Historical urban growth in Europe (1300–1800)

Rafael, González-Val

Universidad de Zaragoza Institut d'Economia de Barcelona (IEB)

3 December 2017

Historical urban growth in Europe

(1300–1800)

Rafael González-Val

Universidad de Zaragoza & Institut d'Economia de Barcelona (IEB)

Abstract: This paper analyses the evolution of the European urban system from a long-term perspective (from 1300 to 1800). Using the method recently proposed by Clauset, Shalizi, and Newman (2009), a Pareto-type city size distribution (power law) is rejected from 1300 to 1600. A power law is a plausible model for the city size distribution only in 1700 and 1800, although the log-normal distribution is another plausible alternative model that we cannot reject. Moreover, the random growth of cities is rejected using parametric and non-parametric methods. The results reveal a clear pattern of convergent growth in all the periods.

Keywords: city size distribution, power law, Pareto distribution, Zipf's law, Gibrat's law.

JEL codes: C12, C14, R11, R12.

1. Introduction

European cities date from ancient times (the medieval age, the Roman Empire or even earlier; see Hohenberg and Lees, 1985). Their importance has increased or decreased over time depending on geographical, economic, and historical forces. The literature usually distinguishes between first- and second-nature forces in determining city size and city growth. The former are characteristics related to the physical landscape of a given location, such as the temperature, rainfall, access to the sea, presence of natural resources, or availability of arable land, while the latter refer to factors relating to human actions and economic incentives, such as economies of scale or knowledge spillovers. A number of recent empirical papers consider the importance of natural amenities in explaining city creation and city growth. For instance, Fernihough and O'Rourke (2014) find that geographical proximity to coal had a strong influence on cities' population; according to their estimates, being close to coal mines explains at least 60% of the growth in European city populations from 1750 to 1900.

As Gabaix and Ioannides (2004) point out, historians have produced fascinating series of urban populations that have not yet been fully explored. Most of the literature on historical city growth focuses on the case of the United States (US) (Kim, 2000; Kim and Margo, 2004; Michaels et al., 2012), while evidence on the European case is scarcer. Nevertheless, there are significant differences between the urbanization processes in Europe and those in the US. First, in many cases European cities have existed for hundreds of years, while the US urban system is relatively young (the first census by the US Census Bureau dates from 1790). Second, European inhabitants usually present low mobility compared to US citizens; Cheshire and Magrini (2006) estimate that the mobility in the US is 15 times higher than that in Europe. Finally, the growth rates of American cities react strongly to industry cycles (Duranton, 2007). Thus, in the second half of the nineteenth century and the early twentieth century, the growing urban population was concentrated in the north-eastern region known as the Manufacturing Belt, while in the second half of the twentieth century the rise of the Sun Belt attracted population to the West Coast area.

De Vries (1984) and Bairoch et al. (1988) report comprehensive historical data sets of European cities for several centuries. To date, just a few studies use these data to analyse urban growth in Europe, focusing on the different factors influencing population growth. De Long and Shleifer (1993) examine the relationship between

political regimes and historical city growth in the largest European cities. Acemoglu et al. (2005) use the European city-level data from Bairoch et al. (1988) to investigate which urban centres were driving demographic and economic growth and to contrast the growth of Atlantic ports with that of other ports and inland cities. Bosker et al. (2013) analyse why, between 800 and 1800, the urban centre of gravity moved from the Islamic world to Europe, unravelling the role of geography and institutions in determining the long-run city development in the two regions. However, our paper is closely related to Dittmar's (2011) study of the evolution of the European city size distribution from 1300 to 1800. He considers Zipf's law (a Pareto distribution of which the exponent is equal to one) as the benchmark for city size distribution, concluding that Zipf's law only emerged in Europe after 1500.

The analysis that we provide differs from that of Dittmar (2011) in four fundamental aspects. First, Dittmar (2011) imposes a minimum population threshold of 5,000 inhabitants in all periods, while we use a new methodology to select the optimal truncation point, which can vary by period. Second, we are not interested in whether Zipf's law holds exactly (although statistical tests exist to address this issue; Urzúa, 2000; Gabaix, 2009). Using a new methodology, our approach is to test whether the Pareto distribution (for which Zipf's law is a particular case) is a good description of the European city data, regardless of the particular value of the Pareto exponent. In cases in which we reject the Pareto distribution, we also consequently reject Zipf's law. Third, we also test whether the log-normal distribution is another plausible model that can explain the European city size distribution. Finally, Dittmar (2011) focuses on Zipf's law and the historical factors that help to explain the emergence of the law, including an interesting theoretical model of city growth, but his empirical evidence of urban growth relies on correlations between city size and growth. In this paper we conduct a parametric and non-parametric analysis of urban growth using growth and kernel regressions.

There has been a revival of interest in city size distributions and Zipf's law in the last few decades among urban economists, especially since Krugman (1996) highlighted the 'mystery of urban hierarchy,' although studies of city size distribution have a long tradition. Over the last 100 years, many papers testing the validity of this law for many different countries have been published, especially the last years (see the reviews by Nitsch, 2005, and Soo, 2005). Starting from the wide empirical literature, some

theoretical models are proposed in recent papers to explain the law, with different economic foundations: productivity or technology shocks (Duranton, 2007; Rossi-Hansberg and Wright, 2007) or local random amenity shocks (Gabaix, 1999). These models justify Zipf's law analytically, associate it directly with an equilibrium situation, and connect it to proportionate city growth (Gibrat's law), another well-known empirical regularity that postulates that the growth rates of cities tend to be independent of their initial sizes.[1] In both the theoretical and the empirical literature, Zipf's law is seen as a steady-state situation.

However, the empirical literature considering historical data and long time periods usually finds significant deviations from both empirical regularities. Recent literature focusing on the US case (Giesen and Südekum, 2014; Sánchez-Vidal et al., 2014; Desmet and Rappaport, 2017) documents episodes of convergence or divergence from a long-term perspective, but both convergence and divergence dissipate over time and Zipf's and Gibrat's laws gradually emerge.

The aim of this paper is to analyse historical urban growth in Europe focusing on these two empirical regularities. Our contribution is threefold. First, we consider the comprehensive historical data set on European cities by Bairoch et al. (1988). As explained above, several studies make use of this data bank, but, with the exception of Dittmar (2011), to our knowledge no study analyses the fulfilment of these empirical regularities from a European perspective, as most of the studies focus on individual countries. Second, we apply a new methodology developed by Clauset, Shalizi, and Newman (2009), which allows us to estimate the optimal truncation point and the distribution parameter as well as carrying out tests to check the goodness of fit of the model. Finally, we also conduct an analysis of urban growth using parametric and non-parametric techniques. The analyses of urban growth usually consider short time periods (one or two centuries at the most), so the results from the present study could help in understanding the urban dynamics from a very long-term perspective. The paper is organized as follows. Section 2 introduces the database that we use. Sections 3 and 4 contain the statistical analysis of the distribution of city sizes and their evolution over time, respectively, and Section 5 concludes.

---

[1] Gibrat (1931) observed that the size distribution (measured by sales or number of employees) of firms tends to be log-normal, and his explanation was that the growth process of firms could be multiplicative and independent of the size of the firm.

2. Data

There are various ways to define a city, and the choice of the spatial unit is not trivial (Rosen and Resnick, 1980). The two basic alternatives are the administratively defined cities (legal cities) and the metropolitan areas. Both units have advantages. Metropolitan areas represent urban agglomerations, covering huge areas that are meant to capture labour markets; thus, they are more natural economic units. Legal cities are political units that usually lie within metropolitan areas, and their boundaries not always make economic sense, although some factors, such as human capital spillovers, are thought to operate at a very local level (Eeckhout, 2004). Defining a city is even more difficult when encompassing many countries because administrative definitions vary considerably across countries. Only in recent years have consistent definitions of functional urban areas been developed for European cities. Schmidheiny and Suedekum (2015) are the first to use the novel data obtained from a collaboration of the European Commission and the Organization for Economic Co-operation and Development (OECD). This data set includes 692 functional urban areas in Europe in the year 2006.

We use the historical data set of European cities collected by Bairoch et al. (1988). They provide information by century on a large set of cities (2,135) from many countries from 800 to 1800. We focus only on Western European cities, from the current Austria, Belgium, Denmark, Finland, France, Germany, Great Britain, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, and Switzerland.[2] Bairoch et al. (1988) define a city as an urban agglomeration of at least 5,000 inhabitants in any period between 800 and 1800. Thus, the term 'city' includes not only the precisely defined city but also the urban districts or suburbs surrounding the centre. Each urban agglomeration is identified with the 'principal name', which is the name of the central city presently used in the country where it is located. They apply this consistent definition to this set of European countries to build their data bank. Many data sources by country are used to estimate the population values by city; Chapter 3 in the study by Bairoch et al. (1988) reports the complete list of data sources by country. Moreover, besides population figures they document some historical information by city, such as the changes in the city name over the centuries and the geographical coordinates of all the cities.

---

[2] Following Bairoch et al. (1988), we use the present-day boundaries of countries.

Bairoch et al. (1988) emphasize that data from before 1300 are less reliable (they even skip the year 1100 due to the lack of information), so we only consider data from 1300 to 1800. Moreover, in many cases observations are missing in some years; like Voigtländer and Voth (2013), we fill these gaps using linear interpolated values.[3] In this way we increase the number of observations, obtaining a better fit of the models. However, we repeat all the analyses using the raw data set from Bairoch et al. (1988) and the results do not change (see the Appendix A).

Some authors criticize the Bairoch et al. (1988) data because of some unrealistic values. For instance, the population estimate for Cordoba (Spain) in 1000 is usually considered to be excessively large.[4] Nevertheless, the data sets of de Vries (1984) and Bairoch et al. (1988) are often considered to be the most reliable for historical European urbanization (Voigtländer and Voth, 2013). Bairoch et al. (1988) carry out some comparisons between the two data banks; for instance, when cities of more than 20,000 inhabitants in the year 1700 are considered (91 cities), the deviation in the total population between both data sources is only 0.6%. Dittmar (2011) also compares the de Vries (1984) and Bairoch et al. (1988) databases, concluding that there is no evidence of systematic shortfalls in the populations that Bairoch et al.'s (1988) data record for large cities.

Table 1 shows the number of cities for each century and the descriptive statistics. The table also reports the difference between raw data and data filled with interpolations. Although Bairoch et al. (1988) include all cities of at least 5,000 inhabitants in any period in their database, cities can have a smaller population in some years; thus the minimum population is 1,000. Some authors (Dittmar, 2011; Bosker et al., 2013) impose a fixed minimum population threshold (5,000 and 10,000 inhabitants, respectively). Nevertheless, the methodology used here selects a different threshold by period to obtain the best fit to the empirical data. The sample reflects the urbanization process that took place over time in Europe. From the first period, there was a rapid increase in the number of cities, while the average size of cities remained stable at around 10,000 inhabitants.

3. City size distribution

---

[3] Values interpolated using the ipolate command in Stata.
[4] We repeat all the analyses excluding Cordoba, and the results do not change.

Let S denote the city size (measured by the population); if this is distributed according to a power law, also known as a Pareto distribution, the density function is

$$p(S) = \frac{a-1}{\underline{S}}\left(\frac{S}{\underline{S}}\right)^{-a} \quad \forall S \geq \underline{S} \text{ and the complementary cumulative density function } P(S)$$

is $P(S) = \left(\frac{S}{\underline{S}}\right)^{-a+1} \quad \forall S \geq \underline{S}$, where $a > 0$ is the Pareto exponent (or the scaling parameter) and $\underline{S}$ is the population of the city at the truncation point, which is the lower bound to the power law behaviour. It is easy to obtain the expression $R = A \cdot S^{-a}$, which relates the empirically observed rank R (1 for the largest city, 2 for the second largest, and so on) to the city size. Zipf's law is an empirical regularity, which appears when Pareto's exponent of the distribution is equal to the unit ($a = 1$). This means that, when ordered from largest to smallest, the population of the second city is half that of the first, the size of the third is a third of the first, and so on.

This expression is applied to the study of very diverse phenomena, such as the distribution of the number of times that different words appear in a book (Zipf, 1949), the length of rivers (Krugman, 1996), the intensity of earthquakes (Kagan, 1997), and the losses caused by floods (Pisarenko, 1998) or forest fires (Roberts and Turcotte, 1998). It is used extensively in urban economics to study city size distribution (see the excellent surveys by Cheshire, 1999, and Gabaix and Ioannides, 2004), although a recent discussion concerns the non-linear behaviour of the distribution when all cities are considered with no size restriction (Eeckhout, 2004). The current state of the art is that, although most of the distribution is non-linear, the Pareto distribution (and Zipf's law) holds for the largest cities. Thus, new distributions are proposed, combining linear and non-linear functions and separating the body of the distribution from the upper-tail behaviour. The reason is that the largest cities represent most of the population of a country, and the behaviour of the upper-tail distribution can be different from that of the entire distribution. Actually, the largest cities follow a Pareto distribution in many cases (Levy, 2009; Ioannides and Skouras, 2013).

Taking natural logarithms, we obtain the linear specification that is usually estimated:

$$\ln R = \ln A - a \ln S + u, \quad\quad\quad (1)$$

where u represents a standard random error ($E(u)=0$ and $Var(u)=\sigma^2$) and $\ln A$ is a constant. The greater the coefficient $\hat{a}$, the smaller the dispersion between cities. Similarly, a small coefficient (less than 1) indicates a heavy-tailed distribution. However, this regression analysis, which is commonly used in the literature, presents some drawbacks. The main problem is that the maximum likelihood (ML) estimator is more efficient if the underlying stochastic process is really a Pareto distribution (Gabaix and Ioannides, 2004; Goldstein et al., 2004). Furthermore, as Gabaix and Ioannides (2004), Nishiyama et al. (2008), and Clauset et al. (2009) emphasize, the estimates of the Pareto exponent are subject to systematic and potentially large errors. Finally, Gabaix and Ibragimov (2011) point out that this procedure is strongly biased in small samples. To overcome this last pitfall, Gabaix and Ibragimov (2011) propose to specify Eq. (1) by subtracting $1/2$ from the rank to obtain an unbiased estimation of the exponent:

$$\ln\left(R - \frac{1}{2}\right) = \ln A - a \ln S + u . \qquad (2)$$

Their numerical results demonstrate the advantage of this approach over the standard OLS estimation procedures, especially in small samples.[5] However, again, if the underlying stochastic process is really a Pareto distribution, the ML estimator is more efficient.[6]

Therefore, to estimate the power laws, we use the innovative method proposed by Clauset, Shalizi, and Newman (2009) (CSN, henceforth). The ML estimator of the Pareto exponent is:

$$\hat{a} = 1 + n\left(\sum_{i=1}^{n} \ln \frac{S_i}{\underline{S}}\right), \quad \forall S_i \geq \underline{S},$$

---

[5] The preliminary results obtained from the OLS estimation of Eq. (2) indicate that the power law provides a very good fit to the real behaviour of our European city data. The estimated R^2 is greater than 0.8 in all cases, and the estimated Pareto exponent increases over time from 0.86 in 1300 to 1.24 in 1800. However, as indicated in the main text, the fit provided by the OLS estimator is overcome by the ML estimator if the underlying stochastic process is really a Pareto distribution (Clauset et al., 2009), and inferences based on R^2 may not be robust because a very wide range of heavy-tailed distributions will produce large R^2 values when their distributions are estimated with OLS.
[6] Although Gabaix and Ibragimov's (2011) preliminary results (not shown in the paper) suggest that their specification is more robust than the ML estimator under deviations from power laws.

where n is the number of observations for $S_i \geq \underline{S}$. Clauset et al. (2009) propose an iterative method to estimate the adequate truncation point ($\underline{S}$). Exponent a is estimated for each $S_i \geq \underline{S}$ using the ML estimator (bootstrapped standard errors are calculated with 500 replications), and then the Kolmogorov–Smirnov (KS) statistic is computed for the data and the fitted model. The $\underline{S}$ lower bound that is finally chosen corresponds to the value of S_i for which the KS statistic is the smallest.[7] The ML estimator also has its own drawbacks. First, it assumes that the underlying stochastic process is really a Pareto distribution. This is a testable hypothesis; we use the CSN methodology not only because it estimates an optimal cut-off but also because they develop methods for characterizing and comparing the model fit based on bootstrapping and a Vuong closeness test. Second, the ML estimator is consistent only in large samples. Clauset et al. (2009) argue that the small-sample bias can be significant but in most practical situations it can be ignored, because it is much smaller than the statistical error of the estimator.[8] They recommend the rule of thumb $n \geq 50$ to obtain reliable parameter estimates, a sample size that is much smaller than our number of cities in all periods.

Furthermore, the literature contains other ways to estimate the population threshold that switches between the body of the distribution and the Pareto upper-tail. Fazio and Modica (2015) compare four different methodologies (including the CSN method), concluding that the Ioannides and Skouras (2013) approach tends to underestimate the truncation point moderately, providing the closest prediction. However, Fazio and Modica (2015) show that the differences between thresholds estimated with different methodologies increase with the sample size by running simulations with sample sizes from 1,000 to 25,000 observations. In this paper our sample sizes are considerably smaller (only in the last two periods are there more than 1,000 cities), so the differences between the results obtained with the different methodologies should be small.

Figure 1 shows the results from 1300 to 1800. The data, plotted as a complementary cumulative distribution function (CCDF), are fitted by a power law, and

[7] The power laws and the statistical tests are estimated using the poweRlaw R package developed by Colin S. Gillespie (based on the R code of Laurent Dubroca and Cosma Shalizi and the Matlab code of Aaron Clauset) and the Stata codes developed by Michal Brzezinski, which are all freely available on their web pages.

[8] For small data sets, biases are present but decay as $O\left(n^{-1}\right)$ for any choice of $\underline{S}$, whereas the statistical error of the estimator decays as $O\left(n^{-1/2}\right)$; see Clauset et al. (2009).

its exponent is estimated using the ML estimator. For illustrative purposes a log-normal distribution is also fitted to the data by maximum likelihood (the blue dotted line). The optimal lower bound for both distributions is estimated using the CSN method; the estimated values are reported in Table 2. The black line shows the power law behaviour of the upper-tail distribution. The estimated Pareto exponents are also shown in Table 2, although we are interested in the fit of the distribution rather than the particular value of the parameter. In Figure 1 important deviations between the data and the fitted power law can be observed in the first centuries (1300–1500), especially for the largest cities. Nevertheless, in the last periods (1600–1800), the fit improves and the power law appears to provide a good description of the behaviour of the distribution. In a similar fashion, the fit of the log-normal distribution improves over time.

However, visual methods can lead to inaccurate conclusions, especially in the upper-tail, because of fluctuations in the empirical distribution, so next we conduct statistical tests on the goodness of fit. Clauset et al. (2009) propose several goodness-of-fit tests, based on the measurement of the 'distance' between the empirical distribution of the data and the hypothesized Pareto distribution. This distance is compared with the distance measurements for comparable synthetic data sets drawn from the hypothesized Pareto distribution, and the p-value is defined as the fraction of the synthetic distances that are larger than the empirical distance. We use the CSN semi-parametric bootstrap approach. The procedure is based on the iterative calculation of the KS statistic for 500 bootstrap data set replications. This method samples from the observed data and checks how often the resulting synthetic distributions fit the actual data as poorly as the ML-estimated power law. Thus, the null hypothesis is the power law behaviour of the original sample for $S_i \geq \underline{S}$. Regardless of the true distribution from which our data were drawn, we can always fit a power law, so Clauset et al. (2009) recommend the conservative choice that the power law is ruled out if the p-value is below 0.1, that is, if there is a probability of 1 in 10 or less that we would obtain data merely by chance that agree as poorly with the model as the data that we have.

Therefore, this procedure only allows us to conclude whether the power law achieves a plausible fit to the data. Table 2 shows the results of the tests; the p-values of the test for periods from 1300 to 1600 are lower than 0.1, rejecting the power law behaviour of data in these centuries. Only in the last two periods, 1700 and 1800, are the

p-values clearly higher than 0.1, indicating that the power law is a plausible approximation to the real behaviour of the data.

Finally, we also compare the linear power law fit with the fit provided by the log-normal distribution (a non-linear distribution), using Vuong's model selection test to compare the power law with the log-normal.[9] The test is based on the normalized log-likelihood ratio; the null hypothesis is that the two distributions are equally far from the true distribution, while the alternative is that one of the test distributions is closer to the true distribution.[10] High p-values indicate that one model cannot be favoured over the other, and this is the conclusion reached in all the periods. Therefore, even for the last two centuries, in which we obtain moderate support for the power law behaviour, the power law is an acceptable fit but there is a plausible alternative as well.

According to these results, the support for a Pareto distribution (and thus for Zipf's law) is weaker than that found by previous studies. Using the test developed by Gabaix (2009), Dittmar (2011) rejects Zipf's law in Western Europe up to 1500, but he cannot reject Zipf's law from 1600 onwards. Here we reject the Pareto distribution in 1600. Although Dittmar (2011) considers the same data set as this study (Bairoch et al., 1988), this different result for 1600 may be caused by the distinct truncation points considered – a fixed minimum population threshold of 5,000 inhabitants in the study by Dittmar (2011) and the optimal population threshold here – and the different methodologies used. However, a result that is common to both studies is that the fit provided by the Pareto distribution improves over time, indicating a transition over time to a more stable city size distribution. Using a more recent data set of 692 functional urban areas in Europe in the year 2006, Schmidheiny and Suedekum (2015) conclude that city sizes in Europe do not follow a power law, because the largest cities are "too small" to follow Zipf's law. They argue that the explanation for the deviations from Zipf's law in Europe is that the area is still much less integrated economically (for instance, individuals are less mobile within and particularly across countries).

---

[9] In Figure 1, the lower bound for the two distributions (log-normal and power law) is the same value, the estimated value corresponding to the power law calculated using the CSN method. Although the lower bounds for the two distributions could be different, to compare the distributions, the threshold must be the same in both cases, so we use the same lower bound to run the test.

[10] This procedure is different from the approach followed by Giesen et al., 2010. These authors use information criteria to discriminate between different statistical city size distributions. Thus, they can conclude which distribution best fits their data (although the information criteria penalize the distribution with more parameters). Here we only test whether the two distributions that we consider are equally far from the true distribution. Nevertheless, Clauset et al., 2009, show that the CSN method generally performs better than the Bayesian information criterion (BIC) approach.

As Schmidheiny and Suedekum (2015) point out, the emergence of Zipf's law in an urban system requires time and evolves in parallel with the general degree of economic integration of that area. As Zipf's law is considered as a steady-state situation, the rejection of a Pareto distribution (and Zipf's law) for the first periods (1300 to 1600) for this pool of European cities from different countries indicates that the European urban system was not integrated in those early times. Moreover, the shape of the overall city size distribution has implications for the city size distributions at the national level. Recent works relate the fulfilment of Zipf's law in city size distributions at the regional and national levels. Gabaix (1999) shows that, if the urban growth in all regions follows Gibrat's law, we should observe the Zipfian upper-tail distribution both at the regional and at the national level (in our case at the national and supranational levels). Giesen and Südekum (2011) test this hypothesis for the German case, finding that Zipf's law is satisfied not only for Germany's national urban hierarchy but also for single German regions. However, here we reject the power law behaviour of city size distribution at the European level; thus, this would indicate that Zipf's law does not hold at the national level either.[11] Following the logic of Giesen and Südekum (2011), the national urban systems of those countries were not integrated. Several studies provide evidence of the internal process of economic integration of European countries, which in some cases finished in the eighteenth or nineteenth century. For instance, Bosker et al. (2008) study the evolution of Italian cities over the period 1300–1861, finding significant differences between the north and the south of Italy in the century-specific effects on city growth. As a consequence, the city size distribution of the northern part of Italy was relatively more stable than that of the southern part. González-Val et al. (2017) analyse the growth of Spanish cities during the period 1860–1960. They find that only changes in the market potential from 1900 had a significant effect on population growth, linking this result to the advances in the economic integration of the national market together with an intense process of industrialization. All in all, each individual European country consolidated its national economy and urban system around the end of our period (1800), which is in line with the improvement in the fit of the Pareto distribution for the pool of European cities in the last periods in our sample.

---

[11] Russell (1972) shows that Zipf's law did not hold at the local level in the high middle ages. Appendix B shows the results by country. Unfortunately, the number of cities in some countries is quite low, so these results should be treated with caution due to the small-sample bias.

4. Urban growth

The previous results show what we consider to be a snapshot of the city size distribution from 1300 to 1800. For each period, we conduct a goodness-of-fit test that indicates the plausibility of the power law model only in the last two centuries. For the remaining periods, the Pareto distribution (and thus Zipf's law) is rejected. The literature that studies the distribution of city sizes usually concludes that a Pareto-type distribution is generated by a random growth process (Gabaix, 1999), although other growth mechanisms could also generate Zipf's law in some cases (Gabaix and Ioannides, 2004). Furthermore, there is another plausible alternative model that we could not reject in the previous empirical analysis, the log-normal distribution. Random growth rates of cities could generate both types of distribution – log-normal and Pareto – if there is a lower bound to the distribution (see Gabaix, 1999).[12] The hypothesis that is usually tested is that the growth of the variable is independent of its initial size (the underlying growth model is a multiplicative process), which is known as Gibrat's law. To check whether this is true for our sample of European cities, we carry out a dynamic analysis of growth rates using parametric and non-parametric tools. Studies of urban growth usually consider short time periods (decades or one or two centuries at the most), but here we analyse city growth from a very long-term perspective.

4.1 Parametric analysis

Let S_{it} be the population of city i at time t and let g be its logarithmic growth rate; then $g_{it} = ln\, S_{it} - ln\, S_{it-1}$. We consider the following general expression of the growth equation:

$$g_{it} = \mu + \beta \ln S_{it-1} + \varphi_j + \delta_t + u_{it},\qquad(3)$$

the well-known expression of unconditional β-convergence (Barro and Sala-i-Martin, 2004). Our baseline equation is similar to that proposed by Black and Henderson (2003) and Henderson and Wang (2007) to explain city growth, although our main explanatory variable is the initial population instead of the city market potential. φ_j are country

---

[12] Only a small change from the log-normal generative process yields a generative process with a power law distribution, that is, a bounded minimum that acts as a lower reflective barrier to the multiplicative model (Gabaix, 1999).

fixed effects, δ_t are time fixed effects, and u_{it} is a random variable representing the random shocks that the growth rate may suffer, which we suppose to be identically and independently distributed for all cities, with $E(u_{it}) = 0$ and $Var(u_{it}) = \sigma^2 \; \forall i, t$. If $\beta = 0$, Gibrat's law holds and we find that growth is independent of the initial size with an average growth rate μ. Thus, if the estimation of β is significantly different from zero, we reject the fulfilment of Gibrat's law. In the case of it being lower than zero, we have convergent growth, because there is a significant negative relationship between growth and initial size. Moreover, the estimated value of β indicates the speed of convergence.

Random population growth rates can generate a Pareto distribution. Thus, we could expect to find, at least in the periods in which the power law is rejected, that random growth is rejected, although the emergence of Zipf's law in an urban system requires time and lags may be crucial. Table 3 presents the results. First, we run the OLS regression for each century and for a pool of 1300–1800 including all the observations. The estimates by century include country fixed effects, while the pool estimate also includes time fixed effects. All the estimations show a significant negative coefficient for the initial population, indicating a convergent growth pattern in all the periods. Thus, the greater the initial city size, the lower the population growth. Furthermore, the value of the estimated coefficient is similar in all the periods, with an average value around -0.2, although in some periods (1500–1600 and 1600–1700) the estimated values of the coefficient are lower, indicating a slower speed of convergence across cities. Finally, to take advantage of the panel dimension of our data set, we run a panel regression for the 1300–1800 period, including time and city fixed effects. The last column in Table 3 shows the results, revealing a stronger convergent pattern.

Columns (1) to (6) also show the estimated coefficients for the country dummies (Finland is the base category) based on the present-day country boundaries, included to control for unobserved country-specific variables. As Finland is the excluded country, all the coefficients are relative to the Finnish case. All the coefficients are significant and negative in the first period (1300–1400), but most of them grow over time to reach positive values in the 1600–1700 period. However, in the last period (1700–1800), most of them turn negative again. The exception is Great Britain, the coefficient of which increases constantly over time, indicating that the population in British cities grew more

than average, especially in the last two periods. However, this general trend observed in the evolution of the country dummies over time is linked to the evolution of the population in the base category country, Finland. Thus, the Finnish demographic and political history is important to understand the changes in the signs of the coefficients over time; for instance, the famine in 1695–1697 in Finland, which killed one-third of the population, may explain the positive coefficients for country dummies in the 1600–1700 period.[13]

4.2 Non-parametric analysis

The literature highlights the advantages of the non-parametric approach over the standard parametric one. Mainly, non-parametric methods do not impose any structure on the underlying relationships, which may be non-linear and may change over time (avoiding the assumption of a stationary relationship). Again we define g_i as the logarithmic growth rate $(\ln S_{it} - \ln S_{it-1})$ and normalize it (by subtracting the contemporary mean and dividing by the standard deviation in the relevant year).[14] First, we perform a non-parametric analysis using kernel regressions (Ioannides and Overman, 2003; Eeckhout, 2004). This consists of taking the following specification:

$$g_i = m(S_i) + \varepsilon_i,$$

where g_i is the normalized growth rate and S_i the logarithm of the ith city's population. Instead of making assumptions about the functional relationship m, $\hat{m}(S)$ is estimated as a local mean around point S and is smoothed using a kernel, which is a symmetrical, weighted, and continuous function in S. Thus, this non-parametric estimate lets growth vary with the initial population over the entire distribution. We run the kernel regression for each period and for a pool of 1300–1800 including 3,798 city size–growth rate pairs.

To estimate $\hat{m}(S)$, the Nadaraya–Watson method is used, as it appears in Härdle (1990, Chapter 3), based on the following expression:

---

[13] The author thanks one anonymous referee for this argument.
[14] The growth rates need to be normalized to be able to consider growth rates from different periods jointly in a pooled dataset.

$$\hat{m}(S) = \frac{n^{-1} \sum_{i=1}^{n} K_h (S - S_i) g_i}{n^{-1} \sum_{i=1}^{n} K_h (S - S_i)},$$

where K_h denotes the dependence of kernel K (in this case an Epanechnikov[15]) on bandwidth h.[16] As the growth rates are normalized, if the growth was independent of the initial population, the non-parametric estimate would be a straight line on the zero value and values different from zero would involve deviations from the mean.

The results by century are shown in Figure 2. The graphs also include the 95% confidence bands. The estimates confirm the negative relationship between size and growth obtained with the growth regressions, although, for the last periods (1500–1600, 1600–1700, and 1700–1800), a U-shaped pattern appears and the cities in the upper-tail distribution also display high growth rates. Thus, we can reject the (Gibrat's law) premise that there is no relationship between size and growth in all the periods. The decreasing pattern is clear: the greater the initial population, the lower the growth rate. This points to a high degree of convergence (mean reversion) across cities, especially for the smallest units. We also build a pool with all the growth rates between two consecutive periods; the last graph in Figure 2 shows the kernel regression of growth for the pool. The estimated mean growth clearly decreases with the city size, indicating a convergent growth pattern throughout the whole period and rejecting random growth. Nevertheless, although random growth is rejected for most of the distribution, for the upper-tail (city sizes greater than 10 in logarithmic scale), there are no significant differences in growth rates across cities, indicating that Gibrat's law holds for many middle-sized and large cities.

This evidence against random growth is in line with previous studies considering historical data and long time periods. Some recent studies focusing on the US case (Giesen and Südekum, 2014; Sánchez-Vidal et al., 2014; Desmet and Rappaport, 2017) find episodes of convergence or divergence from a long-term perspective. Giesen and Südekum (2014) develop a theoretical model of urban growth with the entry of new cities, obtaining a pattern whereby Gibrat's law holds in the long run but young cities

---

[15] Several kernel functions are commonly used: uniform, triangular, biweight, triweight, Epanechnikov, normal, and others. The advantage of the Epanechnikov kernel function is that it is optimal in a mean square error sense.

[16] The bandwidth is set using Silverman's (1986) rule of thumb.

initially grow faster. This model is tested using data on the exact foundation dates of 7,000 American cities for the period 1790 to 2000, confirming that the city size distribution is systematically related to the country's city age distribution. Sánchez-Vidal et al. (2014) obtain similar results using untruncated US incorporated places' data but considering only the twentieth century. They find that young, small cities tend to grow at higher rates, but, as the decades pass, their growth stabilizes or even declines; after the first years of existence, Gibrat's law tends to hold better. Finally, Desmet and Rappaport (2017) use data from US counties and MSAs from 1800 to 2000 to conclude that in earlier periods smaller counties converge and larger ones diverge but both convergence and divergence dissipate and Gibrat's law gradually emerges. The results obtained from our sample of European cities are consistent with this literature.

To accommodate the different growth patterns observed over time, new theories of urban growth have been developed. Cuberes (2011) concludes, using a comprehensive cross-country data set, that historically city growth may have been sequential. If cities grow sequentially, the cities that are initially the largest must represent a large share of the total urban population of the country in the initial periods and a relatively smaller one later. He finds that the average-rank of the fastest-growing cities tends to increase over time, a result that he interprets as evidence in favour of sequential urban growth. The key factor in generating sequential growth is the assumption of irreversible investment in physical capital in a theoretical framework with agglomeration (increasing returns to scale) and dispersion forces (the convex costs associated with the stock of installed capital). Nevertheless, in the long run, Gibrat's law holds, because all cities grow at the same rate (Cuberes, 2009).

5. Conclusions

Zipf's and Gibrat's laws are two stylized facts in urban economics. Researchers from various fields (urban economics, statistical physics and urban geography) have checked these empirical regularities considering different countries and time periods. However, there is a new mainstream in the literature that argues that random growth (or Gibrat's law) and Zipf's law correspond to the steady state (a long-run average) but that, to reach that situation, temporal episodes of different growth patterns across some cities are possible. Quoting Gabaix and Ioannides (2004), "*the casual impression of the authors is*

that in some decades, large cities grow faster than small cities, but in other decades, small cities grow faster". Therefore, the time period considered seems to be crucial. Several studies document episodes of convergence or divergence from a long-term perspective in the US, but both convergence and divergence dissipate over time and Zipf's and Gibrat's laws gradually emerge (Giesen and Südekum, 2014; Sánchez-Vidal et al., 2014; Desmet and Rappaport, 2017).

In this paper we study the evolution of the European city size distribution from a very long-term perspective (from 1300 to 1800) considering the historical data set of Bairoch et al. (1988). Using the CSN method, a Pareto-type city size distribution (power law) is rejected from 1300 to 1600. A power law is a plausible model for the city size distribution only in 1700 and 1800, although the log-normal distribution is another plausible alternative model that we cannot reject. Therefore, the support for a Pareto distribution (and thus for Zipf's law) is weaker than that found in previous papers (Dittmar, 2011). Our explanation is that, in those early periods, neither the European nor the internal national urban systems were integrated. European countries consolidated their national economies and urban systems around the end of our period (1800), which is in line with the improvement in the fit of the Pareto distribution for the pool of European cities in the last periods in our sample.

Finally, random growth of cities is unequivocally rejected using parametric and non-parametric methods. The results reveal a clear pattern of convergent growth in all periods, although, for many middle-sized cities, growth is size-independent. Thus, neither Zipf's law nor Gibrat's law holds from a long-term perspective in European cities, although the last periods (1700 and 1800) show some signs of transition to a more stable city size distribution and a consolidated urban landscape.

References

Acemoglu, D., S. Johnson, and J. Robinson. 2005. "The rise of Europe: Atlantic trade, institutional change, and economic growth." *The American Economic Review* 95(3): 546–579.

Bairoch, P., J. Batou, and P. Chèvre. 1988. *La Population Des Villes Européenes*. Geneva: Droz.

Barro, R. J., and X. Sala-i-Martin. 2004. *Economic Growth*. MIT Press. Cambridge, Massachusettes.

Black, D., and J. V. Henderson. 2003. "Urban evolution in the USA." *Journal of Economic Geography* 3: 343–372.

Bosker, M., S. Brakman, H. Garretsen, H. de Jong, and M. Schramm. 2008. "Ports, plagues and politics: explaining Italian city growth 1300–1861." *European Review of Economic History* 12: 97–131.

Bosker, M., E. Buring, and J. L. Van Zanden. 2013. "From Baghdad to London: unraveling urban development in Europe, the Middle East and North Africa, 800–1800." *Review of Economics and Statistics* 95(4): 1418–1437.

Cheshire, P. 1999. "Trends in sizes and structure of urban areas." In: *Handbook of Regional and Urban Economics*, Vol. 3, edited by P. Cheshire and E. S. Mills, 1339–1373. Amsterdam: Elsevier Science.

Cheshire, P., and S. Magrini. 2006. "Population growth in European Cities: weather matters – but only nationally." *Regional Studies* 40(1): 23–37.

Clauset, A., C. R. Shalizi, and M. E. J. Newman. 2009. "Power-law distributions in empirical data." *SIAM Review* 51(4): 661–703.

Cuberes, D. 2009. "A Model of Sequential City Growth." *The B.E. Journal of Macroeconomics*: Vol. 9: Iss. 1 (Contributions), Article 18.

Cuberes, D. 2011. "Sequential City Growth: Empirical Evidence." *Journal of Urban Economics* 69: 229–239.

De Long, J. B., and A. Shleifer. 1993. "Princes and merchants: European city growth before the industrial revolution." *Journal of Law and Economics* 36(2): 671–702.

De Vries, J. 1984. *European Urbanization 1500–1800*. Cambridge: Harvard University Press.

Desmet, K., and J. Rappaport. 2017. "The settlement of the United States, 1800–2000: the long transition towards Gibrat's law." *Journal of Urban Economics* 98: 50–68.

Dittmar, J. 2011. "Cities, markets, and growth: the emergence of Zipf's law." Working Paper. Mimeo, London School of Economics.

Duranton, G. 2007. "Urban evolutions: the fast, the slow, and the still." *American Economic Review* 97(1): 197–221.

Eeckhout, J. 2004. "Gibrat's Law for (All) Cities." American Economic Review 94(5): 1429–1451.

Fazio, G., and M. Modica. 2015. "Pareto or log-normal? Best fit and truncation in the distribution of all cities." *Journal of Regional Science* 55(5): 736–756.

Fernihough, A., and K. Hjortshøj O'Rourke. 2014. "Coal and the European industrial revolution." NBER Working Paper No. 19802.

Gabaix, X. 1999. "Zipf's law for cities: an explanation." *Quarterly Journal of Economics* 114(3): 739–767.

Gabaix, X. 2009. "Power laws in economics and finance." *Annual Review of Economics* 1: 255–294.

Gabaix, X., and R. Ibragimov. 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., and Y. M. Ioannides. 2004. "The evolution of city size distributions." In: *Handbook of Urban and Regional Economics*, Vol. 4, edited by J. V. Henderson and J. F. Thisse, 2341–2378. Amsterdam: Elsevier Science.

Gibrat, R. 1931. *Les inégalités économiques*. Paris: Librairie du Recueil Sirey.

Giesen, K., and J. Südekum. 2011. "Zipf's law for cities in the regions and the country." *Journal of Economic Geography*, 11(4): 667–686.

Giesen, K., and J. Südekum. 2014. "City age and city size." *European Economic Review* 71: 193–208.

Giesen, K., A. Zimmermann, and J. Suedekum. 2010. "The size distribution across all cities – double Pareto lognormal strikes." *Journal of Urban Economics*, 68: 129–137.

Goldstein, M. L., S. A. Morris, and G. G. Yen. 2004. "Problems with fitting to the power-law distribution." *The European Physical Journal B – Condensed Matter* 41(2): 255–258.

González-Val, R., E. Viladecans-Marsal, and D. A. Tirado-Fabregat. 2017. "Market potential and city growth: Spain 1860-1960." *Cliometrica*, 11(1): 31–61.

Hohenberg, P. M., and L. H. Lees. 1985. *The Making of Urban Europe 1000-1950*. London: Harvard University Press.

Härdle, W. 1990. *Applied Nonparametric Regression*. Econometric Society Monographs. Cambridge: Cambridge University Press.

Henderson, J. V., and H. G. Wang . 2007. "Urbanization and city growth: The role of institutions." *Regional Science and Urban Economics* 37: 283–313.

Ioannides, Y. M., and H. G. Overman. 2003. "Zipf's law for cities: an empirical examination." *Regional Science and Urban Economics* 33: 127–137.

Ioannides, Y. M., and S. Skouras. 2013. "US city size distribution: Robustly Pareto, but only in the tail." *Journal of Urban Economics* 73: 18–29.

Kagan, Y. Y. 1997. "Earthquake size distribution and earthquake insurance." *Communications in Statistics. Stochastic Models* 13(4): 775–797.

Kim, S. 2000. "Urban development in the United States." *Southern Economic Journal* 66, 855–880.

Kim, S., and R. A. Margo. 2004. "Historical perspectives on U.S. economic geography." In: *Handbook of Urban and Regional Economics*, Vol. 4, edited by J. V. Henderson and J. F. Thisse, 2982–3019.

Krugman, P. 1996. "Confronting the Mystery of Urban Hierarchy." *Journal of the Japanese and International Economies* 10: 399–418.

Levy, M. 2009. "Gibrat's Law for (all) Cities: A Comment." *American Economic Review* 99(4): 1672–1675.

Michaels, G., F. Rauch, and S. J. Redding. 2012. "Urbanization and structural transformation." *The Quarterly Journal of Economics* 127(2): 535–586.

Nishiyama, Y., S. Osada, and Y. Sato. 2008. "OLS estimation and the t test revisited in rank-size rule regression." *Journal of Regional Science* 48(4): 691–715.

Nitsch, V. 2005. "Zipf zipped." *Journal of Urban Economics* 57: 86–100.

Pisarenko, V. F. 1998. "Non-linear growth of cumulative flood losses with time." *Hydrological Processes* 12(3): 461–470.

Roberts, D. C., and D. L. Turcotte. 1998. "Fractality and self-organized criticality of wars." *Fractals* 6(4): 351–357.

Rosen, K. T., and M. Resnick. 1980. "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy." *Journal of Urban Economics* 8: 165–186.

Rossi-Hansberg, E., and M. L. J. Wright. 2007. "Urban structure and growth." *Review of Economic Studies* 74: 597–624.

Russell, J. 1972. *Medieval Regions and Their Cities*. Bloomington: Indiana University.

Sánchez-Vidal M., R. González-Val and E. Viladecans-Marsal. 2014. "Sequential city growth in the US: Does age matter?" *Regional Science and Urban Economics* 44(1): 29–37.

Schmidheiny, K., and J. Suedekum. 2015. "The pan-European population distribution across consistently defined functional urban areas." *Economics Letters* 133: 10–13.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. 1st Edition. Chapman and Hall/CRC, Cleveland.

Soo, K. T. 2005. "Zipf's Law for cities: A cross-country investigation." *Regional Science and Urban Economics* 35: 239–263.

Urzúa, C. M. 2000. "A simple and efficient test for Zipf's law." *Economics Letters* 66: 257–260.

Voigtländer, N., and H.-J. Voth. 2013. "The three horsemen of riches: plague, war, and urbanization in early modern Europe." *Review of Economic Studies* 80(2): 774–811.

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

Table 1. Descriptive statistics by year

Year	Cities	Cities (Raw Bairoch et al. (1988) data)	Mean	Standard deviation	Minimum	Maximum
1300	438	416	11,594.03	17,058.14	1,000	150,000
1400	527	339	9,879.57	17,596.57	1,000	275,000
1500	717	538	9,032.961	14,128.17	1,000	225,000
1600	952	762	10,155.5	19,538.77	1,000	300,000
1700	1,180	994	10,055.1	27,877.18	1,000	575,000
1800	1,623	1,623	12,468.88	33,941.05	1,000	948,000

Source: Bairoch et al. (1988).

Table 2. Power law fit

Year	Lower bound $\underline{S}$	Observations $S \geq \underline{S}$	Pareto exponent $\hat{a}$	Standard error	Power law test Statistic	p-value	Power law vs. log-normal Statistic	p-value
1300	9,000	163	2.318	0.103	0.075	0.000	-1.642	0.101
1400	8,000	176	2.361	0.103	0.071	0.006	-1.038	0.299
1500	12,333	116	2.533	0.142	0.067	0.084	-0.811	0.417
1600	10,400	203	2.404	0.099	0.029	0.014	-0.610	0.542
1700	19,000	106	2.400	0.136	0.027	0.732	-0.181	0.856
1800	21,000	145	2.357	0.113	0.026	0.288	-0.280	0.780

Note: The lower bound and the Pareto exponent are estimated using the CSN methodology. The power law test is a goodness-of-fit test. H_0 is that there is power law behaviour for $S_i \geq \underline{S}$. The power law vs. log-normal test is Vuong's model selection test, based on the normalized log-likelihood ratio. H_0 is that the two distributions are equally far from the true distribution, while H_A is that one of the test distributions is closer to the true distribution.

Table 3. Growth and initial population

	(1) 1300–1400	(2) 1400–1500	(3) 1500–1600	(4) 1600–1700	(5) 1700–1800	(6) Pool 1300–1800	(7) Panel 1300–1800
ln(Population)$_{t-1}$	-0.223***	-0.204***	-0.127***	-0.160***	-0.223***	-0.198***	-0.613***
	(0.021)	(0.042)	(0.028)	(0.039)	(0.030)	(0.013)	(0.040)
d_Austria	-0.696***	-0.254***	-0.244***	0.485***	-0.165***	0.033	-
	(0.009)	(0.079)	(0.029)	(0.045)	(0.052)	(0.036)	
d_Belgium	-0.734***	-0.126	-0.551***	0.339***	-0.250***	-0.089**	-
	(0.025)	(0.082)	(0.032)	(0.051)	(0.045)	(0.039)	
d_Denmark	-	-0.138	0.384***	-0.283***	-0.316***	-0.189***	-
		(0.092)	(0.028)	(0.060)	(0.035)	(0.028)	
d_France	-0.920***	-0.138	-0.309***	0.312***	-0.282***	-0.079**	-
	(0.019)	(0.079)	(0.036)	(0.065)	(0.052)	(0.037)	
d_Germany	-0.865***	-0.217***	-0.381***	0.155***	-0.190***	-0.109***	-
	(0.012)	(0.067)	(0.024)	(0.043)	(0.038)	(0.035)	
d_Great Britain	-1.139***	-0.350***	-0.344***	0.339***	0.383***	0.136***	-
	(0.007)	(0.053)	(0.016)	(0.032)	(0.025)	(0.019)	
d_Greece	-0.670***	-0.494***	-0.165***	0.069	-0.485***	-0.152***	-
	(0.032)	(0.114)	(0.044)	(0.083)	(0.062)	(0.049)	
d_Ireland	-0.843***	-0.366***	-0.635***	0.810***	0.313***	0.061	-
	(0.002)	(0.060)	(0.020)	(0.035)	(0.054)	(0.039)	
d_Italy	-1.002***	-0.006	-0.254***	0.236***	-0.219***	-0.053	-
	(0.016)	(0.071)	(0.036)	(0.059)	(0.050)	(0.038)	
d_the Netherlands	-0.840***	0.184**	-0.115***	0.355***	-0.459***	-0.040	-
	(0.009)	(0.072)	(0.034)	(0.071)	(0.052)	(0.034)	
d_Norway	-1.207***	-0.770***	-0.244***	0.379***	-0.142***	-0.146***	-
	(0.009)	(0.056)	(0.008)	(0.018)	(0.022)	(0.020)	
d_Portugal	-0.719***	-0.241**	-0.356***	0.233***	-0.370***	-0.117***	-
	(0.015)	(0.087)	(0.029)	(0.057)	(0.051)	(0.039)	
d_Spain	-0.842***	-0.442***	-0.348***	-0.215***	-0.098**	-0.210***	-
	(0.030)	(0.100)	(0.038)	(0.054)	(0.036)	(0.035)	
d_Sweden	-1.062***	-0.503***	-0.554***	0.526***	-0.402***	-0.206***	-
	(0.015)	(0.020)	(0.008)	(0.005)	(0.028)	(0.028)	
d_Switzerland	-0.668***	-0.235***	-0.438***	0.323***	-0.331***	-0.106***	-
	(0.003)	(0.054)	(0.017)	(0.040)	(0.041)	(0.036)	
Country fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	No
Time fixed-effects	No	No	No	No	No	Yes	Yes
City fixed-effects	No	No	No	No	No	No	Yes
Observations	436	527	711	950	1,174	3,798	3,798
R^2	0.223	0.302	0.104	0.238	0.337	0.240	0.380

Notes: All the models include a constant. Coefficient (robust standard errors). Standard errors clustered by country. Finland is the base category in the country fixed effects. Columns (1)–(5): cross-sectional OLS regressions. Column (6): pool 1300–1800 OLS regression. Column (7): panel 1300–1800 with city fixed effects. Significant at the *10%, ** 5%, and *** 1% level.

Figure 1. European city size distribution from 1300 to 1800
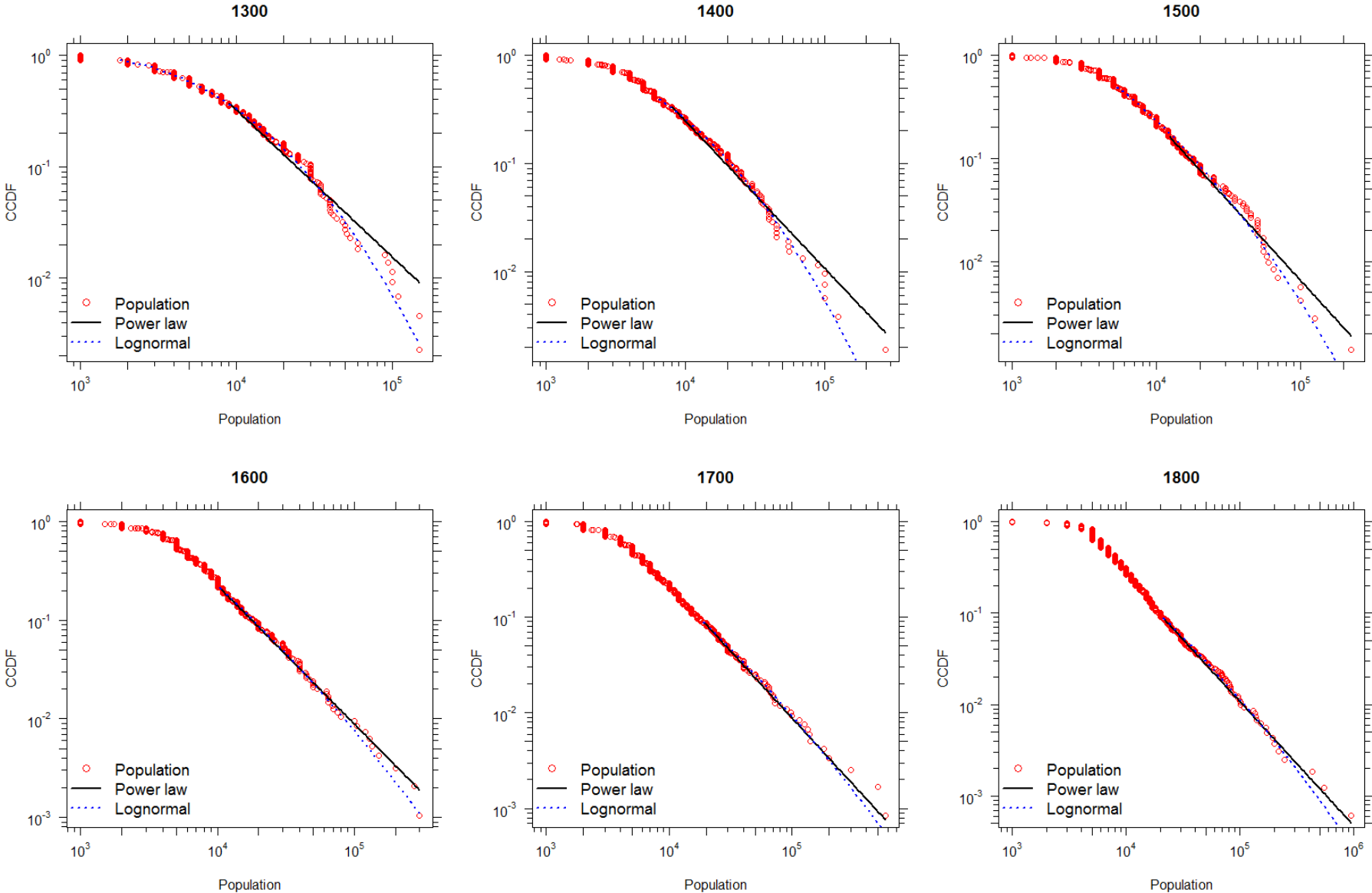
Note: The data are plotted as a complementary cumulative distribution function (CCDF), $\Pr(S \geq \underline{S})$.

Figure 2. Non-parametric estimates of growth by century
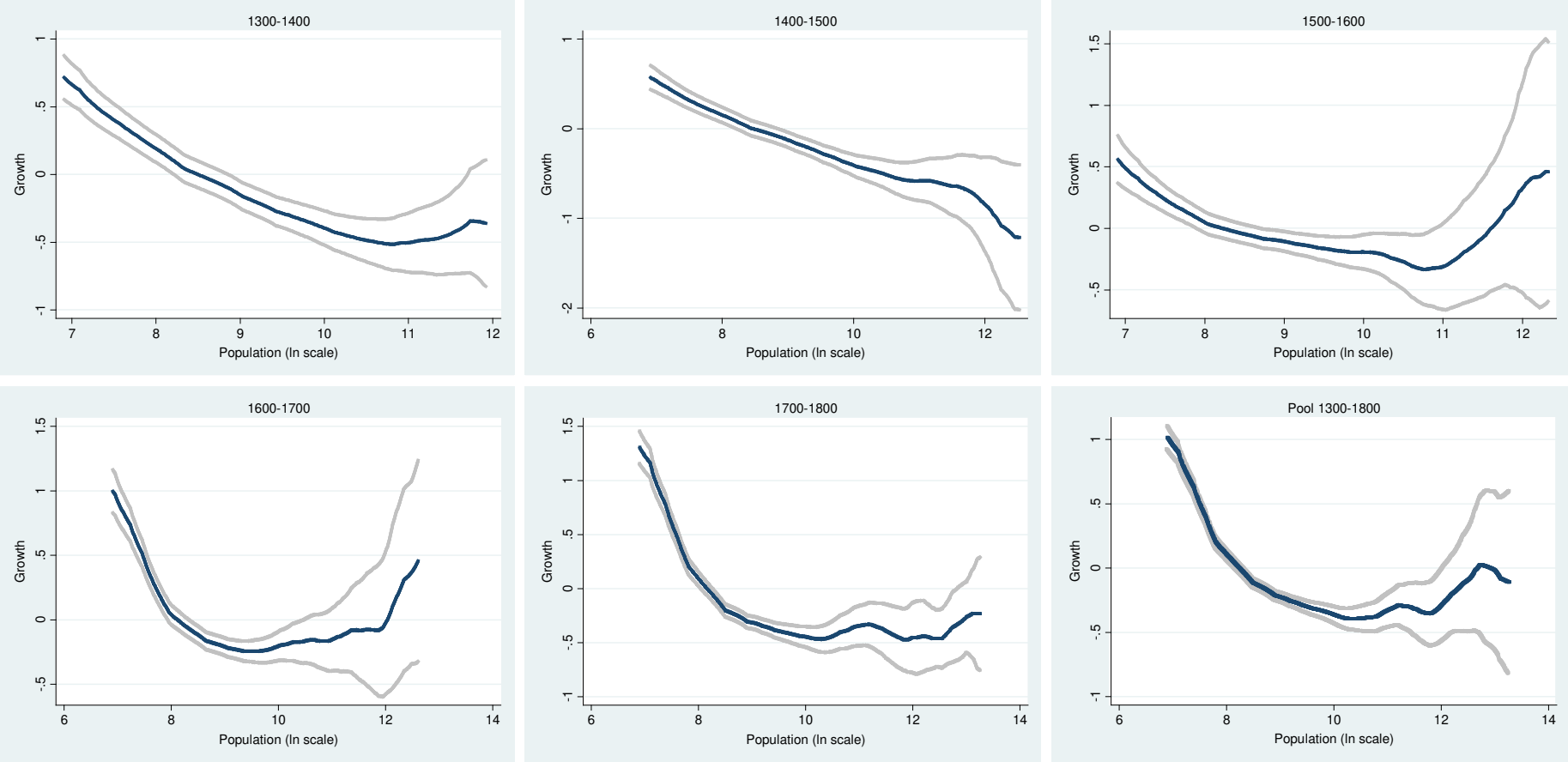
Appendix A: Raw Bairoch et al. (1988) data

Table A1. Descriptive statistics by year

Year	Cities	Mean	Standard deviation	Minimum	Maximum
1300	416	11,855.77	17,360.76	1,000	150,000
1400	339	11,646.02	21,464.81	1,000	275,000
1500	538	10,223.05	16,053.29	1,000	225,000
1600	762	11,393.70	21,622.67	1,000	300,000
1700	994	10,687.12	30,297.65	1,000	575,000
1800	1,623	12,468.88	33,941.05	1,000	948,000

Source: Bairoch et al. (1988).

Table A2. Power law fit

Year	Lower bound $\underline{S}$	Observations $S \geq \underline{S}$	Pareto exponent $\hat{a}$	Standard error	Power law test Statistic	p-value	Power law vs. log-normal Statistic	p-value
1300	9,000	160	2.322	0.105	0.075	0.000	-1.569	0.117
1400	17,000	67	2.592	0.195	0.098	0.020	-0.631	0.528
1500	11,000	130	2.463	0.128	0.074	0.034	-1.009	0.313
1600	13,000	150	2.424	0.116	0.033	0.028	-0.575	0.565
1700	21,000	93	2.413	0.147	0.029	0.770	-0.119	0.906
1800	21,000	145	2.357	0.113	0.026	0.288	-0.280	0.780

Note: The lower bound and the Pareto exponent are estimated using the CSN methodology. The power law test is a goodness-of-fit test. H_0 is that there is power law behaviour for $S_i \geq \underline{S}$. The power law vs. log-normal test is Vuong's model selection test, based on the normalized log-likelihood ratio. H_0 is that the two distributions are equally far from the true distribution, while H_A is that one of the test distributions is closer to the true distribution.

Table A3. Growth and initial population

|	(1) 1300–1400	(2) 1400–1500	(3) 1500–1600	(4) 1600–1700	(5) 1700–1800	(6) Pool 1300–1800	(7) Panel 1300–1800
ln(Population)$_{t-1}$	-0.204***	-0.183***	-0.180***	-0.155***	-0.235***	-0.203***	-0.658***
	(0.034)	(0.022)	(0.042)	(0.030)	(0.024)	(0.014)	(0.062)
d_Austria	-0.711***	-0.144**	-0.111**	0.397***	-0.112**	0.061	-
	(0.064)	(0.066)	(0.043)	(0.042)	(0.044)	(0.038)	
d_Belgium	-0.732***	-0.088*	-0.449***	0.487***	-0.166***	0.001	-
	(0.045)	(0.046)	(0.064)	(0.049)	(0.038)	(0.035)	
d_Denmark	-	-0.185***	0.436***	-0.291***	-0.302***	-0.171***	-
		(0.049)	(0.043)	(0.046)	(0.028)	(0.029)	
d_France	-1.093***	-0.054	-0.050	0.435***	-0.208***	-0.024	-
	(0.048)	(0.049)	(0.075)	(0.063)	(0.043)	(0.025)	
d_Germany	-0.849***	-0.170***	-0.237***	0.076*	-0.020	-0.030	-
	(0.035)	(0.042)	(0.047)	(0.040)	(0.029)	(0.032)	
d_Great Britain	-1.224***	-0.312***	-0.180***	0.401***	0.435***	0.258***	-
	(0.024)	(0.034)	(0.027)	(0.025)	(0.019)	(0.011)	
d_Greece	-0.539***	-1.048***	0.735***	-0.150	-0.172***	0.026	-
	(0.070)	(0.081)	(0.082)	(0.096)	(0.058)	(0.050)	
d_Ireland	-0.523***	-0.825***	-1.298***	1.481***	0.768***	0.549***	-
	(0.044)	(0.060)	(0.035)	(0.013)	(0.043)	(0.023)	
d_Italy	-1.103***	-0.021	-0.129*	0.223***	-0.163***	-0.036	-
	(0.029)	(0.052)	(0.067)	(0.047)	(0.040)	(0.033)	
d_the Netherlands	-0.833***	0.218***	0.090	0.420***	-0.426***	0.005	-
	(0.013)	(0.038)	(0.055)	(0.060)	(0.042)	(0.030)	
d_Norway	-	-	-0.230***	0.377***	-0.133***	-0.014	-
			(0.012)	(0.014)	(0.017)	(0.008)	
d_Portugal	-0.001	-0.103	-0.033	0.210***	-0.053	0.126**	-
	(0.053)	(0.076)	(0.062)	(0.063)	(0.054)	(0.045)	
d_Spain	-0.769***	-0.474***	-0.216***	-0.252***	-0.069**	-0.187***	-
	(0.063)	(0.058)	(0.059)	(0.042)	(0.028)	(0.034)	
d_Sweden	-1.048***	-0.513***	-0.569***	0.526***	-0.390***	-0.202***	-
	(0.025)	(0.010)	(0.012)	(0.004)	(0.023)	(0.028)	
d_Switzerland	-0.494***	-0.188***	-0.415***	0.407***	-0.217***	-0.019	-
	(0.004)	(0.033)	(0.032)	(0.034)	(0.035)	(0.032)	
Country fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	No
Time fixed-effects	No	No	No	No	No	Yes	Yes
City fixed-effects	No	No	No	No	No	No	Yes
Observations	241	243	421	669	988	2,562	2,562
R^2	0.217	0.284	0.191	0.301	0.379	0.295	0.438

Notes: All the models include a constant. Coefficient (robust standard errors). Standard errors clustered by country. Finland is the base category in the country fixed effects. Columns (1)–(5): cross-sectional OLS regressions. Column (6): pool 1300–1800 OLS regression. Column (7): panel 1300–1800 with city fixed effects. Significant at the *10%, ** 5%, and *** 1% level.

Figure A1. Non-parametric estimates of growth
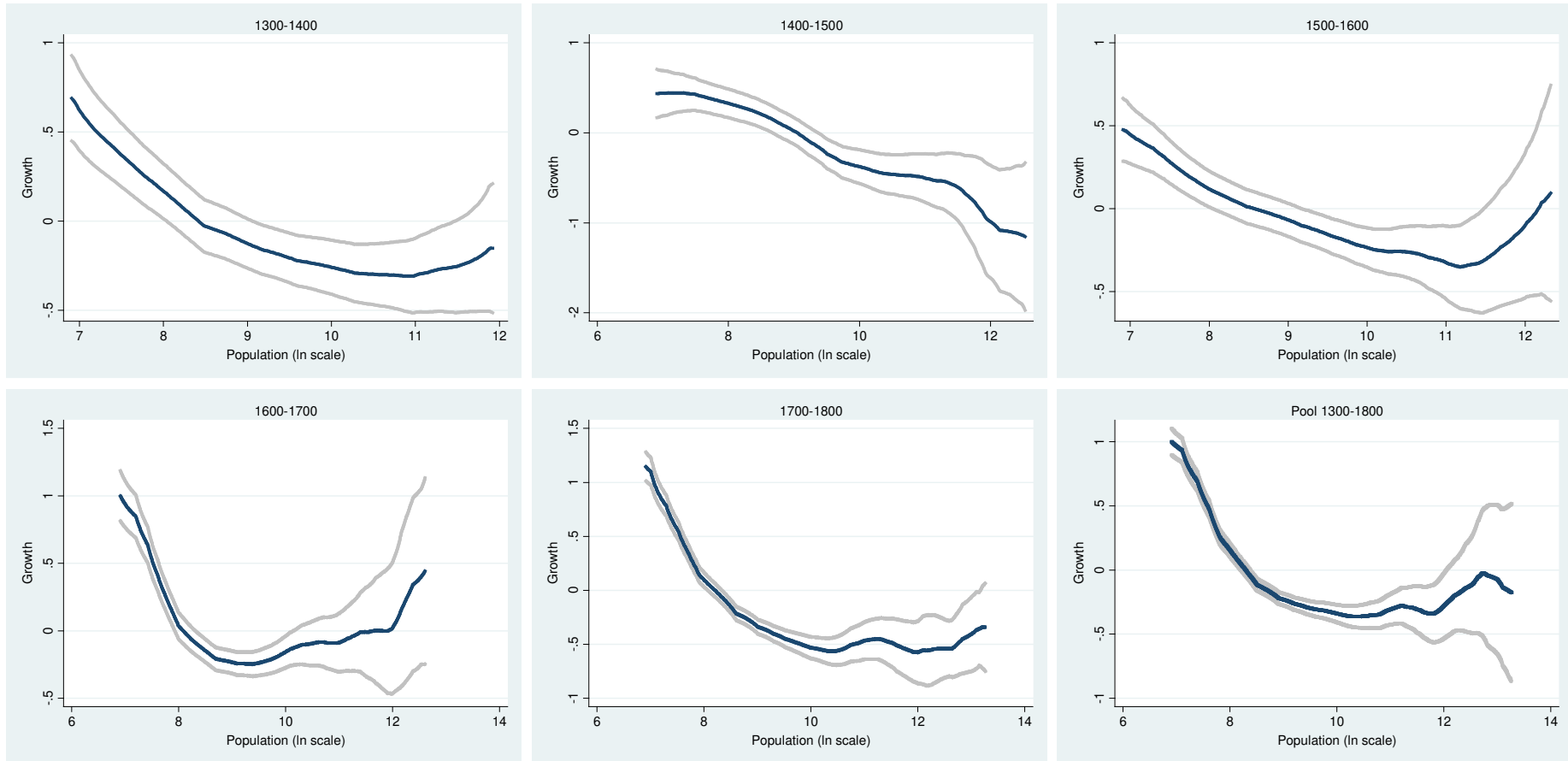
Appendix B: Results by country

Table B1. Cities by country

	1300	1400	1500	1600	1700	1800
Austria	3	4	9	13	13	17
Belgium	18	28	38	42	49	70
Denmark	1	1	2	7	8	10
Finland	-	1	1	3	6	7
France	88	103	116	133	199	329
Germany	78	105	142	172	191	233
Great Britain	33	36	60	76	147	159
Greece	6	6	7	7	8	19
Ireland	14	13	14	14	15	20
Italy	113	129	156	221	252	368
Luxembourg	-	-	-	-	-	1
The Netherlands	15	18	32	33	50	58
Norway	2	2	2	3	6	9
Portugal	10	10	28	26	26	42
Spain	42	47	78	170	174	243
Sweden	8	12	18	18	20	20
Switzerland	7	12	14	14	16	18
Total	438	527	717	952	1180	1623

Note: The samples include linear interpolated values to fill gaps in the data.

Table B2. Estimated lower bounds by country

	1300	1400	1500	1600	1700	1800
Austria	1,000	5,000	4,000	4,000	5,000	7,000
	(3)	(3)	(7)	(10)	(8)	(8)
Belgium	4,000	3,000	5,666	3,500	3,000	9,000
	(18)	(24)	(19)	(33)	(40)	(20)
Denmark	-	-	3,000	6,000	2,000	4,000
			(2)	(3)	(8)	(6)
Finland	-	-	-	1,000	1,000	4,000
				(3)	(6)	(2)
France	4,000	17,500	10,500	7,000	13,000	16,000
	(68)	(17)	(36)	(73)	(48)	(47)
Germany	9,000	9,000	4,750	9,000	7,500	19,000
	(27)	(29)	(71)	(36)	(45)	(24)
Great Britain	7,000	3,500	7,000	4,333	6,000	18,000
	(12)	(20)	(9)	(31)	(32)	(26)
Greece	20,000	8,000	13,600	9,000	12,000	3,000
	(3)	(5)	(3)	(6)	(5)	(18)
Ireland	4,000	3,500	3,800	5,000	5,000	16,000
	(6)	(9)	(9)	(7)	(13)	(7)
Italy	8,666	3,666	6,500	6,500	8,500	18,000
	(48)	(83)	(73)	(118)	(83)	(36)
The Netherlands	3,000	6,000	14,000	4,000	13,000	16,000
	(13)	(9)	(5)	(29)	(15)	(12)
Norway	3,000	2,000	1,000	1,000	4,000	4,000
	(2)	(2)	(2)	(3)	(3)	(6)
Portugal	5,000	3,600	7,000	6,000	6,666	8,000
	(6)	(10)	(7)	(15)	(11)	(11)
Spain	6,000	5,666	8,000	12,000	5,500	17,000
	(38)	(38)	(37)	(27)	(55)	(24)
Sweden	2,000	2,000	3,000	5,000	5,000	9,000
	(3)	(5)	(3)	(2)	(6)	(4)
Switzerland	3,000	2,000	2,000	5,500	3,000	5,000
	(5)	(12)	(13)	(5)	(15)	(13)

Note: The lower bound is estimated using the CSN methodology. The number between parentheses is n (number of cities for $S_i \geq \underline{S}$). Luxembourg is excluded due to data limitations.

Table B3. Pareto exponents by country

	1300	1400	1500	1600	1700	1800
Austria	1.651***	2.741***	2.534***	2.592***	2.104***	2.026***
	(0.376)	(1.005)	(0.580)	(0.503)	(0.390)	(0.363)
Belgium	2.069***	1.922***	2.158***	2.251***	2.032***	2.417***
	(0.252)	(0.188)	(0.266)	(0.218)	(0.163)	(0.317)
Denmark	-	-	2.661***	2.373***	2.340***	2.337***
			(1.175)	(0.793)	(0.474)	(0.546)
Finland	-	-	-	3.162***	4.721***	2.977***
				(1.249)	(1.522)	(1.398)
France	1.970***	3.229***	2.535***	2.318***	2.621***	2.536***
	(0.118)	(0.541)	(0.256)	(0.154)	(0.234)	(0.224)
Germany	3.060***	3.141***	2.637***	2.773***	2.599***	2.868***
	(0.396)	(0.398)	(0.194)	(0.296)	(0.238)	(0.381)
Great Britain	3.517***	3.217***	3.219***	2.649***	2.601***	2.279***
	(0.727)	(0.496)	(0.740)	(0.296)	(0.283)	(0.251)
Greece	3.633**	2.127***	8.871*	2.471***	3.719***	1.822***
	(1.520)	(0.504)	(4.545)	(0.601)	(1.216)	(0.194)
Ireland	2.972***	3.746***	3.802***	4.368***	2.538***	2.267***
	(0.805)	(0.916)	(0.934)	(1.274)	(0.427)	(0.479)
Italy	2.104***	1.979***	2.318***	2.431***	2.464***	2.313***
	(0.159)	(0.107)	(0.154)	(0.131)	(0.161)	(0.219)
The Netherlands	2.435***	5.079***	9.938**	1.926***	2.350***	2.645***
	(0.398)	(1.360)	(3.999)	(0.172)	(0.349)	(0.475)
Norway	3.359**	2.596**	2.027***	2.081***	3.632**	2.838***
	(1.669)	(1.129)	(0.727)	(0.625)	(1.520)	(0.750)
Portugal	2.703***	2.288***	2.687***	2.919***	2.364***	2.330***
	(0.695)	(0.407)	(0.637)	(0.496)	(0.411)	(0.401)
Spain	2.147***	2.163***	2.501***	2.459***	2.297***	2.473***
	(0.186)	(0.189)	(0.247)	(0.281)	(0.175)	(0.301)
Sweden	4.696**	3.343***	3.208**	4.401*	3.284***	2.534***
	(2.136)	(1.048)	(1.275)	(2.406)	(0.933)	(0.767)
Switzerland	3.076***	2.723***	2.392***	3.249***	2.668***	3.270***
	(0.929)	(0.498)	(0.386)	(1.006)	(0.431)	(0.630)

Note: The Pareto exponent (standard error) is estimated using the CSN methodology. The results for data sets with $n < 50$ (see Table B2) should be treated with caution due to the small-sample bias. Luxembourg is excluded due to data limitations. Significant at the *10%, ** 5%, and *** 1% level.

Table B4. Power law test by country

	1300		1400		1500		1600		1700		1800	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Austria	0.272	0.262	0.259	0.832	0.143	0.542	0.239	0.122	0.191	0.996	0.194	0.990
Belgium	0.093	0.028	0.086	0.324	0.102	0.120	0.123	0.060	0.091	0.018	0.119	0.326
Denmark	-	-	-	-	0.365	0.062	0.365	0.658	0.184	0.038	0.162	0.152
Finland	-	-	-	-	-	-	0.365	0.262	0.210	0.000	0.365	0.060
France	0.049	0.000	0.175	0.820	0.100	0.584	0.064	0.276	0.044	0.732	0.100	0.808
Germany	0.088	0.404	0.087	0.088	0.079	0.006	0.099	0.334	0.106	0.434	0.092	0.636
Great Britain	0.081	0.550	0.091	0.144	0.152	0.842	0.143	0.466	0.054	0.442	0.065	0.806
Greece	0.224	0.822	0.217	0.172	0.364	0.652	0.211	0.594	0.319	0.894	0.100	0.028
Ireland	0.222	0.484	0.123	0.976	0.240	0.168	0.195	0.532	0.097	0.942	0.205	0.518
Italy	0.054	0.018	0.060	0.012	0.055	0.442	0.078	0.000	0.056	0.174	0.076	0.328
the Netherlands	0.209	0.160	0.094	0.028	0.172	0.564	0.147	0.000	0.165	0.668	0.212	0.996
Norway	0.365	0.036	0.365	0.066	0.365	0.068	0.365	0.952	0.365	0.822	0.215	0.686
Portugal	0.136	0.958	0.127	0.950	0.251	0.806	0.132	0.324	0.108	0.642	0.122	0.730
Spain	0.077	0.238	0.060	0.300	0.112	0.026	0.124	0.412	0.080	0.156	0.126	0.672
Sweden	0.365	0.000	0.164	0.000	0.316	0.258	0.365	0.018	0.149	0.006	0.224	0.672
Switzerland	0.195	0.254	0.093	0.000	0.183	0.076	0.187	0.662	0.182	0.240	0.140	0.296

Note: The power law test is a goodness-of-fit test. H_0 is that there is power law behaviour for $S_i \geq \underline{S}$. Luxembourg is excluded due to data limitations.

Table B5. Power law vs. log-normal test by country

	1300		1400		1500		1600		1700		1800	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Austria	-0.204	0.839	0.082	0.935	-0.251	0.802	0.625	0.532	0.274	0.784	0.461	0.645
Belgium	-0.424	0.671	-0.381	0.703	-0.427	0.669	-0.215	0.830	-0.331	0.741	-0.178	0.858
Denmark	-	-	-	-	0.015	0.988	0.408	0.683	0.330	0.742	0.400	0.689
Finland	-	-	-	-	-	-	0.299	0.765	2.390	0.017	0.015	0.988
France	-1.482	0.138	1.530	0.126	0.158	0.874	-0.484	0.628	-1.876	0.061	0.302	0.763
Germany	-0.249	0.803	-0.789	0.430	-0.935	0.350	-0.737	0.461	-0.474	0.636	0.247	0.805
Great Britain	0.337	0.736	0.800	0.424	0.289	0.773	0.665	0.506	-50.290	0.000	-32.967	0.000
Greece	0.090	0.928	-0.495	0.620	0.170	0.865	-0.157	0.875	0.197	0.844	-0.993	0.320
Ireland	-0.267	0.789	0.205	0.837	-0.334	0.739	0.670	0.503	0.271	0.787	0.453	0.651
Italy	-1.261	0.207	-1.228	0.219	-0.268	0.788	-1.627	0.104	1.058	0.290	0.161	0.872
the Netherlands	-0.504	0.614	-0.605	0.545	0.362	0.717	-1.516	0.129	-0.130	0.896	-0.593	0.553
Norway	0.016	0.987	0.016	0.987	0.017	0.986	-0.070	0.944	0.085	0.932	-0.217	0.828
Portugal	0.312	0.755	0.349	0.727	0.984	0.325	0.661	0.508	0.887	0.375	0.982	0.326
Spain	-0.382	0.703	-0.506	0.613	-0.552	0.580	-0.149	0.881	-0.248	0.804	-0.185	0.853
Sweden	-0.271	0.786	-0.521	0.602	-0.226	0.821	0.015	0.988	3.131	0.002	1.795	0.073
Switzerland	-0.520	0.603	-0.152	0.879	-0.811	0.417	-0.161	0.872	-0.320	0.749	0.379	0.705

Note: The power law vs. log-normal test is Vuong's model selection test, based on the normalized log-likelihood ratio. H_0 is that the two distributions are equally far from the true distribution, while H_A is that one of the test distributions is closer to the true distribution. Luxembourg is excluded due to data limitations.