainability 11*, 497.
92. Nifo, A. and Vecchione, G. (2014). Do Institutions Play a Role in Skilled Migration? The Case of Italy. *Regional Studies*, Volume 48, Issue 10, 2014, pages 1628-1649.
93. North, D. C. (1990). Institution, Institutional Change and Economic Performance. *Cambridge: Cambridge University Press.*
94. Okun, A.M. (1975). Equality and Efficiency: The Big Trade-Off. *Brookings Institution*, Washington, DC.
95. Our World in Data (2022). https://ourworldindata.org/ [Accessed on 30 October 2022]
96. Ourens, G. (2012). Can the Method of Reflections help predict future growth?. *Documento de Trabajo/FCS-DE*; 17/12.
97. Perotti, R. (1993). Political equilibrium, income distribution, and growth. *Review of Economic Studies 60* (4), 755–776.
98. Perotti, R. (1996). Growth, income distribution, and democracy: what the data say. *Journal of Economic Growth 1 (2)*, 149–187.
99. Persson, T., Tabellini, G. (1994). Is inequality harmful for growth? *American Economic Review 84 (3),* 600–621.
100. Pham, M. H., Truong, H. D. H., Hoang, D. P. (2023). Economic complexity, shadow economy, and income inequality: fresh evidence from panel data.
101. Piketty, T., Saez, E. (2003). Income inequality in the United States, 1913–1998. *Quarterly Journal of Economics*, 118: 1–39.
102. Poncet, S., de Waldemar, F. S. (2013). Economic complexity and growth. *Rev. Écon. 64*, 495–503.
103. Rodrik, D., Subramanian, A., Trebbi, F. (2004). Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development. *Journal of Economic Growth, 9*(2), 131-165.
104. Romero, J. P., Gramkow, C. (2021). Economic complexity and greenhouse gas emissions. *World Dev. 139*, 105317.
105. Sbardella, A., Pugliese, E., Pietronero, L. (2017). Economic development and wage inequality: A complex system analysis. *PloS one*, 12(9), e0182774.
106. Seber, G. A. F. (1984). Multivariate Observations. *New York: Wiley.*
107. Sepehrdoust, H., Tartar, M., Gholizadeh, A. (2021). Economic complexity, scientific productivity and income inequality in developing economies. *Economics of Transition and Institutional Change.*
108. Shin, I. (2012). Income inequality and economic growth. *Economic Modelling*, *29*(5), 2049-2057.

109. Shin, I., Kim, H., Yamamura, E. (2009). Technological Progress and the Future of Kuznets Curve's. *MPRA (Munich Personal RePEc Archive)* Paper No. 18866.

110. Simón, H. (2010). International differences in wage inequality: a new glance with European matched employer–employee data. *British Journal of Industrial Relations*, 48(2), 310-346.

111. Smith, A. (1776). The wealth of nations.

112. Stojkoski, V., Utkovski, Z., Kocarev, L. (2016). The impact of services on economic complexity: service sophistication as route for economic growth. *PLoS ONE 11*, e0161633.

113. Sukiassyan, G. (2007). Inequality and growth: what does the transition economy data say? *Journal of Comparative Economics 35 (1)*, 35–56.

114. Swart, J., Brinkmann, L. (2020). Universities and Sustainable Communities: Meeting the Goals of the Agenda 2030 (eds Filho, L., Tortato, U. Frankenberger, F.) 3–45.

115. Sweet, C.M., Eterovic Maggio, D.S. (2015). Do stronger intellectual property rights increase innovation? World Development 66: 665–677, DOI: 10.1016/J.WORLDDEV.2014.08.025.

116. Tacchella, A., Mazzilli, D., Pietronero, L. (2018). A dynamical systems approach to gross domestic product forecasting. Nat. Phys. 14, 861–865.

117. Tachibanaki, T. (2005). Confronting income inequality in Japan – a comparative analysis of causes, consequences, and reform. *MIT Press*.

118. US Census Bureau (2022). https://www.census.gov/en.html [Accessed on 30 October 2022]

119. Vu, T. V. (2020). Economic complexity and health outcomes: a global perspective. *Soc. Sci. Med. 265*, 113480.

120. Vu, T. V. (2022). Does institutional quality foster economic complexity? The fundamental drivers of productive capabilities. *Empirical Economics*, 1-34.

121. Wang, C., Caminada, K. (2011). Disentangling income inequality and the redistributive effect of social transfers and taxes in 36 LIS countries. *Department of Economics Research Memorandum*, 2011, 1-53.

122. Weil, D.N. (2005). Economic Growth. *Addison-Wesley*.

123. Wooldridge, J.M. (2002). Econometric Analysis of Cross Section and Panel Data. *Cambridge*, MA: MIT Press.

124. Zhu, S., Yu, C., He, C. (2020). Export structures, income inequality and urban-rural divide in China. *Applied Geography*, 115, 102150.