

Modica, Marco

Conference Paper

Does the EU have homogeneous urban structure area? The role of agglomeration and the impact of shocks on urban structure

54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Modica, Marco (2014) : Does the EU have homogeneous urban structure area? The role of agglomeration and the impact of shocks on urban structure, 54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124250>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Does the EU have a homogeneous urban structure area? The role of agglomeration and the impact of (hypothetical) shocks on the EU urban structure

Marco Modica

CNR – CERIS, Via Bassini, 15, 20133, Milan, Italy. Tel: +39 0223699501; e-mail: <m.modica@ceris.cnr.it>

ABSTRACT. The creation of the European Union, the distortions caused by the introduction of a single currency in countries structurally so different and the expansion of mobility of people, capital and services due to the constitution of the so-called Schengen Area from the beginning of '90s might have had some impacts on the dynamics of city populations. This paper provides a study of the hierarchical structure of the cities within the EU Member States with particular attention on agglomeration forces by means of two very well-known empirical regularities: Zipf's law, as a proxy for agglomeration forces, and Gibrat's law as a test for stationarity. We find that the hierarchical structures of Member States is more even than expected. Moreover, the European Union seems to be split in three distinct areas: an area characterized by the validity of Gibrat's law (temporary idiosyncratic shocks might have permanent impacts on the city structure); an area characterized by the presence of mean reversion (any exogenous shock is used up in certain amount of time); a small area where the effects of the shocks is magnified in the big cities. Finally, we find that only the constitution of the Schengen Area and the share of international trade seem to have a weak impact on the hierarchical structures of Member States.

Keywords: City size distributions; European city growth; Zipf's Law; Gibrat's Law

JEL classification: C46, D30, R11, R12

1 Introduction

The urban structures between the Member States of the European Union is very different for historical, geographical, economic reasons (Aldskogius, 2000). However, the population is spread across geographic areas in a way that, although continuously changing, is not possible to define as random. Indeed, countries have faced a strong tendency toward agglomeration, namely population gathers within proper areas like cities, and currently the agglomeration within cities “*is an extremely complex amalgam of incentives and actions taken by millions of individuals, businesses, and organizations*” Eeckhout (2004, p. 1429). Table 1 shows selected characteristics of the Member States in 2011 and gives an interesting picture of the differences between countries in term of urban structure.

Total population and land size vary consistently among countries within the EU, however Member States show big differences also in term of population density, percentage of rural population, people who live in the largest city and also in other variables as agricultural land, road density and rail density.

As an example, Belgium has a population density of 364 people per sq. km and it also shows a low percentage of rural population (2.5 over the entire population) but it presents an agricultural land equal to the 45% of the entire surface. On the contrary, the Netherlands, a country with almost same land size as Belgium, shows a higher population density than Belgium but also a higher rural population and a higher percentage of agricultural land, indicating a different allocation of people between urban and rural areas. This fact is also underlined by the amount of population living in the largest city: 18% for Belgium (Anvers) and 7.7 % for the Netherlands (Amsterdam).

< **Table 1 about here** >

This first rough picture of the differences of urban structures among countries might denote different underlying economic forces of EU's Member States. For instance, the degree of urbanization is strictly related to differences in regional per capita income (Mera, 1975); odds in the population density can lead to different conditions in the labor market (Armington and Acs, 2002) and/or investments in infrastructure (Fay and Yepes, 2003; Randolph et al., 1999); the concentration in the capital cities might denote less democratic societies (Henderson and Wang, 2007).

On the other way round, the impact of a given policy can lead to a change in the urban structure. For instance, Yang (1999) shows that, among the others, urban policies, investments in the urban sector, institutions and financial policies are responsible for the long-term income differential between urban and rural areas. Moreover, urban places expand their influence on a much larger hinterland than rural areas and the difference between urban and rural places goes behind the simple difference in population size, but also in a different concentration of economic activities, people and cultures. (Pacione, 2001).

All these facts underline that the European Union is still a multifaceted entity, at least in terms of urban structure, where the harmonization between countries is still going on and, given those stylized facts, we can ask whether the deeper and deeper integration of European Union affected the hierarchical structure of the city system of the Member States.

Then the creation of the European Union, the distortions caused by the introduction of a single currency in countries structurally very different and the expansion of mobility of people, capital and services due to the constitution of the so-called Schengen Area from the beginning of '90s, with the intent to build a common area, might have had some impacts on the dynamics of city populations.

In this paper we use an alternative approach to examine whether, i) the EU affects the city system of the Member States and ii) EU city system can be seen as an area where exogenous shocks, (i.e. extreme weather conditions, financial crises, inflation) affect the population growth in the same way. Specifically, we use, in an alternative way, two very well known empirical regularities, Zipf's law (Auerback, 1915; Zipf, 1949) and Gibrat's law (Gibrat, 1931). Alternative means that aim of this paper is not to verify if the former laws hold but it means that we use them as a tools to explore the research questions i) and ii).

Zipf's law predicts the degree of hierarchization of a system of cities and it can be seen as an indicator of the strength of the agglomeration forces in a system of cities (Brakman et al., 2001; Singer, 1930). Gibrat's Law suggests that the growth of a city is independent of size, a condition that would hold if temporary shock has a permanent impact on the pattern of growth of the cities (Brakman et al., 2004). The present paper then, takes the moves from the huge literature on Zipf's and Gibrat's law. In particular, this work has its roots on the research of Rosen and Resnick (1989) and Soo (2005) for the part related to Zipf's law and on the works of Brakman et al. (2004), Gabaix (1999) and Giesen and Suedekum (2012) for the part related to Gibrat's law.

Rosen and Resnick (1989) provide an international analysis of Zipf's law over 44 countries. They found the validity of Zipf's law in most of the cases and moreover they provide empirical evidence that the Zipf exponent is positively influenced by per capita GNP, total population and railroad density, and negatively related to land area.¹ Soo (2005) updates these results but concludes for a rejection of Zipf's Law in more than half of cases. Moreover he found that political variables matter more than economic variables in determining the size distribution of cities. In details the total government expenditure and war dummy enter with a very strong positive coefficient. Instead, among economic variables, the degree of scale economies seems to significantly and positively influence the Zipf coefficient, whereas trade as a percentage of GDP weakly and negatively influence it.

Gabaix (1999) and Giesen and Suedekum (2011) address the same issue, i.e. if urban growth in all regions follows Gibrat's law, then we should observe the Zipfian rank-size rule among large cities both at the regional and national level. Gabaix (1999) provides theoretical rationale while Giesen and Suedekum (2011) provide empirical evidence. Finally, Brakman et al. (2004) analyze whether war shock has an impact on post-war German city growth finding that war shock has only a temporary impact on it since German city growth faces a tendency toward mean reversion.

¹Positively means that greater per capita GNP, total population and railroad density are associated with a more even distribution of cities. Instead, negatively means that greater land area is associated with a more agglomerated distribution of cities.

The present paper, then, sets out to do three things: the first is to provide an accurate description of agglomeration. This involves a static analysis, supported by Zipf's law, accounting for the way the population is gathered over different geographic locations. The second is to provide an accurate description of population mobility. This involves a dynamic analysis, supported by Gibrat's law, accounting for the evolution over time of the distribution of the population over different locations. Here we are able to determine if a temporary shock has a permanent impact on the pattern of city growth. Finally, once population mobility is understood, we explore the impact of the underlying economic mechanism and of the constitution of the Euro and the Schengen Area on agglomeration forces.

We find that, generally, the hierarchical structures of Member States of EU is more even than expected. Moreover, EU city system is still far from being an integrated area and in particular, we find that the European Union seems to be split in three distinct areas: an area characterized by the validity of Gibrat's law where large temporary shocks might have permanent impacts on the city structure; another area that is characterized by the presence of mean reversion and where any exogenous shock is used up in a certain amount of time and, finally, a small area where the effects of the shocks are magnified in the big cities. Finally, we find that only the constitution of the Schengen Area and the share of international trade seem to have a weak impact on the hierarchical structures of Member States presenting respectively a positive impact on the Zipf coefficient (more even distribution) and a negative impact (more agglomerated distribution) and then indicating two sources one against and one in favor of agglomeration forces. Those results are in line with previous findings of Soo (2005) and Alperovich (1993).

The next section outlines Zipf's law and Gibrat's law. Section 3 describes the data and Section 4 presents the results. Then Section 5 concludes.

2 Zipf's law, Gibrat's law: definitions, methodologies and economic interpretations

The Member States of the EU have a long history of conflict. Only in the last century Europe has been the main theater of the two World Wars. However, after WWII, European Countries started an integration project that led to the creation of European Union that, in some way, covered up old tensions (Nahoi, 2011). The creation of the European Union with free movement of goods, services, people and capital might have changed the population distribution across geographic areas. Nevertheless, as mentioned before, people tend to concentrate within common restricted areas like cities in a way that is not random (Eeckhout, 2004).

On this regard, Brakman, et al. (2001) affirm that a simple and plausible way to explain birth and growth of cities should be definitively linked to the existence of agglomeration and congestion forces. More specifically, following Kooij (1988), urbanization follows three distinct periods: Pre-industrialization, characterized by high transport cost that discourages agglomeration; Industrialization, where transportation cost starts to decrease and the importance of industrial plants becomes higher since an increases of return to scale; Post-industrialization, where congestion and declining importance of industrial production become established. Zipf’s law, although it doesn’t directly address the characterization presented by Kooij, is often described in term of tension between congestion and agglomeration forces, (see Gabaix and Ioannides, 2004; Sutton, 1997 for a survey) and it provides a proxy for urbanization.

In more detail, Zipf’s law is an empirical regularity stating the proportionality of the cities to their rank. This means that for example, in Italy, the size of Rome (the largest city in Italy) is roughly twice the size of Milan, the second largest city, three times the third largest city, Naples, and so on. Formally, this can be written as:

$$R_i = KP_i^{-q}, \tag{1}$$

Equation (1) is known as the rank-size rule and is usually expressed in logarithmic form, as follows:

$$\ln(R_i) = k - q \ln P_i, \tag{2}$$

where R_i is the rank of the cities assigned as 1 to the largest city, 2 the second largest city and N the smallest one, k is a (ln) constant, P_i is the population size and q is the parameter of interest, the so-called rank-size coefficient that here will be used as a proxy for agglomeration forces.²

Several interpretations of the rank-size coefficient, q , have been given in the literature. The reader can consult Modica et al. (2013) for a brief summary. In this paper we interpret the q -coefficient in the simplest way, that is as an indicator of the strength of the agglomeration forces in a system of cities (Singer, 1930; Brakman et al., 2001). On this vein, the rank-size coefficient might be considered an index of metropolisation, in the sense that the lower the q -coefficient, the more important is the urban land value.

²Notice that we estimate the rank-size coefficient using simple rank-size OLS regressions where, following Gabaix and Ibragimov (2011), the rank is shifted by 0.5 to correct for the potential bias in small samples highlighted by Gabaix and Ioannides (2004), so that the estimating equation is $\ln(R_i - 0.5) = k - q \ln P_i$.

In fact the q -coefficient measures how (un)even the city system is, namely this means that the higher the q -coefficient, the more equally distributed the city system (in the extreme, when $q=\infty$, the system of cities is very even: all cities of the same size; when $q=0$, instead, the system will be composed just by one city hosting the entire population).

In the above interpretation (i.e. the tension between agglomeration and congestion) of Zipf's law the size of cities matters, however, there is another way to explain Zipf's law in which the (relative) size of cities does not matter, this is the case of the use of Gibrat's law in explaining Zipf's law (Gabaix, 1999).

Gibrat's law is a rule stating that the (relative) growth of a given entity (city, firm, income and so on) does not depend on its size.³ This means that, albeit an entity can grow at different rates, it can't be found any systematic behavior between their growth and their size. Then, according to Gibrat (1931), we cannot affirm that larger entities grow faster than smaller ones or vice versa. Analytically, following Steindl (1968):

$$\log P(t) = \log P(0) + \varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(T), \quad (3)$$

where, $P(t)$ is the size of a certain entity at time t , $P(0)$ is the initial size, and $\varepsilon(t)$ is a random variable (indicating random shocks). Equation (3) defines the logarithm of the size of a given entity as the sum of the initial size and past growth rates. Moreover as Gabaix (1999) notices, Gibrat's law implies that the growth process of a given entity presents a common mean and a common variance, or, in other terms, mean and variance of the growth have to be independent from the size of the entity.

Several authors have proposed economic interpretations of Gibrat's law that differ only in some shades. For instance, Black and Henderson (2003), state that a shock affects in the same way big and small entities whereas, Brakman et al. (2004), 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. In this work, we adhere to the latter interpretation and then we will use Gibrat's law to test whether a (hypothetical) shock (will) influence the Member States of EU in the same way or, on the contrary, if it (will) has heterogeneous effects.

Several methodologies have been proposed to test Gibrat's law. However, it is commonly tested looking at the non-stationarity of the growth series. The rationale underlying this test is the following: if the series is stationary it will converge toward a constant mean and, therefore, large values (big entities) will be followed at any further periods by smaller and smaller values, while small values (small entities) will be followed by

³The reader can consult Modica et al. (2013) for a deep discussion on Gibrat's law and its relationship with Zipf's law.

larger and larger ones. In this way, the level of the series at any periods can significantly predict the one of the next period and then the growth of the entities are not independent on the initial state. On the other hand, if the series is integrated, then large values and small values will occur with probabilities that do not depend on the current size of the entities and then the growth of the entities are independent on the initial state.

Finally, it has to be noticed that the studies on Zipf's and Gibrat's law usually consider only the upper tail of the data, i.e. the largest cities, with a sample truncation point that is usually arbitrarily chosen. Here to avoid any sources of arbitrariness we use a method proposed by Clauset et al. (2009) and it will be deep presented in the next section.⁴

To sum, Zipf's law shows the static relationship of the size of a certain entity with its rank. Gibrat's law, instead, presents the dynamic growth process of this entity. We will use both the laws to show how the EU tends toward agglomeration and if it can be seen as a homogeneous area, but first we will introduce the data in the next section.

3 The data

In this paper, we collect city population data from the National Statistics Office of the Member States. The time windows vary according to the availability of the data from any Statistical Office, however when we present the cross-countries analysis we adopt a unique time span namely from 1991 to 2011. This is due because of two main reasons. First of all, the time period of interest for this work is the one covering the period of greater strength of integration among Member States, i.e. between 1991 and 2011. Second, that's the only time span that has observations for all the 27 Member States.

One of the main concern in building the data-set is that on the definition of agglomeration: how many inhabitants are necessary before a given agglomeration can be defined as "urban"? Across countries this definition varies substantially and also within the same countries this definition can change from time to time (United Nations, 1974). Moreover as reported in Manual VIII (United Nations, 1974) it is very difficult if not impossible to prescribe a uniform limit for all countries because qualitative characteristics or even residential densities in settlements of a given size can vary significantly among the countries.

However, European countries are trying to adopt a 'common' definition by which units with 10,000 or more inhabitants are considered as "urban", those with 2,000-9,999 inhabitants as "semi-urban", and those

⁴For a thorough analysis of this issue the reader can consult Gabaix (1999), Fazio and Modica (2012) and Ioannides and Skouras (2013).

with less than 2,000 as "rural". Despite this definition, still a considerable number of Member States use different thresholds and/or rules aiming to discriminate what is rural and what is urban. As an example, Table 6 of the Demographic Yearbook 2009-2010 of the United Nations reports the definition of "urban" across countries over the world. It can be noticed that between EU Member States the definition of urban is not homogenous, i.e. Austria defines urban areas all the localities with a population above 2,000; France adopts a population threshold of 10,000 and Bulgaria gives just a legal definition of what might be considered urban (United Nations, 2011, table 6). Furthermore, the method used to define the borders of the statistical units can change dramatically and, often, the definition of cities changes in relation to the official definition of city boundaries given by the statistical authorities (Soo, 2005). As an example the UK administrative geography is organized as follows: Country -> Region -> County/Unitary Authority -> Local Authority -> Ward and - some case Parish (mainly rural areas). Italy, instead, is based on a legal administrative division still strictly related to historical subdivision in State -> Region -> Province -> Municipality. It should be noticed that UK Local Authority is something slightly different from an Italian Municipality, since the latter, typically, do not include surrounding suburbs, i.e. the places where the workers of a city reside are considered as other municipalities.

Moreover, the debate on the unit of observations is still alive. Since a long time, it has been noticed that defining the unit of study as urban places, legal cities or urban agglomeration may affect the estimates of the rank-size coefficient (Cheshire, 1999; Rosen and Resnick, 1980). On this regard, the literature proposed several solutions: metropolitan areas (among the others Ioannides and Overman, 2003; Dobkins and Ioannides, 2000); economic areas (Berry and Okulicz-Kozaryn, 2012); clustered populated areas (Rozenfeld et al., 2011) and, finally, natural cities (Jing and Jia, 2010).

Due to the availability of data, to overcome this issue we choose the most homogenous unit of observations among Member States we have found, namely "locality" that is the smallest administrative level of aggregation available for any country.⁵

Another main concern related to the data, already mentioned in the previous section, is the choice of the appropriate truncation point for the upper tail. The methods generally adopted on this regards are three: the choice of a fixed number of cities between countries (i.e. the 100 larger cities); the choice of a fixed population threshold (i.e. all the cities with a population above 100,000); all the cities for which the sum

⁵It should be noted that EUROSTAT provides "Nomenclature of territorial units for statistics" (hereinafter referred to as NUTS). Although it might be an alternative and perhaps better measure of agglomeration, it is still based on administrative units within the Member States, where the only discriminant is based on the population size, i.e. a minimum of 150,000 and a maximum of 800,000 inhabitants.

of the population accounts for some given proportion of a country’s population (i.e. all the largest cities for which the sum of the population accounts the 10% of the country’s population), Cheshire (1999).

However, all these proposed methodologies discount a certain amount of arbitrariness (Fazio and Modica 2012; Ioannides and Skouras, 2013). In order to avoid this arbitrariness, we exploit an alternative non-parametric methodology to estimate the appropriate minimum threshold, firstly proposed by Clauset et al. (2009).

The rationale behind this methodology lies on the idea that Zipf’s Law should led to a Pareto distribution with a shape parameter (that is our q) equal to 1 (Eeckhout, 2004). Then, Clauset et al. (2009) propose to test the equality between the theoretical and empirical density functions using a Kolmogorov-Smirnov tests over recursively truncated distributions and choosing the sample showing the best fit to a Pareto distribution (the minimum of the statistics D below), indicating the sample of cities where agglomeration forces are stronger (Fazio and Modica, 2012).

In sum, our estimate population threshold, \hat{P} , is the value of population, P , that minimizes the Kolmogorov-Smirnov statistics, D :

$$D = \sup_{p \geq \underline{p}} |\Phi_p(x) - \Phi(x)|, \tag{4}$$

where $\Phi_p(x)$ is the empirical cumulative density function for any city of population p above \underline{p} , and $\Phi(x)$ is the theoretical cumulative density function of the Pareto distribution. The KS statistic just computes the supremum of the absolute value of the distances between the two. Under the null, the distance between the two is zero and then the sample is drawn from the reference distribution, that in our case is a Pareto.

In this way we are able to collect a unique data-set avoiding some of the common issues, namely we identify an upper-tail strictly related to the agglomeration characteristics of the Member States allowing homogeneity in the data. Table 2 summarizes the main characteristics of the data-set.

< **Table 2 about here** >

Data for the second stage regression, which seeks to find the impact of the constitution of the European Union on the city hierarchical structures, are obtained from the World Bank World Development Indicators, the UNIDO Industrial Statistics Database and EUROSTAT.⁶ In more details, the variable road density is constructed using EUROSTAT data on the length of the roads. GDP per capita in constant US dollars, total land, total population, trade as a percentage of GDP, and number of internet users are from the World

⁶In more details, the variable road density is constructed using EUROSTAT data on the length of the roads. The data for public expenditure are those from the World Bank World Development Indicators with the exception of year 2011. Those data come from the original source that is the single national statistical offices.

Bank World Development Indicators. The data for public expenditure are those from the World Bank World Development Indicators with the exception of year 2011 (still not available) and then they are retrieved from the original source, namely the single national statistical offices. Scale economies are constructed starting from information presented in the UNIDO Industrial Statistics Database.

More details on the data used in the second stage regressions will be presented in the next section as well as the results.

4 Results

In this section we present the results. First we provide an accurate description of agglomeration. This involves a static analysis, supported by Zipf's law, accounting for the way the population is gathered over different geographic locations. Second, we move on the analysis of the homogenous/heterogeneous areas within the EU and we provide the dynamic part of the story. This second part provides the dynamic analysis, supported by Gibrat's law, accounting for the evolution over time of the distribution of the population over different locations. Here we are able to determine if temporary shock has a permanent impact on the pattern of city growths. Finally, once population mobility is understood, we explore the impact of the underlying economic mechanism and of the constitution of the EU, the Euro and the Schengen Area on the agglomeration forces.

4.1 Hierarchical structure and its determinants

In this section, we discuss the results from the following equation (rank-size rule with a Gabaix and Ibragimov, 2011 modification):

$$\ln(R_i - 0.5) = k - q \ln P, \tag{5}$$

where k is a constant and P is the population size. Standard errors are given by $(2/n)^{0.5} \hat{q}$. The parameter \hat{q} is the rank-size coefficient that will be used in the second step. Table 3a and Table 3b present the detailed results of the OLS regressions of Eq. (5). We show just the first and the latter results for any Member States, however full details are available from the author upon request.

The rank-size coefficient, q , is an index of metropolization, in the sense that the lower the q -coefficient, the more important the urban land value. In fact the q -coefficient measures how (un)even the city system

is, namely this means that the higher the q -coefficient, the more equally distributed the city system (in the extreme, when $q=\infty$, the system of cities is very even, all cities have the same size; when $q=0$, instead, the system will be composed just by one city hosting the entire population. Finally, when $q=1$ the city system is said to obey Zipf's law).

< **Table 3a about here** >

< **Table 3b about here** >

The largest value of the rank-size coefficient in 2011 (4.049) is obtained for Malta, followed by Belgium (1.724) results in line with those found by Soo (2005), whereas the lowest value is obtained for Denmark (0.949).

In more details, the rank-size coefficient, q , is significantly greater than 1 for 13 of our 27 countries, while a further 14 observations are significantly not different from 1. These results indicate a situation where the half (in general small countries, both in terms of economic and demographic characteristics) of the EU Member States presents a hierarchical structure strictly following a Zipf's distribution (i.e. the largest city is double of the second largest city, three times of the third and so on) and the other half presents a more even distribution. Indeed, this means that for 13 countries agglomeration forces are stronger than the other 14 Member States (or on the same way, congestion forces are stronger in the other 14 countries), indicating a straight hierarchy. Notice that there is no cases where the rank-size coefficient is lower than 1.

These results are stronger than those found by Soo(2005) who presents 17 Member States with a coefficient significantly greater than 1 and two Member States with a coefficient not different from 1.⁷ The differences in the results, in our opinion, come from two sources: first of all, the data-set in the present paper is shaped to take directly in consideration agglomeration forces using the Clauset et al. (2009) method. This method is able to recognize the cut-off where agglomeration forces are stronger and this can lead to an increase of the cases where the Zipf's law strictly holds. Secondly, we use a modified rank-size rule that solves the problem of the downward biasness in small samples (Gabaix and Ibragimov, 2011).

To sum up, this first picture presents a situation where the European Union is exactly split into two halves. Small countries (in terms of lands) as Cyprus, Denmark, Luxembourg, Netherlands and Portugal, or transition countries as Bulgaria, Czech Republic, Estonia and Hungary are experimenting strong forces toward agglomeration and present a highly hierarchical city system. Instead, big (in terms of land size) countries as Poland, Romania and developed (in economic terms) countries as Belgium, France, Germany,

⁷In Soo(2005) Cyprus, Estonia, Ireland, Latvia, Lithuania, Luxembourg, Malta, Slovenia are not in the international analysis

Italy, Spain and the United Kingdom present a more even distribution of cities where the congestion forces play an important role.

Finally, considering the EU as a whole (country), the estimated rank-size coefficient is 1.40 indicating a quite even city distribution. It is interesting to compare this result with the estimated q coefficient of the aggregation of countries using the Euro. The estimated coefficient is almost the same (1.38), although it should be noticed that the estimation of the upper tail with the method proposed by Clauset et al. (2009) returns a longer upper tail for EMU, 2660, than EU, 1419.

Until now we have presented just a static draw of the European situation. We will explore the impact of the underlying economic mechanism and of the constitution of the EU, the Euro and the Schengen Area on the agglomeration forces in the sub-section 4.3. In the next section, instead, we address the homogeneity issue.

4.2 Is the EU a homogenous Area? The role of hypothetical shocks

In this section we analyze the homogeneity/heterogeneity of city systems in term of growth within the EU using Gibrat's law. Gibrat's law in assessing heterogeneity is not conventional, however Black and Henderson (2003), state that if Gibrat's law holds, then a shock affects in the same way big vs small entities and, Brakman et al. (2004), state that if Gibrat's law holds, then a large temporary shock can have a permanent impact. In other words, this means that hypothetical shocks can change (permanently) the growth path toward another size equilibrium.

In general we can interpret Gibrat's law as follows: "*A variate subject to a process of change is said to obey the law of proportionate effect if the change in the variate at any step of the process is a random proportion of the previous value of the variate*" (Chesher, 1979, p. 403). Given this definition, Gabaix (1999) affirms that a growth process that has a common mean and a common variance follows Gibrat's law or, in other words, that both the mean and variance have to be independent from the initial city size.

Moreover, proposition 2 of Gabaix (1999) affirms that if a country is composed of several regions and Gibrat's law holds in each of those regions, then in the whole country Zipf's law is satisfied both at regional and national level. This means that if the growth processes are (in the upper tail) identical within each region, but not necessarily across regions, then in the whole country the strength of the agglomeration forces in the system of cities is necessarily the same. This issue has been firstly addressed empirically by Giesen and Suedekum (2011) for Germany. In this section, we take the moves from those papers and we provide a country-wide test of Gibrat's law from a EU Member States' perspective.

Operationally, a typical way to assess the validity of Gibrat’s law is using non-parametric analysis (Eeckhout, 2004; Giesen and Suedekum, 2011; Gonzalez-val et al., 2012; Ioannides and Overman, 2003; Modica et al., 2013). In doing that, we use the normalized growth rate, that is the difference between a city’s growth rate and the mean city growth rate, all divided by the standard deviation of growth.

The strength of non-parametric analysis is that we do not impose any relationship between the dependent and independent variables and, moreover, we can see the behavior of mean and variance at any possible truncation, avoiding in this way the concern about the right selection of the truncation point arose by Fazio and Modica (2012). This is also stressed by the following sentence in Cameron and Trivedi (2005, p. 294), “[non-parametric analysis] *let the data show the shape of the relationship*”. In fact, the standard parametric regression methods provide an aggregate relationship between growth and size that holds over the entire support of city sizes. Instead, the non-parametric method allow the growth to vary with size over the support.

In this analysis, we will use the Nadaraya-Watson (NW) method (Nadarya 1964; Watson 1964), where the bandwidths are calculated with an optimal rule of thumb.⁸ If Gibrat’s law holds, the NW method provides stable estimated conditional mean and variance across different population sizes, that on the light of the normalization are respectively 0 and 1. We also calculate 5% bootstrapped confidence interval based on 500 samples (see Hardle, 1990). If Gibrat’s law does not hold, instead, the series of growth rates is stationary and it will converge toward a constant mean. Therefore, large values (big entities) will be followed at any further periods by smaller and smaller values, while small values (small entities) will be followed by larger and larger ones. In this way, the level of the series at any periods can significantly predict the one of the next period.

The figures below provide the non-parametric estimates of conditional mean and variance for all the Member States.

< **Figure 1a about here** >

< **Figure 1b about here** >

< **Figure 1c about here** >

Figures 1a-1c show that, although the EU is still far from being a homegenous area (in terms of city growth), indeed we are able to recognize three different areas. One where exogenous shocks might affect the city structure in the same way (Gibrat’s law holds). In details, this area is composed by the following

⁸We refer readers to their papers for a more detailed description.

Member States: Belgium, Cyprus, Denmark, Estonia, Finland, France, Germany, Ireland, Lithuania, Malta. In all these cases, indeed, the conditional mean and the conditional variance look independent from the city size.

In the rest of the countries Gibrat's law does not hold and in particular the growth series presents mean reversion and stationarity namely big cities are expected to grow less than small ones (Black and Hendreson, 2003) and this fact is characterized by a conditional mean growth and variance that are respectively lower than 0 and 1 for larger size of cities. Those countries are Austria, Czech, Greece, Hungary, Italy, Luxembourg, Poland, Portugal, Slovakia, Slovenia, Spain and UK, and all present mean reversion in the growth process. In this situation, then, any exogenous shock has to be thought as temporary and the countries will converge toward the pre-shock conditions in a certain period of time.

Finally, very few countries show a situation where there exists an "explosive" growth process, that is a situation in which big cities are expected to grow more than small ones. In more details those Member States are Bulgaria, Latvia and Sweden. They present a situation where people tend to moves from the small city to the big ones and it is shown by a conditional mean growth and variance that are respectively greater than 0 and 1 for larger size of cities.⁹

To sum, the European Union seems to be split in three distinct areas: an area characterized by the validity of Gibrat's law where large temporary shocks might have permanent impacts on the city structure; another area that is characterized by the presence of mean reversion and where any exogenous shock is used up in certain amount of time and a small area where the effects of the shocks is magnified in the big cities.

Given the static and dynamic picture of the European Union, it remains to explore the impact of the underlying economic mechanism and of the constitution of the EU, the Euro and the Schengen Area on the agglomeration forces. This is studied in deep in the next sub-section.

4.3 The impact of economic/institutional variables on agglomeration forces

In this sub-section we try to answer to the following question: does the creation of the European Union with free movement of goods, services people and capital, change the population distribution across geographic areas? The rank-size coefficient, indeed, can be seen as a measure of urbanization: the larger the value of the coefficient, the more even the population of cities in the urban system.

There are several potential explanations for variations in its value, one of these can be found in a model of economic geography a la Krugman (1991) and Fujita et al. (1999). These models can be viewed as models

⁹Unfortunately we don't have enough data to explore Gibrat's law in the Netherlands and Romania

of unevenness in the distribution of economic activity and moreover, they state that for certain parameter values, economic activity is agglomerated, while for other parameter values, economic activity is dispersed (i.e. a city system will be more agglomerated the greater are scale economies, the lower are transport costs and the lower the share of international trade in the economy). Henderson and Wang (2007), instead, emphasize on the role of the institutions. The constitution of the EU, then, might had affected the urban structure of the Member States.

In this section we look for both the impact of institutional and economic variables on the level of the hierarchical structure. We also control for other variables that could influence the size distribution of cities, including the size of the country as measured by population, land area or GDP. Following Soo (2005) our estimated equation is:

$$q_{it} = \beta_0 + \beta_1 CONTROL + \beta_2 ECON + \beta_3 POLITIC + \beta_4 DUMMY + u_{it}$$

where q is the rank-size coefficient estimated in the previous step, CONTROL is the vector of controls for the size of the country, including the log of per capita GDP in constant US dollars, the log of the total land of the country, and the log of population. ECON is a group of economic-geography and connectivity variables: scale economies, transport costs (the inverse of road density), trade as a percentage of GDP, number of internet users per 100 people.

Scale economies is the degree of scale economies, and it is constructed following Soo (2005). They are the share of industrial output in high-scale industries. Transport cost is measured as the inverse of road density.

POLITIC is a set of political variables as government expenditure as a share of GDP in education and foreign direct investment as net inflow (% of GDP). Finally, DUMMY indicates the variables strictly related to the creation of European Union (belonging to EU, adhesion to Schengen and interaction between the two).

One potential concern, stressed by Soo (2005), is the fact that using an estimated coefficient from a former stage regression as a dependent variable in a subsequent stage regression it might lead to inefficient estimates and heteroskedasticity in the second stage because of likely measurement error in the first stage, Lewis and Linzer (2005). We could use feasible generalized least squares (FGLS) to overcome this problem. However, Baltagi (1995) and Beck and Katz (1995) show that: i) FGLS yields consistent estimates of the variances only if T goes to infinity and, ii) FGLS tends to underestimate standard errors.

In the present paper, following Soo (2005), we use, first, panel corrected standard errors with OLS, as proposed by Beck and Katz (1995) because it does not underestimate standard errors and second we run time fixed effect and random effect estimations. The regressions are those that are reported below:

< Table 4 about here >

Table 4 presents the results using the OLS estimate of the Pareto exponent as the dependent variable. The number of observations is somewhat less than the full sample because data is not available for all countries in all years. Column (1) is the panel corrected standard errors with OLS model without dummy variables. Column (2) includes all the variables and Column (3) and (4) are, respectively, time fixed effect and random effect models. Two factors seem to have a significant impact on the city hierarchical structures of the countries: the extent of international trade and the adhesion to the Schengen Area. Moreover, population and land size respectively positively and negatively influence the city structure of the Member States.

In more detail international trade enters with a (small) negative impact on the rank-size coefficient, that is something not predicted from the theory since, following Fujita et al. (1999), a greater extent of international trade weakens the force for agglomeration and leads to a more even distribution of economic activity. However this result is in line with previous findings of Soo (2005) and Alperovich (1993).

Column (4) instead shows that the adhesion to the Schengen Area leads to a more even distribution (higher rank-size coefficient). This result can be interpreted as the results of the increased freedom of mobility between Member States that leads to the movements from the big cities (especially in the transition countries) to the small cities.

Comparing our results to previous findings, we find that our results are broadly in line with those of Rosen and Resnick (1980) as they find that the rank-size coefficient is positively related to the total population and negatively related to land area. However, we found lack of significance for most of the economic and political variables presented by Soo (2005).

In conclusion, it seems that the constitution of the European Union *per se* does not have any influence on the change of the population distribution across geographic areas. The only variables that have influenced the city structure of the Member States are those strictly related the increases of the mobility of people, capitals and services (the constitution of the Schengen area) and a variable closed to the idea of globalization, that is the share of international trade.

5 Conclusion

The aim of our research work was to draw the current city system of the Member States of the European Union firstly for any single country and subsequently as a whole state. Given this picture we aim to explore if and how the creation of European Union affects the structure of the system of cities of the Member States

and primarily if EU city system can be seen as an integrated area. The main objective was the study of the agglomeration forces within all the Member States and the EU as a whole and for this reason we used two very well-known empirical regularities that address (indirectly) this issue, namely Zipf's law and Gibrat's law.

For this purpose we used the rank-size exponent, q , as a proxy for agglomeration forces in a system of cities (Singer, 1930; Brakman et al., 2001) since the lower the q -coefficient, the more agglomerated is the city distribution (Soo, 2005). Gibrat's law, instead, is used as a tool to test whether a (hypothetical) shock (will) influence the Member States of EU in the same way or, on the contrary, if it (will) has heterogeneous effects. Indeed, if Gibrat's law holds a large temporary shock might have a permanent and similar impact on the growth path of the cities, this means that a shock can change the growth path toward another size equilibrium.

We started our analysis providing an accurate description of agglomeration. This has involved a static analysis, supported by Zipf's law, accounting for the way the population is gathered over different geographic locations and, a dynamic analysis, supported by Gibrat's law, accounting for the evolution over time of the distribution of the population over different locations.

We showed that the hierarchical structures of Member States of EU is more agglomerated than expected. In more details, the rank-size coefficient, q , is significantly greater than 1 for 13 of our 27 countries, while a further 14 observations are significantly not different from 1. These results indicate a situation where the half of the Member States of EU presents a hierarchical structure strictly following a Zipf's distribution (i.e. the largest city is double of the second largest city, three times of the third and so on) and the other half presents a more even distribution. Indeed, this means that for 13 countries agglomeration forces are stronger than the other 14 Member States (or on the same way, congestion forces are stronger in the other 14 countries), indicating a straight hierarchy. Notice that there is no cases where the rank-size coefficient is lower than 1, so that we did not find any evidence of countries strongly agglomerated.

Given this picture, the dynamic analysis, supported by Gibrat's law, showed that EU city system is still far from being a homeogenous area and in particular, we found that the European Union seems to be split in three distinct areas: an area characterized by the validity of Gibrat's law where large temporary shocks might have permanent impacts on the city structure; another area that is characterized by the presence of mean reversion and where any exogenous shock is used up in certain amount of time and a small area where the effects of the shocks is magnified in the big cities.

Finally, by means of a regression analysis we explored how the creation of European Union affects the

structure of the system of cities and then the agglomeration forces of the Member States. Our results showed that only the constitution of the Schengen Area and the share of international trade seem to have a weak impact on the hierarchical structures of Member States presenting respectively a positive impact on the Zipf coefficient (more even distribution) and a negative impact (more agglomerated distribution).

In conclusion Gibrat's law and Zipf's law are here used to reflect important organized structures in the topology of systems of cities. However this study can be extended in several directions taking into account for instance the persistent existence of socioeconomic disparities between Member States (based, for example, on Gini index) or considering the role of the new communication technologies on the system of cities or finally considering the transport infrastructure in a integrated system. Clearly, the dynamics of such processes deserve due attention.

References

- Aldskogius, G., 2000. Urban policy in the structural policy of the European Union. CERUM Working Paper, Umea University, Faculty of Social Sciences, Centre for Regional Science, Umea.
- Alperovich, G.A., 1993. An explanatory model of city-size distribution: evidence from cross-country data, *Urban Studies*, 30, 1591–1601.
- Armington, C., Acs, Z., 2002. The determinants of regional variation in new firm formation, *Regional Studies*, 36(1): 33–45.
- Auerbach, F., 1915. The law of population concentration, *Scientific American Supplement*, 76: 384.
- Baltagi, B.H., 1995. *Econometric Analysis of Panel Data*, Chichester: Wiley.
- Beck, N., Katz, J.N., 1995. What to do (and not to do) with time-series cross-section data, *American Political Science Review*, 89: 634–647.
- Berry, B.J.L., Okulicz-Kozarin, A., 2012. The city size distribution debate: resolution for US urban regions and megalopolitan areas, *Cities*, 29: 17–23.
- Black, D., Henderson, V., 2003. Urban evolution in the USA, *Journal of Economic Geography*, 3(4): 343–372.
- Brakman, S., Garretsen, H., van Marrewijk, C., 2001. An introduction to geographical economics. Trade, location and growth. Cambridge: Cambridge University Press.
- Brakman, S., Garretsen, H., Schramm, M., 2004. The strategic bombing of German cities during World War II and its impact on city growth, *Journal of Economic Geography*, 4: 201–218.
- Cameron, A.C., Trivedi, P.K., 2005, *Microeconometrics*. Cambridge: Cambridge University Press.
- Chesher, A., 1979. Testing the law of proportionate effect, *Journal of Industrial Economics*, 27: 403–411.
- Cheshire, P., 1999, *Trends in sizes and structures of urban areas*, in a Cheshire, P. and Mills, E.S. (eds.). Handbook of Regional and Urban Economics, vol. 3. Amsterdam: Elsevier, pp. 1339–1372.
- Clauset, A., Shalizi, C.R., Newman, M.E.J., 2007, Power-law distributions in empirical data. arXiv: 0706.1062.
- Dobkins, L., Ioannides, Y.M., 2001. Spatial interactions among U.S. cities: 1900–1990, *Regional Science and Urban Economics*, 31(6):701–731.

- Eeckhout, J., 2004. Gibrat's law for (all) cities, *American Economic Review*, 94(5): 1429–1451.
- Fay, M., Yepes, T., 2003. Investing in infrastructure: what is needed from 2000 to 2010?, Working Paper No. 3102, Policy Research Working Paper Series, The World Bank.
- Fazio, G., Modica, M., 2012. Pareto or log-normal? A recursive truncation approach to the distribution of (all) cities, Working Papers 2012_10, Business School, Economics, University of Glasgow, Glasgow.
- Fujita, M., Krugman, P., Venables, A.J., 1999, *The Spatial Economy: Cities, Regions and International Trade*. Cambridge: MIT Press.
- Gabaix, X., 1999. Zipf's law for cities: an explanation, *The Quarterly Journal of Economics*, 114(3): 739–767.
- Gabaix, X., Ibragimov, R., 2011. Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents, *Journal of Business & Economic Statistics*, 29(1): 24–39.
- Gabaix, X., Ioannides, Y.M., 2004. *The Evolution of City Size Distributions* in a J. V. Henderson & J. F. Thisse (eds.). Handbook of Regional and Urban Economics. Amsterdam: Elsevier, pp. 2341–2378.
- Gibrat, R., 1931. *Les inégalités économiques*. Paris: Libraire du Recueil Siray.
- Giesen, K., Suedekum, J., 2011. Zipf's law for cities in the regions and the country, *Journal of Economic Geography*, 11: 667–686.
- González-Val, R., Lanaspa, L., Sanz, F., 2012. New evidence on Gibrat's law for cities. Working Papers 2012/18. Institut de Economia de Barcelona (IBE), Barcelona.
- Hardle, W., 1992. *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.
- Henderson, J.V., Wang, H.G., 2007. Urbanization and city growth: The role of institutions, *Regional Science and Urban Economics*, 37: 283–313.
- Ioannides, Y.M., Overman, H.G., 2003. Zipf's law for cities: an empirical examination, *Regional Science and Urban Economics*, 33(2): 127–137.
- Ioannides, Y.M., Skouras, S., 2013. US city size distribution: robustly Pareto but only in the tail," *Journal of Urban Economics*, 73(1), 18–29.
- Jiang, B., Jia, T., 2010. Zipf's Law for All the Natural Cities in the United States: A Geospatial Perspective. arXiv: 1006.0814v2.

- Kooij, P., 1988. Peripheral cities and their regions in the Dutch urban system until 1900, *The Journal of Economic History*, 48(2): 357–371.
- Krugman, P., 1991. *Geography and Trade*. Cambridge: MIT Press.
- Lewis, J.B., Linzer, D.A., 2005. Estimating Regression Models in Which the Dependent Variable Is Based on Estimates, *Political Analysis*, 13:345–364.
- Mera, K., 1975. On the Urban Agglomeration and Economic Efficiency, *Economic Development and Cultural Change*, 21(2): 309–324.
- Modica, M., Reggiani, A., Nijkamp, P., 2013. Are Gibrat and Zipf Monozygotic or Heterozygotic Twins? A Comparative Analysis of Means and Variances in Complex Urban Systems. Working Paper.
- Nadaraya, E.A., 1964. Some new estimates for distribution functions, *Theory of Probability & Its Applications*. 9(3): 497-500.
- Nahoi, O., 2011. War in Europe? Not so impossible, *Eurozine*.
- Pacione, M., 2001. *Urban Geography: A Global Perspective*. London: Routledge.
- Randolph, S., Bogetic, Z. Hefley, D., 1999. Determinants of Public Expenditure on Infrastructure: Transportation and Communication. Working Paper No. 1661, Policy Research Working Paper Series, The World Bank.
- Rosen, K.T., Resnick, M., 1980. The size distribution of cities: an examination of the Pareto law and primacy, *Journal of Urban Economics*, 8(2): 165 – 186.
- Rozenfeld, H.D., Rybski, D., Gabaix, X., Makse, H.A., 2011. The area and population of cities: new insights from a different perspective on cities, *American Economic Review*, 101(5): 2205–25.
- Singer, H. W., 1930. "The Courbe Des Populations". A parallel to Pareto's law, *The Economic Journal*, 48(182): 254–263.
- Soo, K.T., 2005. Zipf's Law for cities: a cross-country investigation, *Regional Science and Urban Economics*, 35(3): 239–263.
- Steindl, J., 1968. Size distributions in Economics, in *International Encyclopedia of the Social Sciences*, Vol. 14, Silks, ed. New York: Macmillan and the Free Press.

Sutton, K.T., 1997. Gibrat's legacy, *Journal of Economic Literature*, 35(1): 40–59.

United Nations, 1974. Manual VIII: Methods for projections of urban and rural population. United Nations publication, Sales No. E.74.XIII.3.

United Nations Statistical Division (UNSTAT), 2011. United Nations Demographic Yearbook 2009-2010. United Nations Statistical Division (UNSTAT).

Watson, G.S., 1964. Smooth regression analysis, *The Indian Journal of Statistics, Series A*, 26(4): 359–372.

World Commission on Environment and Development, 1987. Our common future. Oxford: Oxford University Press.

Yang, D.T., 1999. Zipf's Law for cities: a cross-country investigation, *American Economic Review*, 89(2): 306–310.

Zipf, G.K., 1949, *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

Table 1: Spatial Economics Characteristics of the 27 EU Member States in 2009

Country	Population (thousands)	Surface (km^2)	Population		Population in the largest city (% tot urban)*			Road Density (length *1000 km^2)**	Rail density (length *1000 km^2)**
			Density (people * km^2)	Rural Population	the largest city (% tot urban)*	Agricultural land (% tot)*			
Austria	8,419	82,430	102	32.3	30.0	38.4	1,278.9	76.3	
Belgium	11,008	30,280	364	2.5	18.0	45.0	4,987.4	105.9	
Bulgaria	7,476	108,560	69	26.9	22.0	46.3	362.5	37.3	
Cyprus	1,117	9,240	121	29.5	31.4	13.5	1,586.1	0	
Czech Republic	10,546	77,250	137	26.6	15.1	54.9	1,619.4	122.1	
Denmark	5,574	42,430	131	13.1	24.4	62.1	1,698.7	61.9	
Estonia	1,340	42,390	32	30.5	42.8	22.0	1,283.1	26.4	
Finland	5,387	303,900	18	16.3	32.6	7.6	230.9	17.5	
France	65,437	547,660	119	14.3	20.7	53.4	1,870.4	54.5	
Germany	81,726	348,610	234	26.1	5.7	48.4	1,805.2	117.6	
Greece	11,304	128,900	88	38.5	47.1	63.6	890.7	19.5	
Hungary	9,971	90,530	110	30.6	25.0	63.9	2,123.2	99.0	
Ireland	4,487	68,890	65	37.8	39.5	60.8	1,366.6	46.1	
Italy	60,770	294,140	207	31.6	8.2	47.3	1,618.4	67.2	
Latvia	2,220	62,180	36	32.3	46.2	29.5	1,131.4	34.7	
Lithuania	3,203	62,670	51	32.9	24.4	42.9	1,240.9	27.1	
Luxembourg	517	2,590	200	14.6	21.9	50.6	2,018.2	106.2	
Malta	419	320	1309	5.2	50.9	28.1	696	0	
Netherlands	16,696	33,730	495	16.9	7.7	56.8	3,294.7	69.7	
Poland	38,216	304,200	126	39.1	7.3	53.0	1,356.0	62.1	
Portugal	10,637	91,470	116	39.0	44.0	40.3	899.9	36.0	
Romania	21,390	230,060	93	47.2	16.5	58.8	342.8	45.2	
Slovak Republic	5,440	48,090	113	45.2	14.1	40.1	892.5	73.9	
Slovenia	2,052	20,140	102	50.1	26.4	23.2	1,920.3	60.6	
Spain	46,235	498,800	93	22.6	16.2	55.5	1,348.1	30.3	
Sweden	9,453	410,340	23	14.8	16.3	7.5	1,272.2	25.8	
United Kingdom	62,641	241,930	259	20.4	15.5	71.6	1,619.1	67.6	
Euro Area	332,990	2,551,580	131	24.5	15.4	45.5	1,445.1	54.1	
European Union	503,680	4,181,730	120	26.0	15.6	45.1	1,355.7	53.4	

* Data refer to year 2009 ** from Worldstat.info

Table 2: Description of the data and estimated threshold of the 27 EU Member States

Country	Population* (thousands)	N. of Localities*	Average population per locality	Threshold*	Population size at truncation*
Austria	8,431	2,357	3,577.00	654	2,497
Belgium	10,928	589	18,553.48	184	17,691
Bulgaria	7,327	5,059	1,448.31	1,405	571
Cyprus	839	388	2,162.37	67	1,966
Czech Republic	10,335	6,216	1,662.64	1,891	728
Denmark	4,832	1,469	3,289.31	392	1,528
Estonia	1,324	226	5,858.41	187	962
Finland	5,375	336	15,997.02	81	14,067
France	64,304	36,674	1,753.39	3,377	3,052
Germany	81,752	11,421	7,158.04	838	17,164
Greece	10,934	1,034	10,574.49	492	4,638
Hungary	9,986	3,153	3,167.14	1,012	1,526
Ireland	2,318	858	2,701.63	293	945
Italy	59,571	8,092	7,361.72	1,170	10,300
Latvia	2,261	523	4,323.14	372	853
Lithuania	2,171	102	21,284.31	30	11,623
Luxembourg	512	115	4,452.17	90	1242
Malta	412	67	6,149.25	13	10,770
Netherlands	14,432	2,025	7,126.91	175	14,885
Poland	38,200	2,478	15,415.66	1,099	8,293
Portugal	10,132	3,867	2,620.12	1,196	1,485
Romania	19,600	3,182	6,159.65	1,659	2,979
Slovak Republic	5,435	2,888	1,881.93	1,051	938
Slovenia	1,915	3,074	622.97	1,496	222
Spain	47,190	8,115	5,815.16	513	15,851
Sweden	8,003	1,912	4,185.67	249	4,518
United Kingdom	52,518	3,121	16,827.30	206	46,357

* The data is referred to the last observation in the sample

Table 3a: Rank-size coefficients and city thresholds of the 27 EU Member States (first and last observation available)

Country	Period	Rank-size coefficient (first year available)	Upper Tail	Rank-size coefficient (last year available)	Upper Tail
Austria	1981-2011	1.469 (0.0645)	1036	1.440 (0.0796)	654
Belgium	1990-2011	1.691 (0.1713)	195	1.724 (0.1798)	184
Bulgaria	1985-2011	1.114 (0.0383)	1686	1.021** (0.0385)	1405
Cyprus	2001-2011	0.773 (0.0756)	209	0.994** (0.1717)	69
Czech Republic	1996-2011	1.014** (0.0318)	2025	1.066** (0.0347)	1891
Denmark	1976-2011	0.876 (0.0377)	1078	0.949** (0.0676)	392
Estonia	2001-2011	1.096** (0.1142)	184	1.141** (0.1929)	70
Finland	1990-2010	1.226** (0.1380)	158	1.250** (0.1977)	81
France	1975-2009	1.039** (0.0174)	7167	1.200 (0.0292)	3337
Germany	1991-2011	1.284 (0.0619)	859	1.322 (0.0646)	838
Greece	1991-2001	1.182 (0.0061)	544	1.164 (0.0077)	492
Hungary	1980-2011	1.137 (0.0490)	1075	1.110* (0.0493)	1012
Ireland	1991-2011	0.787 (0.0453)	602	0.987** (0.0815)	293
Italy	1991-2011	1.300 (0.0465)	1563	1.400 (0.0579)	1170
Latvia	2001-2009	1.159* (0.0805)	415	1.079** (0.0791)	372
Lithuania	1989-2011	0.904** (0.2506)	26	0.948** (0.2449)	30
Luxembourg	1821-2011	2.139 (0.4322)	49	1.197** (0.1784)	90
Malta	1901-2011	1.084** (0.2395)	41	4.049 (0.2597)	13
Netherlands	2001	-	-	1.199** (0.1282)	175
Poland	1988-2010	1.351 (0.0485)	1553	1.365 (0.0582)	1099
Portugal	2001-2011	1.159 (0.0466)	1239	1.115* (0.0456)	1196
Romania	2011	-	-	1.402 (0.0487)	1659

* Significant at 1% ** significant at 5%; for rank-size coefficient significantly not different from 1

Table 3b: Rank-size coefficients and city thresholds of the 27 EU Member States (first and last observation available)

Country	Period	Rank-size coefficient (first year available)	Upper Tail	Rank-size coefficient (last year available)	Upper Tail
Slovakia	1991-2010	1.135 (0.0478)	1130	1.169 (0.0510)	938
Slovenia	2001-2011	1.203 (0.0420)	1639	1.204 (0.0440)	1496
Spain	1991-2011	1.110** (0.0615)	651	1.198 (0.0748)	513
Sweden	1990-2010	1.136** (0.1026)	245	1.085** (0.0972)	249
United Kingdom	1991-2001	1.459 (0.1636)	159	1.467 (0.1445)	206
Euro Area	2001	-	-	1.377 (0.00528)	2660
European Union (excluded Romania)	2001	-	-	1.401 (0.0526)	1419

* Significant at 1% ** significant at 5%; for rank-size coefficient significantly not different from 1

Table 4: Panel estimation of Eq. (6) (dependent variable = OLS coefficient of q)

Dependent Variable	(1)	(2)	(3)	(4)
	OLS	OLS	FE	RE
Trade (% of GDP)	-.00398** (.00167)	-.00382** (.00157)	-.00228*** (.00134)	-.00173 (.00193)
Scale Economies	-.00153 (.01150)	.01735 (.01336)	-.00215 (.02064)	.00237 (.02807)
Expenditure in education (% of GDP)	-.00944 (.04172)	-.01972 (.04647)	-.04112 (.03071)	.02409 (.05024)
Transport cost	.63581 (1.6369)	.71197 (2.0571)	-.82159 (1.9224)	.789131 (4.06137)
Foreign direct investment	-.00071 (.00083)	-.00113 (.00094)	-.00022 (.00068)	-.00015 (.00093)
Internet users	.00340 (.00222)	.00268 (.00205)	.00379 (.00385)	-.00005 (.00228)
ln (population)	.21367* (.04678)	.19841* (.05319)	.06791 (.04653)	.24237* (.08341)
ln (land area)	-.41339* (.10989)	-.40515* (.10916)	-.13959* (.04491)	-.36686 (.07509)
ln(GDP)	.020184 (.0546245)	-.00061 (.05700)	.03152 (.05631)	-.00688 (.09203)
Dummy Euro		-.23996 (.16537)	-.13857 (.17491)	.00029 (.24136)
Dummy Schengen		.12124 (.16815)	-.13399 (.12181)	.22971*** (.13477)
Interaction		.18028 (.25287)	.31286 (.19222)	-.13661 (.26264)
Constant	2.68183** (1.0473)	3.0277* (1.0788)	1.81866*** (.92899)	1.5775 (1.3881)
R-squares	0.5626	0.5724	0.3616	0.5370
N. obs.	66	66	66	66

* Significant at 1% ** significant at 5% *** significant at 10%

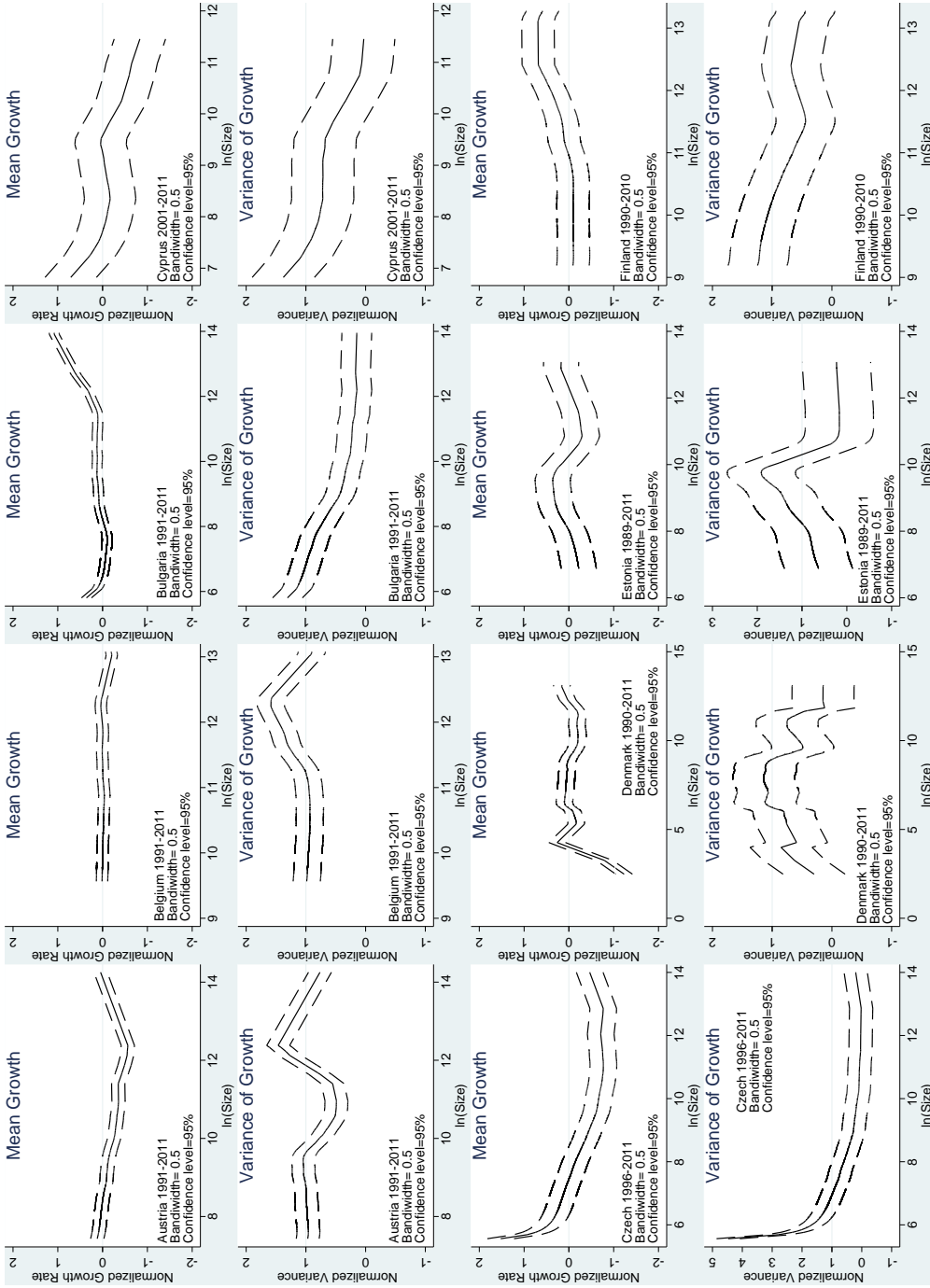


Figure 1a: NW estimator of mean normalized growth rate and normalized variance of the growth rate (different countries and years)

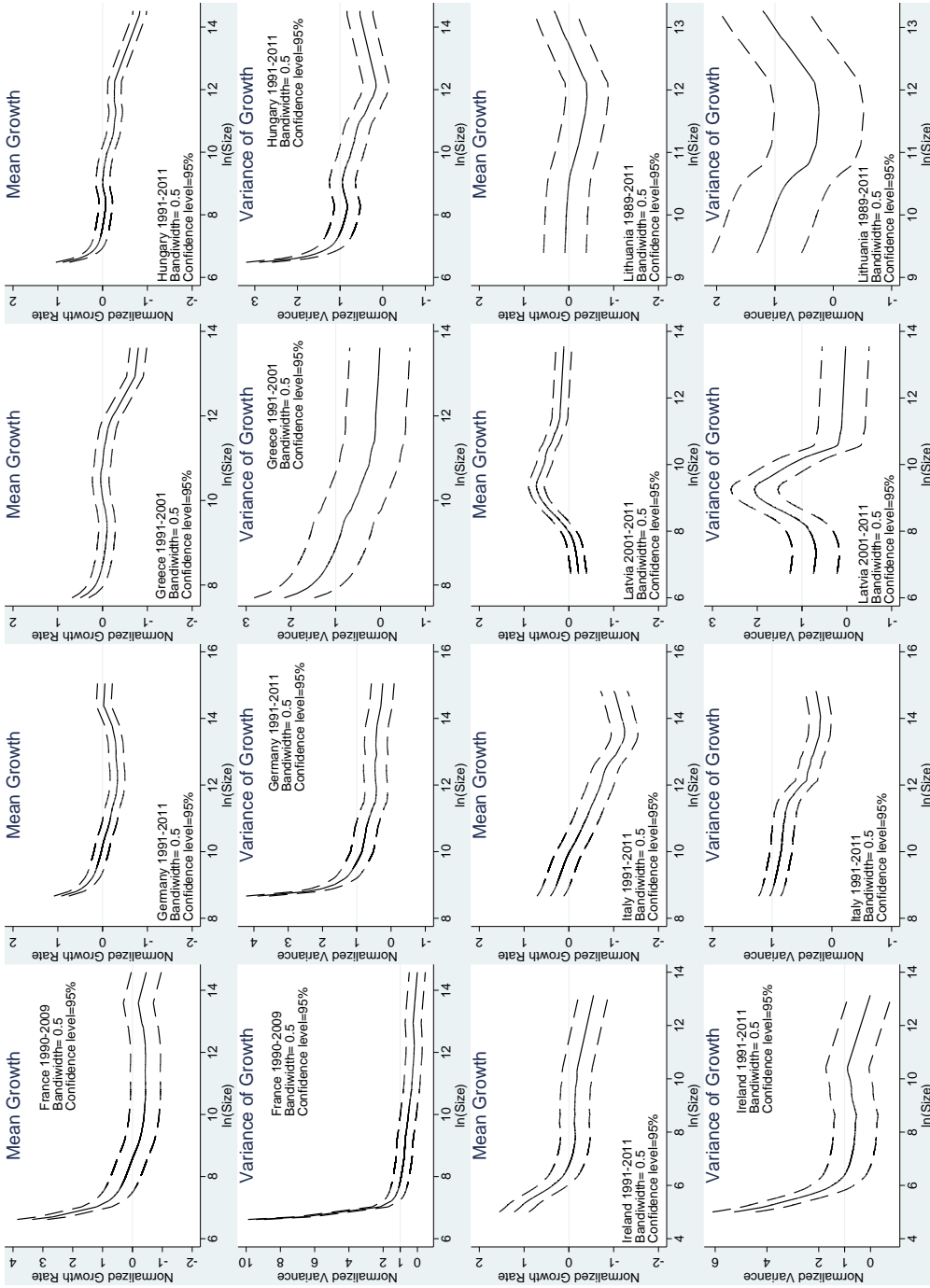


Figure 1b: NW estimator of mean normalized growth rate and normalized variance of the growth rate (different countries and years)

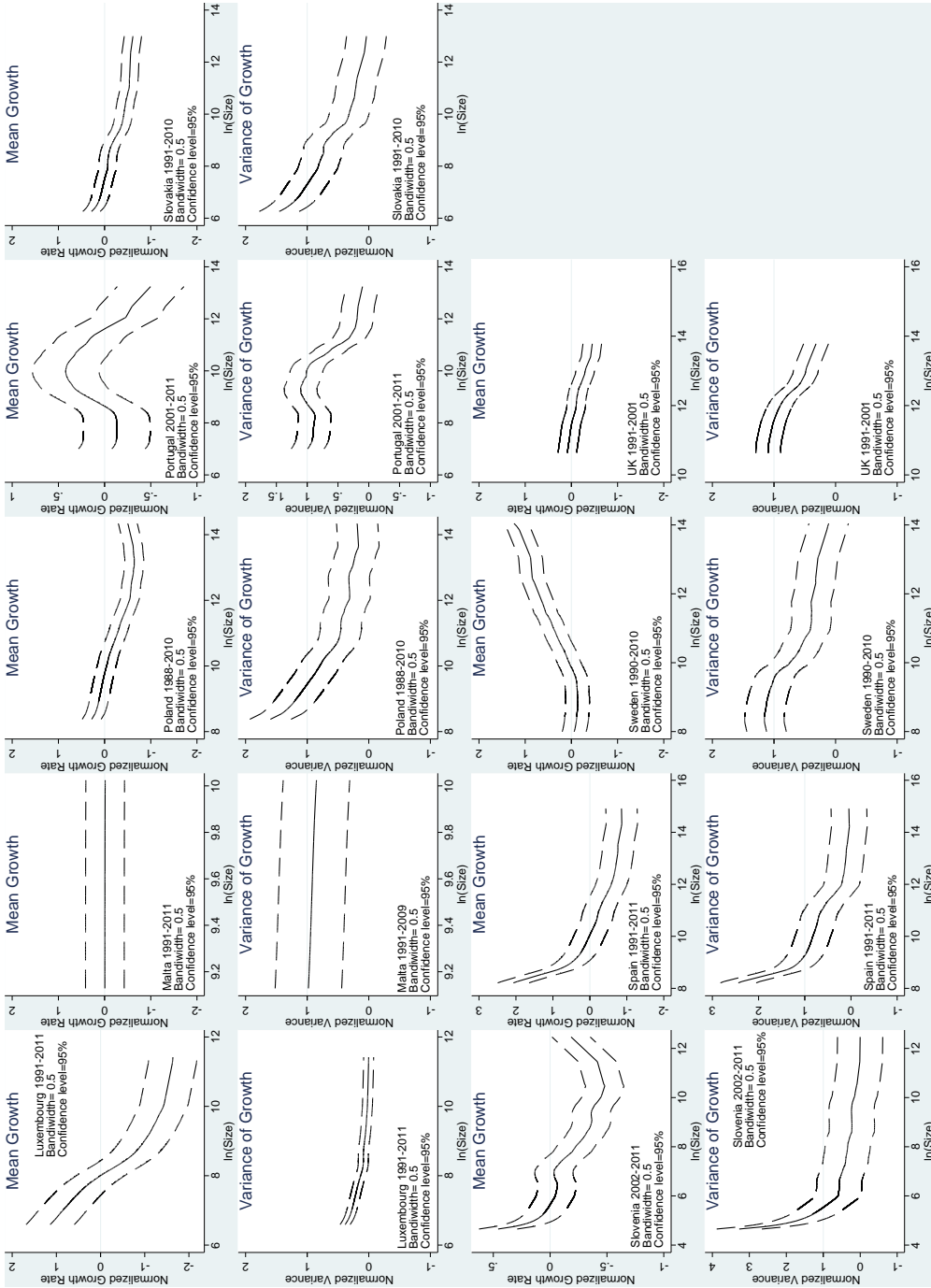


Figure 1c: NW estimator of mean normalized growth rate and normalized variance of the growth rate (different countries and years)