The correlation between trade and inequality

An inter-regional analysis of Italy

Thesis presented by
Stefano Caliò

Supervisor(s)
Prof. Gonzague Vannoorenberghe

Reader(s)
Prof. Fabio Mariani

Academic year 2016/2017

Master 120 en Sciences économiques, Orientation générale, Finalité Specialisé

Ecole d'économie de Louvain/UCL • Place Montesquieu 3 • 1348 Louvain-la-Neuve
Département des Sciences économiques/UNamur • Rempart de la Vierge 8 • 5000 Namur

Acknowledgements

First and foremost, I want to thank my supervisor, professor Gonzague Vannoorenberghe for his invaluable help throughout this whole work. I am indebted to him, as he provided me with countless ideas and suggestions. Section 3.3, which tries to approach the data-gathering phase of research from an innovative angle, originates from an idea of his own and I was not much more than a mere executor.

I thank professor Fabio Mariani for his availability in being the reader for this thesis.

On a more personal note, I want to thank all the friends that accompanied me in this two-year journey, of which this thesis is hopefully the apex. If I have grown more mature, not just as an economist, it was also because of many casual intellectually stimulating chats I had the pleasure to be part of. We helped each other, we had great and tough times, and it surely was a memorable experience. They also taught me that even in loud and vibrant, but rather little known Wallonie, I can found plenty of fellow Italians to share some home feeling.

I want to thank my parents, which provided me with both financial and personal support and love. I will never be able to pay them back in any of these aspects, but I want to express my infinite gratitude for all they have done for me in these years.

Finally, I want to thank Elena, for her unconditional love and support. I like to think I have done, am doing and will do the same for her during her academic and personal life. Without her, I might not have had the strength to move out of home to study at UCL.

They say those who love us light our path. I am very lucky to walk an astonishingly bright road.

Contents

1. Introduction

The link between international trade and inequality has been the object of impressive layman and academic debate over the last decades. On the one hand, worries about possible negative outcomes from opening to trade fostered a great deal of discussion among workers facing increased, cheaper competition from foreign firms; on the other hand, economists have tried to establish the existence of this phenomenon (or lack thereof) and possibly quantify it, to help guide policy towards either a more open, globalised world or a more closed one.
This thesis analyses the correlation between international trade and income inequality, to try and measure the impact of trade on the distribution of income in Italian regions. In order to accomplish this goal, I use econometric techniques such as the standard multiple linear regression and some panel data models, namely fixed and random effects regressions.
The analysis, impacted by lack of data on some variables that could have been candidates for entering either the main specification or alternative measures, outputs an unclear picture of what the final effect is. The results suggest that the link between trade and inequality in the data, if ever present, is very weak and seems to affect only one of the used measures of the Gini index.

This thesis is structured as follows: section 1 introduces the aim of the analysis, section 2 gives some theoretical framework and presents the inherent literature, to understand what line of research previous work followed and which economic (and specifically trade-related) variables are expected to impact inequality figures the most; section 3 presents the sources and features of the dataset used in the following empirical analysis; section 4 shows the econometric strategy that tries and grasp what the actual link between trade and inequality is, using various regression techniques; section 5 presents the results of the empirical strategy; section 6 concludes.

2. Theoretical background and literature review

The last decade of economic research has arguably been one of the most inequality-focused ones, as it is easy to grasp from the impressive body of work that analyses it either as an outcome or as a peculiar component of almost any economic activity: in this sense, it is impossible not to notice the impact that (broadly debated) Piketty (2014) [1] has caused to both the public and academic discussion on the distribution of income and wealth. Piketty surely reignited an important topic, going through the difference between the growth rate of an economy and the return on capital, building his explanation of widening inequalities between percentiles.

Trade theory has naturally tackled the issue from its own, different perspective and virtually any model describing trade flows, no matter how far back in time it was created, either deals with income inequality or presents some mechanism that lets the labour market adjust to the post opening-to-trade situation, therefore ruling out the chance of increasing inequality. Specifically, the "classical" trade models, such as the Ricardian one, tend to assume perfect mobility for workers, implying that opening to trade should not impact wages and/or employment as labour can move across sectors at no cost and adapt to a changing, more open market: in other words, the impact of trade on sectors and industries of an economy that lose market share because of foreign competition (i.e. imports) is null or fairly limited. On the other hand, more recent literature started to challenge the admittedly quite unrealistic assumption of perfect mobility and even empirically test whether opening to trade could generate sizeable unemployment and/or wage inequalities. As an example, the more recent Heckscher-Ohlin model relaxes the aforementioned assumption by allowing for the presence of some friction in the labour market: workers are mobile between different sectors, but only in the long run. However, the model does not go any further in conceding space to inequality issues per se, as it still predicts wage equalisation, even if possibly only after some time. Most importantly linked to this model, though, is the *Stolper-Samuelson effect*: when opening to trade, the real income of the abundant production factor used by the exporting sector (i.e. what the model predicts to be the sector in which the country specialises) increases, while the real income of the scarcer factor decreases. This mechanism evidently allows for possible negative outcomes: trade might then create inequality *within countries*.

Finally, New Trade Theory and, later on, "New new Trade Theory", moved the attention towards discrepancies *within sectors*, i.e. highlighting intra-industry trade and the heterogeneity between firms populating the same market via a ranking mechanism based on the firms' production functions: it predicts how the least competitive firms are kicked out of the market, some more competitive ones only sell in the home market and the most competitive ones are able to efficiently produce and exports their goods. This goes hand in hand with a corresponding difference in the wages paid by those firms, in that exporting firms are able to offer higher pay rates. In other words, the work of economists such as Paul Krugman and Marc Melitz shows how being a player in one of a country's "strongest" markets (i.e., in the words of classic trade theory, having a comparative advantage or a sizeable endowment of the productive factor) does not necessarily lead to economic success and, instead, may still endanger employment and/or wage levels. It also is some of the most recent and arguably advanced theoretical framework trade theory has produced up to now (and has built upon, since its birth) and has been empirically tested and confirmed, among the others, by Eaton, Kortum, and Kramaz (2011) [2] using French firms data to analyse trade levels and market participation, and by Trefler (2004) [3] in the context of the 1987 Canada-US Free Trade Agreement and the consequences on Canadian firms: both papers highlight the role of firms' productivity and efficiency in shaping markets and trade patterns. Trefler (2004) is particularly relevant in the context of this thesis, since it points out how trade shocks affect the labour market, as well, underlining rising unemployment for some firms on the one hand and booming productivity for other firms, on the other.

Having outlined some of the relevant theory, it is interesting to evaluate the impact of the most inherent empirical literature dealing with trade and inequality, in order to better grasp which aspects and contexts are likely to be affected and whether a link between international trade and inequality is present at all.

One of the most cited works is surely Autor, Dorn, and Hanson (2013) [4], which analyses the effect of Chinese competition in the US labour market and highlights a remarkable effect: looking at the exposure of several markets to Chinese competition, the authors find that a quarter of manufacturing's employment fall is explained by competition coming from imports. As far as wages are concerned, competition from Chinese imports reduce wages of both skilled and unskilled workers. Moreover, the paper underlines another social cost related to foreign competition, i.e. the rise of benefits, healthcare and pensions expenditure in the most impacted sectors.

Bosio, Grasseni, and Falzoni (2016, draft) [5] follow the same line of research, looking at the effects of Chinese competition in Italy during the recent crisis period, investigating the impact they have on firms that both are and are not involved in international trade. Their findings highlight a mixed effect on labour markets, largely dependant on the imported goods' grade of substitution or complementarity to local production: in this sense, both positive and detrimental effects on the labour markets are observed, depending on those markets' fluidity (i.e. the ability of workers to move within or across sectors) and the kind of imported goods.

Artuç and McLaren (2012) [6] use US data to estimate the cost of switching job and determine it is quite sizeable, but also that the industry in which a worker is employed is very relevant, surpassing the importance of skill level or occupation. It is an interesting result and it partly contradicts the theoretical mechanisms that I presented above, which considers the general case.

A relatively innovative perspective is the one followed by Dobson and Ramlogan (2009) [7], in a paper that uses the framework of the Kuznets curve: this recent line of trade research uses a relation which is well-diffused in inequality related issues, i.e. the Kuznets curve, and tries to connect it with international trade theory, by creating a link between inequality and openness to trade. The result should then be a single-peaked, concave curve, indicating that very high levels of openness to trade correspond to lower inequality values than average ones. The authors analyse the case of several South American countries, finding relevant evidence that such a relation does exist and that a push for more openness to trade would bring benefit to the area, under the condition that some help is provided in order to tackle the issues encountered in the ascending segment of the curve.

Another analysis using the concept of the Kuznets curve can be found in Jalil (2012) [8], which actually follows Dobson and Ramlogan (2009) and investigates the relationship between trade and the Gini index in China. The paper finds evidence that the analysed patterns do reflect the peculiar functional form of the curve, emphasising how a progressive opening to trade should prove beneficial if accompanied by policies capable of alleviating the initial negative impact.

In a very recent paper, Helpman et al. (2017) [9] look at Brazilian data to test their model, which again points at within-sectors inequality as a sizeable chunk of the overall figure. The authors find that the theoretical framework holds empirically in the case of Brazil and that exporting firms do pay higher wages than non-exporting ones, fostering rises in inequality figures. Importantly, the results confirm an increase in wage inequality following the opening to trade, but also a subsequent decrease in the same figure, as also predicted in Helpman et al. (2010) [10] and resembling the Kuznets curve-type of relationship mentioned above. Specifically, Helpman et al. present a concave path for a Theil index of inequality, increasing while opening to trade and then decreasing after a mid-level threshold.

Somewhat linked to Dobson and Ramlogan (2009), in that it deals with a positive trade shock in a relatively poor country, Verhoogen (2008) [11] looks at the Mexican case through the lenses of quality-upgrading and heterogeneous firms: the market shaping reflects New Trade Theory's predictions, since only the most productive firms are able to export and those very firms tend to pay higher wages compared to less efficient competitors, as well. Within-industry wage inequality arises accordingly and the main engine is the mechanism of quality upgrading for the exported goods, due to the necessity of appealing richer trade partners (in that case, mainly the US) and the consequent need for more expensive and skilled labour in order to produce such high-quality goods.

Another paper looking at a developing country opening to trade is Topalova (2005) [12], that, through a difference-in-differences approach, shows how, after opening to trade, most industrial Indian districts were less alleviated from poverty than others, due to the limited mobility of workers whose firms are most impacted by new competition. Inequality figures, however, were not relevantly impacted.

Concentrating on both India and Brazil, Daumal (2013) [13] conducts a somewhat similar analysis, exploiting trade shocks to analyse the correlation with inequality, at the country level. The results show opposing indications, as on the one hand India has experienced aggravated figures in inequality, while on the other Brazil has seen an improvement after liberalising international trade. As in other cases, the manufacturing sector seems to be the major driving force in moving income inequality.

The interesting contrast between developed and developing countries is surely not only object of the aforementioned papers and it generally focused on the puzzling evidence of rising inequality in both groups of countries, whereas it could be expected that developing regions would be positively affected by a first opening to international trade. Zhu and Trefler (2005) [14] focuses on the distinction between OECD and poorer countries to show how technological change is a big component in explaining this anomaly: both developed and developing countries tend to continuously put effort in implementing better, skill-intensive technology to gain an advantage and catch-up with the other group of countries, respectively. This leads to the already mentioned mechanism of wage increasing linked with a demand for high-skilled workers, creating a within-sector gap.
Lastly, Hartmann et al. (2017) [15] follow a relatively unexplored path, comparing economic complexity and inequality dynamics by analysing the export complexity of several countries in a very large time frame. The authors demonstrate that countries exhibiting a higher level of product complexity in their export figures also show a lower Gini index.

As it is easy to grasp from this relatively short list of citations, the literature investigating the relationship between international trade and inequality is vast and tackles the issue from different standpoints. This thesis parts from those analyses in that it uses regional level data for a developed, European country, whereas most literature focuses on between-countries estimations in areas who are relatively closed to international trade or have opened to it in recent times. This leads to a lack of proper trade shocks to evaluate, as one could find in countries opening to trade, but lets the analysis focus on a more micro perspective, to capture some within-country differences.

3. Data

3.1 Sources and main variables

The data for this analysis come from a univocal source, the Istituto Nazionale di Statistica (ISTAT)'s "I.stat" web database, with the sole exception of the data for education, which come from Eurostat's regional dataset. Merging from these two datasets, I obtain data to analyse the link between trade and inequality in the 20 Italian regions throughout the 2003-2014 period of time.
Unfortunately, Eurostat does not publish a bank of data as complete as ISTAT's, meaning that I lacked access to education data for the Trentino Alto Adige region (for the full period I consider in the analysis) and the 2013-2014 values for all regions: this led to the creation of an unbalanced panel dataset for the empirical analysis.
As for the variables used in the specifications, I use the level of exports as the main independent variable of interest, GDP per capita, the share of population in education and the sectoral composition of workforce as control variables, and the Gini index as dependent variable to measure income inequality. Notably, I use both the available versions of the index: one considers the incidence of rents in the computation, the other does not.
As far as the reasons for choosing these variables are concerned, the Gini index was the first choice for the dependent variable, as it is the epitome of inequality measures, giving a synthetic scale of income dispersion and, moreover, two different measures are available (both with and without the incidence of rents in the computation), opening to a possibility of further debate; as for the main independent variable of interest, there is clearly a variety of plausible options among which it is possible to choose: trade indicators include exports, imports, the trade balance, FDIs and various "custom-made" trade indices that empirical literature shows.

However, the only available indicator for regional-level Italian data in the selected time framework is the level of exports, so I use that alone.

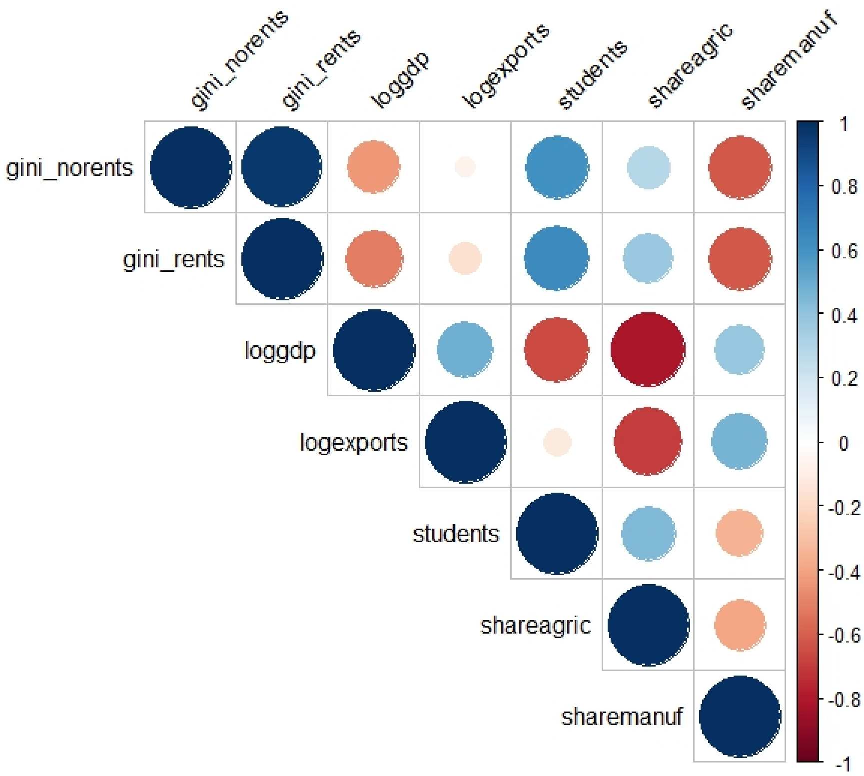
Regarding control variables, GDP per capita is the most relevant one in reflecting variations of income inequality: one popular approach, also mentioned in the literature review, is the one of a Kuznets curve relating GDP level and income inequality, so that an influence of GDP is most likely to be present; as for other regressors able to *explain* the dynamics of a Gini index, I followed, as mentioned in the relative section, the available literature to try and grasp what could be a good indicator to do just that and education proved to be one of the most relevant. The best candidate for entering the empirical model, i.e. a measure of the level of education like the average number of schooling years, was not available, so I fell back onto an admittedly inferior choice like the share of students in the population. Although such measure is not as efficient as the best one in marking the education level in different regions, it should be precise enough in the relevant time frame: given that $T = 12$ and that the Italian public, mandatory education system forces children into school for 10 years, more than a full "generation" of new students entered school (and possibly exited, if they did not continue studying through high school) in the considered period, partly moulding the education level of the different regions. Albeit imperfect, this dynamic gives some power to a measure like the share of students in the population and makes it a relevant choice for an analysis like the one this thesis is based on. As a partial confirmation of the relevance, R-squared values consistently increased after the introduction of this variable into the models.

Finally, sectoral composition of employment, represented by the shares of workers employed in the three sectors of the economy is the last regressor used in the analysis. This partly reflects an effect observed in the literature, by which it is usually not the services sector, but the manufacturing and (to a lesser extent) agriculture ones to be most impacted by trade shocks. Refraining from including all three regressors in the model, in order to avoid plausible issues of multicollinearity, I then include the share of workers in the manufacturing and agricultural sectors in the specification.

3.2 Descriptive statistics

The dataset contains values for the 20 Italian regions throughout a period of 12 years, giving a total of 240 observations (with the above-mentioned restrictions regarding some missing values).
A correlation matrix is useful to grasp how the variables relate to each other.

FIGURE 3.1: Correlation matrix for all variables



Note: both colour and size of the dots show the magnitude of the correlation: the bigger the dot, the larger correlation is in absolute value.

Figure 3.1, which comprises all variables, shows at a glance the correlations one is most interested about in an analysis such as this one and it pictures a rather interesting lack of relation between both Gini measures and export figures. Clearly, this is only a first approach at the issue and regression analysis gives a better explanation of the actual dynamics.

To better understand the evolution of the most important variables, a confrontation of 2003 and 2014 (i.e. begin and end years of the analysed period) values for the Gini index and for exports is very effective.
Figure 3.2 and Figure 3.3 plot a map showing this evolution, for both measures of the Gini index.

FIGURE 3.2: 2003 and 2014 figures for the Gini index excluding rents

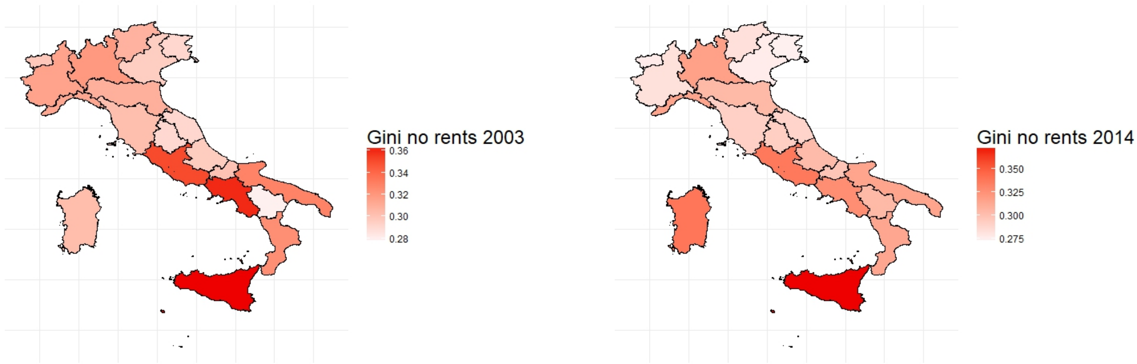
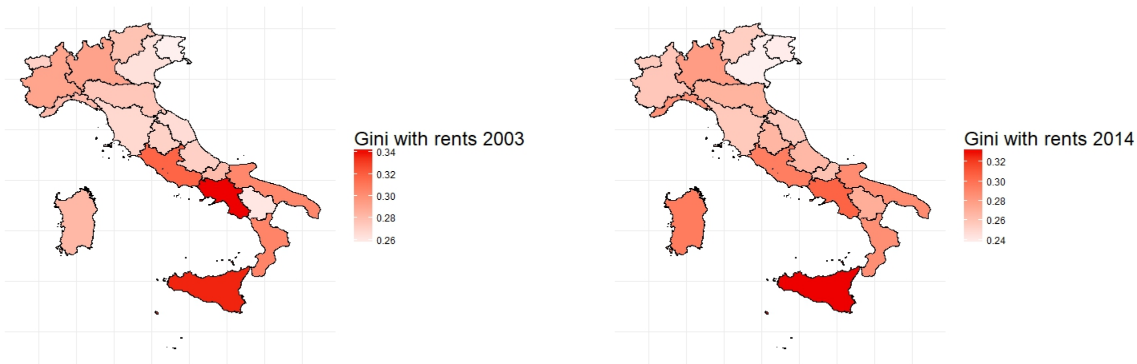


FIGURE 3.3: 2003 and 2014 figures for the Gini index with rents



Note: as for the legend, the colour gradient shows the magnitude of the index (i.e., the darker the colour, the higher inequality is)

As the figures show, there has been a general decrease in regional inequality over the time period (with a few exceptions). Specifically, Table 3.1 shows the mean and standard deviation figures for both measures of the Gini index and both time periods.

Table 3.1: Mean and standard deviation figures for the Gini index

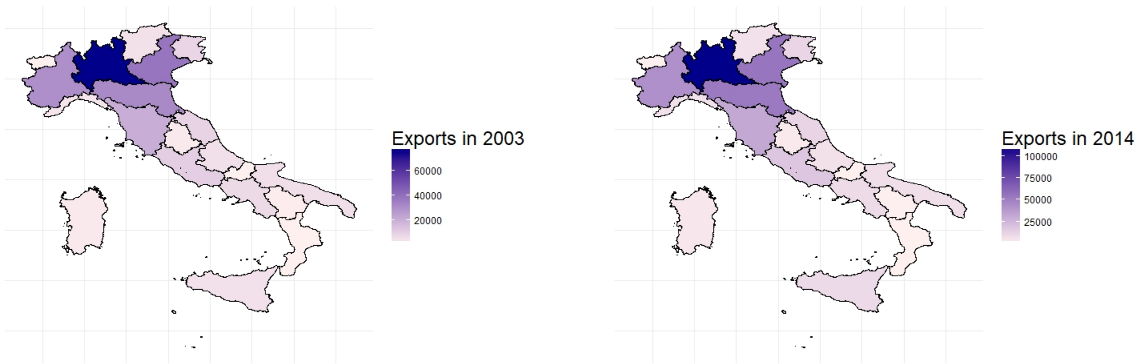
	2003	2014
Mean	.285	.273
Std. Dev.	.244	.230

	2003	2014
Mean	.312	.305
Std. Dev.	.237	.253

Note: the left table shows figures for the index excluding rents, the right one for the measure including them

To complete the overlook, it is possible to plot the same map with export figures: Figure 3.4 does that.

Figure 3.4: 2003 and 2014 figures for exports



Note: as for the legend, the colour gradient shows the magnitude of the exports (i.e., the darker the colour, the higher the level of exports)

This plots show what was an actually remarkable increase in exports levels between the two points in time: it was not only a phenomenon regarding regions lagging behind in 2003, but also the leading ones greatly increased (most of them almost doubled) their exports.
As a side note, patterns seem to follow the standard predictions of trade theory, such as the *Gravity Equation* first proposed by Tinbergen [16] of a positive relation between GDP size and trade activity of a region (cf. Figure 3.5, that presents a very similar picture).
To give a more precise measure, Table 3.2 shows means and standard deviations of exports for both time periods.

TABLE 3.2: Mean and standard deviation figures for exports

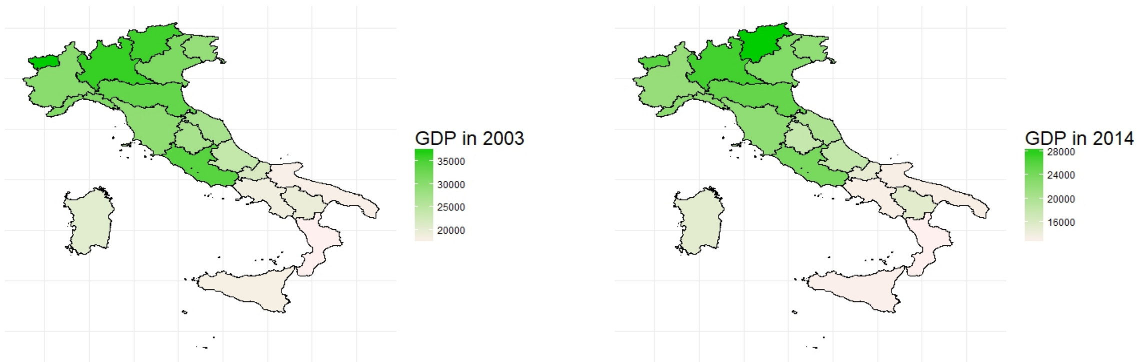
	2003	2014
Mean	13187	19700
Std. Dev.	18565	27104

Note: data in millions of Euro

As also observable in the maps, the standard deviations of export figures are very sizeable, signalling a large variation between regions. Generally speaking, all but two regions saw an increase in export figures during the time period, but as said the increase was so widespread that the starting differences remained quite stable. This is a feature of peculiar interest, because it represents a relatively large variation in trade activity, but it also leaves the pre-"shock", inter-regional differences unaltered.

Lastly, it is interesting to investigate the same patterns for the case of GDP per capita. Figure 3.5 presents the data in map form.

FIGURE 3.5: 2003 and 2014 figures for GDP per capita



Note: as for the legend, the colour gradient shows the level of GDP per capita (i.e., the darker the colour, the higher GDP is)

Again, mean and standard deviation in Table 3.3 help grasping the dynamics of the variable better.

TABLE 3.3: Mean and standard deviation for GDP per capita

	2003	2014
Mean	26949.2	19806.4
Std. Dev.	6804.1	5250.6

Note: data in millions of 2003 Euro

As the maps suggest, GDP per capita has actually decreased over the period of time. Arguably, the final, real 2014 values are greatly impacted by what is now commonly called "Great Recession", i.e. the combination of crises that hit the worldwide economy (2008) and the Eurozone (around 2010), that clearly destroyed a sizeable chunk of Italy's production (on the counterpart, population, i.e. the other aspect of the GDP per capita formula, did not fall considerably in the same period, leading the decline's burden to fall almost solely on economic havoc).

3.3 Alternative approaches to data collection: Google Trends

Throughout the data-gathering process preceding this analysis, I tried to approach the issue of lacking (or fairly inconsistently low-quality) data by the use of Google Trends: this tool gives access to data about online research on the Google search engine at a very detailed, "micro" level. This lets the user browse for any search query and see how frequently it was looked for by Google users. Given the power such a tool provides, it could arguably be a viable option for researchers who find it hard to get consistent data.
The approach in this thesis tries to gather the data about income inequality, using the results that Google Trends outputs for online searches regarding certain brands of groceries, supermarkets, cars (comparing them with the frequency of search for public transport information), and clothing. The rationale behind this kind of analysis is using these data as a proxy for consumption of the same goods, considering the polarisation of the results to either one or the other tail of the distribution of prices as an indicator for inequality. As an example regarding Italy, I considered the differences between searches for "discount", low price supermarkets,

such as LIDL, and the ones for higher tier and price brands, such as Esselunga or Carrefour.
During the process, unfortunately, some issues appeared to be tackling the consistency of such a tool as a viable data-gathering solution:

- it is not the case in which all keywords and queries are widespread in all the analysed regions [1]
- the goodness and consistency of the use of online searches as a proxy for actual consumption is not necessarily valid on all kinds of goods

The second point is especially critical and, at the same time, fairly hard to evaluate: it is easy to infer that most of the Google users who searched for supermarket brands were probably driven to do so by interest in those very brands (i.e., they were willing to learn, for example, about special offers or about the location of the nearest branch, leading to subsequent shopping at that supermarket), strengthening the argument. On the other hand, it is also arguable that online searches for expensive brands of cars are not necessarily as good a proxy as the ones for supermarkets: Google users might have just wanted to look at pictures of attractive or desirable cars, even if they were not willing to buy them. This point is actually very applicable to brands that are definable as "strong", i.e. very popular brands selling products that people might desire, but might not necessarily be able to afford[2].
These drawbacks pushed me to rely on standard data sources only, but the Google Trends tool could be, when used for research in branches of economics which are less prone to suffer from these kinds of problems, an innovative alternative source of data.

[1] In the aforementioned example, both LIDL and Esselunga are not present in some of the Italian regions and this clearly tends to influence the results one expects from Google Trends

[2] This line of reasoning can be applied to some popular, cheap products, but it suits especially well expensive goods that attract large interest because of their marked differentiation from competition. Their peculiarities create strong hype or interest, but they do not necessarily trigger consumption

4. Econometric framework

4.1 Estimation strategy

The baseline specification for the analysis is the multiple regression model:

$$\begin{aligned} Gini_{i,t} = \beta_0 + \beta_1 \log Exports_{i,t} + \beta_2 \log GDPpct_{i,t} + \beta_3 \log GDPpct^2_{i,t} \\ + \beta_4 Students_{i,t} + \beta_5 Manufacturing_{i,t} + \beta_6 Agriculture_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{4.1}$$

with an intercept β_0, $Gini_{i,t}$ being the Gini index (either with or without rents), $\log Exports_{i,t}$ the logarithm of exports, $\log GDPpct_{i,t}$ the logarithm of GDP per capita (I also include its squared term, to reflect its functional form and its concave relationship with the dependant variable), $Students_{i,t}$ the share of population enrolled in education, $Manufacturing_{i,t}$ the share of workers in the industrial sector, $Agriculture_{i,t}$ the share of workers in the agricultural sector and $\varepsilon_{i,t}$ the error term, all for region i at time t, with $i = 1, ..., N$ and $t = 1, ..., T$. As mentioned in Section 3.2, $N = 20$ and $T = 12$.
Given the availability of two different measures of the Gini index, I run here and for all following models two distinct regressions: one using the index excluding rents, the other using its version including them. The only difference, therefore, lies in the dependant variable that the two models present, whereas the regressors are the same.

Given the lack of other control variables to be added, a potential loss of inference power (which is, at least partly, showed in the low R-squared values found in the results) for Equation 4.1 is possible. As for regression theory, variation accruing to missing variables is simply attributable to the error term $\varepsilon_{i,t}$ if those variables are not endogenous in the estimated model: this would hardly impact

the efficiency of the regression, but rather increase the weight of the error term in explaining the dependant variable. However, this condition is not very likely as at least one variable (namely, the level of imports) is both missing and related to one of the regressors (i.e. *Exports*). This leads to facing the omitted variable bias. To improve the estimation and, at least partly, catch some of the mentioned uncontrolled variation, a fixed effects model is the subsequent tested option.
The aim of this regression model is capturing the features that are characteristic of the various regions, so that the changes in variables over time reflect only *non-fixed* movements. In other words, this model allows to grasp the time-led dynamics within the variables, ruling out the unobserved effects that impact the variation in those variables. As an example, it might be the case that a Northern region of Italy benefits from specific economic or social conditions that are likely to affect inequality (say, region- or even province-specific subsidies, different local taxation thresholds and so forth) that a Southern region does not encounter, and vice versa. To control for this heterogeneous, region-specific effects and therefore attribute a more appropriate weight to the regressors, I then estimate the fixed effects model:

$$\begin{aligned} Gini_{i,t} = \alpha_i + \beta_1 \log Exports_{i,t} + \beta_2 \log GDPpct_{i,t} + \beta_3 \log GDPpct^2_{i,t} \\ + \beta_4 Students_{i,t} + \beta_5 Manufacturing_{i,t} + \beta_6 Agriculture_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{4.2}$$

Equation 4.2 uses regional fixed effects to estimate the relation object of the analysis, again using both the Gini index measures. Notation remains identical to the one of the basic specifications, except for the introduction of the regional fixed effects term α_i, with $i = 1, ..., N$.
As a confirmation of the improvements a fixed effects model brings over a standard multiple regression, I run a F-test for individual effects: an extraordinarily low p-value imposes a failure to reject the null hypothesis of lacking individual effects and confirms that this is the better choice among the two models.
At a first glance, however, it is not obvious whether a fixed effects regression would necessarily be the best choice among other panel data models. A first-difference model should be ruled out, as $T > 2$ and, with a sufficiently large N, this is commonly referred to as a condition for which fixed effects should be preferred. There are, though, other techniques which could be implemented: the rationale between fixed effects is the belief that $cov(\alpha_i, x_{i,t,j}) \neq 0$, i.e. assuming that there is some arbitrary correlation between the fixed, individual effects and the independent variables $x_{i,t,j}$ ($j = 1, ..., J$ ordering the regressors in the equation). At

a first glance, this seems a reasonable assumption and, therefore, this should be the right model to make inference from the data. However, with the purpose of checking more viable options, a random effects regression could be estimated, as well. This model makes the opposite assumption of $cov(\alpha_i, x_{i,t,j}) = 0$, therefore implying no correlation between α_i and the regressors in the equation. Moreover, random effects account for the serial correlation given by constant factors, an issue which is likely to be present in a fixed effects regression. Hence, I estimate a random effects model, as well.
A Hausman test is necessary to understand which one between the fixed and random effects models best fits the data: the results of this test output a preference for a random effects regression for the specification estimating the Gini index excluding rents, whereas a fixed effects model is suggested for the model using Gini with rents.

4.2 Potential problems of the estimation

4.2.1 The endogeneity issue

Having presented the estimation strategy, the reported results need to be interpreted with some care given that, as mentioned, an omitted variables bias may be present. This can cause the variation belonging to the omitted variables to be wrongly accruing to the regressors included in the model, returning biased estimators. More formally, being q the omitted variable in the model

$$\begin{aligned} E(Gini_{i,t} \mid \log Exports_{i,t}, \log GDPpct_{i,t}, \log GDPpct_{i,t}^2, Students_{i,t},\\ Manufacturing_{i,t}, Agriculture_{i,t}, q) = \beta_0 + \beta_1 \log Exports_{i,t}\\ + \beta_2 \log GDPpct_{i,t} + \beta_3 \log GDPpct_{i,t}^2 + \beta_4 Students_{i,t}\\ + \beta_5 Manufacturing_{i,t} + \beta_6 Agriculture_{i,t} + \gamma q \end{aligned} \tag{4.3}$$

the omission of a variable such as the imports level implies that the β_1 coefficient measures the partial effect of exports on the Gini index, *ceteris paribus*, i.e. holding fixed, among the others, the omitted imports variable, as well. Rewriting the regression models in order to include q in the error term outputs the following equation:

$$\begin{aligned} Gini_{i,t} = \beta_0 + \beta_1 \log Exports_{i,t} + \beta_2 \log GDPpct_{i,t} + \beta_3 \log GDPpct^2_{i,t} \\ +\beta_4 Students_{i,t} + \beta_5 Manufacturing_{i,t} + \beta_6 Agriculture_{i,t} + u_{i,t} \end{aligned} \tag{4.4}$$

or, most relevantly, in the fixed and random effects models:

$$\begin{aligned} Gini_{i,t} = \alpha_i + \beta_1 \log Exports_{i,t} + \beta_2 \log GDPpct_{i,t} + \beta_3 \log GDPpct^2_{i,t} \\ +\beta_4 Students_{i,t} + \beta_5 Manufacturing_{i,t} + \beta_6 Agriculture_{i,t} + u_{i,t} \end{aligned} \tag{4.5}$$

with $u_{i,t} = \gamma q + \varepsilon_{i,t}$, with $E(\varepsilon_{i,t}) = 0$, $cov(\varepsilon_{i,t}, q) = 0$ and $cov(\varepsilon_{i,t}, X_q) = 0$ (where X_q denotes the set of regressors). An issue of endogeneity arises as long as q is correlated with any of the regressors in the model, since this leads to $E(u_{i,t}) \neq 0$, violating one of the assumptions for an unbiased OLS estimation (Wooldridge, 2010) [17].

Even though establishing whether that is the case in this analysis is not obvious, it is likely, following the aforementioned example, that there might be some form of correlation between exports and the omitted imports. Whether this correlation would be positive or negative is not easy to determine: generally speaking, it would seem reasonable that the two components of the trade balance should go hand in hand, in a framework of "openness to trade". However, in some cases such as the ones of autarky-led regions, the correlation could well be negative. Since this is most probably not the case in any of the Italian regions, it should be safe to assume that the correlation is positive, albeit not necessarily sizeable.

What would be especially hard to evaluate, instead, is the sign of the coefficient linked to the omitted variable, had it been included in the regression: β_1 proved to be negative, very small in size and basically close to a null factor, even tough statistically significant in the fixed effects estimation (cf. Table 5.4). In this sense, determining whether the beta coefficient for another trade indicator such as the level of imports would follow the same line is a tough call. Assuming that would be the case and that the correlation between imports and exports would be positive, then the omitted variable bias would be negative, i.e. the estimate of β_1 in the biased model is lower than the "true" one.

As mentioned, some measures are taken to partly account for the bias, but a lack of viable instrumental variables to use for an IV model, which is usually indicated by econometric theory as the first option to solve the omitted variable bias, means that it was not possible to completely eradicate the issue from the model. One possible workaround of this issue would be using one or more independent variables as an instrument, by lagging them. However, this method can actually cause a loss of efficiency if the model is overidentified, i.e. if too many instruments are introduced in the regression. Although this must not necessarily be the case of every analysis, the instruments generated via this technique are generally weak, therefore implying the use of several ones, inducing the overidentification issue. More importantly, such kind of instruments is at risk of being actually correlated with the error term, violating one of fundamental assumptions of IV modelling: the principle is that a lagged variable is not correlated with a "present" error term, but this is of course not the case when errors are impacted by serial correlation. This would produce a biased estimation with no improvement over the specifications used in the analysis. Lastly, even if such "simulated" IV technique was implementable and produced an unbiased estimation, it would nonetheless be difficult to determine whether this solution is actually valid enough for definitely eradicating the endogeneity issue.

4.2.2 Robustness checks and tests

In order to refine the analysis and find the best data fit, I then proceed with further testing: given the unbalanced nature of the panel dataset, I follow Baltagi and Li (1990) [18] and use the specific Lagrange-multiplier test to determine whether the estimation would benefit from implementing time effects, too. This is a modified version of the standard Breusch-Pagan test and it is specifically designed to suit a situation of unbalanced panel datasets. The results show that a regression model with individual effects only provides the best fit, so I refrain from adding time effects, too.

Another standard assumption for a good OLS estimation is the homoskedasticity of the error term, therefore I first run a Breusch-Pagan test to check for the presence of heteroskedasticity: this test suggests that I fail to reject the null hypothesis of homoskedasticity. However, such a result in the Breusch-Pagan test does not exclude the chance of having a non-linear functional form of heteroskedasticity: to rule out such possibility, the implementation of a White test is necessary. This test has the drawback of being less useful when the number of independent variables is rather high, but since this is not the case in the analysis, it is here a better option than the Breusch-Pagan test. The White test actually confirms the presence of a form of heteroskedasticity for the fixed effects model, whereas the random effects one presents homoskedastic errors.

Furthermore, it is necessary to test for the presence of serial correlation in the data, a rather common issue in a panel dataset. Wooldridge (2010) and, generally speaking, econometric literature, tends to accept that, in a fixed effects model with a preponderance of c_t in the usual composite error $v_{i,t} = u_{i,t} + c_{i,t}$, serial correlation tends to decrease and eventually cease over a large enough time frame. It also suggests, however, that the case of a strong serial correlation of the idiosyncratic error $u_{i,t}$ might lead to a very large error in the fixed effects estimation. To test for such correlation I therefore use what is known as the *Wooldridge's test for serial correlation in short panels*, which is suggested for panels with small T and large N: the residuals of the estimated fixed effects model are lagged and a pooling AR(1) specification:

$$\hat{\varepsilon}_{i,t} = \alpha + \delta \hat{\varepsilon}_{i,t-1} + \eta_{i,t} \tag{4.6}$$

is used (taking, therefore, the last two time periods). Specifically, this tests the hypothesis $H0 : \delta = -1/(T-1)$, where $\delta = corr(\ddot{u}_{i,T-1}, \ddot{u}_{i,T})$ with $\ddot{u}$ denoting, as usual, the time-demeaned error term of fixed effects regressions. The test suggests presence of serial correlation, so I follow Arellano (1987) [19] to compute heteroskedasticity and serial correlation-robust standard errors for the fixed effects model. As a side note, the fact that these clustered errors are smaller in magnitude than some of the originally computed ones suggests that the aforementioned correlation is negative.

As for the random effects model, I follow Baltagi and Li (1995) [20]: this tests the presence of serial correlation in the idiosyncratic error. The test confirms a presence of serial correlation, therefore I follow White (1984) [21] to obtain robust standard errors.

5. Results

Table 5.1 shows the results of the multiple regression estimations using the "no rents" Gini index as outcome variable, increasingly using all variables. Table 5.2, on the other hand, presents the outcome of the same estimation strategy, using the Gini index including rents as dependant variable. Standard errors are noted in brackets. As mentioned in Section 4.2, these results, albeit statistically significant, are particularly subject to bias and the estimation needs a better model to fit the data properly. Hence, the use of fixed and random effects models.

Table 5.3 outputs the results of the fixed effects specifications: similarly to Table 5.1, these regressions use as outcome variable the Gini index excluding rents, whereas Table 5.4 shows the same specifications using the index computed including rents. Again, standard errors are noted in brackets. As mentioned in Section 4.2, the standard errors for regression (4) in Table 5.4 are heteroskedasticity and serial correlation-robust as for Arellano (1987).

Lastly, Table 5.5 and Table 5.6 present the results of the random effects models, using the Gini without and with rents as dependant variable, respectively. Standard errors are noted in brackets and the ones of regression (4) in Table 5.5 are robust to serial correlation, as explained.

TABLE 5.1: Basic specifications, Gini excluding rents

	Gini excluding rents			
	(1)	(2)	(3)	(4)
Exports	−0.001	0.004***	0.002	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
GDP capita		−3.242***	−3.600***	−2.219***
		(0.470)	(0.516)	(0.580)
GDP capita^2		0.158***	0.177***	0.108***
		(0.023)	(0.025)	(0.028)
Students			0.005***	0.003***
			(0.001)	(0.001)
Manufacturing				−0.002***
				(0.0003)
Agriculture				−0.001*
				(0.001)
Constant	0.316***	16.915***	18.500***	11.640***
	(0.009)	(2.372)	(2.620)	(2.960)
N	240	240	189	189
R^2	0.002	0.372	0.498	0.633
Adjusted R^2	-0.002	0.364	0.487	0.621
Residual Std. Error	0.024 (df = 238)	0.019 (df = 236)	0.017 (df = 184)	0.015 (df = 182)
F Statistic	0.460 (df = 1; 238)	46.592*** (df = 3; 236)	45.611*** (df = 4; 184)	52.361*** (df = 6; 182)

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 5.2: Basic specifications, Gini including rents

	Gini including rents			
	(1)	(2)	(3)	(4)
Exports	−0.002**	0.003***	0.0004	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
GDP capita		−3.181***	−3.590***	−2.567***
		(0.452)	(0.493)	(0.576)
GDP capita^2		0.155***	0.176***	0.125***
		(0.022)	(0.024)	(0.028)
Students			0.005***	0.004***
			(0.001)	(0.001)
Manufacturing				−0.001***
				(0.0003)
Agriculture				−0.001**
				(0.001)
Constant	0.296***	16.605***	18.448***	13.396***
	(0.009)	(2.285)	(2.501)	(2.940)
N	240	240	189	189
R^2	0.018	0.415	0.551	0.645
Adjusted R^2	0.014	0.408	0.542	0.633
Residual Std. Error	0.024 (df = 238)	0.019 (df = 236)	0.016 (df = 184)	0.015 (df = 182)
F Statistic	4.400** (df = 1; 238)	55.900*** (df = 3; 236)	56.514*** (df = 4; 184)	55.102*** (df = 6; 182)

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

TABLE 5.3: Fixed effects regressions, Gini excluding rents

	Gini excluding rents			
	(1)	(2)	(3)	(4)
Exports	−0.001	−0.001	−0.004	−0.010
	(0.005)	(0.005)	(0.005)	(0.006)
GDP capita		−0.970	−1.785	−1.195
		(0.677)	(1.196)	(1.240)
GDP capita^2		0.047	0.086	0.059
		(0.034)	(0.059)	(0.061)
Students			−0.001	−0.0002
			(0.002)	(0.002)
Manufacturing				−0.002**
				(0.001)
Agriculture				−0.001
				(0.002)
N	240	240	189	189
R^2	0.0003	0.013	0.058	0.082
Adjusted R^2	-0.091	-0.087	-0.067	-0.053
F Statistic	0.066 (df = 1; 219)	0.966 (df = 3; 217)	2.548** (df = 4; 166)	2.435** (df = 6; 164)

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

TABLE 5.4: Fixed effects regressions, Gini including rents

	Gini including rents			
	(1)	(2)	(3)	(4)
Exports	−0.010**	−0.010**	−0.013***	−0.017***
	(0.004)	(0.004)	(0.005)	[0.006]
GDP capita		0.085	−1.565	−1.188
		(0.636)	(1.112)	[1.236]
GDP capita^2		−0.004	0.074	0.057
		(0.031)	(0.055)	[0.06]
Students			0.003	0.003
			(0.002)	[0.002]
Manufacturing				−0.002
				[0.001]
Agriculture				−0.001
				[0.001]
N	240	240	189	189
R^2	0.023	0.023	0.090	0.101
Adjusted R^2	-0.067	-0.076	-0.031	-0.031
F Statistic	5.070** (df = 1; 219)	1.715 (df = 3; 217)	4.085*** (df = 4; 166)	3.065*** (df = 6; 164)

Note: robust standard errors in squared brackets

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

TABLE 5.5: Random effects regressions, Gini excluding rents

	Gini excluding rents			
	(1)	(2)	(3)	(4)
Exports	−0.001	0.001	0.003	0.002
	(0.003)	(0.002)	(0.002)	[0.006]
GDP capita		−1.430**	−3.107***	−2.417*
		(0.620)	(0.815)	[1.294]
GDP capita^2		0.069**	0.151***	0.118*
		(0.031)	(0.040)	[0.062]
Students			0.002	0.003
			(0.001)	[0.002]
Manufacturing				−0.002
				[0.001]
Agriculture				−0.002
				[0.003]
Constant	0.317***	7.686**	16.180***	12.694***
	(0.024)	(3.128)	(4.115)	(4.272)
N	240	240	189	189
R^2	0.0004	0.056	0.142	0.266
Adjusted R^2	-0.004	0.044	0.124	0.241
F Statistic	0.104 (df = 1; 238)	4.685*** (df = 3; 236)	7.621*** (df = 4; 184)	10.963*** (df = 6; 182)

Note: robust standard errors in squared brackets

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

TABLE 5.6: Random effects regressions, Gini including rents

	Gini excluding rents			
	(1)	(2)	(3)	(4)
Exports	−0.005*	−0.002	−0.001	−0.001
	(0.003)	(0.002)	(0.002)	(0.002)
GDP capita		−0.641	−2.510***	−2.288***
		(0.595)	(0.785)	(0.835)
GDP capita^2		0.030	0.122***	0.111***
		(0.029)	(0.039)	(0.041)
Students			0.004***	0.004***
			(0.001)	(0.001)
Manufacturing				−0.001**
				(0.0005)
Agriculture				−0.002
				(0.001)
Constant	0.321***	3.665	13.080***	11.986***
	(0.024)	(3.005)	(3.965)	(4.239)
N	240	240	189	189
R^2	0.014	0.043	0.163	0.244
Adjusted R^2	0.010	0.030	0.144	0.219
F Statistic	3.484* (df = 1; 238)	3.494** (df = 3; 236)	8.932*** (df = 4; 184)	9.783*** (df = 6; 182)

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

The results give a mixed picture. Following the indications of the Hausman test and focusing on Table 5.4 and Table 5.5, with a particular emphasis on the more comprehensive regressions, it is observable how the Gini index excluding rents does not seem to be affected by the level of exports, as the coefficient is both negligible and, most importantly, not statistically significant. The constant term absorbs a very sizeable part of the explanation power of the model, while the main included variable to produce a significant effect is GDP per capita. This should probably not strike as surprising, given that it is strongly linked with any measure of income inequality (as a matter of fact, GDP per capita and income per capita, albeit not perfectly overlapping, are two very similar concepts). More surprising, instead, is the lack of significance for the other regressors which should be bringing an impact in the Gini index dynamics: both the education measure and the sectoral composition of workforce do not play a role in this model.

As for Table 5.4, the results are quite different, since there is an arguably small, but statistically significant impact of exports suggested by the fixed effects model. As a reminder, the Gini index is defined in the $[0, 1]$ interval, with 0 representing a perfectly equal distribution of income and 1 being complete inequality. Therefore, the negative coefficient in the regression means a contribution in lowering inequality, even if ever so slightly. Moreover, this coefficient proves to be very insensible to the addition of variables to the model, by remaining almost identical throughout the various specifications and actually slightly increasing in magnitude. On the other hand, GDP per capita loses both in the size and the significance of its effect onto income inequality. Remarkably, then, all the FE specifications show a small and significant effect of trade on the Gini index including rents, but fail to attribute any such weight to the other regressors, whose coefficients are small and not significant all across the various specifications.

More surprising seems the lack of an effect coming from the variables regarding sectoral composition of the workforce. This might suggest at least two features which are highly inherent to the territorial dimension of the labour market: a great degree of mobility within sectors and/or a comparable one between sectors. This would imply that it is not necessarily the case in which certain sectors are particularly prone to wage inequality brought by foreign competition, as it is instead vastly showed in the empirical literature.

It is also important to remember, however, a crucial factor induced by using a regression model with individual effects, i.e. that such a technique, however useful, tends to eradicate most of the variation from the estimation: excluding the effect

induced by local features means considering time-led variations only. This aspect should be reflected in the estimation with relatively unimpressive numbers, which are surely already impacted by the mentioned drawbacks coming from both data and regression models.
Albeit not being a perfect tool, an indicator of the goodness-of-fit like the R-squared value signals a certain lack of explicative power in almost all regressions models, except in the case of the simple multiple regression with all variables, which is not, however, the appropriate model for this panel dataset[1]. This is partly true even when taking into account the difference in the reported R-squared, in that the two panel data models report its *within* measure (for the fixed effects models) and its *overall* one (for the random effects regressions).

Overall, the results suggest that no sizeable effect is brought by exports in shaping the Gini index. However, it is necessary to remark how the endogeneity issue arising from the omitted variable bias was not completely eradicated by the unobserved effects models, so the estimated coefficients could be at least slightly biased. This implies that, with the assumptions showed in Section 4.2, the coefficient is probably overestimated (in absolute value), so that the actual impact of exports on income inequality is lower than the results show.
Given all these considerations, the results appear to be in contrast with a large part of the cited literature, which find more significant and sizeable impacts of trade on income inequality. As previously mentioned, research has mostly focused on developing countries (especially those facing an opening to trade) rather than a rich one like Italy, and only a few papers have analysed sub-national variations, so that comparing the outcome of those results with the ones in this thesis is not straightforward, nor was it having a precise *a priori* idea about what my estimation strategy would output as a result. However, the initial correlation matrix in Figure 3.1 showed a very weak degree of relationship between exports and the Gini index in the data (especially in the case of the measure excluding rents). This indication, albeit imperfect, could be considered as a first clue for the almost null estimated impact. Overall, it could be said that an uncertain result might be expected for an analysis like the one in this thesis, considering the micro-level of interest and, most importantly, the absence of shocks in trade patterns.

[1] As a side note, the multiple regressions also display a significant effect of the *Students* variable, which is instead not consistent in the RE and FE models

In this sense, studying a developing country should lead to more marked indications, especially through techniques such as a difference-in-differences regression model, which perfectly suits a "before-after shock" comparison.
An alternative approach to extend the analysis of this thesis, which I was not able to pursue because of the lack of the necessary data, would be extending the sample to a larger number of regions, for example by including all the European ones. This would allow for a bigger sample in the analysis and the possibility to compare more heterogeneous regions, opening to the chance of an interesting debate.

6. Conclusions

A large literature has investigated the effect of trade on inequality from various angles, often coming to the conclusion that an impact, increasing inequality, is present at least in the first period after opening to foreign competition. This seems to happen as both between- and within-sectors mobility of workers might be limited. In this thesis, I estimate the impact of export activity on income inequality, using regional-level data from Italy.

Using panel data models such as fixed and random effects regressions, I find that no such effect is present onto the Gini index excluding rents estimated via random effects, but in fact across all regression models. A small effect is instead estimated to lower the measure of inequality including rents. This is at odd with a large chunk of the literature which, mostly looking at developing countries opening to trade activity, shows a more definite, negative impact of trade on inequality.

Endogeneity arising from the omission of variables such as the level of imports tackle the unbiasedness of the estimation, therefore leading to necessary caution in interpreting the results of the empirical analysis.

Bibliography

[1] Thomas Piketty. *Capital in the 21st Century*. Harvard University Press, 2014. ISBN: 9780674430006. URL: http://www.jstor.org/stable/j.ctt6wpqbc.

[2] Jonathan Eaton, Samuel S. Kortum, and Francis Kramarz. "An Anatomy of International Trade: Evidence From French Firms". In: *Econometrica* 79.5 (2011), pp. 1453–1498. DOI: 10.3982/ECTA8318. URL: http://onlinelibrary.wiley.com/doi/10.3982/ECTA8318/abstract.

[3] Daniel Trefler. "The Long and Short of the Canada-U. S. Free Trade Agreement". In: *American Economic Review* 94.4 (2004), pp. 870–895. DOI: 10.1257/0002828042002633. URL: https://www.aeaweb.org/articles?id=10.1257/0002828042002633.

[4] David H Autor, David Dorn, and Gordon H Hanson. "The China Syndrome : Local Labor Market Effects of Import Competition in the United States". In: *American Economic Review* 103.6 (2013), pp. 2121–2168. DOI: http://dx.doi.org/10.1257/aer.103.6.2121. URL: https://www.jstor.org/stable/42920646.

[5] Giulio Bosio, Mara Grasseni, and Anna Maria Falzoni. "Local labour markets, technology, and Chinese imports' competition: Evidence from Italy during recession". Bergamo, Italy, 2016. URL: http://www.aiel.it/cms/cms-files/submission/all20160615154203.pdf.

[6] Ehran Artuç and John Mclaren. "Trade Policy and Wage Inequality : a structural analysis with occupational and sectoral mobility". Cambridge, MA, 2012. URL: http://www.nber.org/papers/w18503.

[7] Stephen Dobson and Carlyn Ramlogan. "Is there an openness Kuznets curve?" In: *Kyklos* 62.2 (2009), pp. 226–238. DOI: 10.1111/j.1467-6435.2009.00433.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-6435.2009.00433.x.

[8] Abdul Jalil. "Modeling income inequality and openness in the framework of Kuznets curve: New evidence from China". In: *Economic Modelling* 29.2 (2012), pp. 309–315. DOI: 10.1016/j.econmod.2011.10.012. URL: http://www.sciencedirect.com/science/article/pii/S0264999311002483.

[9] Elhanan Helpman et al. "Trade and inequality: From theory to estimation". In: *Review of Economic Studies* 84.1 (2017), pp. 357–405. DOI: 10.1093/restud/rdw025. URL: https://academic.oup.com/restud/article-abstract/84/1/357/2669956.

[10] Elhanan Helpman, Oleg Itskhoki, and Stephen Redding. "Inequality and Unemployment in a Global Economy". In: *Econometrica* 78.4 (2010), pp. 1239–1283. DOI: 10.3982/ECTA8640. URL: http://onlinelibrary.wiley.com.proxy.library.cornell.edu/doi/10.3982/ECTA8640/abstract%7B%5C%%7D5Cnhttp://onlinelibrary.wiley.com.proxy.library.cornell.edu/doi/10.3982/ECTA8640/pdf.

[11] Eric A. Verhoogen. "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector". In: *Quarterly Journal of Economics* 123.2 (2008), pp. 489–530. DOI: https://doi.org/10.1162/qjec.2008.123.2.489. URL: https://ideas.repec.org/p/iza/izadps/dp2913.html.

[12] Petia Topalova. "Trade liberalization, poverty and inequality: evidence from Indian districts". Cambridge, MA, 2005. URL: http://www.nber.org/papers/w11614.

[13] M Daumal. "The Impact of Trade Openness on Regional Inequality: The Cases of India and Brazil". In: *International Trade Journal* 27.3 (2013), pp. 243–280. DOI: 10.1080/08853908.2013.796839. URL: https://ideas.repec.org/a/taf/uitjxx/v27y2013i3p243-280.html.

[14] Susan Chun Zhu and Daniel Trefler. "Trade and inequality in developing countries: A general equilibrium analysis". In: *Journal of International Economics* 65.1 (2005), pp. 21–48. DOI: 10.1016/j.jinteco.2003.11.005. URL: http://www.sciencedirect.com/science/article/pii/S0022199604000200.

[15] Dominik Hartmann et al. “Linking Economic Complexity, Institutions, and Income Inequality”. In: *World Development* 93.September (2017), pp. 75–93. DOI: 10.1016/j.worlddev.2016.12.020. arXiv: 1505.07907. URL: https://arxiv.org/abs/1505.07907.

[16] Jan Tinbergen. “Shaping the World Economy: Suggestions for an International Economic Policy”. In: *The Economic Journal* 76.301 (1966), pp. 92–95. DOI: 10.2307/2229041. URL: http://www.jstor.org/stable/2229041.

[17] Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2010, pp. 310–312. ISBN: 9780262232586. URL: http://www.jstor.org/stable/j.ctt5hhcfr.

[18] Badi H. Baltagi and Qi Li. “A lagrange multiplier test for the error components model with incomplete panels”. In: *Econometric Reviews* 9.1 (1990), pp. 103–107. DOI: 10.1080/07474939008800180. URL: http://www.tandfonline.com/doi/abs/10.1080/07474939008800180.

[19] Felix M. Arellano. “Computing Robust Standard Errors for Within Group Estimators”. In: *Oxford Bulletin of Economics and Statistics* 49.4 (1987), pp. 431–434. DOI: 10.1111/j.1468-0084.1987.mp49004006.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0084.1987.mp49004006.x/abstract.

[20] Badi H. Baltagi and Qi Li. “Testing AR(1) against MA(1) disturbances in an error component model”. In: *Journal of Econometrics* 68.1 (1995), pp. 133–151. DOI: 10.1016/0304-4076(94)01646-H. URL: http://www.sciencedirect.com/science/article/pii/030440769401646H.

[21] Halbert White. “A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity”. In: *Econometrica* 48.4 (1980), pp. 817–838. DOI: 10.2307/1912934. URL: http://www.jstor.org/stable/1912934.

[22] Staffan Burenstam Linder. *An Essay on Trade and Trasformation*. Stockholm: Almqvist & Wicksell, 1961. ISBN: B000WO8EQ6.