IMT School for Advanced Studies Lucca
Lucca, Italy

Essays on Urban and Environmental Economics

PhD Program in Economics, Markets, Institutions
XXVII Cycle

By
Nicola De Vivo
2016

The dissertation of Nicola De Vivo is approved

Programme Coordinator: Massimo, Riccaboni, IMT School for Advanced Studies Lucca, Piazza San Poziano 6, 55100, Lucca, Italy

Advisor: Guido, Caldarelli, IMT School for Advanced Studies Lucca, Piazza San Poziano 6, 55100, Lucca, Italy

Co-Advisor: Marco, Modica, CNR –IRCrES, Research Institute on Sustainable Economic Growth, Via Corti 12, 20133, Milan, Italy

The dissertation of Nicola De Vivo has been reviewed by:

Matthieu, Cristelli, ISC-CNR, Institute for Complex Systems, Rome, Italy
Antonio, Scala, ISC-CNR, Institute for Complex Systems, Rome, Italy

IMT School for Advanced Studies Lucca
2016

To Raffaella and GreenBean

Contents

List of Figures

List of Tables

Acknowledgements

At last, I succeed in reaching my objective: I am going to defend my Ph.D. Thesis. It took much more than expected, but, however, the most important thing is to be here now. In the meanwhile, so many things happened in my life since when I began and I am changed so much, it seems to me that the beginning of my Ph.D program was something belonging to a life ago.

During this long road, so many people hit my road and I would like to thank them all, hope to do not forget anyone. First of all, I would like to thank my supervisor, prof. Guido Caldarelli, for his support and advices during the period in which I wrote my work: it has been an honor for me being one of his Ph.D. students.
Then, I would like to thank all of my co-authors: Valentina Tortolini on the first paper, Marco Modica, Peter Nijkamp and Aura Reggiani on the second one and Giovanni Marin fo the third one: all of them have played a crucial role during the development of this thesis work. Most of all, I would like to spend a few words about Marco Modica (my co-advisor) and Giovanni Marin. I cannot consider them only co-authors because they are, firstly, friends of mine. Their help has been of fundamental importance since the very beginning of this work and it increased as time passed by; they helped me in a lot of different ways, from the simple helping with some Latex or Matlab problems to the advices about the best way to continue a work that seemed to be run aground.
Lastly, I would like to thank IMT Institute for Advanced Studies Lucca and all the people I met during these four years for the opportunity they gave me to approach the wonderful world of actual scientific research (even though, then, I decided to leave it) and to make me feel part of a great international community: from all of the professors and the researchers I have met during these years to the Ph.D. office people, from the canteen people to the facilities office people and all the colleagues from other Ph.D. programs, I want to thank everyone of them because everyone, in a way or another, is part of this thesis. Most of all, I would like to thank two of my colleagues: Niccolò Fiorini and Tiziano Distefano. As I said about Giovanni and Marco, they are much more than colleagues, they are friends. My time spent at IMT would have been much more difficult and not enjoyable at all if they would not have been part of IMT and, in particular, of my class. I will never forget all the hours spent in our study room, studying all together to overtake all the exams, spaced out by our tennis table matches or our ice creams and yogurts.

On the personal side of my life, first of all, I have to thank my wife, Raffaella, and my future son (or daughter, we still do not know), Green-Bean. In few words, Raffaella is simply the center of my life and the only woman I want to grow old with. She gave me the strength to keep going in pursuing this objective when I was demoralized on the possibility to reach it. She sacrificed with me a lot of weekends and holidays during these three years in which I worked on my thesis and I could study only on holidays, as I was working in working days. I have to say, I am always more convinced she believed in me reaching this objective, even more than I believed, even when everything worked against me and I did not see the light at the end of the tunnel. She was always by my side and now, we are going to have one of the greatest joy of all: the birth of our baby. He (or she) is not still born but he is already one of the reasons of my life. When I feel him kicking, I cannot even describe the joy he brings in my heart and that is way I want to thank him (or her).
Then, I would like to thank my parents and my brothers (together with their girlfriends) for always supporting me, not only during my Ph.D. program but during all of my life. They always trusted in me and support my choices (even though, they did not agree with the choices I was making) and I hope I will always be a son and a brother they can be proud of.
Lastly, I would like to thank all of my friends, no one excluded, because my life without them would absolutely not be the same and I am sure it would not be so beautiful and happy as it is at the moment.

Vita and publications

VITA

April 2, 1982	Born
2011	Master's Degree in Physics University of Salerno Salerno, Italy
2012	Course on Quantitative Finance MIP School of Business Milano, Italy

1 Introduction

Urban and Environmental Economics are two branch of Economics that are more and more tightly interconnected and always should be. Trying to study how population is distributed across cities is a key point for several issues, either from a theoretical point of view or from a policy implementation point of view. Among the policies possibly affected by population distribution, policies dealing with climate change are one of the most affected, as people keep attributing a growing importance to the quality of their life, to the protection to natural risks and, then, policy makers have to care about how people are spread across cities. Viceversa, an effective climate policy should try to improve people's life quality and to leave at least unaltered the population distribution, as it can cause, for example, job losses due to company relocation, which can alter in a substantial way the way in which population distributes in cities.
In this thesis work, we aim to provide the international scientific community with new insights on some of the most relevant topics in these two branches: What is the actual distribution of population in cities of a country and what were the processes leading to it? Using different demographics variables or introducing some demographic characteristics (as age structure) could lead us to different results and give us different insights on the way in which people distribute across cities in a country? Could the greenhouse gas emissions behavior of a company be affected by the way in which the company is given the rights to emit a certain amount of carbon dioxide? These are the main topics concerning the three chapters of which this thesis is constituted.

In the first chapter, we tackle a long living research question: in which way people distribute across cities? The only stylized fact recognized by the entire scientific community is that the way in which people agglomerate is not random, but it follows a certain distribution. Three are three main candidates to be the real distribution: power-law, log-normal and Double Pareto log-normal distributions. Scientific community provided empirical evidences supporting all of them, therefore an agreement between the three opponents has not still reached. Approaching to these topics means to get in touch with two empirical regularities: Zipf's Law and Gibrat's Law. Roughly speaking, the first one states that the city size in a country is proportional to its rank and, then, it follows a power-law distribution, whereas the second one affirms that the growth rate of an entity (i.e., firm, city, etc.) is independent of its size. Although the acknowledgement of two empirical regularities could be an help in de-

veloping a theory, this is not the case because the two above mentioned relationships seem to be not unified by a single theory, as Gibrat's Law should lead to a Log-normal distribution, instead of a power law one, as predicted by Zipf's Law.
In the chapter, we study city size distribution for Italian cities, using data from three different censuses. After collecting some empirical result confirming or rejecting the stylized facts collected in the previous literature (and, in particular, about Zipf's Law and Gibrat's Law), we find that, among the different distributions proposed in the literature, the most suitable distribution for Italian cities seems to be the Double Pareto log-normal distribution, as distribution of Italian cities seems to be log-normal in the body and Pareto in both the tails. Then, we try to provide a simple model about the evolution of the city distribution in order to find average parameters for the growth rate distribution (in particular, mean and variance) leading to the actual distribution. As this task was not fulfilled, then, we start to question some of the widely recognized regularities and we find that they have not been always as they seem at the moment.

In the second chapter, instead, we go beyond the idea of finding what the actual distribution of people in cities should be and we look at whether the way in which people spread across cities can vary whether we consider different demographic variables (for example, employment rather than population) or whether we add another parameter to the analysis, as age structure can be. In this chapter, we use yearly data on population and employment for all German cities and towns, from 2001 to 2011, split by age cohorts, to perform Zipf's Law analysis and Gibrat's Law analysis. We found that differentiating for age cohort and using different dependent variables can lead to very different result. For what Zipf's analysis is concerned, we found that employment is much more concentrated with respect to population and younger cohorts show a clear tendency towards agglomeration in larger cities, whereas elder cohorts show the opposite behavior[1]. For what, instead, Gibrat's Law is concerned, we found that the two variable behave in a different way. For example, Gibrat's Law seems to be in operation for employment but not for population, where biggest cities seem to grow more than smallest ones.
All of these new results should be taken into consideration, either in developing new theoretical models of urban growth that want to be more accurate and in structuring new policies that want to be more effective.

[1] Zipf's exponent of Pareto distribution can be seen as a measure of concentration

Then, in the last chapter, we move from urban economics to environmental economics, and we try to analyze emissions behavior of establishments under the EU ETS (Emissions Trading Scheme), a cap-and-trade scheme regarding greenhouse gases emissions. In few words, establishments under this scheme must have permits to emit CO_2: these permits can be grandfathered (in the first two steps of the project, almost all of the permits were released in this way) or they have to be auctioned and then traded among the plants (in the third phase of the project, the amount of grandfathered permits was drastically reduced). Since the beginning, this scheme has attracted a lot of attention among scholars and policy makers, because it represents the central EU policy instrument in order to mitigate climate change and to be compliant to objectives recorded in the Kyoto protocol. The potentially harmful impacts on the competitiveness of European firms subject to the EU ETS (industries have to buy permits for emissions and this increases costs) coupled with the fact that the EU ETS was unilaterally introduced in Europe may induce firms to relocate their carbon-intensive production activities in countries with less stringent regulations for mitigating climate change (this effect is called carbon leakage effect). Carbon leakage has two negative implications for the country (or group of countries) that introduces an unilateral stringent climate mitigation policy. First of all, emissions at the global level are not reduced but only displaced towards other countries. Second, the relocation of carbon-intensive industries has a negative impact on the wealth within the country, as it causes job losses. For this reason, the European Commission has been particularly sensitive about the issue of carbon leakage and in the third phase of the scheme (the phase in which industries are required to auction all of the permits they need) decided to allocate permits for free, only to those sectors that are exposed to high level of carbon leakage risk. So we decided exploit the asymmetry in the allocation mechanisms introduced from the third phase of the EU ETS as a way to evaluate whether different allocation mechanisms are neutral in terms of emission abatement decisions.
We found that, for establishments in the sector of manufacturing, it seems that grandfathering permits leads to an increase in emissions with respect to plant who have to auction or buy all of their permits.
This result is in contradiction with respect to the theoretical prediction of neutrality of allocation mechanisms in cap-and-traade schemes, thus leading to sub-optimal outcomes. These findings should be taken in consideration by policy makers about possible ways of improving ETS-like schemes in order to improve their economic efficiency and correct for po-

tential distortions induced by specific rules for specific case, such as the case of carbon leakage.

2 On the distribution of (all) the Italian cities

2.1 Introduction

An accurate description of the spatial distribution of population is important for a number of theoretical and policy relevant issues, ranging from a better understanding of firms and people localization decisions to the implementation of national and regional policies in terms of incentives and transport infrastructures (as it is reported in Fazio and Modica (2012)[1]). The only fact about the international scientific community agreed is that the way the population is distributed across geographic areas, while continuously changing, is not totally random, but it follows a certain distribution. In fact, there is a strong tendency toward agglomeration, i.e. the concentration of the population within common restricted areas like cities or towns. And while physical geography (rivers, seas, coasts and mountains) may have played a crucial role in early settlements, in the current day and age, the evolution of the population across geographic locations is an extremely complex amalgam of incentives, and actions taken by millions of individuals, businesses and organizations at the same time. Most of the people involved in this kind of research agrees that economic factors are the principal determinant of the dynamics of city population. Unfortunately, the literature is still far from reaching consensus on which kind of distribution should be.
This lack of consensus about city size distribution is essentially due to a puzzle caused by two well-known and well-established empirical regularities: Zipf's law (named after the name of the economist who first proposed it, George Zipf, in his work of 1949[2]: that is, city size distribution must follow a Pareto distribution, at least in the upper tail) and Gibrat law (named after the first economist who proposed it in 1931, Robert Gibrat[3]: that is, city growth size does not depend on the city size). Although the acknowledgement of two empirical regularities could be an help in developing a theory, this is not the case because the two above mentioned relationships cannot be unified by a single theory: Gibrat law, also known as proportionate growth process, gives rise to a log-normal distribution, instead of a Pareto one, as it is supposed by Zipf's Law. On the basis of these results, most of the literature dealing with this topic is based on the debate between scientists providing theoretical and empirical support to the Pareto distribution and scientists opposing this view and providing theoretical and empirical support to the log-normal distribution.
However, more recently, another kind of distribution attracted the atten-

tion of the scientific community: the Double Pareto log-Normal distribution, proposed in the seminal paper by Giesen et al. (2010)[4], that is a distribution that is Pareto in the two tails and log-normal in the body. This is because this distribution seems to be the perfect trait-d'union between the two above mentioned empirical regularities: it arises from a little modification of the proportionate growth process (we will explain it better later on) and it shows Pareto behavior in both of the tails.
So, three specific distributions are the most accredited in the literature: the power-law, the log-normal and the Double-Pareto log-normal. Disentangling between the three has even important implications from a theoretical point of view. For example, a power-law distribution implies that cities are the result of agglomeration forces and industry specific shocks. A log-normal distribution, instead, implies that cities grow proportionally and independently from the initial city size and their distribution results from city-wide rather than industry specific shocks, as explained in Gabaix (1999)[5]. In Reed(2002)[6], we can find the proof that a Double Pareto log-normal distribution arises, instead, from a process in which cities have not all the same age but some are older than others; this process resembles the Yule process, first described in biology (Yule (1925)[7]).
As you can simply see from this brief introduction on the topic, studying the actual city size distribution is still an open question, without an agreement on what kind of distribution it is neither an universally accepted theory to explain it. Our paper aims to provide further empirical evidence to the debate using the most recent data about Italian distribution (2011 census) together with the last two censuses (2001 and 1991) in order to confirm or reject some of the main findings obtained in the literature (explained in details in sec. 2.2). We find that, among the different distributions proposed in the literature, the most suitable distribution for Italian cities seems to be the Double Pareto log-normal distribution, as distribution of Italian cities seems to be log-normal in the body and Pareto in both the tails.
Based on this result, we try to develop a simple model about the evolution of city distribution in order to find the average parameters for the growth rate distribution leading to the actual city size distribution. As we did not succeed in this task, we start questioning the hypothesis of the model, and, in particular, the hypothesis of proportionate growth process, the so-called Gibrat Law.
The rest of the paper is structured in the following way. First, in section 2.2, we give a complete literature review of the topic. Then, in section 2.3, we describe our data. Instead, in section 2.4, we will provide the whole set of our empirical results and in section 2.5 we present the theoretical

results obtained by means of numerical simulations. Lastly, in section 2.6, we question the validity of the law of proportionate growth for the evolution of Italian cities, showing some examples in which it seems to not hold. Section 2.7 concludes.

2.2 Literature review

As we said before, the main substantial difficulty in the description of the population mobility derived from a puzzle caused by two very well known empirical regularities.
The first empirical regularity is that the largest cities satisfy Zipf's Law. As we said before, despite the apparent chaotic evolution of city populations, surprising regularities have been observed in the size distribution of cities. As early as 1682, Alexander Le Maître observed a systematic pattern of the size distribution cities in France. He describes how the size of Paris is related to two groups of cities, each of them proportionally smaller than Paris[9]. But it was not until 1913 that Felix Auerbach[10], and George Kingsley Zipf[2] in 1949, formally established the first empirical regularity. They proposed for the first time that the city size distribution could be closely approximated by a power-law distribution. In particular, city sizes are said to satisfy a peculiar power-law distribution, referred to as Zipf's Law, that is, for large sizes S, we have

$$P(Size > S) = \frac{a}{S^{\zeta}} \tag{2.1}$$

where a is a positive constant and $\zeta = 1$. That is, the size of a city times the percentage of cities with larger size equals a constant.
An approximate way of stating Zipf's Law is the so-called rank-size rule. This is a deterministic rule that follows from the definition: the second largest city is roughly half the size of the largest city, the third largest city is roughly a third the size of the largest, etc. That is, whether we rank cities from largest (rank 1) to smallest (rank n), and denote their sizes $S_1 \geq S_2 \geq ... \geq S_n$, respectively, the rank i for a city of size S_i is proportional to the proportion of cities greater than i. Therefore, by rewriting Eq. (2.1) we have

$$S_i \simeq \frac{k}{i} \tag{2.2}$$

where k is a positive constant.

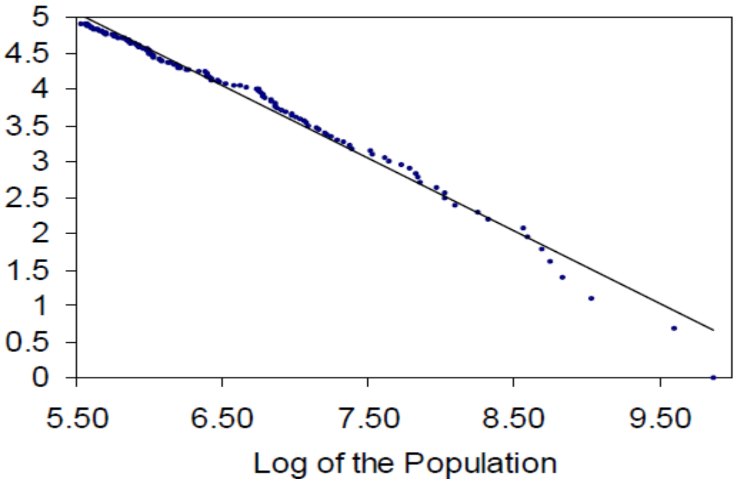


Figure 2.1: Log Size vs Log Rank

This relation can easily be derived in an elementary way. Let us suppose that at each time a person is born in a city, that all cities have the same birth rate and that with very small probability, this person creates a new city. Then, the total population n_0 of cities existing at time t_0 is proportional to t_0: $n_0 \sim t_0$. The rank of the city created at time t is proportional to t: $R \sim t_0$. The ratio between the size of the city and the total population remains the same: $K/n = 1/n_0$. This implies that: $K \sim 1/n_0 \sim 1/t_0 \sim 1/R$. In the end, size is inversely proportional to the rank. It is important to bear in mind that even though Zipf's law holds perfectly, the rank-size rule would hold only approximately.

In fig. (2.1) we can visualize the rank-size rule for cities of the United States. In order to do this, we took a country (in this particular example, the United States), and ordered its cities by population: New York has rank 1, Los Angeles has rank 2, etc. We then draw a graph, known as Zipf's plot: on the y-axis, we place the logarithm of the rank (New York has rank ln1, Los Angeles has rank ln2, etc.); on the x-axis, the logarithm of the population of the corresponding city (which we previously referred to as the size of the city). As we can see, the result is something very close to a straight line. Furthermore, fitting a linear regression yields

$$\ln Rank = 10.53 - 1.005 \ln Size \tag{2.3}$$

we can observe that R^2 is 0.986 (these data have been taken from Gabaix (1999)[5]). As you can see, the slope of the regression line is very close to -1, as it is predicted by the rank-size rule.
Support for this empirical regularity comes from numerous country studies and comparative international evidence. In the field of country studies, we can mention Dobkins and Ioannides (2000[11]), Fujita, Krugman and Venables (1999[12]) and Gabaix (1999[5]) while Rosen and Rosnick (1980[13]), Brakman, Garretsen and van Meerwijk (2001[44]) and Soo (2005[15]) are the most complete empirical international comparative studies. These are typically conducted along the lines given by the following equation:

$$\ln Rank = A - \zeta_n \ln Size_i \tag{2.4}$$

where by ζ_n we denote the Zipf exponent deriving from a sample whose length is n. This procedure is the most used in this kind of studies because it has the main advantage that it yields even a direct visual goodness of fit with the power law.
Dobkins and Ioannides report OLS estimates of ζ, that are obtained along the lines of eq. (2.4) with repeated cross sections of U.S. Census data for metro areas. Their estimates decline from 1.044 in 1900, to 0.949, in 1990. When they use the upper one-half of the sample only, a practice that conforms to some other estimations of Zipf's Law (such as in Fujita, Krugman and Venables (1999[12])), the estimate of ζ declines from 1.212 in 1900, with 56 metro areas in the entire sample, to 0.993 in 1990, with 167 metro areas in the sample. Fig. (2.1) is taken from Gabaix (1999)[5] and reports an estimate equal to 1.005, using the 135 largest metro areas in 1991 as reported in the Statistical Abstract of the United States. Rosen and Resnick (1980[13]) examine city distributions for 44 countries in 1970. The average Zipf's exponent is 1.13 with a standard deviation of 0.19, with almost all countries falling between 0.8 and 1.5. Brakman *et al.* (2001[44]) show that city-proper data are associated with higher Zipf exponent (mean=1.13, standard deviation=0.19, N=42) than urban agglomeration data (mean=1.05, standard deviation=0.22, N=22). Soo (2005[15]) updates these results without altering the basic findings. He finds a Zipf coefficient of 1.105, for cities, but 0.854 for urban agglomerations. As we can simply see, the estimated dispersion in the Zipf exponent is large. Looking at the average of exponent estimates, however, we

see that whether the average value ζ is not exactly equal to 1, it is typically in the range [0.85 1.15]. This results lead the international scientific community to conclude that power laws describe well the empirical regularity, with a Zipf exponent typically around 1. Furthermore, they add that predicting a value in a range [0.8 1.2] may be included in the list of criteria used to judge the success of urban theories (as reported in Gabaix (1999)[5]).

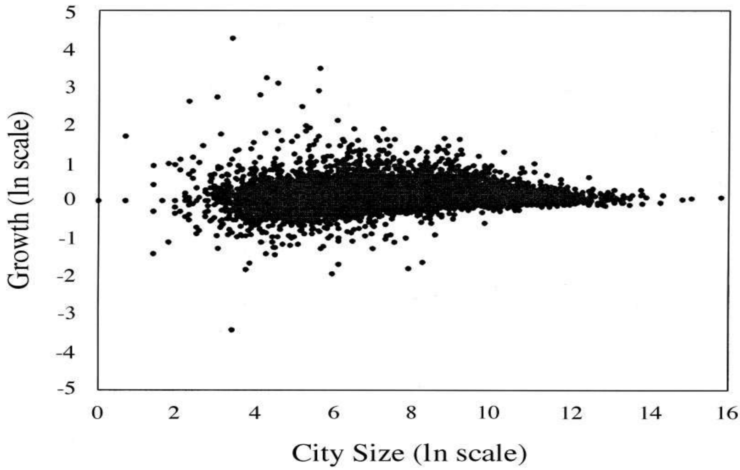


Figure 2.2: Growth rate vs city size

The second most important empirical regularity is that the growth rate of city populations does not depend on the size of the city. Even though growth rates between different cities can even vary substantially, there is no systematic pattern with respect to size, i.e., the underlying stochastic process is the same for all cities (as reported in Gibrat (1931)[3]). This is labeled the proportionate growth process and it is usually referred to as the Gibrat Law. Empirical research (among which we can mention, for example, Eaton and Eckstein (1997[17]), Ioannides and Overman (2003[18]) and Glaeser Scheinkman and Shleifer (1995[47])) has repeatedly shown that city growth is proportionate: larger cities on average do not grow faster or slower than smaller cities. This effect is visually shown in fig. (2.2): in this picture, it is reported a scatter plot in logarithmic scale of city growth rate versus city initial size. As we can simply see, there is no emergency of any trend in the study of growth rates as a function of the city size: the data can be very well fitted by means of an horizontal line meaning that there is no systematic difference between the growth rate of large and small cities.

While it is surprising that such regularities emerge from a highly intricate underlying mechanism, there is also a puzzle: the two regularities cannot easily be reconciled, leading Krugman to write in 1995 "*We have to say that the rank-size rule is a major embarrassment for economic theory: one of the strongest statistical relationship we know, lacking any clear basis in theory*"[20]. In particular, the proportionate growth process gives rise to the log-normal distribution (as reported in Eeckhout (2004[16]) and not to the power-law distribution we have seen in eq. (2.1). This is a well-known proposition established by Gibrat (1931[3]) and originally formulated by the astronomer Jacobus C. Kapteyn (1903[21]): a stochastic growth process that is proportionate gives rise to an asymptotically log-normal distribution. This is not to say that a proportionate growth process plus "something else" cannot give rise to a power-law distribution or another distribution.
There is a long tradition in the economics of income inequality starting with David G. Champernowne (1953[22]) and industrial organization (for example, John Sutton (1997 [23]) and Boyan Jovanovic (1982 [24])) studying the relation between proportionate growth and size distributions different from the log-normal. With respect to the size distribution of cities, Xavier Gabaix (1999[5]) and Aharon Blank and Sorin Solomon (2000[25]) propose a resolution of the puzzle and show that proportionate processes can generate Zipf's Law at the upper tail. In particular, they consider random growth processes with the entry of new cities and apply a process developed in Champernowme (1953[22]) and Kesten (1973[26]). In order to have a steady state, they need a mechanism that prevents the small cities from becoming too small: the clearest version they propose of such a mechanism is given by a "random walk with a lower barrier". They claim this mechanism is necessary because if such a mechanism were not present, the city-size distribution would become degenerate, and in particular they claim the city-size distribution would become a log-normal one, where most cities would have infinitesimal size.
While these process do generate steady states that are power-law distributions, Blank and Solomon (2000[25]) point out that details specifying and enforcing the smallest size of the cities are crucial, as are the rules for creating new entries cities. Whether or not the resulting limiting distribution is power-law with exponent equal to one is very sensitive to this

entry process. Moreover, testing whether the entry process satisfies the exact and detailed requirements for the power-law distribution is a challenging empirical endeavor (for the Metropolitan Areas in the United States, for example, there has been no entry or exit in the set of MAs between 1990 and 2000; the 276 MAs in 1990 are identical to those in 2000) and up to date, no such evidence has been provided.

In contrast to all this literature, Eeckhout (2004[16]) proposes a different solution to this puzzle, exploiting the availability of US Census 2000 data. This dataset is substantially larger than those of earlier censuses, including observations on the entire size distribution of geographic locations, referred to as "places". In particular, in this dataset, there are 25359 places, including cities, towns and villages, ranging in population from 1 to over 8 million while, in the datasets used in the previous literature, only the truncated distribution, i.e., the upper tail of the distribution (the 135 largest cities) was considered, namely the 0.5 percent of the 2000 sample and 30.2 percent of the sample population. Using this new dataset, Eeckhout shows that the size distribution of the entire sample is log-normal and not power-law, as it was claimed by the previous literature. Moreover, for the observations for which it was possible, he computes the growth rate of cities, showing that growth is independent of city size. Thanks to these two results, he can establish that, when we consider the entire distribution and not only the upper tail, Gibrat's prediction concerning the stochastic process holds. The only issue to solve is to explain why there is a so wide confirmation of Zipf's Law in the literature, even though the underlying distribution is log-normal. The two distributions are very different and this is why the distribution should never fit to power-law whether the true distribution is a log-normal one: goodness of fit tests will categorically reject the power-law distribution. Nevertheless, when regressing log rank on log size (even for the entire size distribution), the coefficient comes out significant. However, this regression test merely confirms that there is a relation between rank and size, but it does not provide a test for the linearity of this relation. As such, testing the significance of the linear coefficient is not the equivalent of a goodness of fit test for the power-law distribution.
More important though is that until this work the literature considered the truncated distribution. At the very upper tail of the distribution, there is no dramatic difference between the density function of the log-normal and the power-law because both the power-law and the truncated log-normal distributions are downward sloping and similar (the power-law is slightly more convex). As a result, both the distributions

trace the data relatively closely. The problem is that the estimated coefficient on the power-law distribution is extremely sensitive to the choice of the truncation point: as the truncation point increases, the estimated coefficient increases, while the estimated log-normal coefficients remain unchanged. Moreover, for lower truncation points, the power-law fits the data less and less well. In the paper, it is shown that these observed empirical changes in the estimated power-law coefficient are theoretically consistent with the comparative static of a changing truncation point of the log-normal distribution. This sensitivity of the power-law coefficient to the truncation point has been observed even in the previous literature (as reported in Gabaix and Ioannides (2003[8]) even though they have been explained by different theories.
Levy (2009[27]) criticizes the work by Eeckhout (2004[16] claiming that in the top 0.6 percent range of the largest cities, the size distribution diverges dramatically and systemically from the log-normal, and instead the size distribution is much better described by a power-law. For this reason, he divides the distribution of city size in two different regions: the bottom and the middle ranges where the empirical distribution fits the log-normal, and the top range where the empirical data fit a power-law distribution. While this upper part accounts only for the top 0.6 percent of the cities (about 150 out of 25359), the claim is relevant since these cities account for over 23 percent of the total US population.
Eeckhout (2009[28]) replies to this criticism claiming that Levy is induced in a mistake by the peculiarities of the log-log plot that, according to him, has 4 main caveats:

1. The log scale heavily distorts low rank observations because it blows up bounds and deviations for very few at low ranks

2. The data are heavily concentrated in the middle of the distribution

3. Log-log plots are very uninformative for parts of the distribution with a high density and a large number of observations

4. A biased representation similar to log-log plots occurs with a normal probability plot (one of the evidences brought by Levy), where the deviations in the middle of the distribution are not picked up.

In the end, he concludes claiming that city size distribution is a log-normal one because the upper tail falls into the confidence bands of the log-normal estimates.

Recently, a new distribution arose support from the international scientific community: the Double Pareto Log-Normal distribution[2]. This distribution has a log-normal body in the medium range and exhibits a power-law in both the lower and the upper tail. Empirical evidences on 8 different countries show that the city size distribution in all countries can be well approximated by a log-normal but they show even that a Double Pareto log-normal distribution approximates data even better (as reported in Giesen *et al.* (2010[4]). The authors claim that, although it is quite obvious that a more flexible functional form delivers a better fit, their work is not a theory free one because they rely on a mechanism developed by Reed (2002[6]). In this model, cities grow stochastically as under Gibrat's law, but in every time interval dt there is the probability λdt that a new city emerges as a satellite of an existing one. The initial size of the new city is drawn from a log-normal distribution with mean μ_0 and variance σ_0^2. These new cities then also exhibit proportionate growth. Asymptotically, this process leads to a Double Pareto log-normal distribution, which now comes out from a reasonable evolutionary process and not only as a trick to better fit the data.

2.3 The data

It is important to point out that there is no universally accepted definition of a *city* for statistical purposes. The main distinction is between the *proper city* and the so-called *Metropolitan Area (MA)*. A metropolitan area is a region consisting of a densely populated urban core and its less-populated surrounding territories, sharing industry, infrastructure, and housing. A metropolitan area usually comprises multiple jurisdictions and municipalities: neighborhoods, townships, cities, exurbs, counties, and even states. As a general definition, a metropolitan area combines an urban agglomeration (the contiguous, built-up area) with zones that are not necessarily urban in character, but closely bound to the center by employment or other kinds of commerce. These outlying zones are sometimes known as a commuter belt, and may extend well beyond the urban zone, to other political entities. In practice, the parameters of metropolitan areas, in both official and unofficial usage, are not consistent. Sometimes they are little different from an urban area, and in other cases they cover broad regions that have little relation to a single urban settlement; comparative statistics for metropolitan area should

[2]Pareto distribution is another name given to the power-law distribution

take this into account. Population figures given for one metro area can vary by millions. There has been no significant change in the basic concept of metropolitan areas since its adoption in 1950, although significant changes in geographic distributions have occurred since then, and more are expected. Because of the fluidity of the term ″metropolitan statistical area,″ the term used colloquially is more often ″metro service area,″ ″metro area,″ or ″MSA″ taken to include not only a city, but also surrounding suburban, extra urban and sometimes rural areas, all which it is presumed to influence. Moreover, always because of the fluidity of the term, each country has its own unique definition. For example, in the United States (the most used country in the previous literature), the Office of Management and Budget defines Core Based Statistical Areas (CBSA) used for statistics purposes among federal agencies. Each CBSA is based on a core urban area and is composed of the counties which comprise that core as well as any surrounding counties that are tightly socially or economically integrated with it. These areas are designated as either metropolitan or micropolitan statistical areas, based on population size; a ″metro″ area has an urban core of at least 50,000 residents, while a ″micro″ area has fewer than 50,000 but at least 10,000.

To perform this work, instead, we have data about all the Italian cities (″comuni″ in Italian, more or less the equivalent of municipalities) recorded in the last three censuses (1991, 2001 and 2011). We choose this kind of data because, in despite of the fact that all the previous literature uses data about the upper tail of the Metropolitan Areas distribution, we believe whether one wants to explain how people distributes in cities, he has to rely on the entire distribution and to rely on proper city data. This is due to the fact that, as we said before, it does not exist a unique definition of Metropolitan Areas (each country has its own definition). Furthermore, all the existing definitions are given for statistical purposes: in some way, this fact can affect the distribution derived from this kind of data. Moreover, Metropolitan Areas data do not cover the whole population. On the opposite, proper city definition is based only on administrative criteria and they take into account the whole distribution of the population, leading to a natural study on the people distribution.

Let us have a look to the data. All the three data-sets provide us with the population for each municipality: for these data, in tab. (2.1) we report some basic descriptive statistics. As one can simply see, whereas the minimum size, the maximum size and the standard deviation show an alternating behavior from 1991 to 2011, the average population per municipality is constantly increasing from 1991 to 2011. Obviously, these

Table 2.1: Some descriptive statistics

	1991 Census	2001 Census	2011 Census
Minimum	31	33	30
Maximum	2733908	2546804	2663666
Average	7008.68	7035.643	7361.663
St. Dev.	42087.29	39326.61	40262.28
Median	2315	2345	2443
Skewness	42.85704	42.22574	43.72075
Kurtosis	2415.146	2374.504	2546.574

consideration should be read carefully, since we are dealing only with three historical observations.

Two interesting descriptive statistics are those that are presented in the two last rows of tab. (2.1), namely skewness and kurtosis.
In probability theory and statistics, skewness is a measure of the asymmetry of a distribution. The skewness value can be positive or negative. Qualitatively, a negative skew indicates that the tail on the left side of the probability density function is longer than the right side and the bulk of the values lie to the right of the mean. A positive skew indicates that the tail on the right side is longer than the left side and the bulk of the values lie to the left of the mean. A zero value indicates that the values are relatively evenly distributed on both sides of the mean, typically implying a symmetric distribution. According to this definition, the city-size distribution for Italian cities is strongly positively skewed (with a skewness value around 43 in all the three observations): this means that the most part of the Italian cities are smaller than the average. This observation can be confirmed by looking at the percentile distribution[3] which tells us that more than the 75% of the whole population lies at the left of the average (the 75% being around cities with 5000/6000 inhabitants in all the three observations). A further confirmation to this observation can be given by looking at the comparison between the average and the median of the distribution. As one can simply see, median and average differ a lot for all the three distribution. On average, the median has a value around 2000 while the average has a value around 7000. In a symmetric (not skewed) distribution, this two values should, more or less, coincide. As for the kurtosis, this is another descriptor of the shape os a probability distribution. In particular, this measure indicates the fatness of the tails

[3]we decided to not report this distribution, for the sake of conciseness

of the distribution. Usually, a way to interpret this statistics is to compare the value for the required distribution with the standard value for a Gaussian distribution that is equal to 3. If a distribution has a kurtosis value greater than 3, this distribution is called leptokurtic distribution: this indicates that distribution's tails are fatter than Gaussian's ones. If a distribution has a kurtosis value smaller than 3, this distribution is called platykurtic: this indicates that distribution's tails are thinner than the Gaussian's ones. According to these definitions, the city-size distribution for Italian cities is strongly leptokurtic, having a value around 2400 that is much bigger than 3. This means that there is a much higher probability to find extreme cities (very small or very big cities) than it would be predicted if the cities distribution was a Gaussian one.

Beyond this kind of data, our data-sets come providing us with some further data. For example, 1991 and 2001 censuses data provide the difference in population for each municipality (either in value or in percentage rate) and the population density per square kilometer, whereas the 2011 census data provides some more information: for each municipality, it provides data on men and women populations, number of family, residing population in family, average number of family components and residing population in cohabitation. However, we are interested only in understanding the spatial distribution of population, so we will use only some of those additional data in order to control some of our results.

2.4 Empirical results

The first empirical result we want to show is the rank-size relation for the entire distribution of Italian cities. Let us recall briefly the way in which we can build up this relation. We sort the distribution by number of inhabitants in descending order. In this way, the largest city (Rome, in our case) has rank 1, the second largest has rank 2 (Milan, in our case) and so on and so forth. So, we have two columns: one for cities population and one for ranks. We take the base-n logarithm (natural logarithm) of the two column. Then, we plot on the x-axis the base-n logarithm of rank and on the y-axis the base-n logarithm of the population. In fig. (2.3), we report the obtained result for all the three census for which we have data.

There are two main features one can notice from this plot. First, over the last twenty years (from 1991 to 2011) the city distribution showed absolutely no change at all in her shape, neither in the upper tail (the range

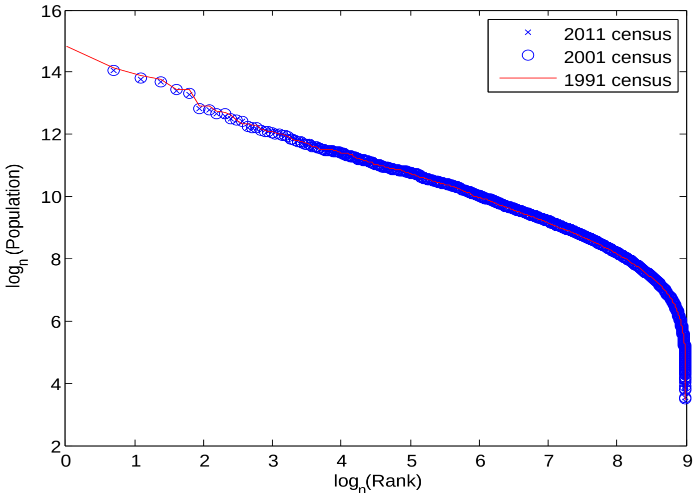


Figure 2.3: Rank-size relation

in which we have the largest cities) or in the lower tail (the range in which we have the smallest cities): the three distributions overlap in a perfect way. Obviously, this graph does not tell us anything whether there has been some changes in the rank position of the cities even though, at least for the upper tail, we can check that no changes have been looking at the data: Rome has been the largest city in all the twenty years we are looking at, Milan has always been the second and so on an so forth.
Another simple observation we can draw from this graph is that Zipf's Law (or, more properly, a linear relation between rank and size) does not hold for the lower tail of the distribution. Indeed, at a certain point, the linear relation between rank and size starts to deviate in the lower tail since it becomes almost a vertical line. This result is essentially due to the very high concentration of very small cities: as we can see from tab. (2.1), the 50 percent of the distribution (then, about 4000 cities out of 8000, composing our population) has a size lower than 2000 inhabitants and the lowest has about 30 inhabitants. So, even though in logarithmic term there is a huge change in city size, the rank does not change so much (always speaking in logarithmic terms).

However, in order to verify whether at least largest cities satisfy Zipf's Law, we have to check the linear coefficient in the regression eq. (2.4). We have to say that, with respect to eq. (2.4), we perform the inverse regres-

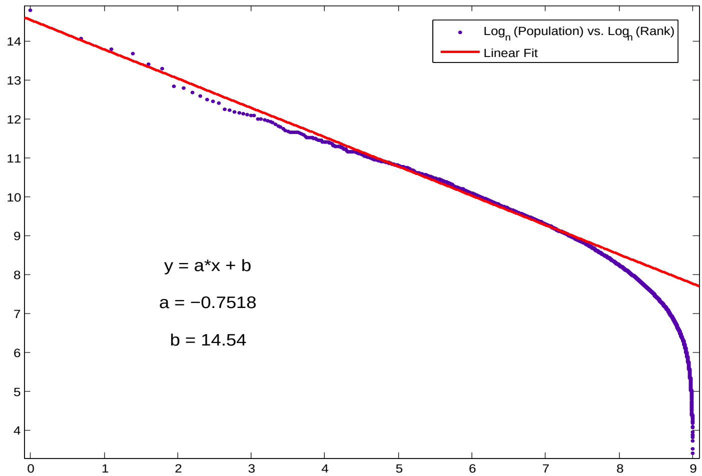


Figure 2.4: Fitted rank-size relation

sion but this would absolutely not affect our results whether Zipf's Law is satisfied. Indeed, according to Zipf's Law, both the regressions should show a coefficient $\zeta = 1$. Since this coefficient would be equal to 1 in both the regressions, there is no particular reason to choose one choice of the axis with respect to the other one. If, instead, the Zipf's Law will be not satisfied by our data, the result will be different in both the possible choices so, again, there is no reason to prefer one choice to the other one. Obviously, since it is glaring that the whole distribution cannot be fitted by a linear fit, we have to exclude some points from the regression, in order to have approximately a line on which we are going to perform our regression.
One of our results is shown in fig. (2.4), from which we can see that the coefficient is quite far from 1, being about 0.75 with a confidence interval ranging from 0.7495 to 0.7544. However, we found that this coefficient is very sensitive to the truncation point, ranging from 0.68 (CI [0.6774,0.6897]) to 0.78 (CI [0.7788,0.7844]). This observation leads us directly to another consideration driven to our mind by the work by Eeckhout (2004[16]): the linear coefficient in the regression changes depending on the truncation point of the upper tail of the distribution, whether the underlying distribution is not a power-law, as clearly it is in our case.

However, as we said before, we know that even though Zipf's Law holds perfectly (then with an exponent equal to 1), the rank-size relation could hold only in an approximate way. So, before claiming that, for Italian city size distribution, Zipf's Law does not hold, we have to check it on the counter cumulative distribution function, namely the probability that the size of a city is greater than a specified size, for all the possible sizes ($P(S > \overline{S})$). In order to perform this regression, we compute the cumulative distribution function of our distribution and then we work on the counter cumulative by means of the transformation

$$CCDF = 1 - CDF \tag{2.5}$$

where, obviously, CCDF stands for Counter Cumulative Distribution Function and CDF stands for Cumulative Distribution Function.

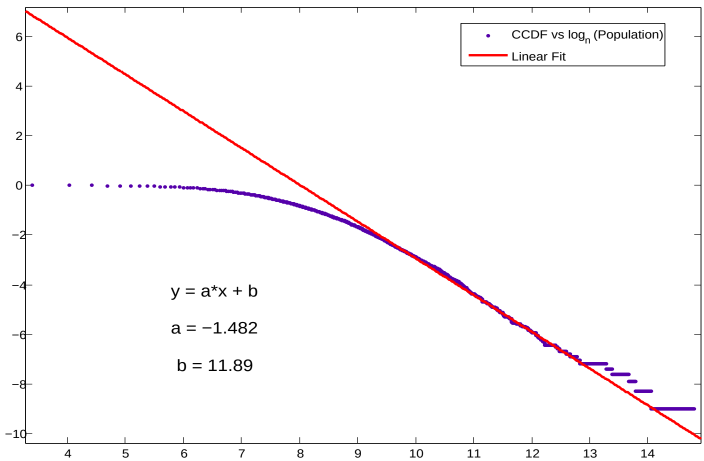


Figure 2.5: Fitting of CCDF

The result of our fitting is reported in fig. (2.5). As you can simply see, the linear coefficient (1.48 with CI [1.479,1.483] is very different from the previous fit (fig. (2.4)). This is because, in this case, we have a natural choice regarding the two axis: on the x-axis we put the base-n logarithm of the population and on the y-axis we put the value of the counter cumulative distribution. In a rank-size relation this corresponds to the inverse of the choice we adopted before but we think in that case the most natural choice is to put the population on the y-axis and the rank on the x-axis, as we do. To compare the two results, it is enough to invert one of the two results and compare the two. If we perform this computation, we see that the two results are quite different (1.33 against 1.48), however, we have to remember that these coefficients are very sensitive to the choice about excluded points and, furthermore, that the two relations are not properly the same. However, what we are interested in is the fact that the two coefficients are both quite far from 1, meaning that we can exclude that Zipf's Law could hold for the Italian cities distribution.

So far, we were only able to exclude that Italian cities can follow a power-law distribution. So, according to all the literature, we have to check whether the distribution could be a log-normal one or a Double-Pareto log-normal one.
Let us test whether our empirical data could be best fitted by a log-normal distribution. First, we present the histogram of the base-n logarithm of our distribution.

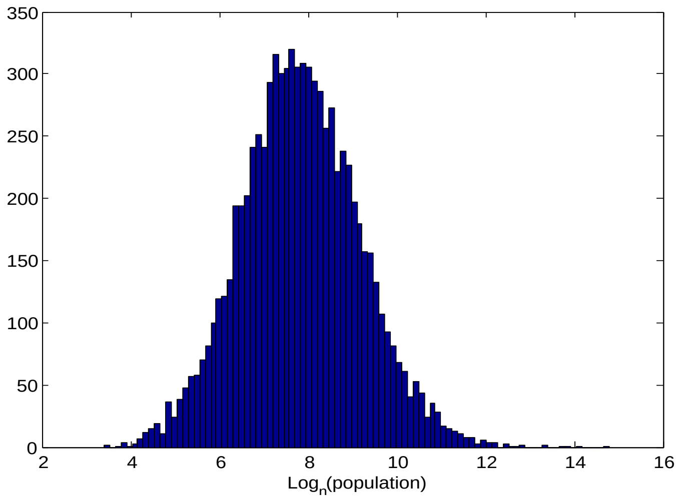


Figure 2.6: Histogram of the base-n logarithm of the distribution

Doing this, whether the distribution is a log-normal one, we would observe a Gaussian distribution. Our result is sketched in fig. (2.6). By simply looking at the graph, we can see there is some similarity between our histogram and a Gaussian distribution but we have to go deeper in checking. The first thing we can do is to estimate mean and standard deviation for a normal distribution from the logarithm of our empirical data and then generate a theoretical normal distribution with the same parameters. Then, we can easily plot the rank-size relation for both the distributions.

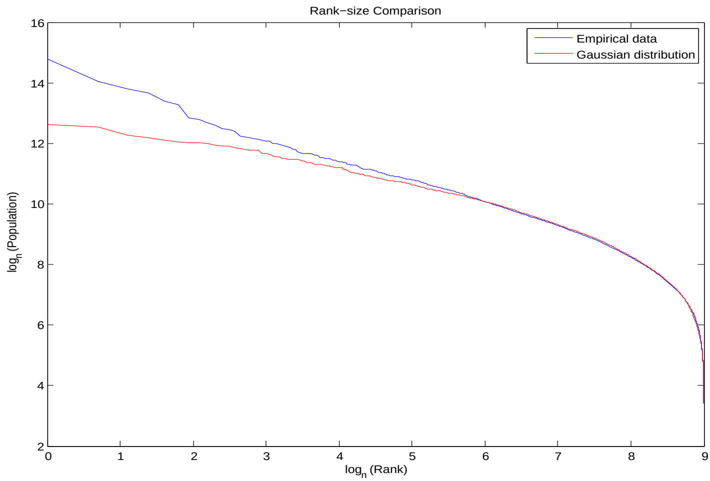


Figure 2.7: Rank-size comparison.

The results for this kind of test are shown in fig. (2.7) where we show a comparison between rank-size relation for the empirical and the theoretical Gaussian distribution. As we can see, there is high overlapping in the body and in the left tail of the two distributions, but this overlapping is totally lost in the right tail of the distribution: this would mean that the log-normal distribution fails in describing the behavior of the biggest cities. However, before pronouncing such a definitive sentence, let us perform some other comparisons. What we want to do now is to perform the fit of a log-normal distribution with some simple function. To do this, we have to recall that, in a log-log plot, a log-normal distribution should look like a parabola. So, we can plot our empirical data on a log-log scale and the try to fit them with a quadratic fit. The results are reported in fig. 2.8.

This comparison, like the previous one, tells us that the distribution of our empirical data is very close to a log-normal distribution in the body but this tends to not to be true in both the tails: biggest and smallest cities in our data are not well described from what a log-normal distribution predicts.
Lastly, we perform a quantile-quantile plot of the natural logarithm of

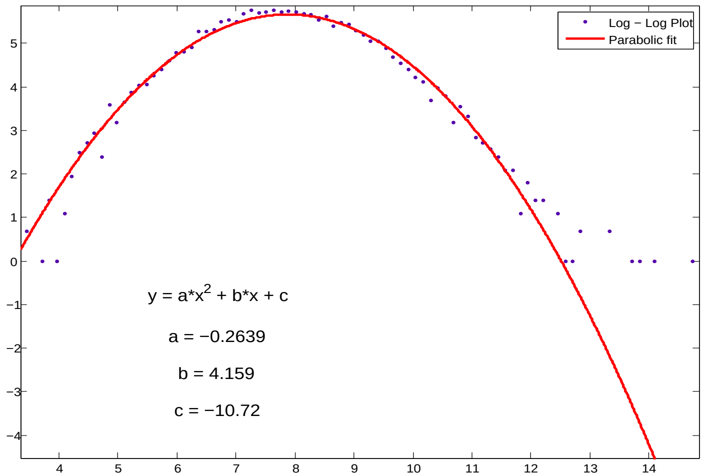


Figure 2.8: Parabolic fit.

our empirical data against a normal distribution. A Q–Q plot (where "Q" stands for quantile) is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. First, the set of intervals for the quantiles is chosen. A point (x,y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate). Thus the line is a parametric curve with the parameter which is the (number of the) interval for the quantile. If the two distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line $y = x$. If the distributions are linearly related, the points in the Q–Q plot will approximately lie on a line, but not necessarily on the line $y = x$.

The results are reported in fig. 2.9. If the data followed a log-normal distribution, this plot would result in a straight line ($y = x$, as we said before) and it would completely overlap the red dotted line. Once again, there is a perfect overlapping between the two distributions in the body but there is not in both the tails. It is worth notice that the left tail seems more little than the right one, that is, there is a smaller number of smallest cities showing a Pareto behavior with respect to those biggest cities showing the same behavior: we will find this result again in the following.

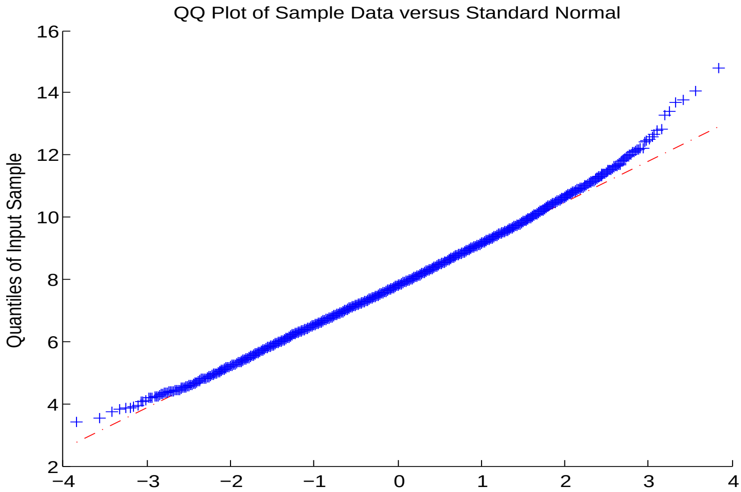


Figure 2.9: Quantile-quantile plot

All of these results lead us to the hypothesis that the log-normal distribution could not be the distribution we were looking for, even though, at least for the main body of the distribution, this could be a very good approximation.
Next step to perform is to move towards a more sophisticated distribution that could give us a better fit of our empirical data and that can be derived by a realistic hypothesis on random growth process. In order to find this distribution, it comes in our help the work by Giesen et al. (2010[4]) in which they suggest a Double Pareto log-normal distribution as a possible solution to our puzzle.
So far, we showed in several ways that the body of our distribution follows a log-normal distribution. Then, we have to show that both the tails show a Pareto behavior so we can argue that the Double Pareto log-Normal distribution could be a very good approximation for our empirical data. To perform this check, we will use the test developed by Clauset et al. in (2009[29]). The results, for the two tails, are shown in fig. 2.10: on the left, we have the right tail of the distribution (biggest cities) while on the right we have the left tail of the distribution (smallest cities).

As we can simply see, both the tails are well described by means of a Pareto distribution. The right tail shows a Pareto behavior in the first 1200 cities, with a cutoff point (the point where the distribution loses its Pareto behavior) set around a population of 10269 inhabitants and an ex-

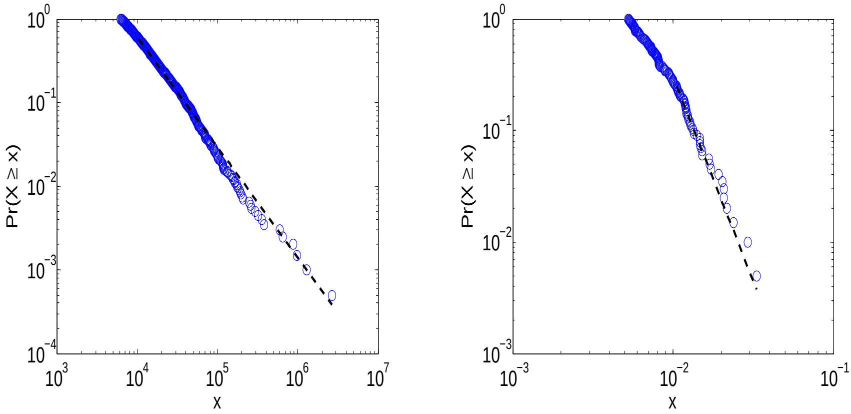


Figure 2.10: Pareto test for tails.

ponent equal to 1.32 (a value that is in the range we established with our fits). The left tail, instead, shows a Pareto behavior for the last 50 cities with a cutoff point around 100 inhabitants and an exponent equal to 3.65. These results confirmed something we already said in the previous: the left tail is much smaller than the right one but the left tail is much steeper than the right one (remember, the left tail seemed to be almost a vertical line). In terms of population, this represents mathematically what we noticed before: in the left tail, there is a high concentration of cities and then, even though there is huge change in city size, there is a small change in city rank.

2.5 The model

So far, we showed that the Double Pareto log-Normal seems to be the most suitable distribution to describe the distribution of all Italian cities. Now, what we want to do is to simulate the process on which the Double Pareto log-Normal distribution grounds, in order to find what are those historical average parameters in terms of cities growth rate mean and standard deviation that could describe the actual Italian cities distribution. The question we would like to answer is: "If we try to average out all the exogenous shocks in Italian history, what would be the average and the standard deviation of the cities growth rate giving rise to the actual Italian cities distribution?".

The model that we are going to simulate has been firstly derived by Reed (2002[6]) and it is based on very simple probabilistic assumptions about the formation and the growth of cities, which reflect the intrinsic variability in these two processes. This model is essentially mathematic since it is a consequence of a stochastic process which has certain characteristics. However, this does not mean that geographic and economic factors are not important in determining the growth and eventual size of any city. Rather it means that when we look at the distribution of cities, the effects of the variation in these factors can be modeled effectively by stochastic processes.
This stochastic model is composed by two main components, one for the foundation of new cities and the other for their evolution after foundation. Cities grow in different and varying ways. The proportional rate of growth size in a given year will vary from city to city and for a given city will likely vary from year to year, depending on economic, demographic factors, etc. This variability can be modeled mathematically by assuming that the logarithm of population size for any city constitutes a realization of a random walk (Gibrat (1931[3])). So, the size $X(t)$ of a city will be assumed to follow Geometric Brownian Motion (GBM) governed by the Itô stochastic differential equation

$$dX = \mu X dt + \sigma X d\omega \tag{2.6}$$

where $d\omega$ is white noise (the random increment of a Wiener process in time dt). The parameter μ is the mean proportional growth rate over all cities and all times (the so-called drift), and σ is a parameter reflecting the variability in this growth rate. If the initial size of the city is X_0, then, under the GBM model, the size X_T of the city T time units after will be a log-normal random variable.
It is possible that the distribution of starting sizes has changed over time (for example, agricultural cities likely were initially smaller than industrial ones). One can easily accommodate this by assuming that X_0 also evolves as a GBM.
Another important characteristics of the model is the way in which we can model the foundation of new cities. Of course, this can depend on many factors but we assume that all these factors can be modeled by a simple stochastic process.
The simplest stochastic model that one could assume is that foundations

occurred in a Poisson process, i.e. they occurred randomly and independently at a constant average time. However, an even more realistic model results from assuming that in the time interval $(t, t+dt)$, any existing cities can form a new satellite settlement with probability λdt. This is a so-called Yule process first proposed by Yule[7] as a model for the creation of new biological species, a process that is similar in many respects to the foundation of new cities. For this model one can show that the distribution of the time T since foundation of a city currently in existence, is of the form of an exponential distribution truncated at τ (the age of the first city), with an atom of probability (reflecting the probability that the given city is the oldest) of size $\frac{\lambda\tau e^{-\lambda\tau}}{1-e^{-\lambda\tau}}$ at the point τ. In most cases it is probably reasonable to assume that τ is large, thus to consider the limiting distribution as $\tau \to \infty$. This is exponential distribution with density $\lambda e^{-\lambda t}$ for $t > 0$.
Under the Yule process for the foundation of cities and the GBM model for their subsequent growth, the distribution of the current size can be computed yielding a probability density of the form

$$f(x) = \frac{\alpha\beta}{\alpha+\beta}\left[x^{-\alpha-1}\exp\left\{\alpha\mu_0 + \frac{\alpha^2\sigma_0^2}{2}\right\}\Phi\left(\frac{\ln(x)-\mu_0-\alpha\sigma_0^2}{\sigma_0}\right) + x^{\beta-1}\exp\left\{-\beta\mu_0+\frac{\beta^2\sigma_0^2}{2}\right\}\Phi^c\left(\frac{\ln(x)-\mu_0-\beta\sigma_0^2}{\sigma_0}\right)\right] \tag{2.7}$$

on $x > 0$ where Φ is the cumulative distribution function of the standard normal distribution, Φ^c is the counter cumulative distribution function, and α and β are the roots of a characteristic quadratic equation. This distribution is called Double Pareto log-normal (DPLN).

In order to see whether we can obtain what are the historical average parameters giving rise to the actual city distribution, we simulate this Yule process in Matlab. However, to simulate the evolution, we had only parameters coming from the nowaday distributions (either of population and growth rate of population) but these parameters for sure do not reflect the situation as it was during all Italian history. To provide for this lack of information, we multiply all the parameters we needed in our simulation for some adjustment coefficients in order to take into account the fact that distributions could have been very different during the evolution.

We start from a sample of n cities, normally distributed with mean given by the actual mean of the distribution divided by 1.5 (it corresponds to an average population of 200 inhabitants, as it is known that new born cities were smaller in the past) and standard deviation given by the actual distribution. At each time step (we simulate 2800 time steps, as the first settlements we can refer to as city are from the 8th century BC), each city receives a shock in its population[4]: these shocks are normally distributed with mean a times the actual mean and standard deviation b times that of the actual growth rate distribution. Our goal is to find the parameters a and b that give rise to the distribution best approximating our empirical data and, for this reason, we performed cycles over a wide range of values, in order to find the best possible parameters. At each time step, with a fixed probability λ[5], several new cities are born with a size normally distributed with the same mean and standard deviation as the initial ones. To avoid the possibility of too small cities or even a city with negative population, we set a lower barrier at the process: if, during the evolution, a city became smaller than the nowaday smallest city, we erased it from our simulations. To make all the results less sensitive to the process of random number generation, for each couple of parameter, we simulate 1000 evolutions and then we take the average.
The best result we obtain is shown in fig. (2.11), where we report the rank-size relation obtained either for our empirical data and for the distribution arising from our simulations.

As we can simply see, there is a very good approximation for the body of the two distribution, whereas there is no agreement in the two tails: in particular, the theoretical size of the biggest and of the smallest cities seems to be smaller than the empirical one.
To try to overcome these problems and, hence, to improve our model and the adherence to the empirical data, we applied a series of modification to our initial hypothesis:

- **Sine wave varying probability of creation of new cities**. Creation of new cities is a process that is more likely and more favourite when the state of economy is glowing rather than when the eco-

[4] To avoid lack of generality, we performed several tests varying each of these parameters (number of initial cities, average size of initial cities, number of time steps): none of these parameters can affect the shape of the final distribution. Each of these parameters affect only the final parameter a and b that approximate the empirical distribution in the best way.

[5] Even this parameter underwent several changes to avoid lack of generality: no effect was seen on the shape of distribution, there was an effect only on the values of the best fitting parameters.

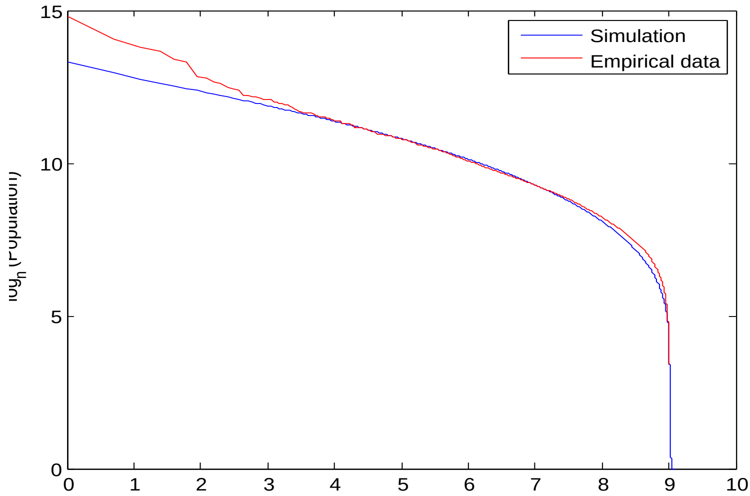


Figure 2.11: Comparison between empirical data and simulation

nomy is not. With this modification, we try to capture the possible influence economic cycles can have on the probability of creation of new cities.

- **Increasing average size of new cities**. It is well known that, as time goes by, average size of new cities becomes bigger (for example, industrial cities are bigger than agricultural ones). So, we set the average size for the new cities directly proportional to our variable representing years gone by.

- **Decreasing probability of creation of new cities**. As years pass by, creation of new cities is always less likely, as the increasing concentration, due to the increasing number and size of cities, does not allow for the creation of new cities (for example, we saw that there is no creation of new cities in Italy between 2001 and 2011 censuses).

We tried to include all of these new features in our simulations but none of these made our results improve. The only way these modifications affected our results is in a change in the value of the best parameters but the shape of the distribution, and, hence, the drawbacks of our theoretical distribution, were totally unaffected.

2.6 A change in the paradigm

Once we attested we were not able to overcome these bad results by modifying our simulations, we started to question whether the hypothesis of the model were correct.
In particular, the main hypothesis of our model (and of all the literature dealing with city size distribution) is the law of proportionate growth, also known as Gibrat Law, stating the independence of city growth rate on city size. First, let us give a look to the city growth rate as a function of the logarithm of the city size at time $t-1$.

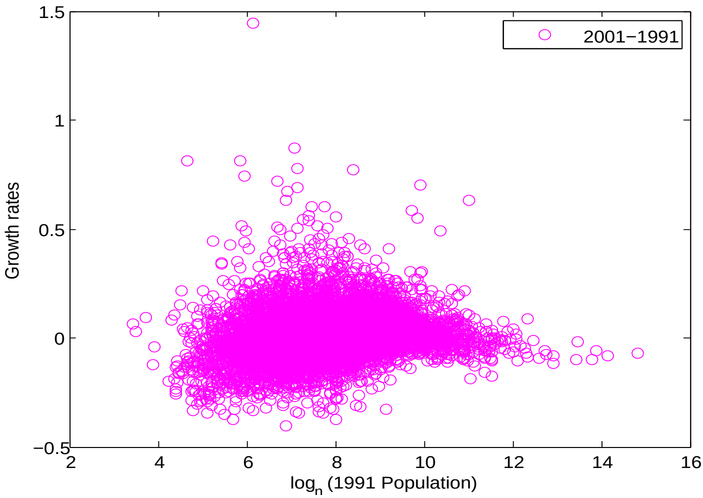


Figure 2.12: Growth-size relation

Results are sketched in fig. (2.12). In this graph, growth rate are computed as percentage returns, hence:

$$g = \frac{S(2001) - S(1991)}{S(1991)} \tag{2.8}$$

where S(1991) and S(2001) are, respectively, the city size as resulting from the 1991 census and the 2001 census. The graph has the same shape, even though we consider the next period (2001-2011) or whether we move to another growth rates computation method (for example logarithmic returns, $g = \ln\left(\frac{S(t+1)}{S(t)}\right)$). For sake of conciseness, we do not

RHS var.	1	2
Const.	-0.143 ***	-0.510***
	(0.008)	(0.030)
Log(Pop$_{t-1}$)	0.022***	0.117***
	(0.001)	(0.008)
Log^2(Pop$_{t-1}$)		-0.006***
		(.000)
# obs.	8085	8085
R^2	0.0622	0.0804

Table 2.2: Results of the regressions

report all the plots we have done, showing, more or less, the same behavior.
To state that Gibrat's law holds for Italian cities, we have to test the equation

$$g = \beta_0 + \beta_1 \cdot log(S_{t-1}) \tag{2.9}$$

where with S_{t-1} we denote the city size at $t-1$, i.e. the city size in the year from which we start computing the growth rate g. If Gibrat Law holds, we should obtain that the regression coefficient β_1 is not statistically significant. Furthermore, as from the graph it seems to emerge a parabolic pattern (despite the noise due to the high concentration of points), we are going to test even the equation

$$g = \beta_0 + \beta_1 \cdot log(S_{t-1}) + \beta_2 \cdot log^2(S_{t-1}) \tag{2.10}$$

Results of these regressions are sketched in tab. (2.2): in this table, the dependent variable is city growth rate in percentage returns, standard errors are reported in brackets and *** means 1% statistical significance (in column 1 we report results for regression based on eq. (2.9) whereas in column 2, we report results for regression based on eq. (2.10). As you can simply see, either the first order and the second order coefficient are highly statistically significant, leading us to the conclusion that Gibrat law does not hold for Italian cities. However, one may assert that our result is driven by the high noise in points distribution. So, in order to reduce the noise, we perform the same graph on average, i.e. we take the average population and the average growth rate of a certain range of cities. In this way, we can erase much of the noise due to the high number of observation. In order to check the previous results, we divide the whole range of our cities in N bins[6]. For each bin, we take the average population and growth rate of all the cities falling in the bin, i.e. for each bin, we compute the average population of the cities falling in the bin

$$\overline{S} = \frac{1}{N_b} \sum_{i=1}^{N_b} S_i \tag{2.11}$$

and the average growth rate of the cities falling in the bin

$$\overline{g} = \frac{1}{N_b} \sum_{i=1}^{N_b} g_i \tag{2.12}$$

where N_b is the number of cities falling in each bin.

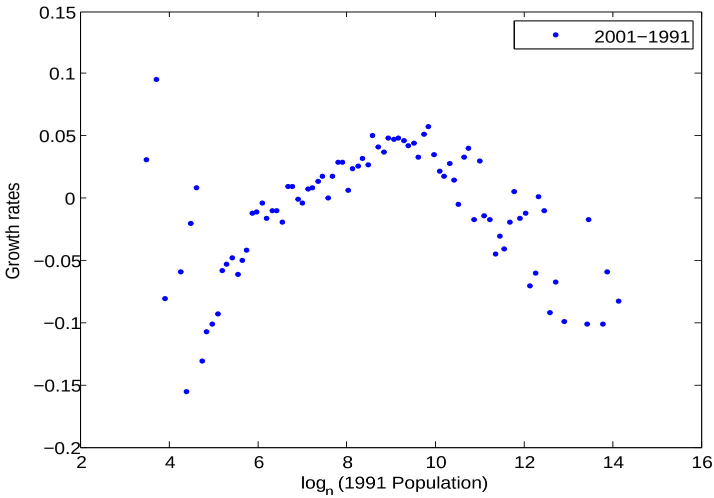


Figure 2.13: Growth size relation for averages

The results we obtained are shown in fig. (2.13): each point represents the average population and the average growth rate of the cities

[6] As for the other parameters, we perform several tests with different values of N but the results were not affected: in all the graphs, we report the result for a number of bin equal more or less to the square root of the number of observations.

falling in a bin. Now, there is no need of a regression to see that there is a parabolic behavior in the distribution of city growth rate with respect to city size. This could imply that, in the distribution, there is a sort of "ideal" number of inhabitants for the city size that is around 10^4: city whose size is around this ideal value can attract people while cities with a number of inhabitants above or under that threshold make people run away from them. This result is confirmed even if we perform the same computation not over the whole Italian cities but on a single region (in Italian, 'Regioni'): we found the same behavior (not reported) for almost all the regions.
One could think that this result could be due to the particular way in which we built our bins. For this reason, we built our bins in another way. Instead of having bins with the same length, we have bins with the same number of cities in each bin and then we perform all the computations: we found exactly the same results, either on a national level and on a regional level.

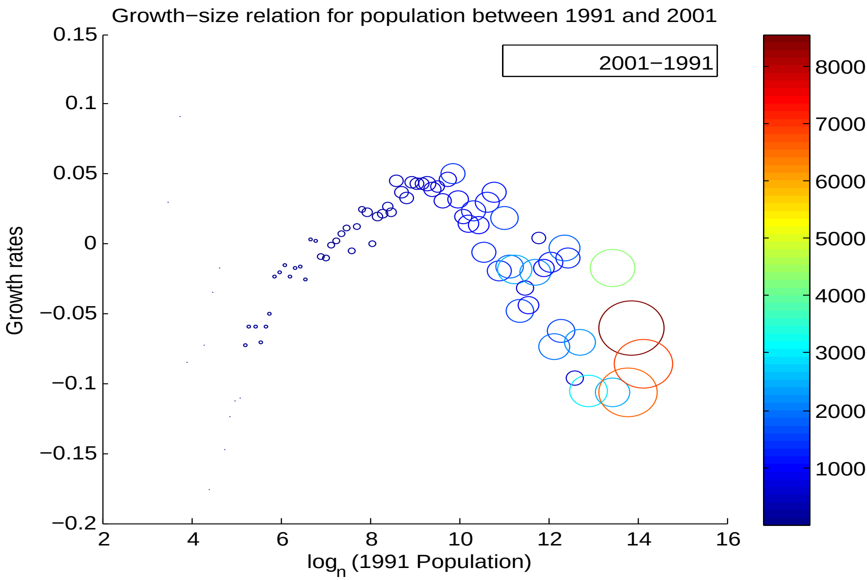


Figure 2.14: Growth size relation for averages

The last check we made on this result is whether this seeming escape from small and big cities could depend (at least for big cities) on inhabitants' density (number of inhabitants per square kilometer). Our results are sketched in fig. (2.14), where we plotted the same result as fig. (2.13) but now the circles' size is proportional to the average population density of the cities falling in the bin. As we can simply see, there is not a huge change in city density, unless we move to very big cities (Rome, Milan etc.), as the most of the cities ranges from 1000 to 3000 thousands persons for square kilometer. So, this parabolic behavior does not seem to be density driven.

We did not try to adapt this new result to our model in simulations because our hypothesis is that this seeming escape from small and big cities is not a structural feature of the growth rate distribution but only something that emerged in the last years as a result of the fact that people are always more careful about life quality and medium size cities offer the best compromise in terms of life quality: this kind of city is not so much crowded (as biggest cities are) but, at the same time, medium size cities offer a good range of services and leisure activities (differently from smallest cities, that, usually, offer poor services). To prove our hypothesis, we obtained census data from 1951 and 1961 censuses and we perform the same exercise as before.

For Italy, that was the period of the so-called Economic miracle: in this period, the country was transformed from a poor, mainly rural nation into a major industrial power. Moreover, it was also a period of momentous change in society and culture. The results of this new exercise are shown in fig. (2.15): as before, each point represents the average population and the average growth rate of the cities falling in a bin. As we can simply see, although a simple graph of growth size relation would result in very similar graph to the previous one (not reported for the sake of conciseness), when we go to average out growth rates and population, we obtain a result that is very different from what we obtained before. In that period, biggest cities grew at a rate that is, on average, very much bigger than the one of small and medium: cities growth rate grew almost linearly with city size.

These results, on the one hand, suggest a possible justification to the scarcity of the results of our simulation, as we applied a proportionate growth hypothesis to the entire evolution process, and, doing so, we did not take into account any different possible behavior. On the other hand, more importantly, they could (and they should, in our opinion) suggest a discontinuous change in the way to approach the search for the right city

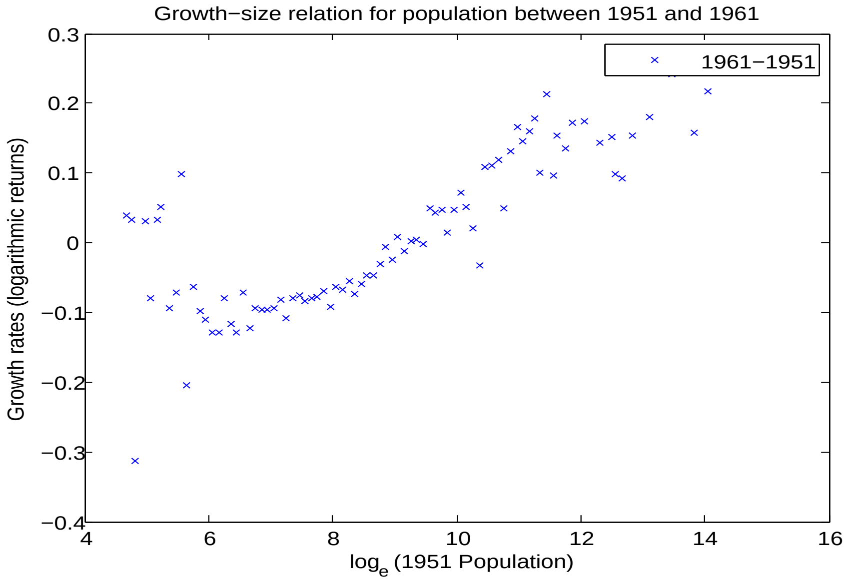


Figure 2.15: Growth size relation for averages

size distribution. When we go to average out city growth rate and size, Gibrat law does not seem to hold anymore and, moreover, the relation coming out does not seem to be constant over time. Thus, on average over the whole history, growth size relation could be very different from what Gibrat law told us up to now.

2.7 Conclusions

The way in which population is distributed among cities and/or Metropolitan Areas is still an open question, in which either scholars and policy makers are interested. Giving a globally accepted answer to this question can be of fundamental importance to better understand even firms localization decisions or to implement policies regarding transport at all the levels. Unfortunately, only two regularities are widely accepted by the international scientific community: Zipf's Law for the upper tail of the distribution and Gibrat law for the growth rates of cities. Moreover, it seems these two empirical regularities cannot be unified by a single theory.

In this work, we studied the distribution of all the Italian cities, exploiting data from the last three censuses. We obtained several important results. First, we proved that the upper tail of the distribution follows a power-law distribution, even though the linear coefficient in a log-log plot is quite far away from what it is expected to be. According to Zipf's Law, the linear coefficient should be equal to one, while from our fits we found at most a coefficient around 0.78. This lead us to claim that, according to all the previous literature, there were two possible curves fitting our data: log-normal and Double Pareto log-normal distribution. From our fits and our tests, we found that a log-normal distribution fits very well the body of our distribution but the agreement gets a bit lost either in the lower and in the upper tail. Thus, we skipped to the only distribution that could be according to the literature on the topic: the Double Pareto log-Normal distribution. As we obtained several confirmation about the fact that the body of our distribution is a log-Normal one, we checked the tails in order to see if they are Pareto. The tests gave a positive result, so that we can affirm that, among the proposed distributions in the literature, the Double Pareto log-Normal seems to be the most suitable one to describe the 2011 distribution of Italian cities.
At this point, we moved away from detecting what is the actual city size distribution for Italian cities and we tried to simulate the Yule process leading to a steady state represented by the Double Pareto log-normal distribution in order to find out what are the parameters, in terms of average and standard deviation of growth rate distribution, that best represent the historical evolution of Italian cities. The question we would like to give an answer to was: "If we try to average out all the exogenous shocks in Italian history, what would be the average and the standard deviation of the cities growth rate giving rise to the actual Italian city size distribution?" These simulations gave us an incredibly high overlapping with the distribution of our empirical data, for what the body of the distribution is concerned, but very scarce results for the tails; in particular, in both the tails, theoretical predicted cities were smaller than empirical ones. We tried our best to improve the model and, hence, the agreement between theoretical and empirical distribution but we did not succeed in obtaining it. This lead us to question the goodness of model hypothesis, and, in particular, the goodness of proportionate growth hypothesis (Gibrat law) to represent the average historical growth size relation. We found out that, averaging over N bins either cities growth rate and city size, we do not obtain a straight horizontal line (as it is supposed to be if Gibrat law held) but a functional form depending on the particular historical period: for example, we obtained a parabola in the last 20 years and an almost straight line (not horizontal) if we go back to the '50s.

Is it the Gibrat law the correct way to model the growth size relation? Is the evolution of cities a proportional growth process? In our opinion, we should start to go behind Gibrat law, at least for Italian cities, and to propose new ways of relating cities growth and size

3 Ageing and Labor Market: Testing Gibrat's Law for Germany

3.1 Introduction

Gibrat's law and Zipf's law are two very well known empirical regularities, currently very debated among scholars and policy makers. Very briefly, Gibrat's law affirms that the growth rate of an entity (i.e. firm, city) is independent of its size, meaning that no systematic behavior exists between its growth rate and its size. In the field of city size distribution, this implies that, although cities can grow at different rates, it is not possible to affirm that larger cities grow faster than smaller ones or vice versa.

Zipf's law, instead, states that the city size distribution could be closely approximated by a power-law distribution, at least in its upper tail. An approximate way of stating Zipf's Law is the so-called rank-size rule that states the size of a city in a country is proportional to its rank. This means that, for example, the largest city of a country is twice the second larger city, three times the third one and so on. In fact, Zipf's law (by means of this rank-size rule) is able to measure how unequal the city distribution is. According to Gabaix (1999[5]), Zipf's law is directly linked to Gibrat's law because it is an outcome of Gibrat's law, at least in the upper tail.

Many studies have focused on the analysis of these two regularities, driven by the idea that a truthful description of the actual distribution of people along the space might be very important for policy relevant issues and also for defining more precise theoretical models. These issues can range from a better understanding of firms and people localization choices to the implementation of national and regional policies, for instance, in terms of incentives and transport infrastructures (Fazio and Modica (2015[30])). Indeed, Gabaix (1999[5]), provides a theoretical model leading to a population growth process that follows the Gibrat's law (i.e a random growth process described by a common mean and variance) driven by migration which, in turn, "forces utility-adjusted wages to equate at the margin" (Berry and Okulicz-Kozaryn, (2011[31]) p. S18). At the same time, Eeckhout (2004[16]), providing empirical evidence about Gibrat's and Zipf's law validity, models a city growth process characterized by two main driving forces: random productivity process of local economies and the perfect mobility of workers. Both these models (but literature provides other similar models, see, for instance, Cordoba, (2008[32])) underline how local labor market characteristics and migration are important factors that concur in the 'materialization' of both Gibrat's and Zipf's law. However, there might be also idiosyncratic reasons why individuals decide to localize in a given city or choose to move between cities that typically are not considered in this literature. For instance, Peri (2001[33]) shows how young educated workers prefer working in larger cities. Hunt (2006[34]) provides evidence of different localization behavior between old and young people, finding that young German people are more mobile and they show more sensitivity to source region wages and relative insensitivity to source unemployment, in comparison with elders. Moreover, the migration of young people is seen as labour-force related in densely populated urban areas (Bures, (1998[35]); Frey and Speare (1988[36]), Longino et al., (1984[37])), while elderly migrants prefer less densely and amenity areas (Bures, (1998[35]); Longino et al., (1984[37]); Scott and Storper, (2003[38])). Furthermore, many countries and especially Europe will become an ageing society in the decades to come. Indeed, *"as Europe continues to age, the historical shape of its age pyramid has moved away from a triangle (associated with an expanding population) and has been reshaped, with a smaller proportion of children and young people and an increased share of elderly persons.* "(Eurostat (2015[39], being young in Europe today).

This shift in the old-age dependency ratio will predictably have massive impacts on many socio-economic phenomena, such as, for example, the housing market, the labour market, the demand for goods and services and so forth. Consequently, ageing has become an important source of serious research and policy concern, among scholars and policy makers. However, it should be noted that the multitude of ageing effects will not show a uniform pattern across and within countries. Most likely, the spatial (urban and regional) variations will be significant, as a result of different local circumstances, region-specific labour market participation, differences in regional in- and outmigration patterns etc.

The above mentioned demographic developments will certainly have deep consequences for the functioning and evolution of local and regional labour markets. They will most likely exhibit varying developmental profiles, in terms of labour force participation, exit rates, productivity impacts and the like; ageing is not a neutral generic phenomenon, but may create significant socio-economic disparities across cities and regions in the same country, especially over a longer time horizon. A country like Germany, for example, – with a rapidly rising ageing profile – will most likely witness an unprecedented dynamics on its regional labour markets.

Aim of the paper is then to provide a truthful description of the distribution of people along the space taking into account the demographic differences between people. To provide this analysis, we will focus both on population (in terms of place of residence) and employment (in terms of place of work) data: we will use annual observations regarding population and employment for all German towns and cities in the period from 2001 to 2011. Then, the questions we want to answer are the following: does the growth rate of employment/population depend on the size of the cohort? What is the level of employment and population concentration/deconcentration? Are there any differences between these two variables? Are there any differences if we differentiate for age cohorts? In answering all of these questions, we will first show the results of Zipf's analysis and we move then to Gibrat analysis. The main idea is then to verify whether the size of a city (either in terms of population and in terms of employment) according to 5-years age classes and its growth rate are independent. To the best of our knowledge, this is the first study attempting to introduce demographic characteristics (and in particular, age structure) into the Zipf's and Gibrat's analysis. The novelty introduced by the present study is then given by: the introduction in the analysis of demographic characteristic and the use of different measures for city size other than population for place of residence (e.g. number of employees for place of work). Furthermore, this study will definitively provide new information for modeling a city growth process more in line with the reality, incorporating idiosyncratic factors of the population and the different localization preferences of young and elders. Moreover, it will provide a better picture of the local and regional labour markets able to support the implementation of national and regional employment and urban policies.

The paper is structured in the following way. In section 3.2, we will provide the readers with a very brief reminder about the theory about Zipf's Law and Gibrat's Law and the literature dealing with these two empirical regularities. In section 3.3, we will describe the data used to perform our analysis. In section 3.4, we will show the methods we used to obtain our results and the results themselves. Then, in section 3.5, we will briefly conclude.

3.2 Gibrat's Law and Zipf's Law: A reminder

Gibrat's Law, also known as proportionate growth process,(Gibrat, (1931[3])) may explain in stochastic terms the systematically skewed pattern of the distributions of cities' size (Santarelli et al., (2004[40])). Indeed, since the first study on Gibrat's law (Gibrat, (1931[3])), it has been

observed that the size distribution of firms (in first instance) and cities are well approximated by a log-normal distribution (even though, at least for cities, there is no global consensus among scholars on which distribution should be, see, for example, Giesen et al. (2010[4])). The main reason why this happens can be always related to the initial arguments provided by Kapteyn (1903[21]), affirming that, if a variable is generated by a stochastic growth process that is proportionate, it gives rise to an asymptotically lognormal distribution. This process states that "*the change in the variate at any step of the process is a random proportion of the previous value of the variate*" (Chesher, (1979[41]), p. 403), then "*the probability of a given proportionate change in size during a specified period is the same for all firms - regardless of their size at the beginning of the period*" (Mansfield, (1962[42]), p. 1031).
Although many efforts have been devoted to the theoretical implication of Gibrat's law and on the mechanisms that can lead to the fulfillment of the law, only recently, economic interpretations of the law have been explored. For instance, according to the city size distribution literature, several authors have suggested a fair number of economic interpretations, nonetheless differing only in some shades. To our opinion, two studies are remarkable for the aim of this paper (reader interested in a more complete description can consult Modica et al., (2015[30])). In a short run view, Black and Henderson (2003[43]) state that a shock affects in the same way big and small entities (or in the other way round, both small and big entities have the same relative growth rate, independently by the initial size of the entity). In a long run view, Brakman et al. (2004[44]), state that a large temporary shock can have a permanent impact. This means that a shock can change the growth path toward another size equilibrium.
According to Gabaix (1999[5]), Gibrat's law explains well the so-called Zipf's law. This law can be formalized as follows:

$$P(S_i > S) = \frac{a}{S^\sigma} \tag{3.1}$$

where a is a positive constant and $\sigma = 1$. That is, the size of a city times the percentage of cities with larger size equals to a constant or, equivalently, the probability that size of city i is greater than S is inversely proportional to S. As we said before, there is also an approximate way of stating the Zipf's Law: the so-called rank-size rule. This is a deterministic rule that follows from the definition: the second largest city is roughly half the size of the largest, the third largest city is roughly a third the size of the largest, ecc. That is, whether we rank cities from largest (rank 1) to smallest (rank n) and denote their population $P_1 \geq P_2 \geq P_3 \geq ... \geq P_n$, respectively, the rank R_i for a city of population P_i is proportional to the proportion of cities greater than *i*. Therefore, we can rewrite the previous equation in the following way:

$$P_i = KR_i^{-q} \tag{3.2}$$

Equation (3.2) is also known as the rank-size rule and is usually expressed in logarithmic form, as follows:

$$\log(P_i) = log(K) - q\log(R_i) \tag{3.3}$$

where P_i is the population of city i, R_i is the rank of the ith-city and K is a constant. Zipf's law is said to hold precisely, when the coefficient q is equal to one[7].

Also for Zipf's law, in the recent years, they have been proposed several economic interpretations (see, among others: Gabaix and Ioannides, (2004[8])). Very roughly speaking, the Zipf's coefficient, q, can be seen as a proxy for the hierarchical degree of a system of cities (e.g , if q is exactly equal to 0, all the cities are of the same size; on the contrary, if q tends to infinite the city system is composed by an enormous city where all the population is condensed). Another interesting interpretation is the one proposed by Nijkamp and Reggiani (2015[45]), considering the urban structure as a network. In this way, they are able to compare the Zipf's law with the connectivity degree distribution, as in Barabasi and Oltvai (2004[46]), meaning that the smaller is the value of the connectivity degree and the higher the number of the connections of the hubs.

In the field of spatial economics, these two regularities have given rise to an increasing number of empirical and theoretical studies. Empirical studies try, in different ways, to test the (non)validity of both these laws, singularly or in combination, the following partial list cover the last works:

[7]This relation can be easily derived in an elementary way. Let us suppose that at each time a person is born in a city, that all cities have the same birth rate and that, with a very small probability, this person creates a new city. Then, the total population n_0 of cities existing at time t_0 is proportional to t_0: $n_0 \sim t_0$. The rank of the city created at time t is proportional to t: $R \sim t_0$. The ratio between the size of the city and the total population remains the same: $K/n = 1/n_0$. This implies that: $K \sim 1/n_0 \sim 1/t_0 \sim 1/R$. In the end, size is inversely proportional to the rank. It is important to bear in mind that, even though Zipf's Law holds perfectly, the rank-size rule would hold only approximately.

for example, see, among the others, Black and Henderson (2003[43]); Glaeser, Ponzetto, and Tobio (2014[47]); Gonzalez-Val (2012[48]); Guerin-Pace (1995[49]); Rosen and Resnick (1980[13]); Soo (2005[15]); and Storper (2010[50]).
In most studies, Zipf's law and Gibrat's law are both generally accepted; among all the studies we can suggest: Giesen and Suedekum (2011[51]), Eeckhout (2004[16]), Gonzalez-Val (2012[48]), Ioannides and Overman (2003[18]), Gabaix and Ioannides (2004[8]), Black and Henderson (2003[43]) and Gonzalez-Val et al. (2013[52]).
Typically, all the above mentioned studies differentiate cities only by their population size, without taking into account any demographic difference (for example, age structure) and without considering any regularities on the employment size. However, it might be useful, both to build more realistic random population growth model, to delineate more precise policy implications in terms of population ageing and to consider appropriate labor polices, to take into consideration the demographic characteristics of the population and of the employment size at municipalities level. In contrast to the relevant literature, then, this paper set out to focus on trends in geographic concentration of employment across municipalities with a particular focus on the demographic characteristics of the population.

3.3 Data

The data set for this study has been provided by the German Institute for Employment Research (IAB, Institut für Arbeitsmarkt -und Berufsforschung). It contains annual data on the number of inhabitants and the number of employees for all German cities, even very small towns, covering the time period from 2001 to 2011. Data cover not only the total amount for employment and population but are available also for cohorts of 5 years (i.e 20-24, 25-29 and so on), for both employment and population, giving us the possibility to analyze in depth the two above mentioned empirical regularities, even on the age structure of the country. Descriptive statistics for the data are summarized in Table 3.1. We have been provided with a sample of annual observations in the period between 2001 and 2011 for more than 11,000 towns and cities. It is interesting to notice that the average growth rate of employment and population in the period taken in consideration shows different patterns, either at aggregate level and at cohort level. The total employment shows an average growth rate of 9.1% while the total population, instead, decreases of 3.5%. We can observe the same pattern in the youngest cohorts: the employment of young cohorts (for example, from 20 to 29 years) shows a positive trend (an increase of about 7%) while the population of the same cohorts decreases (almost 2%). This decrease in the population of younger cohorts can even be seen as a first sign of population ageing. This ageing process is even more underlined by the huge increases we can observe in the population of the elder cohorts, that is also reflected by the increases of the employment for those cohorts[8]. The only cohort in which this pattern breaks up is the cohort 60 – 64, in which we can observe a decreasing population growth, even though we can observe a huge increase in employment.

The correlation between the two variables population and employment (either at a global level and on a cohort level), however, is high (from 0.94 to 0.98, as you can see in Table 3.2), indicating a good source of comparability between the results.

3.4 Methods and Results

In some papers, especially in the first reappraisal of Gibrat's and Zipf's law, the interest for small towns and cities was very low (see for example Giesen and Suedekum (2011[51]; Soo (2005[15]); Overman and Ioannides (2003[18])). This is due to the fact that Zip's law holds only in the upper tail of the distribution and it deviates in a substantial way when it comes to the body of the city size distribution: because of this reason, the studies exploring Zipf's law focus mainly on the upper tail of the distribution. However, more recently, the interest in the distribution of all the cities becomes higher. This is thanks to the seminal works of Gabaix(1999[5]) and Eeeckhout (2004[16]), that focus on the theoretical relation between Zipf's and Gibrat's law (the former) and the empirical implication (the latter) that may lead to mispecification and wrong results (for instance, Eeckhout (2004[16]) shows that the estimated OLS

[8]Notice that the huge increase of the employment in the elder cohorts can also be due to the reform of labor market that took place in Germany between 2003 and 2005, also known as the Hartz package (Hartz I-IV). In a few word, the first three stages of the reforms sought to improve job search efficiency and employment flexibility. They included deregulation of the temporary work sector to give individual employers more flexibility to vary employment levels without incurring hiring or firing costs, as well as a restructuring of the federal labor agency in order to improve training and matching efficiency of job searchers. The final set of reforms entailed a major restructuring of the unemployment and social assistance system that considerably reduced the size and duration of the unemployment benefits and made them conditional on tighter rules for job search and acceptance.

Table 3.1: Descriptive statistics in 2011

Variable	Mean	Standard Deviation	Max	Avg Growth Rate
Total Emp	2585.95	19800.52	1151344	9.1%
Emp 20-24	257.36	1709.77	92707	7.3%
Emp 25-29	305.45	2491.82	140546	7.4%
Emp 30-34	309.59	2610.99	142905	-21.6%
Emp 35-39	298.19	2344.96	122459	-30.5%
Emp 40-44	391.46	2899.82	154350	5.3%
Emp 45-49	428.00	3053.74	176104	40.8%
Emp 50-54	362.35	2441.65	138754	52.8%
Emp 55-59	284.68	1907.36	110110	78.4%
Emp 60-64	144.36	1001.87	54978	120.6%
Total Pop	7252.44	47194.49	3501872	-3.5%
Pop 20-24	439.51	3158.61	231178	0.1%
Pop 25-29	442.23	3772.01	284687	-1.9%
Pop 30-34	438.01	3706.80	276989	-29.2%
Pop 35-39	424.03	3185.06	232551	-37.6%
Pop 40-44	563.86	3691.97	270099	-4.9%
Pop 45-49	632.49	4019.15	307041	26.6%
Pop 50-54	569.16	3362.41	253595	37.3%
Pop 55-59	491.81	2854.18	215153	58.7%
Pop 60-64	434.05	2639.21	200495	-7.8%

Table 3.2: Correlation among variables

	Tot Emp	Emp 20-24	Emp 25-29	Emp 30-34	... ...	Emp 50-54	Emp 55-59	Emp 60-64
Tot pop	0.97	0.97	0.97	0.96	...	0.97	0.98	0.98
Pop 20-24	0.97	0.97	0.97	0.96	...	0.97	0.98	0.98
Pop 25-29	0.97	0.97	0.97	0.97	...	0.97	0.97	0.98
Pop 30-34	0.97	0.97	0.98	0.97	...	0.97	0.97	0.98
Pop 35-39	0.98	0.97	0.98	0.97	...	0.98	0.98	0.98
Pop 40-44	0.97	0.97	0.97	0.97	...	0.98	0.98	0.98
Pop 45-49	0.96	0.96	0.96	0.95	...	0.97	0.97	0.97
Pop 50-54	0.96	0.96	0.96	0.95	...	0.97	0.97	0.97
Pop 55-59	0.96	0.96	0.96	0.95	...	0.97	0.97	0.97
Pop 60-64	0.96	0.96	0.96	0.95	...	0.97	0.97	0.97

coefficient of the rank-size rule varies depending on the truncation point in the city size distribution). This has lead to an always greater interest to the definition of statistical methods providing robust tool for the right truncation of the sample (see Fazio and Modica, (2015[30]) for a comparison of the methods).
For our purpose, we keep all the data for exploring Gibrat's law while we truncate the sample for Zipf's law, using the method proposed by Ioannides and Skouras (2013[53]). They provide an approach that estimates the truncation point as a parameter of a Pareto-lognormal distribution, $h(\cdot)$, by means of maximum likelihood estimation:

$$\max_{\mu,\sigma,\tau,q} \quad \ln h(P;\mu,\sigma,\tau,q),$$
$$s.t. \quad \tau > \exp(\mu)$$

where the Pareto-lognormal distribution has density:

$$h(P;\mu,\sigma,\tau,q) = \begin{cases} b(\mu,\sigma,\tau,q)f(P;\mu,\sigma), & \tau > P > 0 \\ a(\mu,\sigma,\tau,q)b(\mu,\sigma,\tau,q)g(P;q,\tau) & P \geq \tau \end{cases}$$

μ and σ are the parameters of the lognormal density function $f(\cdot)$ while q and τ are the parameters of the Pareto density function with τ the truncation parameter. Finally, $a(\cdot)$ is a continuity condition for $h(\cdot)$ and $b(\cdot)$ are conditions that ensure that $h(\cdot)$ is a density (See Ioannides and Skouras, (2013[53]) for more details).

In this analysis, depending on whether we use population data or employment data, we get different results. The first different result is that the population sample is truncated after 300 cities while the employment sample is truncated after only 125 cities, even though we apply exactly the same method. We then apply the Eq. (3.3)(that is also known as the rank-size rule, expressed in logarithmic terms), even though we use a slight modification of the equation; we estimate the q-coefficients in the rank-size rule by means of a modification proposed by Gabaix and Ibragimov (2011[54])[9] , that is expressed by the following:

[9]In the above mentioned paper, the authors show the estimator obtained by means of the usual regression is a biased one. They also show that this bias could be minimized if they subtract 0.5 from the rank value.

Table 3.3: Zipf's coefficient for employment

	Tot Emp	Emp 20-24	Emp 25-29	Emp 30-34	...	Emp 50-54	Emp 55-59	Emp 60-64
2001	0.84	0.85	0.87	0.86	...	0.85	0.87	0.88
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2002	0.84	0.85	0.88	0.86	...	0.84	0.87	0.88
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2003	0.84	0.84	0.88	0.86	...	0.84	0.87	0.88
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2004	0.84	0.84	0.88	0.86	...	0.84	0.86	0.88
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2005	0.84	0.83	0.88	0.86	...	0.83	0.86	0.88
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2006	0.84	0.83	0.88	0.87	...	0.83	0.85	0.88
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2007	0.84	0.82	0.88	0.87	...	0.83	0.84	0.87
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2008	0.84	0.82	0.88	0.87	...	0.83	0.84	0.86
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2009	0.84	0.83	0.88	0.88	...	0.82	0.84	0.85
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2010	0.84	0.83	0.89	0.88	...	0.82	0.83	0.84
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01
2011	0.84	0.82	0.88	0.88	...	0.82	0.83	0.84
	0.00	0.01	0.01	0.01	...	0.01	0.01	0.01

$$\log(P_i) = \log(K) - q\log(R_i - 0.5) \tag{3.4}$$

where P_i is the population of the city i, R_i is the rank of the ith city and K is a constant. Zipf's law is said to hold precisely, when the coefficient q is equal to one.

Results are shown in Table (3.3) and Table (3.4)[10]. According to Soo (2005[15]), the estimated Zipf coefficient might be thought as a measure of inequality; this is because, as we said before, if we assume that $q = \infty$ then, we could argue that all the population will be agglomerated in only one big city. On the other way round, if we assume that $q = 0$, then all

[10]Complete table are provided up to request.

Table 3.4: Zipf's coefficient for population

	Tot Emp	Pop 20-24	Pop 25-29	Pop 30-34	...	Pop 50-54	Pop 55-59	Pop 60-64
2001	0.75	0.73	0.79	0.79	...	0.75	0.77	0.74
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2002	0.75	0.74	0.79	0.79	...	0.74	0.77	0.75
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2003	0.75	0.75	0.80	0.80	...	0.74	0.76	0.76
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2004	0.75	0.75	0.80	0.80	...	0.74	0.76	0.76
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2005	0.75	0.75	0.81	0.81	...	0.74	0.75	0.76
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2006	0.75	0.75	0.81	0.81	...	0.74	0.75	0.76
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2007	0.76	0.76	0.82	0.83	...	0.74	0.74	0.76
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2008	0.76	0.76	0.82	0.83	...	0.73	0.74	0.76
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2009	0.76	0.76	0.82	0.84	...	0.73	0.73	0.75
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2010	0.76	0.76	0.83	0.84	...	0.73	0.73	0.74
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01
2011	0.76	0.77	0.83	0.85	...	0.74	0.73	0.74
	0.01	0.02	0.02	0.01	...	0.01	0.01	0.01

the cities will have same size. According to that, our results show that the total employment results much more concentrated with respect to total population for all the samples we use. Moreover, these results hold if we move our attention to the different cohorts: employment results always much more concentrated with respect to population for every cohort. However, we do not find any substantial differences between the concentration of employment of young people and elder ones (the estimated rank-size coefficient for cohort 25-29, 0.87, for example, is almost the same the cohort 60-64, 0.88). On the contrary, if we look at the estimated results for population, young people tend to concentrate more than eldest people (if we take into consideration the same cohorts we took before, we can see that the estimated rank size coefficient is 0.82 in cohort 25 – 29 while it is 0.75 in cohort 60 - 64). These first results show that agglomeration is higher when we look at the employment: this result was, in a sense, quite expected, as we can suppose places of work are not spread uniformly across all cities and tend to agglomerate in some cities more than others. On the population side, instead, we can notice that young people (cohorts going from 20 to 34 years old) show a greater agglomeration in larger cities than eldest ones, that is young people prefer living in larger cities, while elderly people are more spread across cities. This could be explained, for example, by the fact that, usually, larger cities have a wider range of leisure activities (for example, more gyms, wider range of restaurants, concerts, theatrical plays and so on), usually more intended for younger people with respect to elder ones. Or, on the other hand, young people can prefer living in biggest cities because, usually, biggest companies locate in and around biggest cities and these companies could be more attractive for young people than for older ones.

Finally, if we look at how the estimated coefficients change during the time, we do not recognize any significant difference in the estimated coefficients regarding the employment: aggregation level remains, more or less, the same, either on an aggregate level and on a cohort level. In a sense, this result is supporting our previous explanation: usually, companies do not relocate their headquarters or their offices so often, even because cities with a great number of headquarters offer a wide series of infrastructures, that, in principle, could be not offered by any city.

The result is different when we move our attention on the historical pattern for population. On an aggregate level, we cannot notice any sensible change in agglomeration level but this result comes from a compensation effect between the pattern of younger people and elder people. Indeed, during the ten year period taken in consideration, we can observe a substantial increase of the estimated coefficient for the younger cohorts (from 0.75 to 0.81, if we consider an average among the level of cohorts from 20 to 39 years old) and a substantial decrease in agglomeration level for elder cohorts (from 0.75 to 0.73, if we consider an average among the level from 50 to 64 years old): these results highlight a tendency towards agglomeration in larger cities for younger cohorts, whereas elder cohorts tend to spread more across little cities and towns.

If we instead move our focus to Gibrat's law, we make use of non-parametric analysis, strictly based on Ioannides and Overman (2003[18]). We use the normalized growth rate, namely the difference between a city's growth rate and the mean city growth rate, all divided by the standard deviation of growth rate. The strength of this non-parametric estimation is that we do not impose any relationship between the dependent and independent variables, i.e., we do not make any supposition about which the relation should be. According to Cameron and Trivedi (2005[55], p. 294), we: "let the data show the shape of the relationship"; this is an especially convenient approach when we do not know a priori the correct distribution of the data. In our analysis, we will use the Nadaraya-Watson (NW) method (Nadaraya (1964[56]); Watson (1964[57])), where the bandwidths are calculated with an optimal rule of thumb.
If Gibrat's law holds, the non-parametric estimation of the conditional mean and variance should be stable across different population sizes. Furthermore, because of normalization, we expect the conditional mean growth to be equal to zero, and the conditional variance of growth equal to one. It should be noted that, while the standard parametric regression methods provide only an aggregate relationship between growth and size, which is constrained to hold over the entire distribution of city sizes, the non-parametric estimates allow the growth to vary with size over the distribution.

In Figures from (3.1) to (3.6), we show the NW estimator for conditional mean growth (upper panels) and variance (lower panels) for the entire city size distribution and for their age cohorts: in particular, from fig. (3.1) to (3.3) we report results for employment and its age cohorts, whereas from fig. (3.4) to (3.6), we report the same results for population and its age cohorts. Following Cordoba (2003[58]), the independence of the expected conditional growth rate always has to be satisfied, while the variance can be affected by the city size. In general, smaller cities face a faster growth than larger ones. However, very quickly (in most cases), the conditional mean appears to become stable. This evidence is consis-

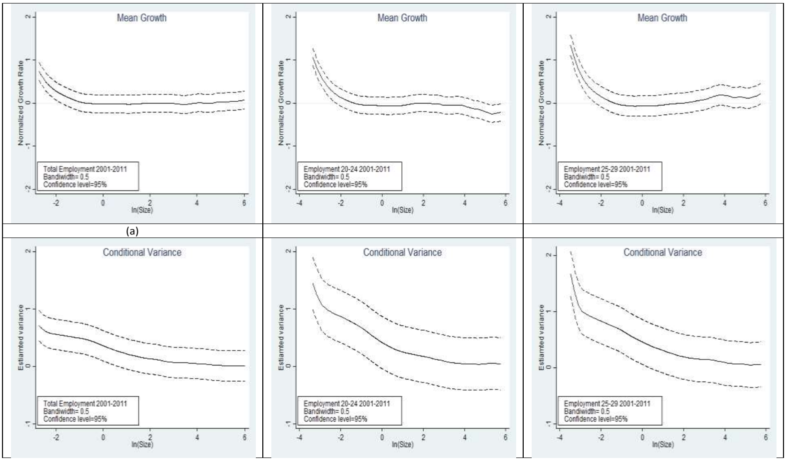


Figure 3.1: Conditional mean and conditional variance for employment

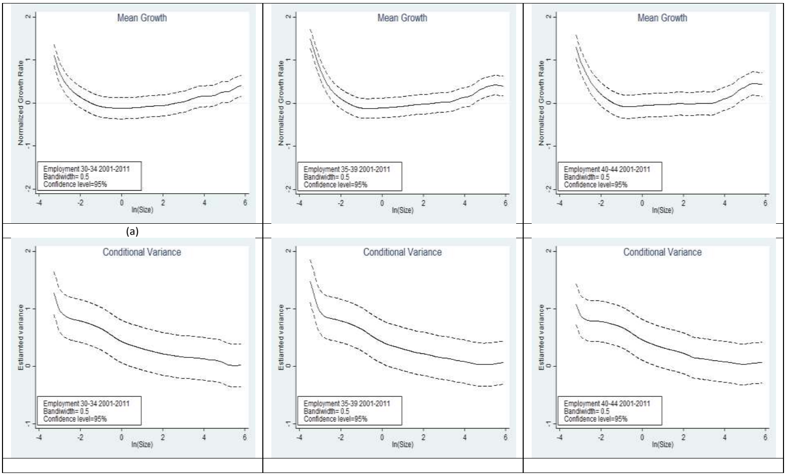


Figure 3.2: Conditional mean and conditional variance for employment

tent with the model proposed by Gabaix (1999[5]), where a truncation concerning the small cities is necessary to have stationarity.
Considering all the sample, without differentiating for cohort, we find

that Gibrat's law behave differently according to the choice of the analyzed variable. For instance, for employment, Gibrat's law is in operation, namely the growth rate of employment of a city is independent from the initial size, when we consider all the sample (Fig. 3.1, panel a) while, looking at the population, we can affirm that larger cities grow more than smaller ones (Fig. 3.4, panel a). This evidence might denote the possibility of a break-down in the choice of where to live with the availability of work (i.e. I will choice to live in a large city close to the city where I work, because a larger city can offer more in terms of leisure activities, as we said before or, because in a big city there are more job opportunities and, then, more career opportunities).

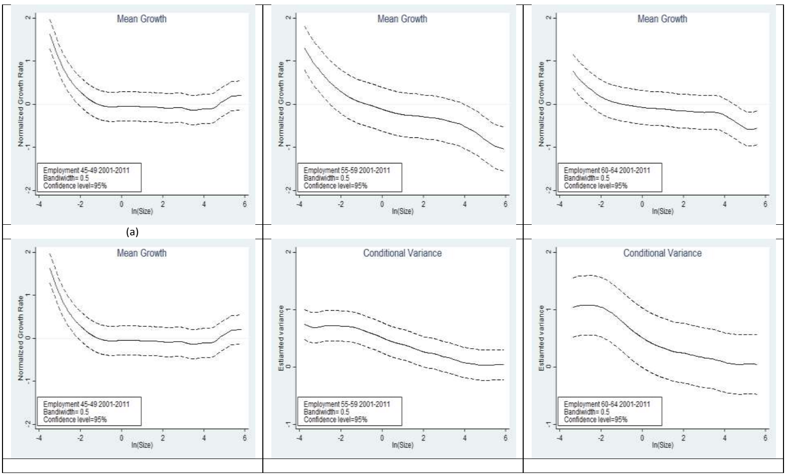


Figure 3.3: Conditional mean and conditional variance for employment

This result is corroborate by the different behavior of the Gibrat's law when we differentiate for ageing cohorts. Young cohorts (20 - 29) show a different pattern between employment and population: Gibrat's law is verified for cohorts 20 - 24 and 25 - 29 when we talk about employment but it is not verified when we talk about population (see Figures 3.1, panel b, 3.1, panel c and Figures 3.4, panel b, 3.4, panel c): in particular, we can see that growth rate increases sensibly, as the city size increases. Cohorts in the range 30 - 49 (See Figure 3.2, Figure 3.3, panel a, Figure 3.5 and Figure 3.6, panel a) instead show same results, both for employment and population: we can clearly see that growth rate of employment and

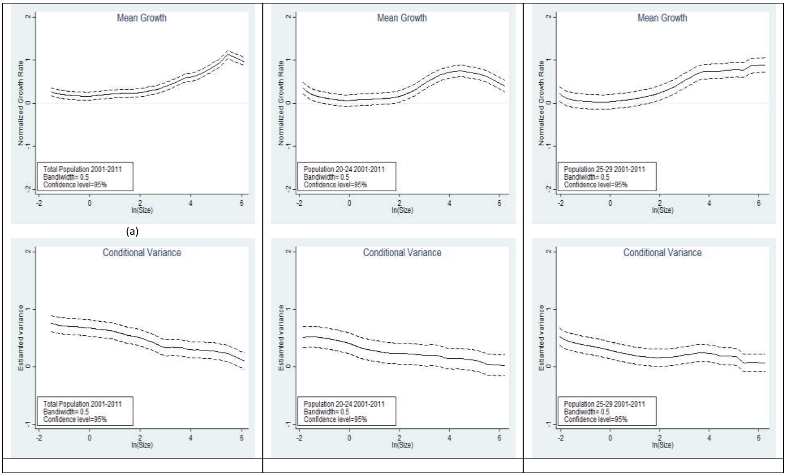


Figure 3.4: Conditional mean and conditional variance for population

population depends on the initial size, more in details larger cities show higher growth rate. For older cohorts (50 - 65) we obtain opposite results (see Figure 3.3, panel b and c and Figure 3.6, panel b and c): indeed, larger cities show lower growth rate both in population and in employment, indicating a situation where older people tend to concentrate in smaller cities.

3.5 Conclusion

Understanding the actual distribution of people in cities and the dynamics leading to this distribution could be of fundamental importance, either for policy relevant issues and also for developing more precise theoretical models. When one approaches these topics, without doubts one have to face with the two most recognized empirical regularities: Zipf's law and Gibrat's law. In an approximate way, the first one states that the city size in a country is proportional to its rank, whereas the second one affirms that the growth rate of an entity (cities, for example, but it can be applied even to firms) is independent of its size.
Scholars have dedicated a lot of studies to the analysis of these two empirical regularities, either in supporting them or in denying them, and no universal consensus has been reached, at the moment. However, com-

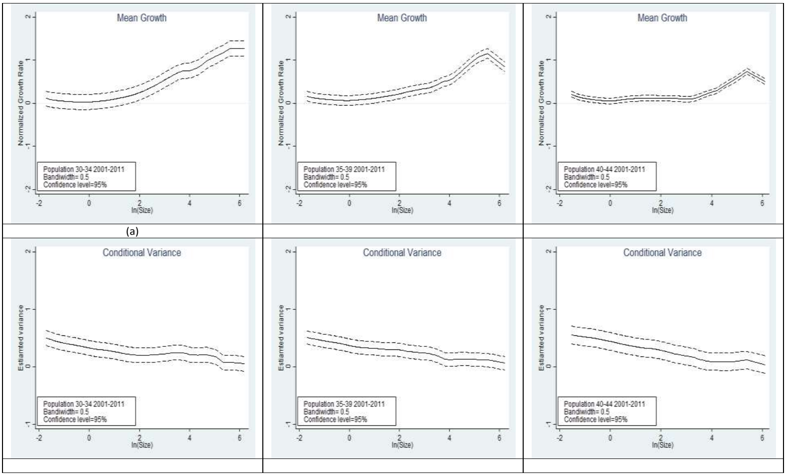


Figure 3.5: Conditional mean and conditional variance for population

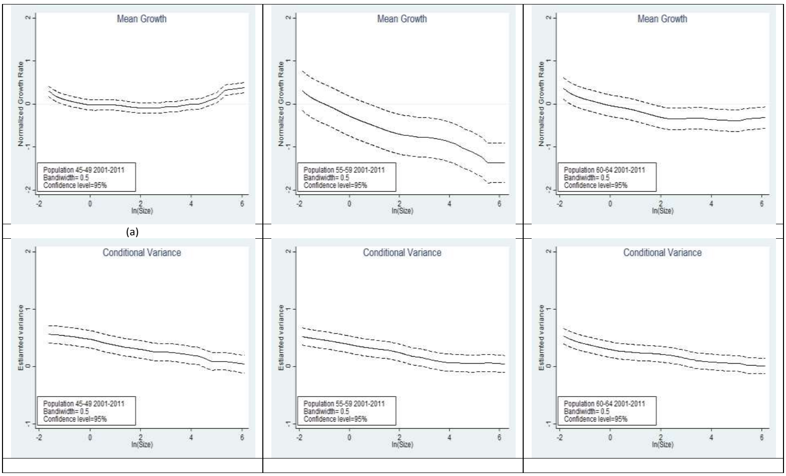


Figure 3.6: Conditional mean and conditional variance for population

mon literature provides analysis of Gibrat's law and Zipf's law without taking into account any demographic characteristics. In this paper, we go beyond the usual literature, focusing on population (and its age structure) and employment (and its age structure). In this way, we obtained a twofold novelty because first, we consider a demographic variable (employment) usually not considered in Gibrat and Zipf literature, then we introduce the focusing on demographic characteristics, in particular age structure. We made this by using a data set with annual observation of these two variables for all German cities and towns.

What we have shown is how differentiating for age cohort and using different dependent variables can lead to very different result. In particular, on an aggregate level, we found that employment is much more concentrated with respect to population (and this could be quite expected, as places of work are, usually, not spread across all cities), and, then, tends to concentrate in some cities rather than others. When we moved our attention to age structure, we found that younger cohorts show a clear tendency towards agglomeration in larger cities, whereas elder cohorts show the opposite behavior, i.e. spread across cities and towns. All of these results can be explained by different needs people with different age can have. For example, larger cities, usually, offer a wider range of leisure activities that are, usually, intended for younger people more than for elder people, or large cities offer more career opportunities, usually more attractive for younger people.

All of these new results should be taken into consideration, either in developing new theoretical models of urban growth that want to be more accurate and in structuring new policies that want to be more effective.

4 How neutral is the choice of the allocation mechanism in cap-and-trade schemes? Evidence from the EU-ETS

4.1 Introduction

Since the very beginning, the European Emission Trading Scheme (EU ETS) has attracted a lot of attention among international scholars and policy makers, because it represents the central European Union (EU from here on) policy instrument in order to mitigate climate change and to be compliant with the target agreed in the Kyoto protocol. The EU ETS represents the first attempt to develop a transboundary system of emissions trading and, therefore, it represents a prototype to other ETSs spreading around the world (see Borghesi et al, (2016[59])). The attractiveness of a cap-and-trade scheme like the EU ETS is due to the fact that, under certain conditions, it allows to attain a certain environmental target exogenously defined in an efficient way as an homogeneous price for emission permits across all the participants to the scheme will induce the equalization of marginal abatement costs.

By setting a price for carbon emissions, the EU ETS adds a constraint to firms, thus reducing (at least from a static point of view) expected profits with respect to a no-policy scenario. Many recent papers (see Martin et al, (2015[60]) for a comprehensive review) have looked at the potential impact of the EU ETS on the economic performance of treated firms, finding mixed evidence on a large variety of measures of economic performance (productivity, turnover, employment, etc). Carbon pricing, however, also induces investments in low carbon technologies aimed at reducing the cost of complying with the regulation. Calel and Dechezlepretre (2016[61]) found a significant positive inducement effect of the EU ETS on the development of low-carbon technologies (measured with new patent applications) for EU ETS firms.

The potentially harmful impacts on the competitiveness of European firms subject to the EU ETS coupled with the fact that the EU ETS was unilaterally introduced in Europe may induce firms to relocate their carbon-intensive production activities in countries with less stringent regulations for mitigating climate change (this effect is called carbon leakage effect). Carbon leakage has a two negative implications for the country (or group of countries) that introduces an unilateral stringent climate mitigation policy. First, from an environmental point of view, emissions at the global level are not reduced but only displaced towards other countries. As GreenHouse Gases (GHG from here on) emissions are global externalities (i.e. they contribute to global warming no matter where they are emitted), a simple displacement of carbon intensive activities undermines the environmental objective of the regulation. Second, from an economic and a political point of view, the relocation of carbon-intensive industries has a negative impact on the wealth and jobs created within the country. For this reason, the European Commission has been particularly sensitive about the issue of carbon leakage. In the first two phases of the EU ETS, emissions permits were allocated for free (i.e. grandfathering) while starting from the third phase (2013-2020) an increasing share of permits will be sold in auctions (generating revenues for public budgets) just for those sectors that are not exposed to carbon leakage risks.

According to the seminal paper by Coase (1960)[62], the level of emissions in equilibrium does not depend on the assignment of property rights over the emissions. If we apply this statement to the case of the EU ETS that means that the same result in terms of the distribution of abatement choice should be attained independently on the choice of the allocation mechanism. Different allocation mechanisms would only generate a different distribution of net benefits across firms or between firms and the government. This nice theoretical results implies that cap-and-trade schemes attain the equalization of marginal abatement costs across firms and thus a socially efficient distribution of the burden for climate policy. On the other hand, if for any reason the allocation mechanism is no longer neutral, a cap-and-trade scheme does not necessarily represent a first-best solution.

In this paper we exploit the asymmetry in the allocation mechanisms introduced from the third phase of the EU ETS as a way to evaluate whether different allocation mechanisms are neutral in terms of emission abatement decisions. The paper is structured as follows. In section 4.2 we describe in detail the European Emission Trading Scheme. Section 4.3 provides a theoretical discussion of the issues at stake, with a focus on the role played by carbon leakage considerations within the EU ETS and the possible reasons that may explain non-neutrality of the allocation mechanism. Section 4.4 provides an overview of recent trends in allocated and verified emissions for the EU ETS. Section 4.5 provides empirical evidence on whether verified emissions are independent on the allocation mechanism. Section 4.6 concludes.

4.2 The EU ETS

The EU ETS was introduced by the Directive 2003/87/EC[11] as the main initiative of the European Union climate change mitigation policy to reach the Kyoto targets and in order to be compliant with other current and future regional and international targets. In a few words, this is a cap-and-trade scheme put in place in order to keep CO_2 emissions under control: in this scheme, emissions permits are exogenously capped and then are allocated to the participants at the beginning of each period, either for free (with a method that is also known as emissions grandfathering) or auctioned. At the end of each period, all the participants are required not to have debt on emissions, that is they are required to return an amount of emissions permits at least corresponding to the actual amount of verified emissions. In the meantime, permits can be traded on a market, that is they can be transferred between participants at a price per ton of CO_2 that, in equilibrium, should be equal to the marginal abatement cost, leading to an efficient distribution of abatement across participants. Within the EU ETS, the penalty for non-complying (that is, a participant is not able to return a sufficient number of emissions permits at the end of the compliance period) was set to 40 euros per ton in the pilot phase (2005-2007) and to 100 euros per ton for the period 2008-2012.

This type of regulation was set in place with the intention to reach a double objective: first, reducing the overall abatement costs of carbon emissions and, second, providing the economic incentives to induce firms to invest in and so to develop low carbon technologies. This latter goal is pursued on the basis of the fact that the political acceptance of the regulation is likely to be higher if high induced innovation effects are expected.

In this scheme, three main periods can be identified: the period 2005-2007, in which the system was set up, and it represented a pilot phase. No banking was allowed between the pilot phase and subsequent EU ETS phases. The first commitment period (2008-2012), that was even the period leading to the Kyoto commitment period (2012), extended the scope of the scheme to aviation (2012). Finally, the second commitment period (2013-2020) introduced a single cap valid for all the EU for total emissions and a rising use of auctioning in the allocation of permits, with some exception for some selected sectors.

[11] Emended by the Directives 2004/101/EC and 2008/101/EC, the Regulation 219/2009 and the Directive 2009/29/EC

The EU ETS covers, now, all EU28 countries plus Norway, Iceland and Liechtenstein. As it is characterised by substantial sunk and fixed costs (including administrative and monitoring costs for participants and governments), the Commission decided to include in the scheme only the biggest emitters of CO_2. These emitters are identified by their sector of operation (or type of activity) and by the size of the plant in terms of production capacity. The scheme currently covers about 11,000 plants in Europe that contribute to around 45 percent of overall European greenhouse gases emissions[12]. The sectors and thresholds involved in the ETS scheme are reported in Annex I of the Directive and have been emended twice since 2003[13].

However, some exemptions were made in order to reduce the possible carbon leakage effect, that is the phenomenon for which firms may relocate part of the production in countries where this kind of regulation is not in place: this effect may hinder the policy effectiveness of the regulation. In this light, a major amendment to the Directive concerned the differentiation of the allocation scheme across sectors for the second EU ETS commitment period (2013-2020) according to the criteria described in the new articles 10 *bis* and 10 *ter* (Directive 2009/29/EC). The Decision of the European Commission 2010/2/EU '*Determining, pursuant to Directive 2003/87/EC of the European Parliament and of the Council, a list of sectors and subsectors which are deemed to be exposed to a significant risk of carbon leakage*' provided a list of 4-digit NACE sectors for which permits could be grandfathered rather than auctioned also in the second commitment period due to potentially relevant risks of off-shoring of these production activities deriving from the EU ETS. These sectors were identified through qualitative and quantitative analysis on the importance of potential carbon leakage and, to some extent, through a political negoti-

[12] http://ec.europa.eu/clima/policies/ets/index_en.htm

[13] The 2003 Directive refers to the following activities (with corresponding capacity thresholds - Annex I of the Directive 2003/87/EC): Combustion installations with a rated thermal input exceeding 20 MW (except hazardous or municipal waste installations); Mineral oil refineries; Coke ovens; Production and processing of ferrous metals; Metal ore (including sulphide ore) roasting or sintering installations;Installations for the production of pig iron or steel (primary or secondary fusion), including continuous casting, with a capacity exceeding 50 tonnes per day or in other furnaces with a production capacity exceeding 50 tonnes per day; Installations for the manufacture of glass including glass fibre with a melting capacity exceeding 20 tonnes per day; Installations for the manufacture of ceramic products by firing, in particular roofing tiles, bricks, refractory bricks, tiles, stoneware or porcelain, with a production capacity exceeding 75 tonnes per day, and/or with a kiln capacity exceeding 4 m^3 and with a setting density per kiln exceeding 300 kg/m^3; Industrial plants for the production of (a) pulp from timber or other fibrous materials (b) paper and board with a production capacity exceeding 20 tonnes per day. The list has been further extended to other sectors (refer to the consolidated version of the Directive 2003/87/EC

ation. Three main criteria were included in the amendment to identify the list of sectors to be exempted from auctioning[14]:

- The first is a 'trade-based' criterion according to which industries (4-digit NACE) having a non-EU trade intensity (import plus export over domestic production) greater than 30% are exempted from auctioning (trade criterion);
- The second refers to those industries (4-digit NACE) that are expected to experience additional (either direct and indirect) costs as a consequence of the implementation of the ETS Directive greater than 30% of their gross value added (emission criterion);
- The last criterion concerns industries (4-digit NACE) having, at the same time, moderate trade intensity and implementation costs (trade intensity greater than 10% and costs greater than 5% of gross value added[15]

The list was subsequently further emended to add other sectors with the decisions of the European Commission 2012/498/EU (that added sector 2614 'Manufacture of glass fibres') and 2014/9/EU (that added sector 2653 'Manufacture of plaster' and sector 2662 'Manufacture of plaster products for construction purposes'). However, the practice of exempting specific sectors from existing regulations is not uncommon: as Martin et al. (2014[63]) recall, since the introduction of carbon taxes back in the '90s, most of the countries involved grant some sort of exemptions to energy intensive firms to avoid their relocation.

4.3 Theoretical framework

As a free allocation scheme can potentially have distorting effects on the effectiveness and the working of an emissions cap-and-trade system, this issue has increasingly attracted interest from research and policy communities. This is because the absence of distortionary effects of free allocations can be seen as a necessary condition for the cost-effectiveness of a cap-and-trade scheme. Whether such distortions occur or not is, obviously, of particular interest, especially for those cap-and-trade schemes

[14]A fourth criterion refer to qualitative assessment (Art. 10bis.17) of the likely impact of EU ETS on production costs, investments and profit margins.

[15]These criteria are thoroughly discussed in the following document: http://ec.europa.eu/clima/policies/ets/cap/leakage/documentation_en.htm

in which a large portion (or even the most) of the total allowance allocation occurs practically free of charge: to all of the affected installations, it is granted an annual endowment, usually based on the installation's behavior in the past. If we were in an idealized world, allocations and emissions had to be totally independent: in this way, any arbitrary distributions of property rights would not affect outcomes, either on the trading side and on the emissions side (Coase, 1960[62]). The occurrence of this independence property in operational cap-and-trade schemes could be a very high advantage, especially from a policy and from a policy-maker perspective, as this give to the policy-makers themselves the ability to use free allocation of allowances in a political way, letting the cap-and-trade system do not suffer any negative consequence, especially in terms of cost effectiveness (Hahn and Stavins, 2011[64]). However, in a real-world cap-and-trade system, there are a lot of way in which annual endowments of free allowance allocation can distort and affect emissions outcomes, for example the presence of transaction costs (Coase, 1960[62]; Stavins, 1995[66]) or behavioral anomalies (Kahneman et al., 1991[69]). So, ensuring in a rigorous way if different allocation mechanisms affect emission outcomes becomes of major relevance.

The insight that, in the absence of any significant friction, optimal emissions at the unit level are invariant with respect to the initial allocation of property rights dates back to Coase (1960[62]). Hahn and Stavins (2011[64]) termed this invariance, the independence property in cap-and-trade systems. It has been shown that this independence (or invariance) property holds in a frictionless cap-and-trade system, as long as allocation occurs in a lump-sum fashion (Montgomery, 1972[68]). However, as we said before, there is a number of reasons why this independence property could fail in a real-world cap-and-trade scheme, when installations receive some endowments of allowance allocations free of charge, even when the allocation is a lump-sum one. Transaction costs (Stavins, 1995[66]) or imperfect competition (Hahn, 1984[65]) can lead to some kind of distortions in installation-level emission and abatement outcomes. As we said before, this independence property can fail, even due to some behavioral anomalies. It has been shown in the behavioural literature that in experimental settings, subjects value their allowance in a different way, depending on their allocation status, leading to under-trading and a loss in cost-effectiveness in the cap-and-trade system (Kahneman et al, 1990[69]; Murphy and Stranlund, 2007[70]).

However, using empirical analysis, it is still challenging to evaluate in a rigorous way whether it exists a casual relationship between installation-level allocation and emissions. The main difficulty is due to the endogeneity of allocations, which are typically set based on historical plant-level emissions, that usually are not observed. Therefore, in order to identify an actual causal effect of allocation on emissions, we require an exogenous source of variation in allocations. This is the major reason why the empirical literature focusing on the causal relationship between emissions and initial allocations in cap-and-trade systems is very poor, consisting of only two papers. Fowlie and Perloff (2013[71]) investigate this question, using the context of California's RECLAIM program[16] and using an instrumental variable approach to identification. Reguant and Ellerman (2008[72]), instead, investigate the same question for Spanish thermal power plants regulated under the EU ETS during ETS Phase I (2005-2008) and exploit a non-linearity in the national allocation rule for identification. Both of these papers did not find a significant endowment effect.

Why in a cap-and-trade scheme there should be a free allocation system? When a government intervenes in a marketplace, this intervention is always intended as a mean to increase net social welfare. Increasing social welfare by regulation could impose a cost on some industries for being compliant to the new regulation and then a government could use part of the revenue to partially compensate industries. This distributional effects of the new regulation could have, in principle, consequences on policy design. Let us suppose industries are offered no compensation at all, then the same industries have strong incentives to spend money on supporting political parties that are against the regulation, or to push to have exemption clauses, that could weaken the policy's effectiveness. Even worse, new regulation could push industries to relocate to unregulated countries, and this is a threat for politicians of all stripes, as job losses could affect re-election probability.

When we are dealing with policy about climate change, this threat

[16] REgional CLean Air Incentives Market is an emission trading program operating in the state of California since 1994. This program was imposed by SCAQMD (South Coast Air Quality Management District) in order to replace a series of more than 40 prescriptive rules, which had been opposed by the industry. The main goal of the program was to make so that hundreds of polluting facilities cut their emissions of nitrogen oxides (NOx) and Sulphur Oxides (SOx).

At its inception, in 1994, RECLAIM included 392 facilities whose combined NOx emissions accounted for over 65% of the region's stationary NOx emissions. Almost all facilities in SCAQMD with annual NOx or SOx emissions of four tons or more were included in the program. A RECLAIM trading credit (RTC) confers the right to emit one pound of emissions in a twelve month period. At the outset of the program, facilities were informed of how many permits they would be allocated for free each year through 2010. NOx emissions permitted under RECLAIM were reduce by over the 70 percent over the first ten years of the program.

about relocation is more aggravated by the so-called 'carbon leakage effect', that is the industrial relocation does not only shifts jobs to other country, but even GHG emissions: in this way, the policy does not only cause job losses in the country (or countries, in case of trans-national regulation) but does not reach the goal in environmental terms. The threat of relocation, then, gave rise to the three exemption criteria, we alerady discussed in section 4.2. Nonetheless, the importance of carbon leakage in climate policy design, there is little empirical evidence about the link the defined EU criteria and a sector's vulnerability to carbon leakage. All the existing studies are ex-post ones and they find no evidence of possible strong adverse impacts of the new regulation on competitiveness indicators, in case allocations are given for free: see, for example, Anger and Oberndorfer (2008) [73]; Abrell et al., (2011) [74]; Bushnell et al. (2013) [75]; Chan et al. (2013) [76]; Commins et al. (2011) [77]; Petrick and Wagner (2014) [78]; Wagner et al. (2013) [79]; Borghesi et al. (2016) [80]. Instead, if we move to theoretical and simulation-based studies, we find a negative impact of the new cap-and-trade scheme on production in most manufacturing industries: see, for example, Reinaud (2005) [81]; Demailly and Quirion (2006, 2008) [82, 83]; McKinsey and Ecofys (2006a, 2006b) [84, 85]. They also show that the free allowances allocation compensate negative profit impacts in most industries and can even lead to overcompensation, as shown in Smale et al. (2006) [86].

4.4 Trends in emissions within the EU ETS

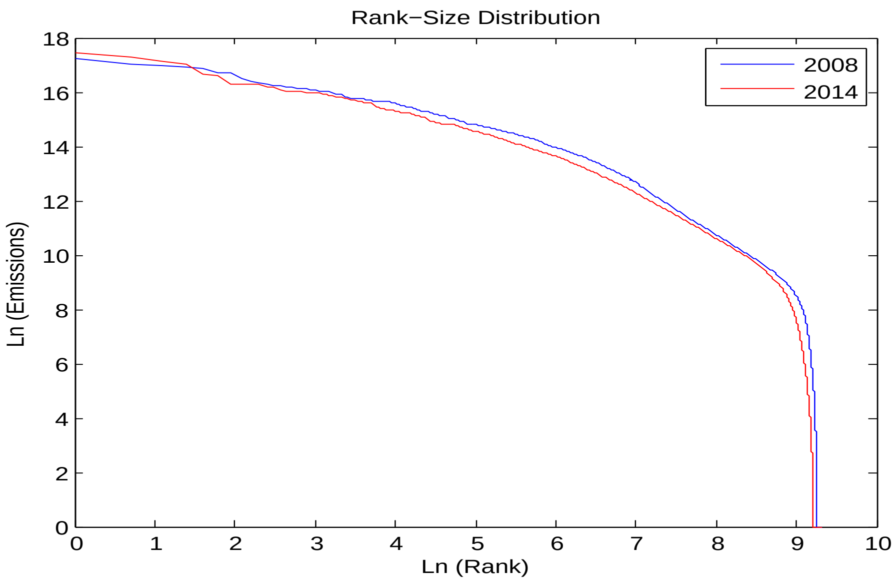
In this section, we analyse the behaviour of emission distribution, either considering our whole data-set and dividing it in different subsets. We employ yearly emissions data at the establishment level available from the European Union Registry[17] for the period 2008-2014.

As a first step, we plot the rank-size distribution, reported in Figure (4.1)

The plot is built in the following way: we sort all the plants in our dataset in a descending order on the basis of their verified emissions and we rank them from 1 to n, then we plot the logarithm of emissions and the logarithm of rank. The distribution has essentially the same shape for the first and last year of our series (2008 and 2014, respectively). What is worth noticing is that the distribution in 2014, with few exceptions, lies below the distribution for 2008 (only the highest ranks behave in

[17] http://ec.europa.eu/clima/policies/ets/registry/documentation_en.htm

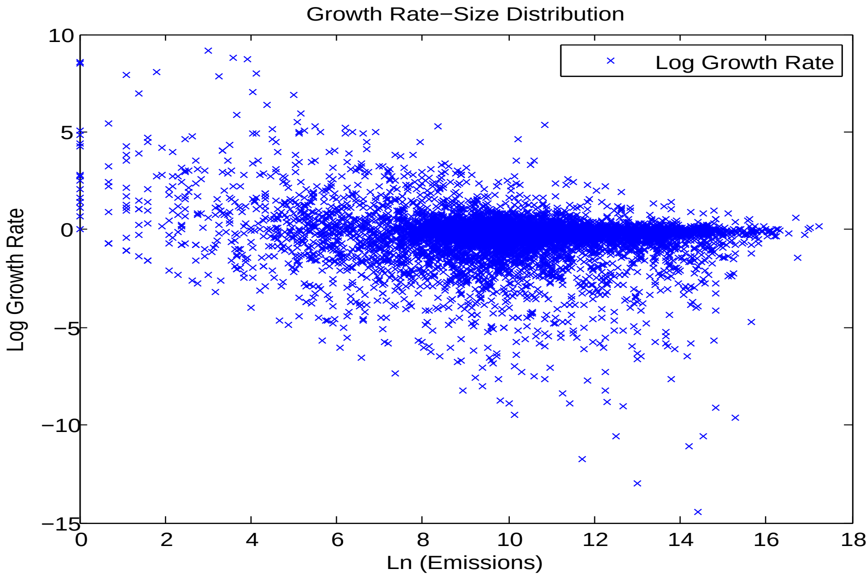
Figure 4.1: Log Emissions vs Log Rank



the opposite way, having an amount of verified emissions that is greater with respect to the 2008) signalling that, in those 6 years, it seems that emissions have been lowered. However, this plot gives us only a broad picture of what happened during the 6 years taken in consideration: plotting only size and rank does not tell us anything about the dynamics of emissions, because ranks could, in principle, even change drastically in the period, leading us to compare very different plants from one year to another. That is why we consider another plot, reported in Figure (4.2)

In this plot, we reported, for each plant for which we have data both in 2008 and in 2014, the plant's verified emission growth rate (computed as logarithmic return: $\ln(emissions_{(2014)}/emissions_{(2008)})$) against the logarithm of the initial emission size (verified emissions in 2008). The plot suggests that there has been a decrease in the growth rate as the emissions' size increase, i.e. the smallest plants have an higher growth rate with respect to the biggest ones. As there is large variance in the growth rate in verified emissions across plants, we decide to split the sample in smaller quantiles of verified emissions and take the mean logarithmic growth rate within each quantile. This plot is much clearer than the first one, confirming our first impression of growth rate decreasing as a

Figure 4.2: Log Growth Rate vs Log (Emissions)

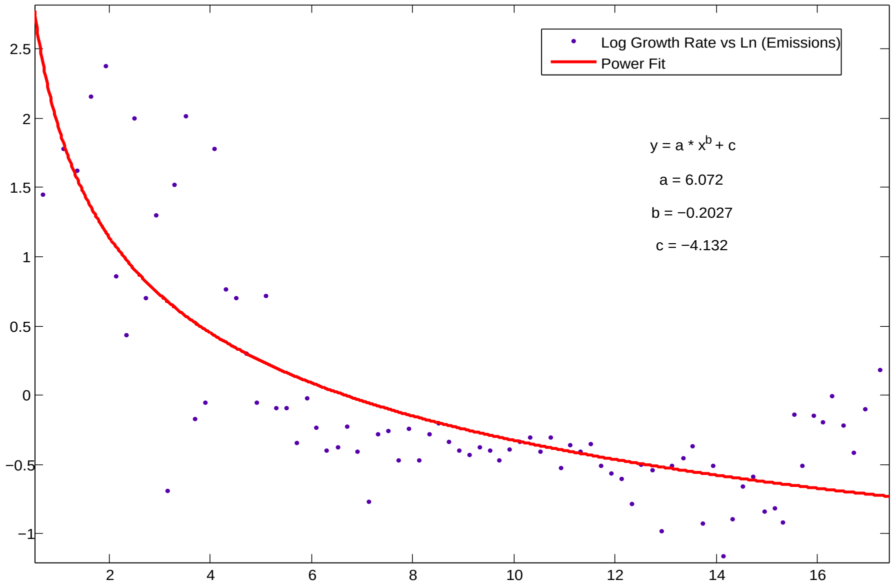


function of initial emissions' size. Moreover, as we found a very regular behaviour, we tried to fit it, obtaining a very good fit with a power law distribution (R^2 equal to 0.63, not reported), as we show in Figure (4.3).

The decreasing relationship seems to be quite glaring: whereas the largest emitters decreased (on average) their amount of emissions, the lowest emitters, on average, increased a lot their amount of emissions. This result seems to go against the evidence discussed for Figure (4.1). However, this is not the case because, as we said before, in Figure (4.1) we do not have any hint about which plant was in the highest ranks in 2008 and which in 2014 whereas in Figure (4.3) on the x-axis the ranking in 2008 is, in a way, considered and we so can observe how it changed throughout those 6 years. Moreover, whereas in Figure (4.1) there is a complete representation of all our data, in Figure (4.3) each point represents an average of plants that fall in the same bin.

So far, we considered the longest period for which we have data (from 2008 to 2014). However, as discussed in the previous sections, during this period, the regulation moved from the second phase (2008-2012) to the third phase (from 2013 onwards), in which emissions were not any more

Figure 4.3: Fitted Log Growth Rate vs Log (Emissions)

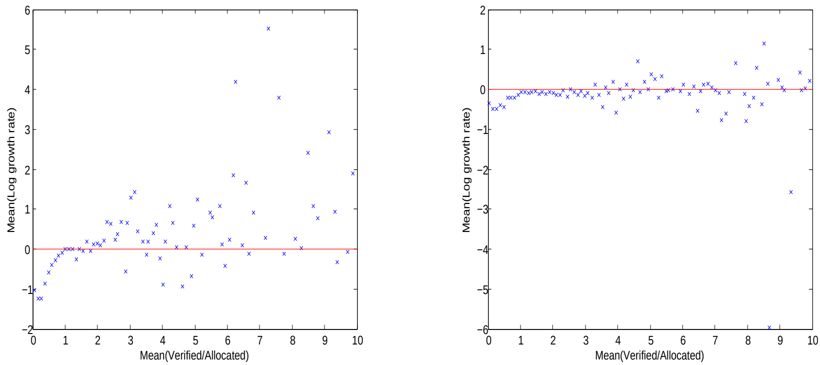


freely allocated (grandfathering) but increasingly auctioned, with the exclusion of leakage-exposed sectors. As discussed in the theoretical part, this change in the regulation should not have effect on the polluting behaviour of the plants as at the margin the opportunity cost of emitting or abating should remain the same. As a first step to evaluate this hypothesis, we look at the logarithmic growth rate of emissions as a function of the ratio between verified and allocated emissions: this could represent a measure of how much impact could have the change of normative on plants' emissions, as plants with a ratio greater than 1 should reduce their emissions or increase their number of permits to be compliant. The plots for the two periods are reported in Figure (4.4)[18].

The plot is realized in the following way. On the y-axis, as we did before, we report the logarithmic growth rate of emissions, whereas, on the x-axis, we report the ratio between verified and allocated emissions. Each point in the plot represent the mean logarithmic growth rate and the mean ratio of verified and allocated emissions of several plants falling in the same bin. On the left, we find the plot for the period 2008-2012

[18] In this case, we directly report the mean plot

Figure 4.4: Logarithmic emissions' growth rate as a function of the ration between verified and allocated emissions.

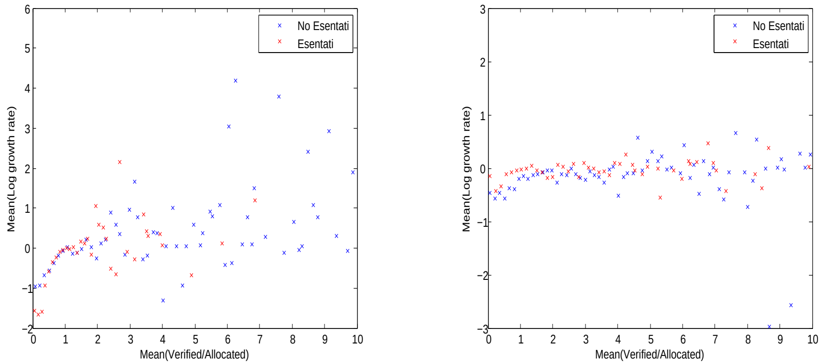


whereas, on the right, we can find the plot for the period 2013-2014. The red line is a zero-level line in order to distinguish plants who increased their emissions from plants who decreased their emissions.

In this way, we can distinguish different behaviour for plants who needed to buy allowances on the market to cover their verified emissions (that is, $Mean(Verified/Allocated) > 1$) and plants who could sell allowances on the market because they were endowed with more than needed ($Mean(Verified/Allocated) < 1$). As we can see from the plot, there was a huge change in the polluting behaviour from the first period (2008-2012, first ETS and Kyoto commitment period) to the second period (2013-2014). There is a huge difference in polluting behaviour between the two period, especially for plants whose ratio between verified and allocated emissions is greater than 1. Whereas, in the first period, the logarithmic growth rate for those plants is very spread, ranging from -1 to almost 6, in the second period, the range is narrower, ranging from -1 to 1.

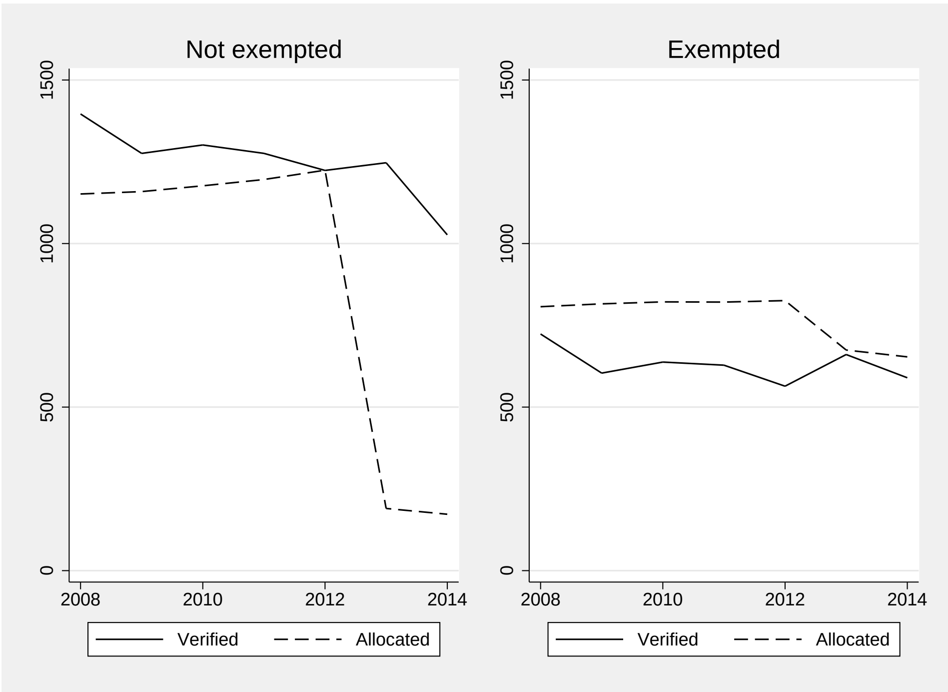
Lastly, we wanted to check if this different polluting behaviour is different between exempted and not exempted sectors, in order to see whether the above mentioned exemption from auctioning could have had an impact on the behaviour in the two periods. To do so, we repeat the plot of fig. (4.4), dividing our sample in two further subsets, one for plants in sectors who were exempted (*Esentati*) and one for plants in sectors who were not exempted (*No Esentati*): the results are reported in fig. (4.5). On the left, we find the plot for the period 2008-2012 whereas, on the right, we can find the plot for the period 2013-2014. In red, we find plants in exempted sectors whereas, in blue, we find plants in not exempted sectors.

Figure 4.5: Logarithmic emissions' growth rate as a function of the ration between verified and allocated emissions.



There does not seem to be a systematic differences in emission growth in the two period between exempted and non-exempted plants for different levels of $Mean(Verified/Allocated)$. When looking at trends in verified and allocated emissions (fig. (4.6)) we observe that, even before the exemption, the amount of permit allocated to sectors more exposed to trade systematically exceeded the amount of verified emissions while the opposite occurred for sectors not exposed to leakage. This fact is a signal that even before the change in regulation sectors exposed to leakage were either over-endowed of permits or they were under-abating with respect to other sectors. As expected, the amount of allocated permits dropped substantially (almost to zero) for establishments in non-exempted sectors from 2013 onwards, while the drop was much smaller for establishments in exempted sectors. Interestingly, in 2013 (first year after the reform) verified emissions of sector exempted from auctioning increased substantially while verified emissions of other sectors (that had to buy permits) experienced only a moderate growth. This is a first indication that the initial endowment of permits matters for the distribution of pollution abatement.

Figure 4.6: Trend in verified and allocated emissions for exempted and non-exempted plants



4.5 Regression analysis

The detailed descriptive evidence highlighted so far suggests that plants in those sectors that were forced to purchase pollution permits through auctions from 2013 onwards reduced their emissions relatively more than plants in sectors exempted from auction. This may suggest that the allocation mechanism chosen by the regulator is not neutral in terms of abatement choice of firms and that the coexistence of different allocation mechanisms within the same cap-and-trade scheme may generate some distortions. To better identify whether this descriptive evidence is not the result of other factors that influence emissions and abatement choice we employ a state of the art econometric approach to evaluate whether this change in regulation influenced abatement choices.

A recent paper by Zaklan (2016[87]) evaluates the impact of the same change in the allocation mechanism of the EU-ETS only for plants that operate in the power sector. Their identification is based on the fact that for 8 EU countries (with a GDP per capita below 60% of the EU average) obtained from the European Commission a postponement of the allocation through auction for plants in the power sector up to year 2020 to ease the modernization of the power sector. Exploiting this asymmetry in regulation, the authors find little support for significant impacts of heterogeneous allocation mechanisms on verified emissions in the power sector.

It should be noted, however, that the features of the power sector (i.e. large firms, non-tradability of the output and inelastic demand for electricity) are likely to attenuate the expected impact of changes in allocation mechanisms on abatement behaviour. Companies in the power sector can easily pass-through increases in production costs to consumers, creating little incentives, at least in the short term, to change their abatement behaviour. On the contrary, plants in tradable industries such as manufacturing industries, exposed to international competition, have much smaller possibility to pass through higher upfront costs for complying with the regulation and may decide to change their abatement behaviour even in the shorter run. For this reason, our focus from now onwards is on the manufacturing industry only.

To evaluate how the distribution of pollution abatement changed after the reform of the EU-ETS that exempts plants in specific sector from auctioning we estimate a simple econometric model. The idea is that in absence of exemption, verified emissions and allocation of permits would have evolved in the same way for both treated and control plants.

We estimate the following equation:

$$log(Verified\_emiss_{it}) = \beta Exempt_s \times Post2013_t + X'\gamma + \tau_t \alpha_i + \varepsilon_{it} \quad (4.1)$$

where $Verified\_emiss_{it}$ represents verified emissions for establishment i in year t, $Exempt_s$ is a dummy variable (time-invariant) that equals one for those sectors exempted from auctioning from 2013 onwards, $Post2013_t$ equals one for years 2013 and 2014 and zero otherwise, $X'\gamma$ is a vector of control variables (EU28-level trends in production by 4-digit NACE sector in log from PRODCOM, country-specific linear trends and main activity-specific linear trends), τ_t are time dummies and α_i is the plant fixed effect.

Our parameter of interest is β which describes the estimated increase in verified emissions in establishments that are exempted from auctioning. This is a simple difference-in-difference approach. The identification assumption is that treated and untreated individuals would have followed the same trend in absence of the treatment. As the assignment to treatment is not random (i.e. it is based on the sector of operation of the plant), there are many possible reasons that may give rise to different trends in emissions even in absence of the treatment. We already partly account for these confounding factors with the inclusion of the establishment fixed effect (that account for time-invariant differences across plants), the sectoral trend in production (that account for the dynamics of demand), country-specific trends and activity-specific trends. This may not be enough as other systematic differences between plants in different sectors may give rise to different trends. For this reason, we try to further reduce the heterogeneity between treated and control plants by matching controls to treated by means of the propensity score. We estimate the probability of belonging to exempted sectors as a function of trends in production for the 4-digit NACE sector over 2005-2009 (to account for possible difference in output growth), average ratio between verified and allocated emissions in 2008-2009 for the 4-digit NACE sector to account for pre-treatment systematic over- or under-endowment in the sector and the log of verified emissions in the plant for 2008 to account for differences in the size of plants. As the number of treated (531) is smaller than the number of potential controls (2344) each selected control will be employed as a counterfactual for multiple treated. We employ kernel matching to exploit as much information as possible about controls in a flexible way. In this way, in fact, the counterfactual is a weighted average of different controls, with weights being specific for each treated

Table 4.1: Propensity score and balancing

Propensity score	PS	Average treated	Average controls (all)	t-test on the difference	Average matched controls	t-test on the difference
Growth in production (4-digit NACE) for 2005-2009	0.0102 (0.0063)	2.52	1.41	4.01	2.26	1.15
Average verified/allocated emissions (4-digit NACE) for 2008-2009	-1.787*** (0.2598)	0.755	0.876	-2.59	0.718	10.13
log(verified emissions of the plant, 2008)	0.234*** (0.0201)	-3.413	-4.268	11.33	3.8561	10.51

N=2875 plants in manufacturing sectors (balanced panel 2008-2014). Matching based on kernel algorithm.

and inversely proportional to the distance in terms of estimated propensity score.

Results of the propensity score estimate and of the balancing (before and after matching) are reported in Table (4.1). The propensity score is successful in reducing the heterogeneity in terms of long term trends in demand between treated and controls. For the other matching variables, even though the tests on the difference in observable features post-matching do not allow to reject the null hypothesis of insignificant difference for over-allocation and establishment size, the magnitude of the difference in averages between treated and controls is much smaller after matching, leading to a more credible counterfactual.

Figure (4.7) reports the distribution of the estimated propensity score for treated, all controls and matched controls. The density function for treated and controls is very similar, suggesting that the two groups are rather homogeneous after matching. The dark grey dots, however, describe those plants that were not matched due to absence of common support in the propensity score: these are 35 plants with either very low or very high probability of being treated.

Average treatment effect on the treated, where the outcome variable is the growth in emissions for the years 2009 to 2014 with respect to emissions in 2008, are reported in table (4.2). These are simple difference-in-differences estimates, with no additional control variable. We observe that the estimated difference in growth rates between treated and controls is small and insignificant until 2012 and becomes large and significant from 2013 onwards, that is the year in which we observe the change in regulation. Overall, the estimated difference is around 14 log points in 2013 and reaches 19 log points in 2014, pointing to a large estimated effect. The positive sign means that those plants that continued to receive their allowances for free have increased their emissions with respect to the ones that had to purchase them in auction, leading to non-neutrality of abatement choices with respect to allocation mechanisms.

Figure 4.7: Distribution of estimated propensity score

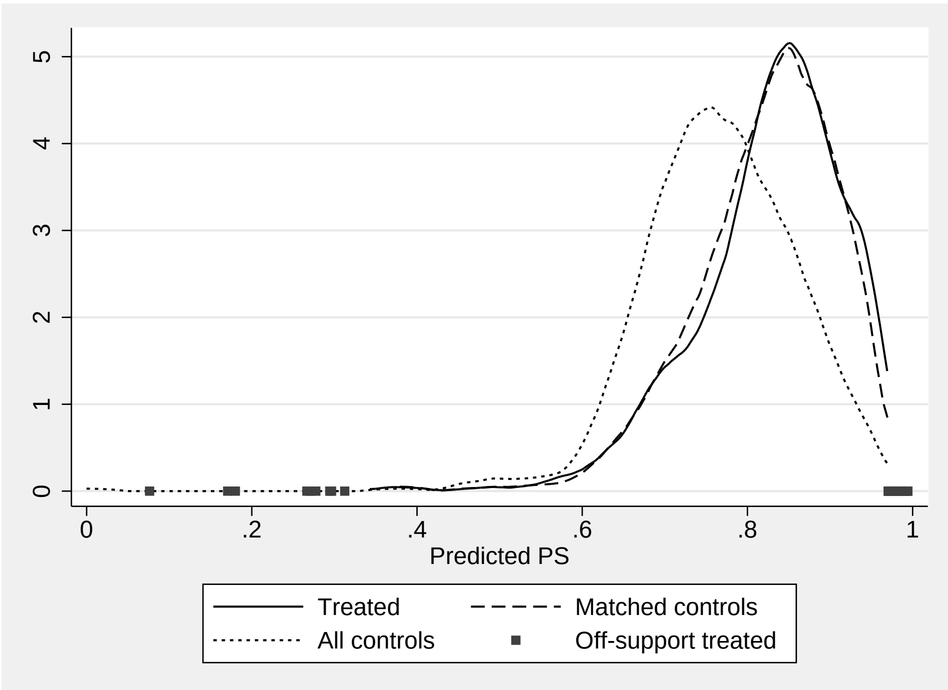


Table 4.2: Average treatment effect (simple difference in differences on matched plants)

Treatment effect		Difference	SE	t-test
Change in log verified emissions 2008-2009	Unmatched	-0.0437	(0.0220)	-1.99
	ATT	0.0255	(0.0292)	0.87
Change in log verified emissions 2008-2010	Unmatched	-0.0119	(0.0240)	-0.49
	ATT	0.0496	(0.0318)	1.56
Change in log verified emissions 2008-2011	Unmatched	0.0094	(0.0309)	0.30
	ATT	0.0754	(0.0394)	1.91*
Change in log verified emissions 2008-2012	Unmatched	-0.0232	(0.0331)	-0.70
	ATT	0.0552	(0.0407)	1.36
Change in log verified emissions 2008-2013	Unmatched	0.0419	(0.0424)	0.99
	ATT	0.1365	(0.0453)	3.01***
Change in log verified emissions 2008-2014	Unmatched	0.0667	(0.0482)	1.38
	ATT	0.1908	(0.0556)	3.43***

Table 4.3: Difference-in-differences with matching and additional controls

log(verified emissions)	(1)	(2)	(3)	(4)	(5)	(6)
Exempted x anticipation (2010-2012)	0.0133	0.0473*	0.0445*	0.0431*	0.0686**	0.0438
	(0.0211)	(0.0270)	(0.0269)	(0.0228)	(0.0281)	(0.0277)
Exempted x post 2013	0.0762**	0.151***	0.168***	0.136***	0.194***	0.169***
	(0.0353)	(0.0453)	(0.0451)	(0.0383)	(0.0441)	(0.0437)
log(Production, NACE 4-digit)			0.509***			0.567***
			(0.0753)			(0.0957)
Matched on PS	No	Yes	Yes	No	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country trends	No	No	No	Yes	Yes	Yes
Main activity trends	No	No	No	Yes	Yes	Yes
N	20125	19810	19810	20125	19810	19810

Our baseline estimates (that also account for a series of control variables) are reported in table (4.3). Column 1 and 4 report estimates for the full sample with no matching of treated with similar controls while in the other columns we weight plants according to the weights estimated in the matching phase. Column 6 is our favourite estimate, in which we match treated with similar control, add country and main activity trends and control for the trend in production at the 4-digit level. The treatment effect is positive, significant and around 17 log points. We also account for possible anticipation effects as the exemption was already agreed upon in December 2009. Overall, there is some little evidence of anticipation which turns out to be insignificant in our preferred estimate.

We also try to differentiate the effect for plants that belong to different categories of exemption from auctioning, namely moderate trade and emission intensity (criterion A), high emission intensity (criterion B), high trade intensity (criterion C) and sector exempted according to qualitative criteria (art 17). These criteria may coexist within the same sector. Results are reported in table (4.4). These results, mostly descriptive in nature,, suggest that the effect was driven by those sector more exposed to trade (criteria A and C) while a negative effect was found for

Table 4.4: Effect for different exemption criteria

log(verified emissions)	(1)	(2)
Criterion A x anticipation (2010-2012)	0.0632***	0.0660**
	(0.0239)	(0.0287)
Criterion A x post 2013	0.222***	0.227***
	(0.0418)	(0.0476)
Criterion B x anticipation (2010-2012)	-0.00743	-0.340***
	(0.0304)	(0.0638)
Criterion B x post 2013	-0.0324	-0.697***
	(0.0625)	(0.126)
Criterion C x anticipation (2010-2012)	0.0619***	0.0577**
	(0.0236)	(0.0266)
Criterion C x post 2013	0.150***	0.142***
	(0.0423)	(0.0508)
Criterion Art 17 x anticipation (2010-2012)	-0.0642*	-0.00911
	(0.0349)	(0.0370)
Criterion Art 17 x post 2013	-0.0443	0.0659
	(0.0603)	(0.0590)
Matched on PS	Yes	Yes
Year dummies	Yes	Yes
Country trends	No	Yes
Main activity trends	No	Yes
N	19810	19810

emission-intensive sectors.[19] These results point to the fact that when international competition is though, even small increases in the amount of resources that should be spent for complying with environmental regulation induces large changes abatement behaviours even in the short run.

4.6 Concluding remarks

The paper propose an empirical evaluation of the neutrality (or absence thereof) of the allocation mechanism for abatement decisions within cap-and-trade schemes. Our analysis is based on data on emissions of the European ETS. This scheme, the largest in the world in terms of amount of emissions and number of involved establishments, is particularly suitable to test the neutrality of allocation mechanisms as it experienced a change in the allocation mechanism in recent years. The move from grandfathering to partial (i.e. with exemption) auctioning allows to estimate whether the way permits are allocated has an influence on abatement decisions, against the prediction of the Coase theorem.

After providing a comprehensive descriptive evidence on recent trends in verified emissions and allocation of permits, we evaluate whether the change in regulation, with an exemption from auctioning granted to leakage-exposed sectors, influenced abatement behaviours of firms.

[19] These include, in the manufacturing sectors, only NACE codes 23.51 (manufacture of cement, 176 plants) and 23.52 (manufacturing of lime and plaster, 148 plants).

Focusing on manufacturing establishment, our preferred estimate suggests an increase in emission of about 17 log points for plants that are exempted from auctioning with respect to the ones that should buy permits through auctions. This contradicts the theoretical prediction about the neutrality of allocation mechanisms in cap-and-trade schemes, thus leading to sub-optimal outcomes. These findings should inform policy makers about possible ways of improving ETS-like schemes in order to improve their economic efficiency and correct for potential distortions induced by specific rules for specific case such as the case of carbon leakage.

Even though these findings already represent an useful contribution for the policy debate, further research is needed to understand which are the more important mechanisms that induce changes in abatement choices as a consequence of changes in allocation mechanism. This additional research should consider both theoretical reasoning about the non-neutrality and empirical validation of these theoretical hypothesis.

References

[1] **Fazio, Giorgio and Modica, Marco**, *"Pareto or Log-normal? A recursive-truncation approach to the distribution of (all) the cities"*, 2012, Working Papers, Business School - Economics, University of Glasgow

[2] **Zipf, George K.**, *"Human Behavior and the Principle of Last Effort"*, Cambridge, MA: Addison-Wesley, 1949

[3] **Gibrat, Robert**, *"Les inégalités économiques: applications: aux inégalités des richesses, à la concentration des enterprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel"* Paris: Librairie du Recueil Sirey, 1931

[4] **Giesen, Kristian, Zimmermann, Arndt and Suedekum, Jens**, *" The size distribution across all cities - Double Pareto lognormal strikes"*, Journal of Urban Economics, 2010, 68, pp. 129-137

[5] **Gabaix, Xavier**, *"Zipf's Law for Cities: An Explanation"*, The Quarterly Journal of Economics, 1999, 114(3), pp. 739-767

[6] **Reed, William J.**,, *"On the rank-size distribution for human settlements"*, Journal of regional sciences, 2002, 42, pp. 1-17

[7] **Yule, Udny G.**, *"A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, F.R.S."*, Philosopical Transactions B, 1925, 203, pp. 21-87

[8] **Gabaix, Xavier and Ioannides, Yannes M.**, *"The Evolution of City Size Distributions"*, Handbook of Regional and Urban Economics, 2003, 4(53), pp. 2341-2378

[9] **Le Maître, Alexandre B.**, *"La metropolitée, ou de l'établissement des villes capitales, de leur utilité passive & active, de l'union de leurs parties & de leur anatomie, de leur commerce"*, Publisher: Amsterdam, Ches B. Boekhold, Pour J. Van Grop, 1682

[10] **Auerbach, Feliz**, *"Das Gesetz der Bevölkerungskonzentration"*, Petermanns Geographische Mitteilungen LIX, 1913, 73-76

[11] **Dobkins, Linda H. and Ioannides, Yannes M.**, *"Dynamic evolution of the U.S. city size distribution"*, in: Huriot J., Thisse J. eds., The Economics of Cities, Theoretical Perspectives, cambridge University Press, Cambridge, 2000, 217-260

[12] **Fujita, Masahisa, Krugman, Paul and Venables, Anthony J.**, *"The Spatial Economy"*, MIT press, Cambridge, MA, 1999

[13] **Rosen, Kenneth T. and Resnick, Mitchel**, *"The size distribution of cities: an examination of the Pareto Law and primacy"*, Journal of Urban Economics, 1980, 8, 165-186

[14] **Brakman, Steven, Garretsen, Harry and van Marrewijk, Charles**, *"An Introduction to Geographical Economics"*, Cambridge University Press, Cambridge and New York, 2001

[15] **Soo, Kwok Tong**, *"Zipf's Law for cities: a cross country investigation"*, Regional Science and Urban Economics, 2005, 35, 239–263

[16] **Eeckhout, Jan**, *"Gibrat's Law for (All) Cities"*, American Economic Review, 2004, 94(5), 1429-1451

[17] **Eaton, Jonathan and Eckstein, Zvi**, *"Cities and Growth: Theory and Evidence from France and Japan"*, Regional Science and Urban Economics, 1997, 27(4-5), 443-474

[18] **Ioannides, Yannes M. and Overman, Henry G.**, *"Zipf's Law for Cities: An empirical Examination"*, Regional Science and Urban Economics, 2003, 33(2), 127-137

[19] **Glaeser, Edward L., Scheinkman, Jose A. and Shleifer, Andrei**, *"Economic Growth in a Cross-Section of Cities"*, Journal of Monetary Economics, 1995, 36(1), 117-143

[20] **Krugman, Paul**, *"Development, geography, and economic theory"*, Cambridge, MA: MIT Press, 1995

[21] **Kapteyn, Jacobus C.**, *"Skew frequency curves in biology and statistics"*, Astronomical Laboratory, Groningen: Noordhoff, 1903

[22] **Champernowne, David G.**, *"A Model for Income Distribution"*, Economic Journal, 1953, 63(250), 318-475

[23] **Sutton, John**, *"Gibrat's Legacy"*, Journal of Economic Literature, 1997, 35(1), 40-59

[24] **Jovanovic, Boyan**, *"Selection and the Evolution of Industry"*, Econometrica, 1982, 50(3), 649-670

[25] **Blank, Aharon and Solomon, Sorin**, *"Power Laws in Cities' Population, Financial Markets and Internet Sites (Scaling in Systems with a Variable Number of Components"*, Physica A, 2000, 287(1-2), 279-288

[26] **Kesten, Harry**, *"Random Difference Equations as Renewal Theory for Products of Random Matrices"*, Acta Mathematica, 1973, 131, 207-248

[27] **Levy, Moshe**, *"Gibrat's Law for (All) Cities: Comment"*, American Economic Review, 2009, 99(4), pp. 1672-1675

[28] **Eeckhout, Jan**, *"Gibrat's Law for (All) Cities: Reply"*, American Economic Review, 2009, 99(4), pp. 1676-1683

[29] **Clauset, Aaron, Shalizi, Cosma R. and Newman, Mark E.J.**, *"Power-law distributions in empirical data"*, SIAM Review, 51 (2009), pp. 661-703

[30] **Fazio, Giorgio and Modica, Marco**, *"Pareto or Log-Normal? Best Fit and Truncationin the Distribution of all Cities"*, Journal of Regional Sciences, 2015, 55, pp. 736-756

[31] **Berry, Brian J.L. and Okulicz-Kozaryn, Adam**, *"An Urban - Rural Happiness Gradient"*, Urban Geography, 2001, 32, pp. 871-883

[32] **Cordoba, Juan Carlos**, *"A Generalized Gibrat's Law"*, International Economic Review, 2008, 49, pp. 1463-1468

[33] **Peri, Giovanni**, *"Young People, Skills and Cities"*, CESifo working papers, 2001, 610, CESifo Group Munich

[34] **Hunt, Gary L.**, *"Population-Employment Models: Stationarity, Cointegration and Dynamic Adjustment"*, Journal of Regional Sciences, 2006, 46, pp. 205-244

[35] **Bures, Regina M.**, *"Migration and the Life Course: is there a retirement transition?"*, International Journal of Population Geography, 1998, 3, pp. 109-119

[36] **Frey, William H. and Speare, Alden Jr**, *"Regional and Metropolitan Growth and Decline in the United States: A 1980 Census Monograph"*, New YOrk Russell Sage Foundation, 1988

[37] **Longino, Charles F., Jr, Biggar, J. C., Flynn, C. B., and Wiseman, R. F.**, *"The retirement migration project (Final report to the National Institute on Aging)"*, Coral Gables, FL: University of Miami, Center for Social Research on Aging, 1984

[38] **Scott, Allen J. and Storper, Michael**, *"Regions, Globalization, Development"*, Regional Studies, 2003, 37, pp. 579-593

[39] **EuroStat**, *"Being young in Europe today - demographic trends"*, EuroStat - Statistics Explained, 2015

[40] **Santarelli, Enrico, Audretsch, David B., Klomp, Luuk, and Thurik, Roy A.**, *"Gibrat's Law: Are the Services Different?"*, Review of Industrial Organization, 2004, 24, pp. 301-324

[41] **Chesher, Andrew**, *"Testing the Law of Proportionate Effect"*, Journal of Industrial Economics, 1979, 27, pp. 403-411

[42] **Mansfield, Edwin**, *"Entry, Gibrat's Law, Innovation, and the Growth of Firms"*, American Economic Review, 1962, *52*, pp. 1023-1051

[43] **Black, Duncan and Henderson, Vernon**, *"Urban Evolution in the USA"*, Journal of Economic Geography, 2003, 3, pp. 343-372

[44] **Brakman, Steven, Garretsen, Harry and Schramm, Marc**, *"The strategic Bombing of German Cities during WWII and its impact on City Growth"*, Journal of Economic Growth, 2004, 4, pp. 201-218

[45] **Nijkamp, Peter and Reggiani, Aura**, *"Did Zipf Anticipate Spatial Connectivity Structures?"*, Environment and Planning B: Planning and Design, 2015, 42, pp. 468-489

[46] **Barabasi, Albert L. and Oltvai, Zoltan N.**, *"Network Biology: Understanding the Cell's Functional Organization"*, Nature Review Genetics, 2004, 5, pp. 101-113

[47] **Glaeser, Edward L., Ponzetto, Giacomo A. M. and Tobio, Kristina**, *"Cities, Skills and Regional Chance"*, Regional Studies, 2014, 48, pp. 7-43

[48] **Gonzalez-Val, Rafael**, *"A Nonparametric Estimation of the Local Zipf Exponent for all US Cities"*, Environment and Planning B: Planning and Design, 2012, 39, pp. 1119-1130

[49] **Guerin-Pace, France**, *"Rank-Size Distribution and the Process of Urban Growth"*, Urban Studies, 1995, 32, pp. 551-562

[50] **Storper, Michael**, *"Agglomeration, Trade, and Spatial Development: Bringing Dynamics Back in"*, Journal of Regional Science, 2010, 50, pp. 313-342

[51] **Giesen, Kristian and Suedekum, Jens**, *"Zipf's Law for Cities in the Regions and the Country"*, Journal of Economic Geography, 2011, 11, pp. 667-686

[52] **Gonzalez-Val, Rafael, Lanaspa, Luis and Sanz, Fernando**, *"Gibrat's Law for Cities, growth regression and sample size"*, Economic Letters, 2013, 118, pp. 367-369

[53] **Ioannides, Yannis M. and Skouras, Spyros**, *"US city size distribution : Robustly Pareto, but only in the tail"*, Journal of Urban Economics, 2013, 73, pp. 18-29

[54] **Gabaix, Xavier and Ibragimov, Rustan**, *"Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents"*, Journal of Business and Economic Statistics, 2011, 29, pp. 24-39

[55] **Cameron, Colin and Trivedi, Pravin K.**, *"Microecnometrics: Methods and Applications"*, Cambridge University Press, 2005

[56] **Nadaraya, E. A**, *"On estimating regression"*, Theory of probability & its applications, 1964, 10, 186-190

[57] **Watson, G. S.**, *"Smooth regression analysis"*, Sankhya Series A, 1964, 26, 359-372

[58] **Cordoba, Juan Carlos**, *"On the Distribution of City Sizes"*, Journal of Urban Economics, 2003, 63, 177-197

[59] **Borghesi, Simone, Montini, Massimiliano and Barreca, Alessandra**, *"The European Emission Trading System and Its Followers. Comparative Analysis and Linking Perspectives."*, SpringerBriefs in Environmental Science, Springer, 2016

[60] **Martin, Ralf, Muuls, Mirabelle and Wagner, Ulrich J.**, *The Impact of the European Union Emissions Trading Scheme on Regulated Firms: What Is the Evidence after Ten Years?*, Review of Environmental Economics and Policy, 2015, 10 (1), pp. 129-148.

[61] **Calel, Rafael, Dechezleprêtre, Antoine**, *"Environmental policy and directed technological change: Evidence from European carbon market"*, Review of Economics and Statistics, 2016, forthcoming.

[62] **Coase, Ronald H.**, *"The problem of social cost"*, Journal of Social Economics, 1960, 3, pp. 1-44

[63] **Martin, Ralf, Muuls, Mirabelle, de Preux, Laure B. and Wagner, Ulrich J.**, *"Industry compensation under relocation risk: A firm-level analysis of the EU Emissions Trading Scheme"*, American Economic Review, 2014, 104(8), pp. 2482-2508.

[64] **Hahn, Robert W. and Stavins, Robert N.**, *"The Effect of Allowance Allocations on Cap-and-Trade System Performance"*, Journal of Law and Economics, 2011, 54, pp. 267-294

[65] **Hahn, Robert W.**, *"Market Power and Transferable Property Rights"*, Quarterly Journal of Economics, 1984, 99, pp. 753-765

[66] **Stavins, Robert N.**, *"Transaction Costs and Tradeable Permits"*, Journal of Environmental Economics and Management, 1995, 29, pp. 133-148.

[67] **Kahneman, Daniel, Knetsch, Jack L. and Thaler Richard H.**, *"Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias"*, Journal of Economic Perspectives, 1991, 5, pp. 193-206

[68] **Montgomery, David W.**, *"Markets in Licenses and Efficient Pollution Control Programs"*, Journal of Economic Theory, 1972, 5, pp. 395-418.

[69] **Kahneman, Daniel, Knetsch, Jack L. and Thaler Richard H.**, *"Experimental Tests of the Endowment Effect and the Coase Theorem"*, Journal of Political Economy, 1990, 98(6), pp. 1325-1348.

[70] **Murphy, James J. and Stranlund John K.**, *"A Laboratory Investigation of Compliance Behavior under Tradable Emissions Rights: Implications for Targeted Enforcement"* Journal of Environmental Economics and Management, 2007, 53, pp. 196-212.

[71] **Fowlie, Meredith and Perhoff, Jeffrey M.**, *"Distributing pollution rights in cap-and-trade programs: are outcomes independent of allocations?"*, Review of Eocnomics and Statistics, 2013, 95, pp. 1640-1652

[72] **Reguant, Mar and Ellerman, Denny**, *"Grandfathering and the Endowment Effect:An Assessment in the Context of the Spanish National Allocation Plan"* MIT CEEPR Working Paper, 2008, 08-018

[73] **Anger, Niels and Oberndorfer, Ulrich**, *"Firm performance and employment in the EU emissions trading scheme: An empirical assessment for Germany"* Energy Policy, 2008, 36(1), pp. 12–22.

[74] **Abrell, Jan, Ndoye, Anta and Zachmann, Georg**, *"Assessing the impact of the EU ETS using firm level data"* Bruegel Working Paper 2011/08, Brussels, Belgium, 2011

[75] **Bushnell, James B., Chong, Howard and Mansur, Erin T.**, *"Profiting from Regulation: An Event Study of the European Carbon Market"* Dartmouth College Working Paper, Hanover, NH, 2011

[76] **Chan Hei Sing, Li Shanjun and Zhang Fan**, *"Firm competitiveness and the European Union Emissions Trading Scheme"*, Energy Policy, 2013, 63, pp. 1056-1064

[77] **Commins, Nicola, Lyons, Sean, Schiffbauer, Marc and Tol, Richard S. J.**, *"Climate Policy and Corporate Behavior"*. The Energy Journal, 2011, 32(4), pp. 51-68

[78] **Petrick, Sebastian and Wagner, Ulrich J.**, *"The Impact of Carbon Trading on Industry: Evidence from German Manufacturing Firms"* Kiel Working Paper, 1912, 2014

[79] **Wagner, Ulrich J., Muuls, Mirabelle, Martin, Ralf and Colmer, Jonathan**, *"The causal effect of the European Union Emission Trading Scheme: Evidence from French manufacturing plants"*, Unpublished, 2013

[80] **Borghesi, Simone, Franco, Chiara and Marin, Giovanni**, *"Outward Foreign Direct Investments Patterns of Italian Firms in the EU ETS"*, SEEDS Working Paper Series, No 2016-01, 2015

[81] **Reinaud, Julien**, *"Industrial competitiveness Trading Scheme"*. IEA Information Paper, International Energy Agency, Paris, France, 2005

[82] **Demailly, Damien and Quirion, Philippe**, *"CO2 abatement, competitiveness and leakage in the European cement industry under the EU ETS: grandfathering versus output-based allocation"* Climate Policy, 2006, 1, pp. 93–113.

[83] **Demailly, Damien and Quirion, Philippe**, *"European Emission Trading Scheme and competitiveness: A case study on the iron and steel industry"* Energy Economics, 2008, 30(4), pp. 2009–2027.

[84] **McKinsey and Ecofys**, *"EU ETS REVIEW Report on International Competitiveness"*, Tech. rep., European Commission Directorate General for Environment, Brussels, 2006a

[85] **McKinsey and Ecofys**, *"Review of Emissions Trading Scheme: Survey highlights"*. Tech. rep., European Commission Directorate General for Environment, Brussels, 2006b

[86] **Smale, Robin, Hartley, Murray, Hepburn, Cameron J., Ward, John and Grubb, Michael**, *"The impact of CO2 emissions trading on firm profits and market prices"* Climate Policy, (2006), 6(1), pp. 31–48.

[87] **Zaklan, Aleksandar**, *"Free allocation and the endowment effect in cap-and-trade systems: evidence from the European electricity sector"* EAERE Conference 2016, Zurich, 2016

[88] **Greenstone, Michael**, *"The Impact of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures"*, Journal of Political Economy, 2002, 110(6), pp. 1175-1219

[89] **Levinson, Arik**, *"Environmental Regulations and Manufacturers' Location Choices: Evidence from the Census of Manufactures"* Journal of Public Economics, 1996, 62(1-2), pp. 5-29