Zipf's Law for Cities and the Double Pareto Lognormal Distribution

Von der Mercator School of Management
- Fakultät für Betriebswirtschaftslehre - der Universität Duisburg-Essen
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft (Dr. rer. oec.)
genehmigte Dissertation

von

Kristian Rüdiger Giesen

aus

Mülheim an der Ruhr

Referent:	Prof. Dr. Südekum
Korreferent:	Prof. Dr. Prinz
Tag der mündlichen Prüfung:	30.07.2012

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Introduction

In 1913, the German geographer Felix Auerbach made a considerable observation. He discovered that for the largest cities of a country, the product of their size and their rank in the urban hierarchy roughly equals a constant. This regularity was picked up by the linguist George Zipf in 1949, who showed that the usage of words within a text follows the same mechanism and can be transferred to a variety of different objects like the sizes of firms or incomes. Zipf reformulated this regularity in terms of the rank-size rule, which indicates that at least for the largest cities, their size is inversely proportional to their rank. If the rank-size rule holds, a city's size S_i is determined by its rank R_i in the urban hierarchy according to $S_i/S_j = R_j/R_i$. This implies the country's largest city being approximately twice as large as the second largest city, or the third largest city being 5/3 times as large as the fifth largest city etc. Using the insights of probability theory, the rank-size rule means that the largest cities follow a Pareto distribution with a slope coefficient of unity. Those two statements about city sizes, being Pareto and the unity slope coefficient, have attracted the attention of generations of scientists. Today, this phenomenon is one of the best-studied topics in Urban Economics and Regional Science. Over the last few decades, dozens of studies have analyzed city size distributions in almost every country. The empirical evidence is so overwhelming that it achieved the status of an empirical law: Zipf's law.

This doctoral thesis contributes to Zipf's law and city size distributions in general, theoretically and empirically in various aspects. From the empirical side, this thesis discovers various regularities across space, which are shown to be crucial for understanding cities with respect to their sizes and their growth process. From the theoretical side, this thesis provides a micro-founded model of urban growth and endogenous city creation. The model is able to solve various unsettled disputes and controversies in the surrounding literature and, most importantly, explains in detail why city size distributions across countries share the same functional form.

To introduce the reader to the topic, figure 1 shows the typical appearance of Zipf's law. The figure shows rank and size of the largest 150 metropolitan areas of the United States (US) in 2010.[1] The rank-size plot is the typical tool to analyze city size distribution and can be understood as being similar to a counter-cumulative distribution plot. In the left picture, the axes are in normal scales and cities show a hyperbolic shape, which already indicates a Pareto distribution. In the right picture, the axes are in log scales, where the linear shape reconfirms the Pareto distribution hypothesis. The lit-

[1] The data is taken from the US Census 2010.

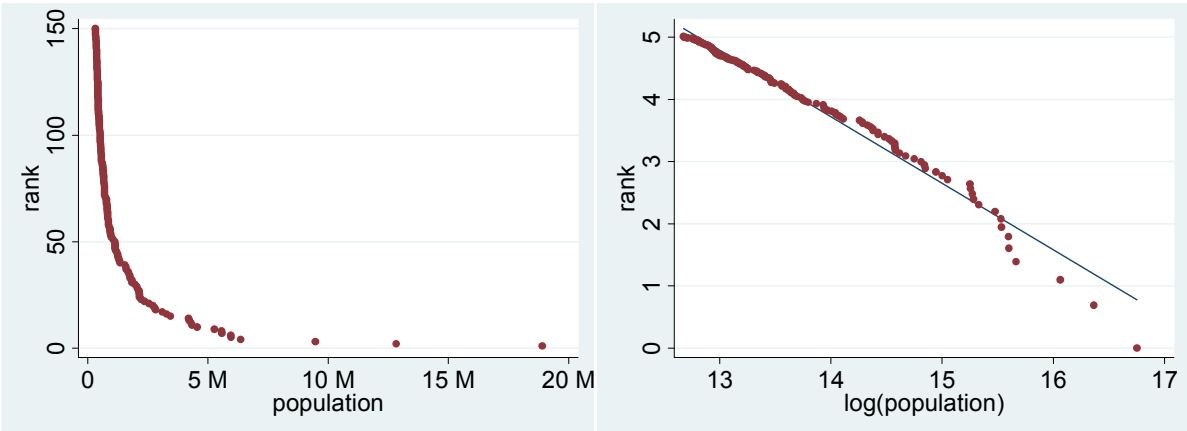

Figure 1.1: The upper tail US city size distribution in 2000 (metropolitan data).

erature provides various econometric methods to estimate, measure or test the validity and implications of Zipf's law (for a discussion of the various techniques see Gabaix and Ioannides (2004) or Gabaix and Ibragimov (2011)). One less sophisticated but effective method to measure the slope coefficient and confirm the hypothesized Pareto distribution is to use an ordinary least square regression. The result for the US case is the following:

$$log(Rank) = 18.68 - 1.069\, log(Size). \tag{1.1}$$

The linear shape of the scatterplot, and thereby the Pareto distribution, is supported by a $R^2 > 0.97$ of the linear regression. Furthermore, the slope coefficient is extremely close to unity. With a standard error of 0.13 the unity coefficient of Zipf's law cannot formally be rejected. Comprehensive studies on Zipf's law by Rosen and Resnick (1980), Soo (2005) and Nitsch (2005) confirm the Pareto distribution to be a good approximation for the size distribution of the largest cities, and the Zipf coefficient to be at least close to unity, in almost all countries. The exceptional fit of Zipf's law, leaves some researchers utterly impressed, for example Krugman (1996): "We are unused to seeing regularities this exact in economics. It is so exact that I find it spooky".

But still, Zipf's law is, and has always been, subject to controversy. Many authors, like Gabaix and Ioannides (2004) or Eeckhout (2004) have argued that there are significant deviations from Zipf's law. One of them is the typical deviation of the very largest cities from the perfect Zipf fit, as also confirmed by the right picture in figure 1. Moreover, like this thesis shows, evidence for or against Zipf's law can easily be manipulated and accordingly some researchers refuse to accept the systematic behavior of cities across countries as an empirical law, e.g. Henderson (1995): "In general and on average, the rank-size rule simply does not hold".

Besides the unsettled empirical dispute on whether Zipf's law holds or not, there is a huge theoretical literature on the question of *why* Zipf's law should hold. Various at-

tempts have been made to formulate economic models able to predict a Zipfian-Pareto distribution with the unity slope coefficient. Fujita et al. (1999) note: "We must acknowledge that it poses a real intellectual challenge to our understanding of cities [...] nobody has come up with a plausible story about the process that generates the rank-size rule [...]".

While it is unclear whether the theoretical literature has settled yet on a unifying explanation, the prevailing, most promising attempts rely on stochastic processes and random growth to replicate the empirical findings. It turns out that plausible explanations are based on a size-invariant random urban growth process, which is known as Gibrat's law. Gibrat's law postulates that cities grow randomly with the same expected growth rate and the same expected variance. Under Gibrat's law, a city might grow or shrink in one year and do the opposite in the other. There might be some cities that grow or shrink with a high pace, some with a low pace and some might not even change their size at all. The requirement, in order to comply with Gibrat's law, is just that there will be no systematic growth pattern across sizes. Empirical evidence for Gibrat's law is provided by many studies and for many countries, for example Eaton and Eckstein (1997) for France and Japan, Ioannides and Overman (2003) and Eeckhout (2004) for the US.

Building on Champernowne (1953) and Levy and Solomon (1996), Gabaix (1999) provides a stochastic proof that explains Zipf's law as the outcome of an urban growth process that resembles a modified version of Gibrat's law. In detail, he poses two propositions on the coherence between Zipf's and Gibrat's law. The first proposition postulates that Zipf's law results from Gibrat's law. The second proposition says that if the single regions within a country follow Gibrat's law, possibly with different parameters, then Zipf's law is not only satisfied for each region but automatically for the country as well. Nowadays, those propositions, along with their proof, are widely accepted in the literature. Moreover, most economic models with a focus on Zipf's law, like Gabaix (1999) or Rossi-Hansberg and Wright (2007), aim at replicating Gibrat's law and generalized versions thereof.

Chapter 2 of this thesis addresses the theoretical findings from Gabaix (1999).[2] This chapter is an empirical examination and analyses whether the two propositions on Gibrat's and Zipf's law are satisfied for Western Germany and its regions. The raw city size data are supplied by the German Federal Statistical Office and covers the years 1975-1997. The study uses three different concepts in order to define a region. The first kinds of regions are the federal states, determined by their legal boundaries. Those regions are characterized by the fact that their cities are spatially adjacent and share the same administrative boundary. The second kinds of regions are the *spatial clubs*, which consist of cities that are spatially connected but do not necessarily belong to the same Federal state. They are

[2]This chapter is joined work with Prof. Dr. Jens Suedekum and is published as Giesen and Suedekum (2011).

constructed by building an algorithm that randomly selects a city and includes all cities within a certain radius. The third kinds of regions are the *random regions*. Those are collections of cities that are randomly drawn from the pool of large cities, disregarding any spatial, political or economic connection. The first finding is that Zipf's law is not only satisfied for Western Germany as a whole, but also tends to hold across each definition for regions. This finding is novel, since the question of whether Zipf's law holds at the regional level has, to the best of my knowledge, never been addressed before. The second finding is that the propositions by Gabaix (1999) are confirmed for Western Germany. By using parametric and non-parametric techniques we show that Gibrat's law as well is satisfied for Western Germany and for its regions, again, regardless of which definition of a region is chosen. This finding is novel as well because it is the first study to address both propositions by Gabaix (1999) empirically. The chapter concludes that Gibrat's law and Zipf's law tend to hold basically everywhere in space in Western Germany.

A well-known drawback of the Zipf's law literature, however, is its focus on the largest cities within the urban hierarchy. This focus was not by choice but forced by limited data availability; data was only available for the largest cities. Since the year 2000, several national statistical offices started providing more precise city size data. This better data availability enabled the analysis of the overall size distribution across all settlements, including even villages with a population below 100 inhabitants. This more detailed data lead to a severe paradigm shift in the literature about Zipf's law. The first study to use this kind of data and focus on a country's overall distribution is by Eeckhout (2004). Using Census 2000 data, he discovers that the overall US city size distribution is not Pareto, but lognormal (LN) distributed. Figure 2 illustrates this point. The solid line shows a kernel density estimate of the overall city size distribution of the United States in 2000, using the same data as Eeckhout (2004). The dashed line represents the best-fit LN distribution, estimated via maximum-likelihood. The kernel density estimate and the fitted LN distribution clearly show that the overall distribution is far from being Pareto distributed. However, the upper tail of the distribution exhibits the typical hyperbolic pattern of the Pareto distribution. Eeckhout (2004) claims this to be the reason why so many studies have found evidence for Zipf's law: they misperceived the upper tail of the LN for the Pareto distribution. Since the LN does not feature a Pareto distribution in the upper tail, the LN implicitly declares Zipf's law as an illusion. It is therefore ambiguous if Zipf's law is really meaningful.

Moreover, as this thesis shows, the existence of an LN distribution seems plausible; under the pure and simplest form of Gibrat's law the sizes of cities must be LN distributed. This was already discovered by Gibrat (1931) and can be easily proven by using the central limit theorem. But, the insights by Eeckhout about the upper tail are questioned by several authors and the literature is unsure whether Eeckhout's findings have really invalidated Zipf's law. Levy (2009), Malevergne et al. (2011) and this thesis clearly

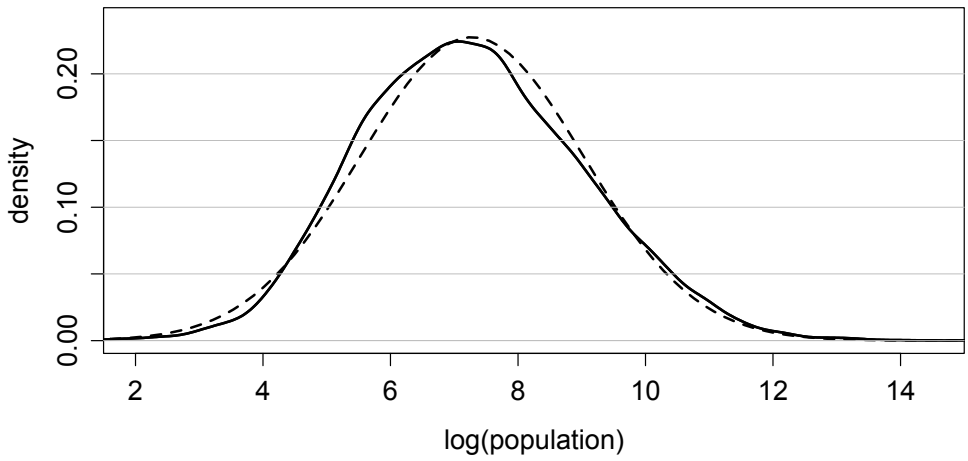

Figure 1.2: The overall US city size distribution in 2000 (administrative data).

show that there is indeed a Pareto distribution in the upper tail, while the body of the distribution is lognormal. This points to a possible existence of Zipf's law. These findings together raise the important question of whether there is a way to reconcile the findings of a lognormal body with a Pareto distribution in the upper tail. There are studies, like Ioannides and Skouras (2009) and Malevergne et al. (2011) that propose a mixture distribution, which switches from a lognormal to a Pareto distribution. However, their functional form is ad-hoc without a stochastic or theoretical motivation.

This thesis provides a unifying solution and argues that city sizes are not LN distributed. The argument is that city sizes instead follow a more general form, the double Pareto lognormal distribution (DPLN) first proposed by Reed (2002). The DPLN is a four-parameter distribution, which results if cities grow according to Gibrat's law and are created in different points of time so that age heterogeneity among cities prevails. The DPLN is therefore not an ad-hoc functional form, which is chosen because it has an impressive data fit, instead it has an explicit stochastic foundation. The DPLN exhibits power law behavior in both the upper and the lower tail, while it has a lognormal body in the range of medium city sizes. Moreover, the DPLN nests the LN. Those features enable the DPLN distribution to be consistent with the mature literature on Zipf's law on one side and with Eeckhout's finding of a lognormal body on the other side.

Chapter 3 of this thesis confirms the statement that city sizes do not follow the LN but the DPLN.[3] The study provides the first multi-country analysis to provide detailed and thorough evidence for the empirical validity of the DPLN and the superior fit compared to the LN, based on data from the national statistical offices of Brazil, the Czech Republic, France, Germany, Hungary, Italy, Switzerland and the United States. For all countries the DPLN has lower absolute deviations from the empirical data than the LN. Even by using model selection tests, like the Akaike information criterion, the Bayesian information

[3]This chapter is joint work with Prof. Dr. Jens Suedekum and Arndt Zimmermann, published as Giesen, Zimmermann and Suedekum (2010).

criterion, Bayes Factors along with Jeffrey's scale and the Log-ratio test, which penalize the DPLN for having more free parameters, the DPLN is indicated to be the preferred model for all countries but Switzerland. This study therefore re-establishes the Pareto distribution for the upper tail and therefore Zipf's law. Furthermore, the study provides indirect evidence that urban growth is characterized by the more general form of Gibrat's law than the simple form of Gibrat's law as proposed by Eeckhout (2004).

Chapter 4 is similar to chapter 3 and also discriminates between the LN and the DPLN.[4] The aim is to concentrate on and analyze in more detail one single country. The chapter therefore selects France because very detailed data on the French overall city size distribution is provided by the French national statistical office (INSEE). The data is for the year 2008 and covers 36,682 French municipalities (communes), which includes almost the whole French population. The data is administratively defined and city sizes range from 2,211,297 inhabitants for Paris to 1 inhabitant for Rochefourchat, a tiny "city" in the Rhône-Alpes region. The first contribution of this chapter is to show that the performance of Zipf's law for the French data is highly sensitive to the definition of the upper tail, i.e. how to truncate the available data from above. This also shows how the focus on solely the upper tail might lead to biased results. The second contribution is to show that Zipf's law holds almost perfectly when data for the French metropolitan areas is used, instead of the administratively defined municipalities. This has been claimed before for other countries by Rosen and Resnick (1980), Soo (2005) and Nitsch (2005). Using data on the major 247 French urban agglomerations, also provided by the INSEE for the year 2008, the slope coefficient is very close to unity at about -1.0075 with a standard error of $\sigma = 0.131$ and a $R^2 = 0.983$. Surprisingly, those estimates do not really react to the truncation point. The third contribution is to confirm that the DPLN provides a superior fit compared to the LN distribution.

The pure observation of Zipf's law, the LN or the DPLN distribution is not of immediate interest to economists per se. However, city sizes are the result of residential location choice, which in turn is driven by economic forces. It is therefore the responsibility of economists to identify and model the economic forces that shape real world city size distributions. Unfortunately, they fail at doing so.

Economic models on urban systems, urban growth, city sizes or location choice can be divided into two categories: deterministic and random models. The very early deterministic models, like Christaller (1933) and Lösch (1940), build on central place theory and explain cities as the consequence of economies of scale and transportation costs. Those models yield great insights, especially for the coherence between large cities, towns and hinterlands, but they lack economic content and are accused of being closer to "geography" than economics. A second generation of deterministic models, such as Henderson

[4] This chapter is joint work with Prof. Dr. Jens Suedekum also referred to as Giesen and Suedekum (2012b).

(1974), predict that city sizes are the result of counteracting agglomeration and dispersion forces and predict a single optimal city size. Those models mainly focus on identifying the determinants of urban growth and the reasons for the existence of cities. Scale economies and human capital externalities are identified as the main driving forces for cities, and congestion costs to limit city size. Putting the exact functional form of the Pareto, LN or DPLN distribution aside, the empirical patterns show that a single optimal city size is far from reality. The failure of those economic models is demonstrated by the stand-alone fact that city sizes range over a wide horizon.

A third generation of deterministic models, such as Eaton and Eckstein (1997), Glaeser (1999) or Black and Henderson (2008), is able to explain the wide dispersion in city sizes but fails at explaining the shape of the city size distribution. An exception is Duranton and Puga (2001), who are able to explain the Pareto distribution in the upper tail but not the remaining distribution. In summary, the deterministic urban growth models fail to generate distributions close to actual city size distributions, as highlighted by Duranton (2011).

In contrast, there are the models of random urban growth, which share the common feature to model some city or industry specific variable via stochastic growth processes. The usual approach of those models is that there are shocks at the city level, which translate into migration and therefore to changes in city sizes. Their main advantage, in contrast to the deterministic models, is their ability to come close to empirical city size distributions without being implausible. Models like Gabaix (1999) or Rossi-Hansberg and Wright (2007), for instance, rely on Gibrat's law and are able to provide simple, but also comprehensive, explanations for why Zipf's law might hold. While Gibrat's law implies a multiplicative random growth process, there are also models with additive random growth process, like Duranton (2007). This study does not rely on Gibrat's law and produces a convex upper tail city size distribution. In addition, he shows that his model is very close to the upper tail city size distribution of the US and France. Findeisen and Suedekum (2008) confirm the latter for the case of Western Germany.

However, the above-mentioned random urban growth models share in common to aim at replicating only the upper tail city size distribution. Their limitation, therefore, is not being able to generate the remainder of the distribution. In contrast, the Eeckhout (2004) model aims especially at the overall distribution. Eeckhout builds a model of local externalities in which cities grow according to a simple version of Gibrat's law. It is simple in the respect that all cities are equally old and have the same initial size. Using those assumptions, he shows that the overall city size distribution is LN. But despite its good fit in the body of the empirical distribution, the LN is not able to feature a Pareto distribution in the upper tail, as mentioned above.

In summary, even the models of random urban growth are not able to provide an explanation for the body and the upper tail of city size distributions at the same time.

They are either not able to explain the Paretian power law in the upper tail, like Eeckhout (2004), not to mention the unity slope coefficient, or they are not able to account for the lognormal body, like Gabaix (1999). This motivates the fifth chapter of this thesis to provide a remedy. This chapter builds on the findings of chapters 3 and 4 and proposes the DPLN to be the *true* city size distribution. This chapter builds a dynamic model of random urban growth, with endogenous city creation that generates a DPLN city size distribution.

The proposed model merges the insights by various economic studies. Using the local externality model from Eeckhout (2004) as a benchmark, several realistic features are incorporated. The first step is to enrich this model by simple population growth, as suggested in Black and Henderson (1999). The second step is to incorporate technological progress as proposed by an overwhelming number of studies in economics. Thirdly, it incorporates city creation as in Rossi-Hansberg and Wright (2007) or Henderson (1974). However, the key feature of the model is age heterogeneity among cities. The simple argument is that in order to explain the size distribution of cities, it is necessary to account for the fact that cities are founded in different periods of time. The large literature on the spatial distribution of population and city size distributions has almost completely neglected this fact, even though the importance is supported by several studies like Henderson and Wang (2007), Dobkins and Ioannides (2001) and Bairoch (1988). An exception is Rossi-Hansberg and Wright (2007), whose model features age heterogeneity but unfortunately they do not pay attention to it.

The model of this chapter is in continuous time, and city sizes result from counteracting agglomeration and dispersion forces. Each city has an idiosyncratic productivity parameter, which evolves according to a geometric Brownian motion. Older cities are more attractive on average, as they had a longer time to develop. As a consequence they host a larger population share than younger cities. City growth follows a random multiplicative growth process that replicates Gibrat's law. Population grows at a constant exogenous rate and distributes endogenously over a rising number of cities. The rate at which cities are born is determined by a social planner, who aims at maximizing the discounted infinite stream of overall welfare. This setup establishes an economic model that exactly replicates the stochastic forces that are responsible for the genesis of the DPLN. Empirically, the proposed model provides the best fit to empirical city size distributions, as shown throughout this thesis. Theoretically, it is the only micro-founded model that accounts for the lognormal body and the Paretian power law in the upper tail.

Chapter 2

Zipf's law for cities in the regions and the country

2.1 Abstract

The salient rank-size rule for city sizes known as *Zipf's law* is not only satisfied for Germany's national urban hierarchy, but also in single German regions. To analyze this phenomenon, we build on the theory by Gabaix (1999) that Zipf's law follows (under certain conditions) from a stochastic urban growth process. In particular, Gabaix shows that if urban growth in all regions follows Gibrat's law, we should observe the Zipfian rank-size rule among large cities both at the regional and at the national level. This theory has never been addressed empirically. Using non-parametric techniques and various definitions of a "region", we find that Gibrat's law holds at the regional level. Consistently we find that city size distributions at the national and at the regional level tend to follow a Zipfian power law.

2.2 Introduction

Dozens of studies during the past decades have addressed the salient rank-size rule for city sizes known as Zipf's law. As is well known, Zipf's law states that the upper tail of the city size distribution within an area (say, the US) and at any point in time can be described by a Pareto distribution with shape parameter equal to -1. This implies a unique rank-size relationship according to which the area's largest city (New York) is approximately twice as large as the second-largest city (Los Angeles), three times as large as the third-largest city (Chicago), and so on.[1] Virtually all existing studies are concerned with the city size distribution of entire countries, however. The starting point of our paper is the observation that Zipf's law for city sizes can also hold in single regions of a country.

Figure 2.1 illustrates this observation with an example from Germany. We provide a standard Zipf plot for the 20 largest cities from one German Federal State only, the State of Hessen. More precisely, we plot the log of the city's rank in Hessen's urban hierarchy (#1 for Frankfurt, #2 for Wiesbaden, #3 for Kassel, and so on) against log city size measured by the number of inhabitants in the year 1997. When running a standard Zipf regression of the type $\log(Rank) = \log(a) - \zeta \log(Size)$ we estimate $\zeta = 1.027$ with standard error 0.325 and $R^2 = 0.99$. Consistent with Zipf's law, we find that the rank-size relationship in log scales can be approximated very accurately by a linear curve. Secondly, the slope of this linear curve is very close to -1, which is also what Zipf's law suggests. As shown below, that picture looks similar for other German regions, including Germany as a whole.

[1] The seminal contributions of this literature are due to Auerbach (1913) and Zipf (1949), comprehensive cross-country studies are provided by Rosen and Resnick (1980) and Soo (2005). Nitsch (2005) conducts a meta-analysis, and Gabaix and Ioannides (2004) present a survey.

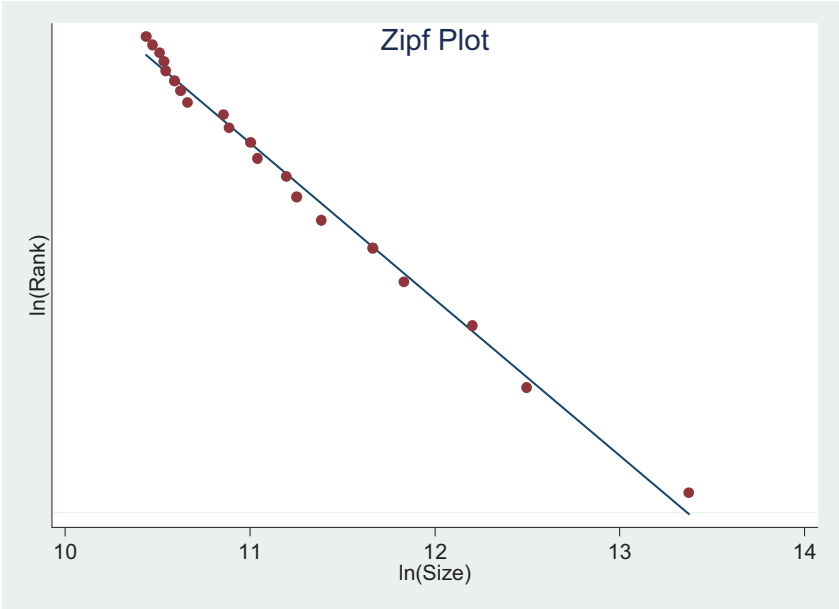

Figure 2.1: Zipf regression for Hessen.

Legend: OLS regression: ln(Rank)= 13.60 - 1.027 ln(Size), R^2=0.9920, Gabaix-Ibragimov corrected standard error: 0.325, p-value < 0.01, Obs: 20.
The standard errors are corrected according to the procedure suggested in Gabaix and Ibragimov (2011). Gabaix-Ibragimov corrected standard errors are calculated according to $\sqrt{\frac{2}{n}}\hat{\zeta}$. Gabaix and Ibragimov (2011) show that simple OLS standard errors are biased. The simple OLS standard error for this regression would be 0.026.

The main aim of this paper is to shed light on the question *why* Zipf's law tends to hold on a regional level.[2] The contribution by Gabaix (1999) is central to us in this respect. In proposition 1 of that paper Gabaix proves that if cities grow stochastically with the same expected growth rate and same variance (a property that is known as *Gibrat's law*), and if one introduces an additional small friction by imposing a lower bound for city sizes, then a steady-state city size distribution is implied that obeys Zipf's law.[3] In other words, Gabaix establishes a version of Gibrat's law as a statistical explanation for Zipf's law, thereby opening new avenues for theoretical and empirical research on the rank-size rule.

[2]We do not claim to be the first to ever estimate a regional Zipf coefficient, but the more recent literature has completely neglected this dimension. Neither Gabaix and Ioannides (2004) nor Nitsch (2005) mention any study on intra-national city size distributions. A more recent exception is Garmestani et al. (2007), who analyze Gibrat's law for the south-eastern part of the US. However, to the best of our knowledge there has been no systematic study on the relationship between Gibrat's law and Zipf's law from a regional perspective for an entire country.

[3]Eeckhout (2004) shows that the pure form of Gibrat's law generates a lognormal steady-state distribution, not the Pareto, and he finds that the lognormal provides a good fit to the empirical size distribution across *all* settlements in the US. This overall size distribution is still consistent with Zipf's law among large cities, since the properties of the lognormal can become virtually indistinguishable from the Pareto shape in the top range (see Eeckhout, 2009). In this paper we follow the lion's share of the literature and concentrate on large cities in order to address the salient Zipf's law from a regional perspective. For an analysis of the *overall* German city size distribution, see Giesen et al. (2010).

Theoretical papers on the economic microfoundations of Zipf's law often aim at theories about Gibrat's law or generalized versions thereof (see, e.g., Rossi-Hansberg and Wright, 2007; Eaton and Eckstein, 1997). On the empirical side, Gabaix's proposition 1 is the basis for the influential study by Ioannides and Overman (2003), who test Zipf's law indirectly by investigating if Gibrat's law holds at the national level in the US.

An even more important insight for our purpose is proposition 2 of Gabaix (1999). There he shows that if a country is composed of several regions, and if Gibrat's law holds in each of those regions, then Zipf's law is satisfied for all regional and also for the national city size distribution.[4] This theoretical insight has never been addressed empirically. In this paper we provide a first country-wide test of Gibrat's law from a regional perspective.

Except for the fact that a *region* is a subset of the *country*, there is nothing specific in Gabaix's theory as to what characterizes a "region". In our empirical analysis we therefore contemplate three different concepts. Most naturally we analyze the German Federal States (Länder), which can be thought of as well-defined clubs of cities that are both spatially adjacent and that share an important administrative commonality. Secondly, we analyze *random regions*, i.e., random draws from the population of large German cities, where the resulting samples of cities need not be adjacent to one another. Thirdly, we build spatial clubs of cities that are geographically adjacent, but that need not belong to the same Federal State. Like the Länder, these spatial clubs represent *non-random* samples from the population of large German cities, yet they do not coincide with administrative borders.

It turns out that Gibrat's law not only holds at the national level in Germany, but it tends to hold in almost *each* region regardless of which type (Federal State, random region, spatial club). What does this finding imply? Firstly, it suggests that urban growth among large cities is scale independent in Germany, and this property prevails not only in the aggregate or in random draws, but also in non-random samples of cities. This is consistent with theories of proportional urban growth (such as Eaton and Eckstein, 1997), where all cities regardless of initial size and location within the country grow with the same rate.

Secondly, according to Gabaix (1999), Zipf's law should then also be valid within the various types of regions and in the country - and in fact, this appears to be true. These findings thus empirically confirm the theory on the close correspondence of random city growth at the regional and the Zipfian rank-size rule at the regional and the national level. This does not mean, however, that Zipf's law holds by definition for every possible combination of cities. If all cities grow with the same expected rate, then the national urban system converges to a steady-state distribution which satisfies Zipf at the national level. Zipf's and Gibrat's law also hold in *random* and certain economically meaning-

[4]Note that the reverse need not be true; Zipf's law may hold at the national level without Gibrat's law being satisfied in all regions, as long as Gibrat is satisfied in the national aggregate.

ful *non-random* samples of cities. But we show that it is still possible to deliberately construct groups of cities such that Zipf's and/or Gibrat's law breaks down.

2.3 Gibrat's law and Zipf's law: A reminder

Gibrat's law states that all cities, regardless of initial size, grow randomly with the same expected rate and same variance. Gabaix (1999) shows that under certain conditions this growth process converges to a Pareto distribution with exponent equal to -1: Zipf's law. This requires some additional frictions, since the pure form of Gibrat's law generates a lognormal city size distribution, not the Pareto (see Eeckhout, 2004). This is not to say, however, that a proportionate growth process plus *something else* cannot give rise to the Pareto distribution, and in fact Gabaix (1999) considers *something else*: a lower bound on city sizes. The idea is that cities follow a growth process of the type

$$dS_{it}/S_{it} = \mu dt + \sigma dB_{it}. \tag{2.1}$$

dS_{it}/S_{it} is the percentage change of population size in city i at time period t. μ reflects the expected growth in normalized sizes: $\mu = \gamma(S) - \bar{\gamma}$, where $\gamma(S)$ is the normalized growth rate for cities with size S, $\bar{\gamma}$ is the mean growth rate, σ is the variance of city growth rates, and B_{it} is a Brownian motion.

The lower bound S_{min} is introduced by considering a reflected geometric Brownian motion, which specifies that a city which is larger than S_{min} will follow the process as given in (2.1), whereas a city with $S_t \leq S_{min}$ will follow $dS_t = S_t\, max(\mu dt + \sigma dB_t, 0)$ where the parameter $\mu < 0$ is a negative drift. In other words, a city that has "walked" below the threshold S_{min} is not able to become smaller or to disappear. Gabaix (1999) proves that this process converges to the countercumulative distribution function $G(S) = a/S^{\zeta}$ where S is the normalized city size, $a \geq S$ is the normalized size of the largest city (rank 1), and where the exponent ζ is given by

$$\zeta = \frac{1}{1 - S_{min}/\bar{S}}, \tag{2.2}$$

with $\bar{S}$ denoting the mean city size. Thus, in the limit as S_{min} goes to zero an exponent $\zeta = 1$ is implied.[5] For values $S_{min} > 0$ the exponent ζ would not be equal to one. This case would not be Zipf's law exactly, but a closely related form of a power law distribution (see Brakman et al., 1999 on this distinction).

Now, consider a country that is composed of R regions. The Gibrat growth process specified above holds within each region, thus $G(S) = a_r/S^{\zeta}$ describes the regional city

[5]See Blank and Solomon (2000) for a discussion of the growth process as specified in Gabaix (1999).

size distribution, where $\zeta \sim 1$ as S_{min} goes to zero. The national city size distribution is then characterized by (see proposition 2 in Gabaix, 1999):

$$G(S) = a/S^{\zeta} \quad \text{where } a = \sum_{r=1}^{R} \lambda_r a_r, \quad \sum_{r=1}^{R} \lambda_r = 1 \tag{2.3}$$

and again $\zeta \sim 1$. In words, if Gibrat's law and hence Zipf's law holds within each region, then Zipf's law also holds for the country. This is true in the case when all cities (across all regions) have the same expected growth rate, but even if regions differ in their average growth rate as long as urban growth is scale independent within each region. This theoretical result is the basis for our empirical analysis in Section 2.6.

2.4 Data

The data set for this study is provided by the German Federal Statistical Office (Statistisches Bundesamt). It contains the area size (in square kilometers) and the number of inhabitants for 2143 German cities, covering the time period from 1975 to 1997. This data is quite exhaustive and includes even very small towns. For our purpose these small towns are of little interest, however, as Zipf's law concentrates on the upper tail of the city size distribution (Eeckhout, 2009). By truncating the data one has to define what the "upper tail" precisely is. Our benchmark estimations for the national level rely on the 71 largest German cities with more than 100,000 inhabitants in 1997, which represent about 46 per cent of the population in the data set. This is a standard definition for the cutoff that has been widely used in the literature (see Rosen and Resnick, 1980; Chesire, 1999; Soo, 2005). For the analysis at the State level we stick to this cutoff wherever possible, but we require a minimum number of observations of $N = 20$ for each Federal State. This implies an effective cutoff below 100,000 inhabitants in most cases (see Table 1 below).[6]

A city is classified by the administratively defined boundaries, i.e., our data follows the "city proper" concept. The alternative would be to use "urban agglomeration" data, which aggregates main cities and suburban areas that often form own administrative units into metropolitan areas (MAs). Both classification types have certain advantages and disadvantages, as discussed for example in Cuberes (2011). One shortcoming of city proper data is that administrative boundaries are sometimes arbitrary and lack economic content. Unfortunately, the Federal Statistical Office does not provide consistent MA data for a sufficiently long time period to analyze urban growth at this level. We therefore have

[6]For Saarland, the smallest German State, we have to suffice with $N = 17$ since there are in total only 17 cities from this State in the data set. The sum of the 167 cities that we have used for the analysis at the level of the Federal States plus the three "city States" Hamburg, Berlin and Bremen (which consist of a single city) account for roughly 50 per cent of the total population in the data set and for roughly 36 per cent of the overall (urban+rural) population in Western Germany.

to stick to city proper data, but below we do consider MA data for the cross-sectional analysis of Zipf's law for a single year.

There are two further data issues that we have to deal with. Firstly, Eastern Germany appears in the data only from 1990 onwards. For our analysis of growth rates we have therefore focused on the former Western German cities, including only former West Berlin. For the analysis of Zipf's law at the national level, which is for the year 1997, we have however considered the entire city of Berlin (Germany's largest city). Secondly, over the time period 1975-1997 there have been some other re-classifications of city boundaries or city mergers. These changes led to some unreasonable jumps of area or population, but none of those affected the large German cities on which we concentrate in this paper.

2.5 The national level: Germany

Previous analyses on Gibrat's and Zipf's law have usually been conducted at the national level. In this section we first follow this typical approach and address Gibrat and Zipf for Germany as a whole, before turning to an analysis at the regional level in the next section.

2.5.1 Gibrat's law

For the analysis of Gibrat's law we use the 71 largest German cities with population size above 100,000 inhabitants in 1997. We then follow these cities back in time until 1975 and compute 22 annual growth rates for each of those cities.[7] Our analysis rests on *normalized* city growth rates, which are constructed as follows: From the annual population growth rate of city i in year t, $(pop_{i,t} - pop_{i,t-1})/(pop_{i,t-1})$, we subtract the mean and divide this by the standard deviation of growth rates of the respective reference group (in this section the 71 largest German cities) in the average across all years. Under the null hypothesis of scale independent urban growth we would thus expect that all cities, regardless of their size, have mean normalized growth rate equal to zero and variance equal to one.

In Figure 2.2 we take a first look at this hypothesis and non-parametrically estimate a stochastic kernel, i.e., a three-dimensional graphical representation of the distribution of city growth rates as a function of city size. The kernel was constructed by dividing the data into percentiles. For each one we estimate the distribution of growth rates via density smoothing. The kernel represents the distribution of growth rates conditional

[7] These cities thus represent the ranks 1 to 71 in the national urban hierarchy for the year 1997. By following those cities back in time, we cannot be sure that they have also been the largest cities in all preceding years. However, our results are little affected by this problem of "panel attrition". Results would be similar if we followed the 69 largest cities with population size above 100,000 in the initial year 1975 over time, even if a city dropped out of the group, or if we arrange our sample by collecting the largest cities separately for each year.

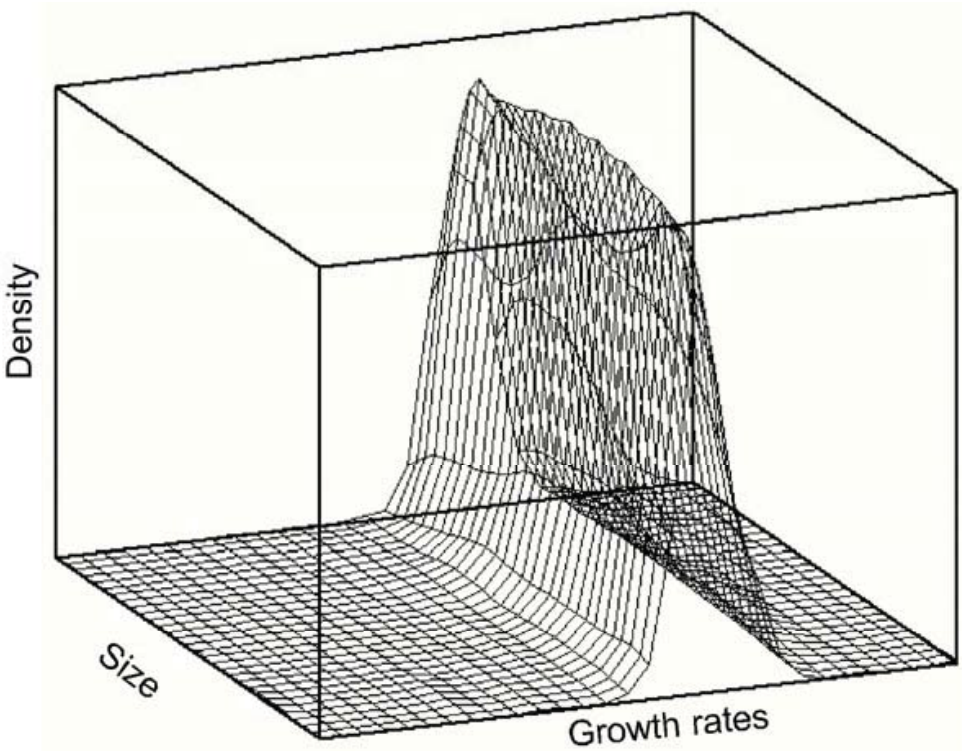

Figure 2.2: Stochastic kernel for German growth rates.

Legend: This kernel shows the distribution of normalized growth rates conditional on city size. This was done by calculating the density of growth rates within each percentile.

on size, and yields a first impression that urban growth appears to be very similarly distributed across different city size classes.

In Figures 2.3a and 2.3b we provide non-parametric estimates for the conditional means and variances of city growth rates. The estimation was performed using the Nadaraya-Watson (NW) technique (see Nadaraya 1964; Watson 1964; Haerdle 1992) which estimates the expectation of growth conditional on size. The underlying regression equation for the NW-estimator is

$$g_i = m(S_i) + \epsilon_i \quad \text{with} \quad m(S_i) = E[g|S_i]. \tag{2.4}$$

The unknown conditional mean $\hat{m}_h(s)$, and the conditional variance of growth rates, $\hat{\sigma}_h^2(s)$, are estimated according to

$$\hat{m}_h(s) = \frac{n^{-1}\sum_{i=1}^{n} K_h(s - S_i) g_i}{n^{-1}\sum_{i=1}^{n} K_h(s - S_i)}. \tag{2.5}$$

$$\hat{\sigma}_h^2(s) = \frac{n^{-1}\sum_{i=1}^{n} K_h(s - S_i)\left(g_i - \hat{m}(s)\right)^2}{n^{-1}\sum_{i=1}^{n} K_h(s - S_i)}. \tag{2.6}$$

In this estimation $\hat{m}_h(s)$ is a locally weighted average, where the kernel K_h (in our case

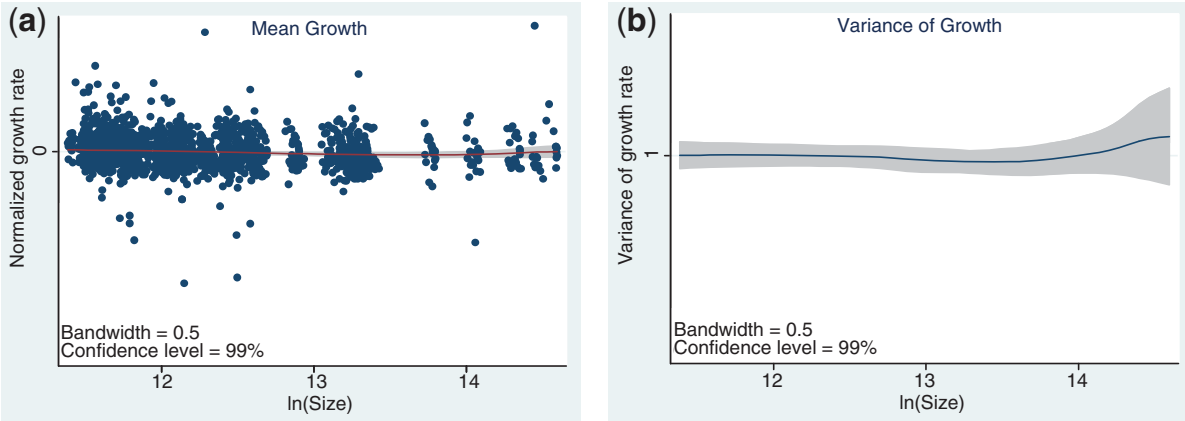

Figure 2.3: Nadaraya-Watson Estimator (a) mean normalized growth rate and (b) normalized variance of the growth rate.

the Epanechnikov kernel) is the weighting function with bandwidth h. We use bandwidth $h = 0.5$ as our benchmark.[8] This non-parametric test for Gibrat's law has been previously used by Ioannides and Overman (2003). The advantage of this approach as compared to parametric regressions is its flexibility with respect to the underlying functional forms. As Haerdle (1992) puts it: "[...] it does not project the observed data into a Procrustean bed of a fixed parametrization". It thereby also does not narrow down the inference about Gibrat's law to a single test statistic, but allows for a detailed inspection in which range of city sizes we may encounter a deviation from the null of scale independent urban growth.

In Figure 2.3a we indicate the single observations for conditional mean growth rates by the dots, whereas in Figure 2.3b we drop this scattering and only focus on the estimated conditional variance and the respective confidence band. As can be seen, the normalized mean growth rate of zero and the normalized variance of one fall inside the 99% pointwise confidence bands throughout the entire range of city sizes. The confidence band for the conditional variance (Figure 2.3b) tends to widen at larger city sizes, which is due to the fact that there are only few observations in this range so that standard errors increase. Moreover, the NW-estimator for conditional mean growth rates appears to be very slightly downward-sloping in the lower range of city sizes where most of the distribution mass is concentrated. Statistically, however, we cannot formally reject Gibrat's law for Germany. This result is consistent with Ioannides and Overman (2003) who found support for Gibrat's law at the national level in the US with a similar methodology.

These findings should also be set into perspective to Bosker et al. (2008), who address the stability of the German city size distribution over the period 1925-1999. They argue that Gibrat held perfectly in Germany prior to WWII, but not in the post-war time

[8] We have also considered the "optimal bandwidth" developed by Silverman (1986) for the smoothing of the non-parametric estimators. Results have been very similar to those reported in the paper.

period 1945-1999 where they found that small cities gained population relative to large ones. This would imply that the German city size distribution was permanently affected by the WWII-shock.[9] Our findings indicate that the transition from the pre- to the post-war distribution has been completed until 1975. That is, if Gibrat's law did not hold in Germany in the first decades after the war, it seems to hold again in the more recent period. The slight negative slope in Figure 2.3a still seems to be consistent with Bosker et al. (2008), but we cannot reject the hypothesis that city growth is independent of city size in our sample.

2.5.2 Zipf's law

As Gibrat's law holds at the national level in Germany, how about Zipf's law? In Figure 2.4a we plot the log of the city's rank in Germany's national urban hierarchy (#1 for Berlin, #2 for Hamburg, #3 for München, and so on) against the log of city size for the year 1997 (∼3.5m for Berlin, ∼1.7m for Hamburg, ∼1.2m for München, and so on) using the city proper data.[10] By inspection the relationship is linear with only small outliers, and in fact, linear regressions yield R^2-levels beyond 0.98.[11] For the estimation we have considered two alternative methods that both draw on the recent contribution by Gabaix and Ibragimov (2011), who show that ordinary least squares estimation yields biased results in rank-size regressions: (i) a log rank-log size regression with corrected standard errors (see legend of Figure 2.1 for details), (ii) a refined OLS regression which uses the log of (rank-1/2) as the left-hand side variable (also with corrected standard errors). Under both approaches we obtain highly significant estimates for the slope coefficient ζ. However, even though we cannot formally reject $\zeta = 1$, there are deviations of the point estimates from unity.

Recall from above that a value of ζ different from one does in principle not conflict with the growth process specified by Gabaix (1999), provided the urban hierarchy actually follows a power law (also see Brakman et al., 1999 on this). The perfect fit of Zipf's law is only achieved in the limit, while deviations from $\zeta = 1$ in empirical applications can result from small sample sizes. Furthermore, it is known from the previous literature

[9] In this respect Germany would be a special case, since the city size distribution of other countries (such as the US, France or Japan) was found to be remarkably stable over time, see Black and Henderson (1999, 2003), Eaton and Eckstein (1997) or Duranton (2007). Recently, however, Michaels et al. (2008) have also argued that Gibrat's law does not perform well in the US once small towns are included.

[10] For the cross-sectional analysis of Zipf's law we focus on the final year 1997, but results would be similar for all other years of the observation period. This is again consistent with the insight by Black and Henderson (1999, 2003) or Duranton (2007) that city size distributions are enormously stable over time.

[11] It is known from Monte Carlo studies that R^2 levels tend to be generically high in Zipf regressions as a result of the ordering of cities by rank (see Gan et al., 2006). This extraordinarily high R^2 level suggests, however, that we do not pick up a spurious relationship, but that German city sizes actually follow a power law in the upper tail. This conclusion is supported by a Kolmogorov-Smirnov test, which clearly cannot reject the hypothesis of Pareto-distributed city sizes.

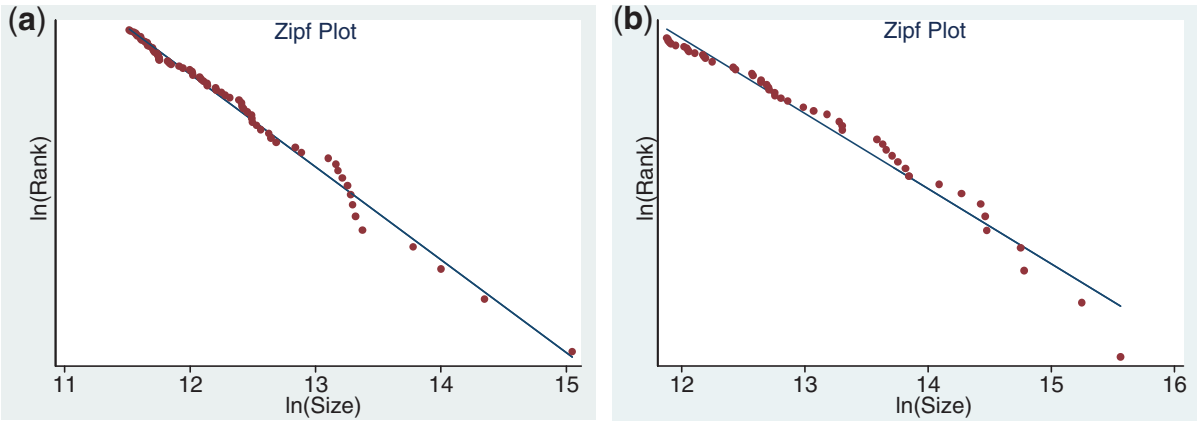

Figure 2.4: Zipf Regression for Western Germany.

Legend: **(a)** city proper data: OLS regression: ln(Rank)=18.45-1.23 ln(Size), R^2=0.9899, Gabaix–Ibragimov corrected standard error: 0.206, Obs: 71; Gabaix–Ibragimov regression: ln(Rank-1/2)=19.53-1.32 ln(Size), R^2=0.9863, Gabaix–Ibragimov corrected standard error: 0.222, Obs: 71 **(b)** urban agglomeration data, OLS regression: ln(Rank)=15.56-0.96 ln(Size), R^2=0.9739, Gabaix–Ibragimov corrected standard error: 0.192, Obs : 50; Gabaix–Ibragimov regression: ln(Rank-1/2)= 16.46-1.03 ln(Size), R^2=0.9532, Gabaix–Ibragimov corrected standard error : 0.206, Obs : 50.

that Zipf's law performs better with urban agglomeration than with city proper data (see Rosen and Resnick, 1980), and the slope coefficient for Western Germany in fact moves considerably closer to $\zeta = 1$ when using that type of data.

This is shown in Figure 2.4b where we provide an analogous Zipf plot for the 50 largest metropolitan areas in the year 1997. The very largest MAs (#1 is now the Ruhrgebiet with 5.7m inhabitants, #2 is Berlin, and so on) appear to be a bit "too small".[12] Yet, the Zipf regression still yields a R^2-level of about 0.97 and a highly significant slope coefficient $\zeta = 0.96$ with the first and $\zeta = 1.03$ with the second estimation approach, which are both very close to the perfect Zipf fit. In sum, we conclude that both Gibrat's law and Zipf's law hold at the national level in Germany.

2.6 Gibrat's law and Zipf's law on a regional level

We now turn to the analysis of Gibrat's law and Zipf's law at the *regional* level, which is the novel conceptual contribution of this paper.

[12]This is in line with the findings by Bosker et al. (2008) who attribute this fact to the asymmetric impact of WWII on the largest urban areas.

2.6.1 Three types of regions

We contemplate three different concepts of a "region": 1) random regions, 2) the German Federal States, and 3) spatial clubs of cities. For the first type of region, we *randomly* draw N observations from the population of large German cities, using $N = 20$ as our benchmark. For each city we compute the rank in the group's urban hierarchy (#1 for the largest, #2 for the second-largest, and so on) in the year 1997 and relate the log of this rank to the log of city size. Furthermore, for each of these N cities we observe 22 annual population growth rates from 1975 to 1997, which we normalize with the mean and standard deviation of growth rates in the random region. These 20×22 normalized growth rates are then used to estimate conditional mean and variance with the same methodology (the NW-estimator) as before. Note that the cities within one random draw need not be spatially adjacent, and also need not belong to the same Federal State.

Secondly, we consider the Federal States which is probably the most natural definition of a German "region". The Länder are *non-random* draws from the population of large German cities. Unlike the random regions they form spatially adjacent clubs of cities. Moreover, these cities share a common State administration and a common history, which in some cases is longer than the history of Germany as a nation. We therefore estimate Zipf's law and Gibrat's law separately for each of the 8 Western German Länder, leaving out the States of Hamburg, Bremen and Berlin which are composed of a single city.

Finally, we construct spatial clubs of cities. Our benchmark definition of a spatial club is as follows: Using the 71 largest German cities we compute for each observation the club of those cities with distance less than $d = 200$ kilometers to the respective "central member".[13] For each club we then use the standard procedures to determine the intra-group urban hierarchy in order to estimate Zipf's law, and to observe 22 annual growth rates for each city in order to estimate Gibrat's law separately for each club. Note that the number of members differs across clubs, that one city can belong to more than one club, and that the respective central member does not have to be the largest city in its club. Note further that the cities within one club can belong to more than one Federal State.[14] The 71 spatial clubs therefore do not have political or administrative content, but this definition of a region may capture the fact that larger urban agglomerations can extend across State borders (as, for example, the Rhein-Main-area). These regions are thus also economically meaningful *non-random* draws from the population of large German cities.

[13] We have constructed a 71×71 distance matrix for this exercise using standard route planning software. Below we also consider two slightly different definitions of a spatial club.

[14] Across the 71 spatial clubs we find that the average number of involved Federal States is 5.3, the minimum is 2 and the maximum is 7. Little surprising, these numbers are lower than for random regions because we impose spatial adjacency of the cities within one club. When considering 71 draws of random regions with size $N = 20$ cities for each, we find that the cities within one random region on average come from 6.5 different Federal States, with a minimum equal to 4 and a maximum equal to 9 Länder.

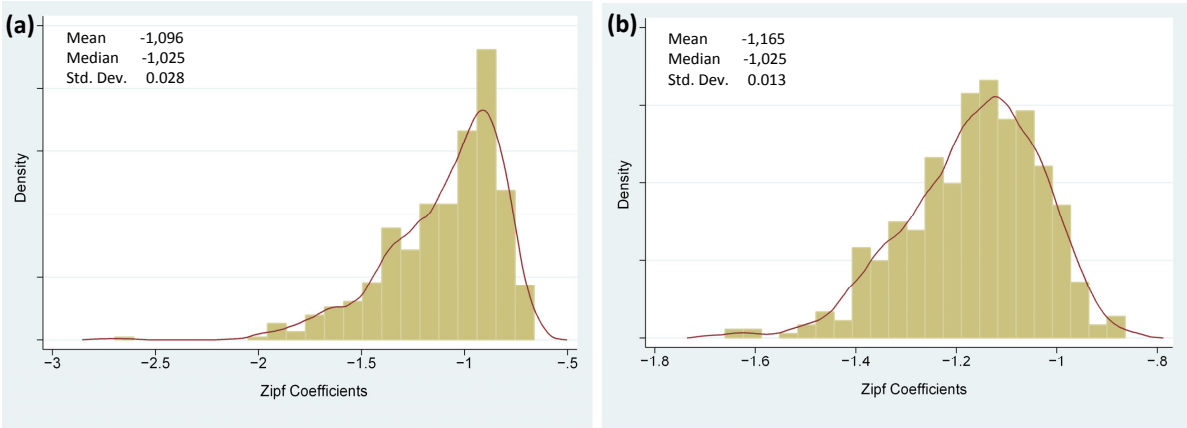

Figure 2.5: Distribution of Zipf coefficients for random city samples: (a) N=20 and (b) N=100.

2.6.2 Results

For the **random regions** we clearly expect Gibrat's law to hold. Since urban growth is scale invariant in the total population of large cities, as shown in Figure 2.3, this property should also be satisfied (except for sampling error) in random draws from this population. This can indeed be verified. The NW-estimation for single random regions yields pictures that in the vast majority look very similar as in Figure 2.3 (the plots are omitted for brevity). The interesting question is whether the validity of Gibrat's law implies Zipf's law in these random groups of cities, as Gabaix's theory would suggest. In Figure 2.5a we summarize the distribution of the OLS estimates of ζ (with negative sign) across 500 draws of size $N = 20$. As can be seen, the bulk of the estimated coefficients is clustered around values between 1 and 1.1, with mean 1.09 and median 1.02. The R^2 levels are consistently very high (typically beyond 0.9), suggesting that a linear function fits the data very well. Results remain robust when changing the number of cities in a random region. In Figure 2.5b we show the distribution of ζ across 500 draws of size $N = 100$.[15] The mean value of the estimated Zipf coefficient now becomes 1.16, whereas the median remains almost unchanged. The slightly worse fit of the exact Zipf's law in the larger random regions is quite intuitive, as the random draws with size $N = 100$ now also include smaller towns among which Zipf's law is known to perform worse (Eeckhout, 2004). The power law shape of the distribution is, however, robust across all random regions. Our findings thus support Gabaix's theory: The validity of Gibrat's law in the random groups of cities comes along with Pareto-distributed city sizes that are close to the Zipfian rank-size relationship.

Turning now to the **German Länder**, note that the scale invariance of urban growth

[15] For the construction of these regions we obviously did not stick to the population of the 71 largest cities but we have included also smaller towns in the population of cities from which to draw. Specifically, we now include about 1,500 towns with minimum population size equal to 6,000.

	Number of cities	Minimum city size	Maximum city size	Zipf coefficient	R^2
Schleswig-Holstein	20	19,94	240,516	**1.068 (0.388)**	**0.9680**
Niedersachsen	20	49,814	520,67	**1.299 (0.411)**	**0.9858**
Nordrhein-Westfalen	30	103,872	964,311	**1.365 (0.352)**	**0.9503**
Hessen	20	34,128	643,469	**1.027 (0.325)**	**0.9920**
Rheinland-Pfalz	20	20,224	186,136	**1.229 (0.389)**	**0.9458**
Baden-Württemberg	20	56,781	585,274	**1.26 (0.398)**	**0.9881**
Bayern	20	43,707	1,205,923	**0.929 (0.294)**	**0.9574**
Saarland	17	11,946	186,402	**1.21 (0.383)**	**0.9320**

Table 2.1: Federal German States (West).

Legend: Mean Zipf coefficient across Federal States: 1.173, median Zipf coefficient across Federal States: 1.220. Gabaix-Ibragimov corrected errors of the Zipf coefficients are shown in parentheses.

in the population of large German cities does not automatically imply that Gibrat's law holds in non-random samples of this population, such as the Federal States. In principle, it is conceivable that urban growth in the Länder is not scale independent, whereas Gibrat's law does hold in the aggregate where regional differences are averaged out.[16] Empirically, however, this is not the case in Germany. We rather find that Gibrat's law holds in *each* Federal State. Over the entire range of city sizes that we analyze, and for all States, we cannot reject the null of zero normalized growth rates and constant variance equal to one (see Figure 2.6), the only slight exception being the small cities in Rheinland-Pfalz. In Table 2.1 we summarize the results for the Zipf regressions for the Länder. The Zipf coefficients range from $\zeta = 0.93$ to $\zeta = 1.37$ (all highly significant), and we obtain R^2 levels beyond 0.9 throughout. Mean and median of ζ across the Länder are similar to the national Zipf coefficient (using city proper data) that we have reported in Section 2.5.[17] Hence, we find that city size distributions on average follow a similar power law pattern in the Länder as in the national aggregate, that is for Western Germany as a whole.

Finally, we investigate if the validity of Gibrat's law and Zipf's law is confined to areas within administrative State boundaries. We do this by analyzing the non-random **spatial clubs** of cities. Focussing at first on Gibrat's law, we depict the NW-estimates for conditional city growth for 8 of the 71 spatial clubs in Figure 2.7, where one club consist of all cities with distance below $d = 200$ kilometers around the respective central

[16] Such a configuration could result if urban growth exhibits mean reversion (small cities growing faster) in some States, but agglomeration effects (divergent growth) in others. In that case Gibrat's law would fail to hold within each State, but it may still hold at the national level.

[17] The deviations of the German regional Zipf coefficients from $\zeta = 1$ are well in the range known from cross-country studies on Zipf's law, if not even a bit smaller. In Rosen and Resnick (1980) the estimates for ζ across 44 countries range from 0.81 in Morocco to 1.96 in Australia. In Soo (2005) the range across 73 countries goes from 0.73 to 1.72.

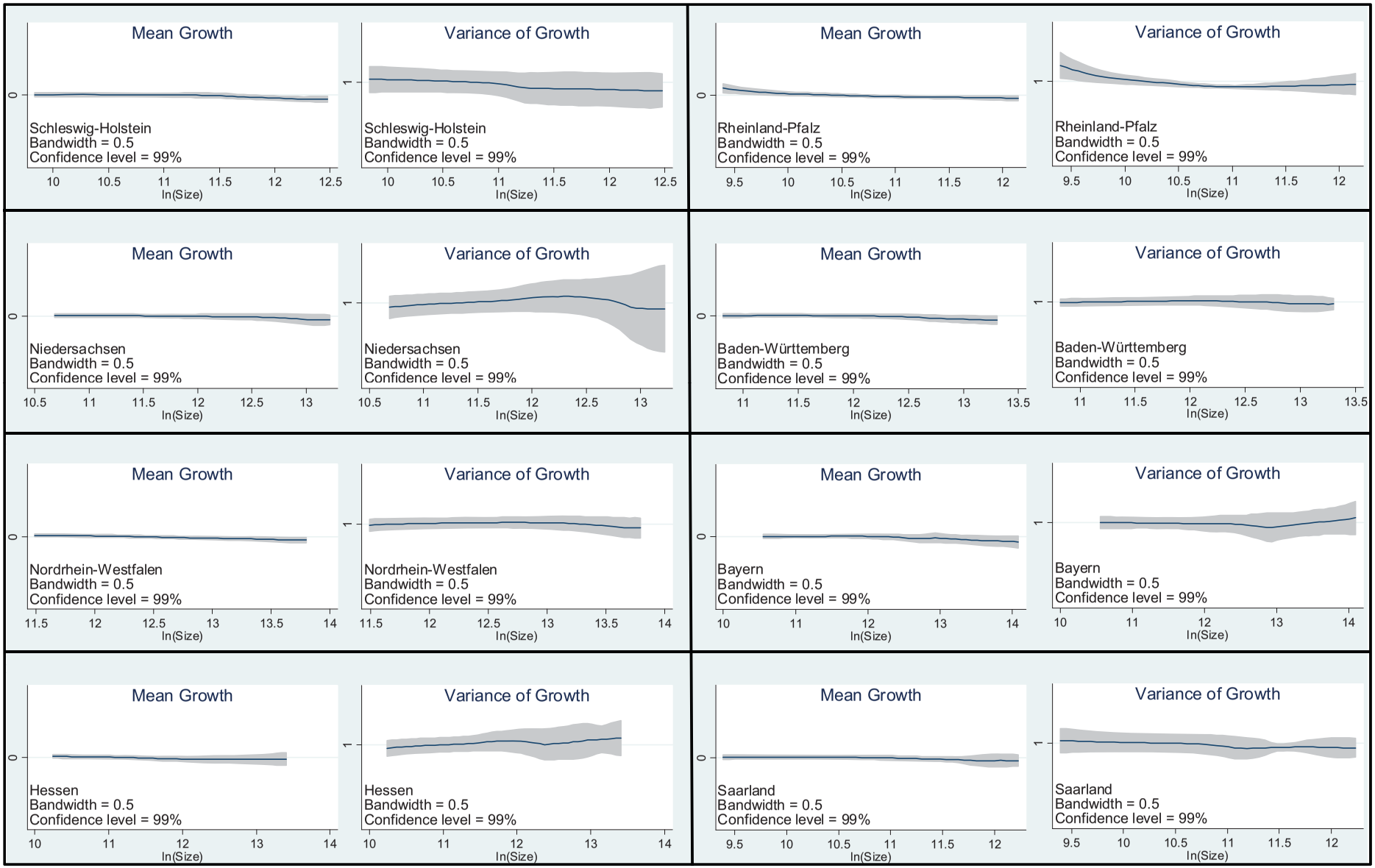

Figure 2.6: Nadaraya-Watson regression for the Federal German States.

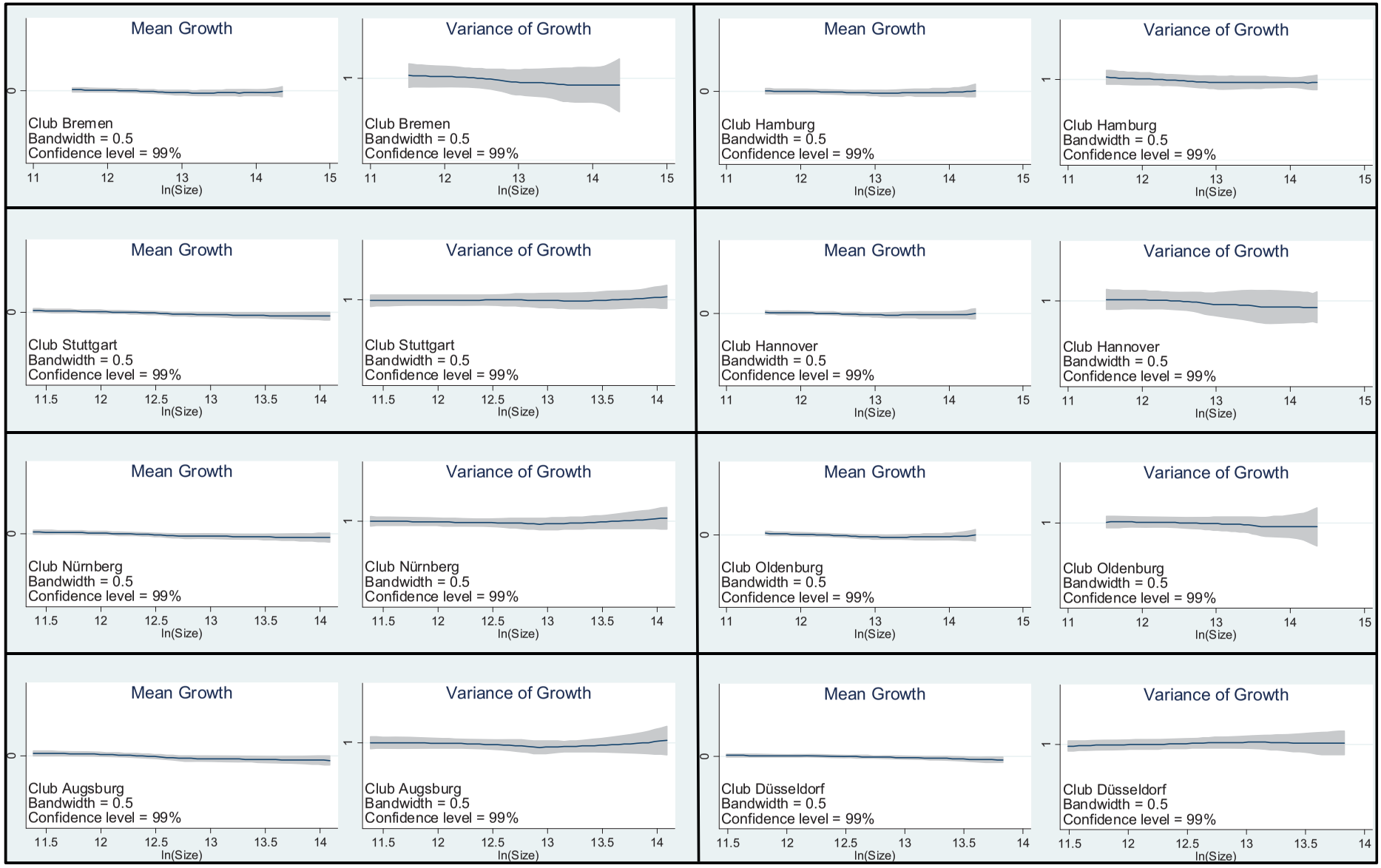

Figure 2.7: Nadaraya-Watson regression for the spatial clubs of cities.

member (see Section 2.6.1 for the detailed description). The picture that emerges from Figure 2.7 is that urban growth is also scale independent in each of these non-random regions. Similar conclusions follow when we construct the spatial clubs in a different way. Firstly, instead of imposing a fixed maximum distance d we have also assumed a spatial club to be the collection of the $M = 20$ large cities with the shortest distance to the respective central member. This approach balances the number of members across, but implies different geographical sizes of the clubs. Secondly, as a compromise between the two former definitions, we have drawn larger circles with $d = 300$ kilometers around each of the 71 large German cities and defined a club to be the $M = 20$ largest cities inside that circle. We have dropped clubs with less than 20 members, which led to a total of 68 clubs.

Regardless of how we construct the spatial clubs, we find that Gibrat's law continues to hold. The null of scale independent urban growth can, basically, not be rejected in any range of city sizes.[18] In some rare cases we find that the confidence band does not cover the value of zero for the mean normalized growth rate, or the value of one for the variance. But such cases are exceptions. The typical pattern is qualitatively similar to that reported in Figure 2.7. Turning to the intra-club city size distributions for the latest year 1997 we again find linear rank-size relationships with R^2 levels beyond 0.9. Figures 2.8a-2.8c summarize the distributions of ζ across all clubs for the three different definitions. As can be seen, the Zipf coefficients are on average similar to the national Zipf coefficient using city proper data ($\zeta = 1.23$), particularly for the second and third definition.

2.6.3 Are Gibrat's law and Zipf's law satisfied by definition?

The results presented in this section have shown that Gibrat's law not only holds at the national level in Western Germany, or in random draws from the population of large cities. It is also satisfied for certain non-random samples of cities, namely the Länder and the cross-state clubs of spatially adjacent cities. How meaningful are these results?

A first concern is that the non-parametric NW-estimation may be such that Gibrat's law can, by construction, almost never be rejected even if urban growth is actually *not* scale independent. Figure 2.9 addresses this issue. Here we deliberately put together some relatively small cities from Baden-Württemberg which exhibited long-run growth rates above the average, and some relatively large but slowly growing cities from Niedersachsen. As can be seen, the NW-estimator now clearly rejects the hypothesis of scale independent urban growth. Figure 2.9 thus suggests that the NW-estimator is a sensible test and does not automatically support Gibrat's law.[19] When looking at economically meaningful

[18] The NW-plots for the two alternative definitions of spatial clubs are omitted for brevity, but all plots for all types of regions can be made available upon request.

[19] Note that the estimated Zipf coefficient for this case is quite far away from $\zeta = 1$, see the left column

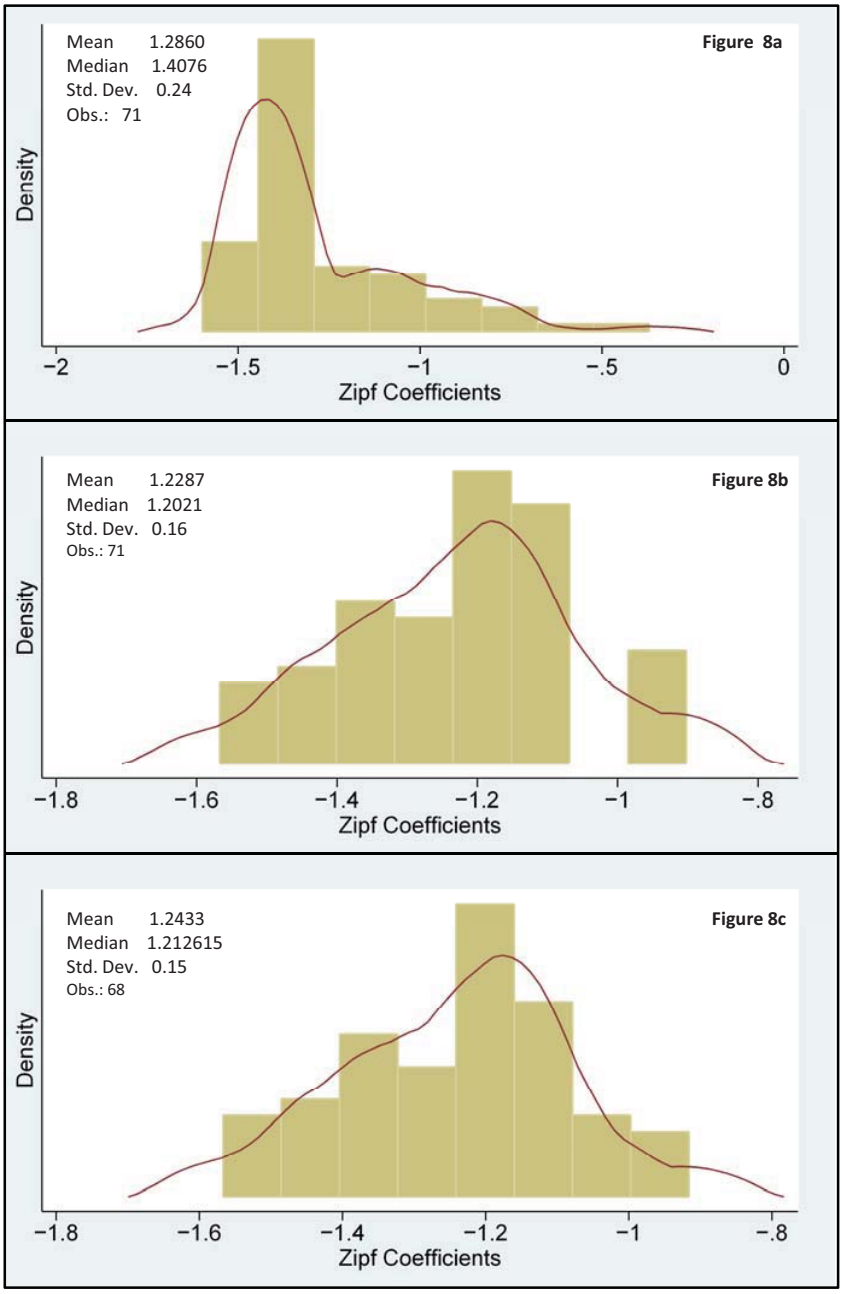

Figure 2.8: Distribution of Zipf coefficients for spatial clubs of cities.

Legend: The clubs were constructed from the population of the 71 largest West German cities with population size above 100,000 inhabitants. Three different rules were used to define a club: Rule 1: Select a city and include all cities within a radius of 200km (a). Rule 2: Select a city and include the 20 cities with the smallest distances to the selected one (b). Rule 3: Select a city and include the largest 20 cities within a radius of 300km (c). (a–c) report the distribution of the slope coefficients $-\zeta$ across the spatial clubs. All estimated slope coefficients were highly statistically significant using Gabaix-Ibragimov corrected std. errors.

samples, however, we cannot reject Gibrat at the regional level. The fact that it seems to hold in almost each of the various types of regions actually suggests that all large cities have grown with the same expected rate over the period 1975-1997, i.e., that Gibrat's law holds basically everywhere in space in Western Germany.

Secondly, we have shown that the city size distributions in the various types of regions can be well characterized by a Zipfian power law. It is important to note, however, that this does *not* imply that Zipf is satisfied by definition for every possible non-random combination of cities. That point is illustrated in Figure 2.10. In the first line we show that Zipf's law (and Gibrat's law) holds nicely for the 100 largest German cities (rank 1-100 in the national urban hierarchy), which is consistent with our results from Section 2.5.2. This is also true for the 50 largest German cities (see the second line), but Zipf clearly breaks down when considering the cities ranked 50-100 (see the third line of Figure 2.10).

The reason is simple. If all large German cities grow with the same expected rate over the long run, we would expect the national urban hierarchy to resemble a Pareto distribution in the upper tail. It is well known that a left-truncation does not change the power law properties of this distribution, i.e., if the 100 largest cities in Western Germany follow a Zipfian power law, so should the 50 largest cities. This is not true for a right-truncted Pareto, however, so there should be no power law for the cities ranked 50-100. Figure 2.10 verifies this point empirically. In other words, it is possible to construct groups of cities such that Zipf's law does not hold, even if Gibrat's law is satisfied. The non-random regions that we have analyzed in this paper are no simple left- or right-trunctions of the national urban hierarchy, however, but economically meaningful and spatially contiguous samples of cities. For those cases it is thus fully consistent with the theory by Gabaix (1999) that we not only find scale independent urban growth, but also a Zipfian power law for the regional city size distributions.

2.7 Conclusions

Our empirical results lead to two main conclusions. Firstly, we have shown in this paper that urban growth among large cities is scale indepedent basically *everywhere* in space in Western Germany. Gibrat's law is thus satisfied not only in the national aggregate but also in each of the various types of regions. Secondly, we have shown that city size distributions within the economically meaningful regions exhibit a strikingly linear rank-size relationship. This is consistent with Gabaix's (1999) theoretical insight that stochastic urban growth at the regional level implies a Zipfian power law shape of the regional and the national city size distributions.

of the figure.

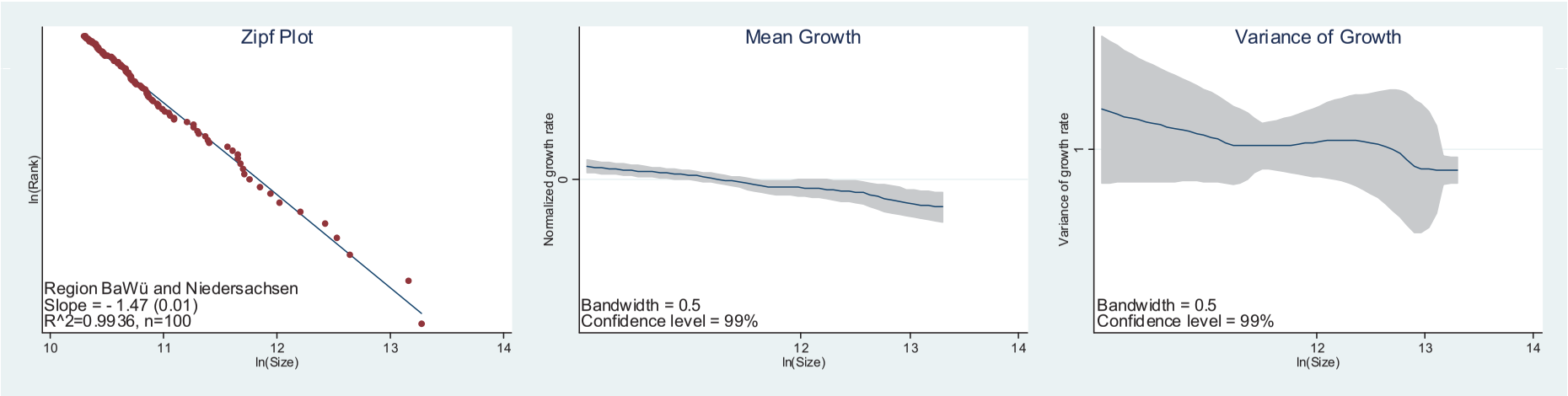

Figure 2.9: A counterexample: Small and fast-growing cities and large and slow-growing cities from two different states.

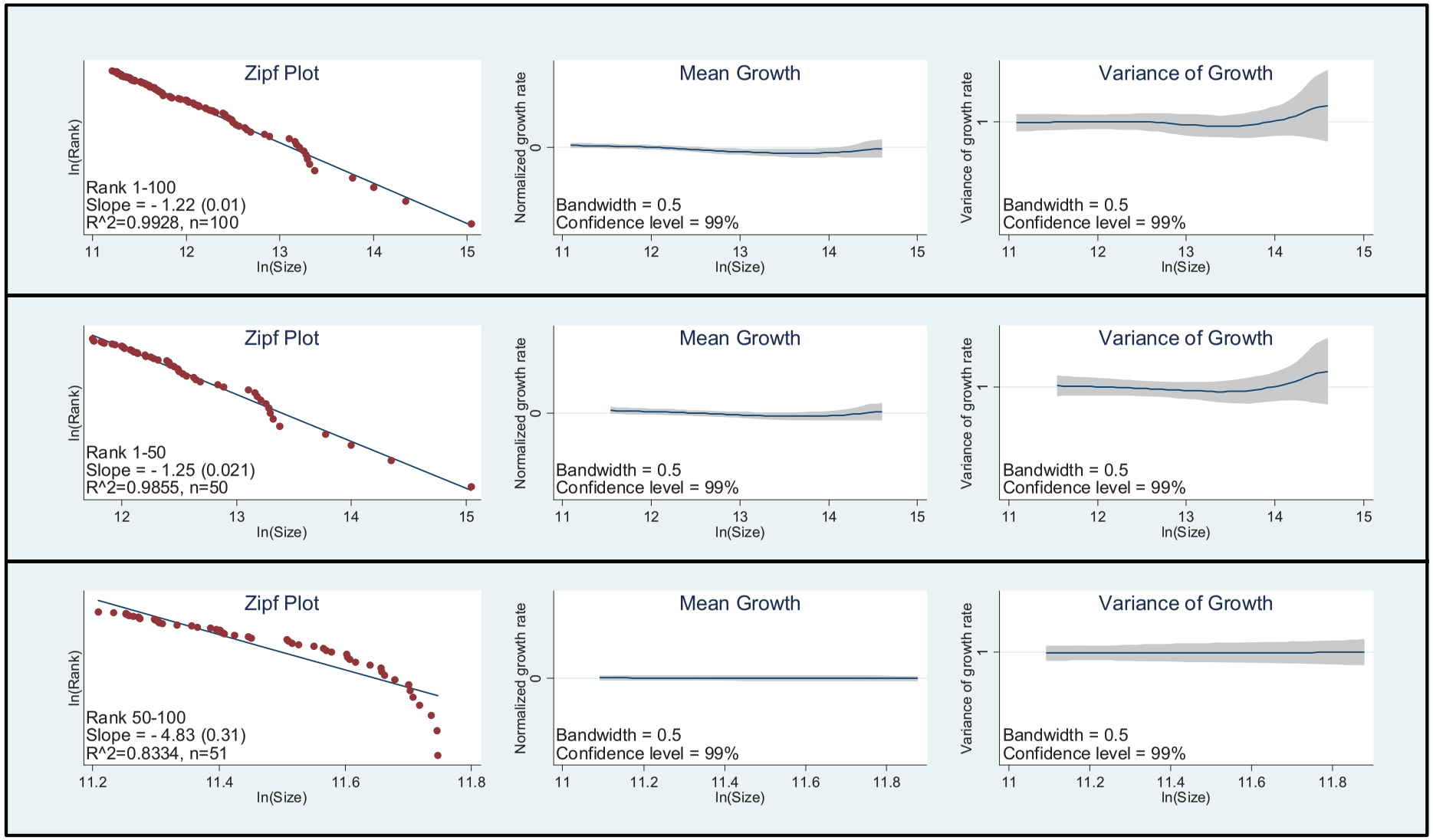

Figure 2.10: Truncation of the national urban hierachy.

Legend: Each line represents a Zipf plot with the corresponding NW plots. Line 1 is calculated for the largest 100 cities in Germany. Line 2 is calculated for the largest 50 cities in Germany. Line 3 is calculated for the cities with rank 50–100 in Germany.

In the literature there has constantly been scepticism about Zipf's law. Is it really a meaningful economic relationship, or only a statistical artefact that holds almost by definition? A well-known example of this awe is the statement by Krugman (1996), who calls it "spooky". We believe that our results point at an actual economic substance behind the rank-size rule. It is intimately entangled with proportionate urban growth, even on a low geographical level, but it is still possible to construct cases where the Zipfian power law breaks down despite the validity of Gibrat's law.

It would be interesting for future work to examine regional city size distributions and urban growth processes for further countries. How does Zipf perform on a regional level in small countries, or in regions with only few cities? Are the results for Germany representative for other developed or less-developed countries? Finally, it seems worthwhile to extend the analysis to *overall* city size distributions at the regional level. In this paper we have focussed on the large cities, but Zipf is a feature that pertains to the upper tail only (Eeckhout 2004, 2009; Giesen et al. 2010). Furthermore, Gibrat's law seems to be sensitive to the inclusion of smaller towns (see Michaels et al. 2008). An extended analysis at the regional level would therefore be useful, in order to gain a better understanding of urban growth processes and urban hierarchies across all types of settlements.

Chapter 3

The size distribution across all cities - double Pareto lognormal strikes!

3.1 Abstract

Using un-truncated settlement size data from eight countries, we show that the "double Pareto lognormal" (DPLN) distribution provides a better fit to actual city sizes than the simple lognormal (LN) distribution. The DPLN has a lognormal body and features a power law in both the lower and the upper tail. It emerges in the steady-state of a stochastic urban growth process with random city formation. Our findings reconcile a recent debate on the Zipfian rank-size rule for city sizes.

3.2 Introduction

Recently there has been an intensive debate about city size distributions. Dozens of older studies have argued that city sizes follow a Pareto distribution, or even adhere exactly to the famous rank-size rule known as Zipf's law.[1] This evidence is problematic, however, since those studies have worked with truncated samples and focussed only on large cities. In an influential article, Eeckhout (2004) has shown that the Pareto does *not* hold when taking into account *all* settlements of a country. Figures 3.1-3.3 below illustrate this point. The solid lines depict, respectively, the entire city size distribution in Germany, the United States, and France (in logarithmic scale). It is immediately obvious that these are no Pareto distributions, i.e., that Zipf's law does not hold across all cities in these countries. This raises three important questions. First, if not the Pareto, what is the appropriate parameterization for city sizes? Second, what can we learn from city size distributions about the underlying urban growth process? And third, has Eeckhout (2004) invalidated the entire old Zipf literature, or is there a way to reconcile them?

Eeckhout (2004) addresses all three questions. In his model cities grow stochastically, and this growth process - the pure form of Gibrat's law - asymptotically generates a lognormal (LN) size distribution. Eeckhout then shows that the LN delivers a good fit to actual city sizes in the US. This may answer the first two, but has delicate implications for the third question. As a matter of fact, the LN does not feature a power law in the upper tail and, hence, it is strictly speaking not compatible with Pareto and Zipf. Why have so many previous studies, including the recent one by Levy (2009) who uses an un-truncated sample, then provided evidence for a Zipfian power law among large cities? The reason according to Eeckhout (2004, 2009) is that the LN and the Pareto distribution have similar properties in the upper tail and can become virtually indistinguishable. In other words, his answer to the third question is that Zipf can be observed among large cities in practice, because the Pareto closely resembles the true size distribution (the LN)

[1] Zipf's law states that city sizes are Pareto-distributed with shape parameter equal to minus one. This implies that city sizes follow a particular power law such that the country's largest city is twice as large as the second-largest, three times as large as the third-largest city, and so on. See Soo (2005) and Nitsch (2005) for comprehensive analyses.

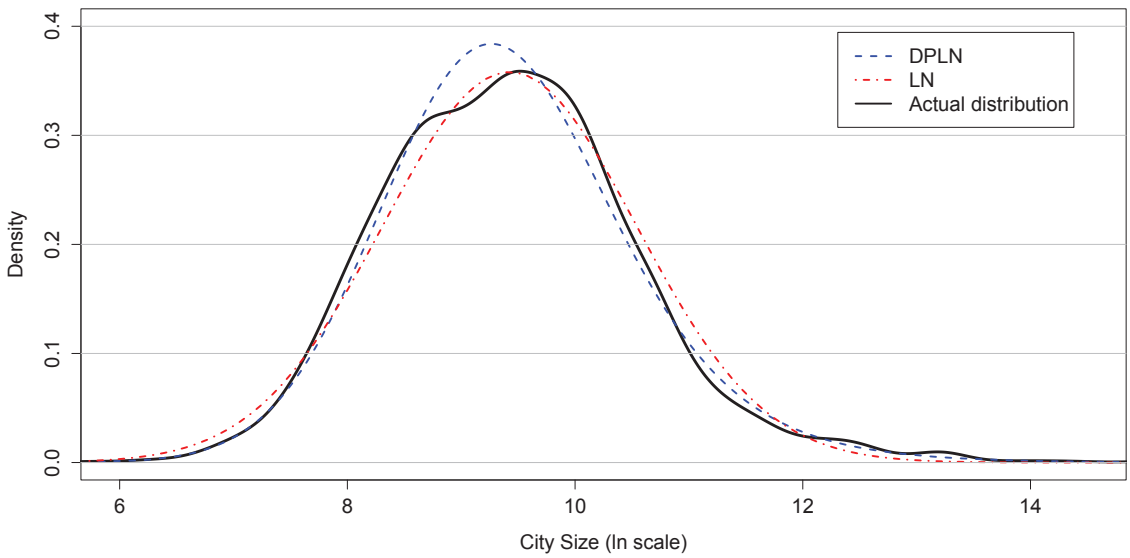

Figure 3.1: City size distribution, Germany 2006.

in the top range.[2]

In this paper we take a fresh look at these issues by using un-truncated settlement size data from eight countries. We confirm that the city size distribution in all countries can indeed be well approximated by a LN. However, that does not mean that there can be no other distribution which is even more successful in fitting the data. In fact, using various methods we show that the "double Pareto lognormal" (DPLN) distribution consistently provides an even closer fit. That distribution has a lognormal body in the medium range and exhibits a power law in both the lower and the upper tail. Taken by itself it is not surprising that one can find a distribution with a more flexible functional form that delivers a better fit, but we also show that the DPLN is the preferred model according to several statistical selection criteria that penalize it for having more parameters than the LN. Furthermore, the important difference between the DPLN and an arbitrary flexible distribution is that it has a theoretical foundation in terms of the underlying urban growth process. Reed (2002) has shown that the DPLN emerges in the steady-state of an evolutionary process which can be thought of as a generalization of Gibrat's law. Put differently, it is not the intention of this paper to fit just any arbitrary distribution in a theory-free manner. We rather provide an alternative answer to the first two questions by pointing at an urban growth theory (developed in Reed 2002) which generates a size distribution that is even closer to the actual data than the LN.

Finally, our findings may be particularly useful because we can provide a more satisfactory answer to the third question and reconcile the recent debate about city size

[2]Also see Mitzenmacher (2004), who shows that the density or the countercumulative distribution function of the LN generate a "nearly straight" line in logarithmic plots when the variance is large. A power law (Pareto) would generate exactly a straight line in such plots. For the Zipfian rank-size rule to hold, such a straight line is required.

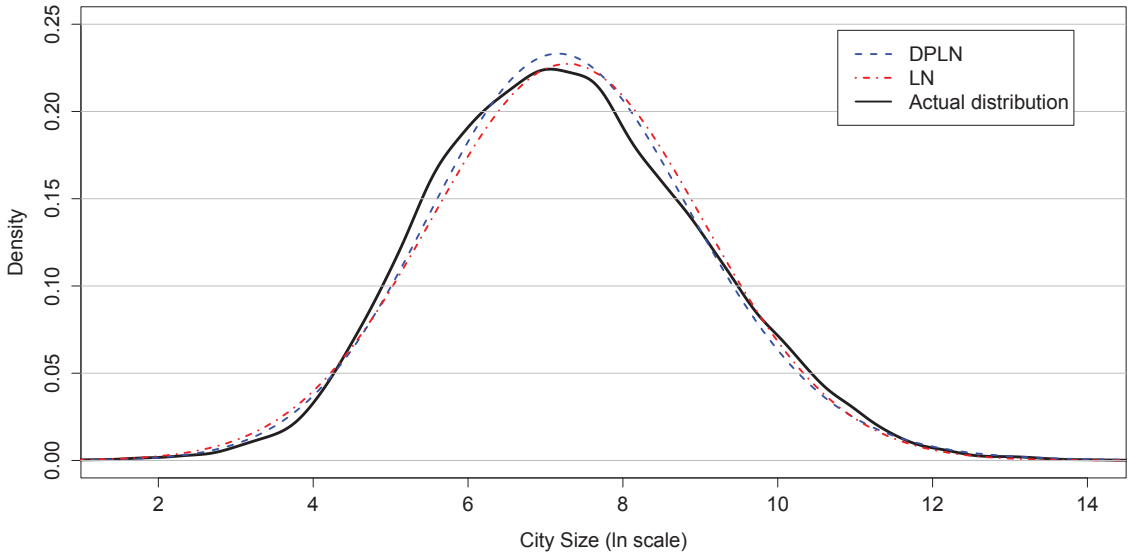

Figure 3.2: City size distribution, USA 2000.

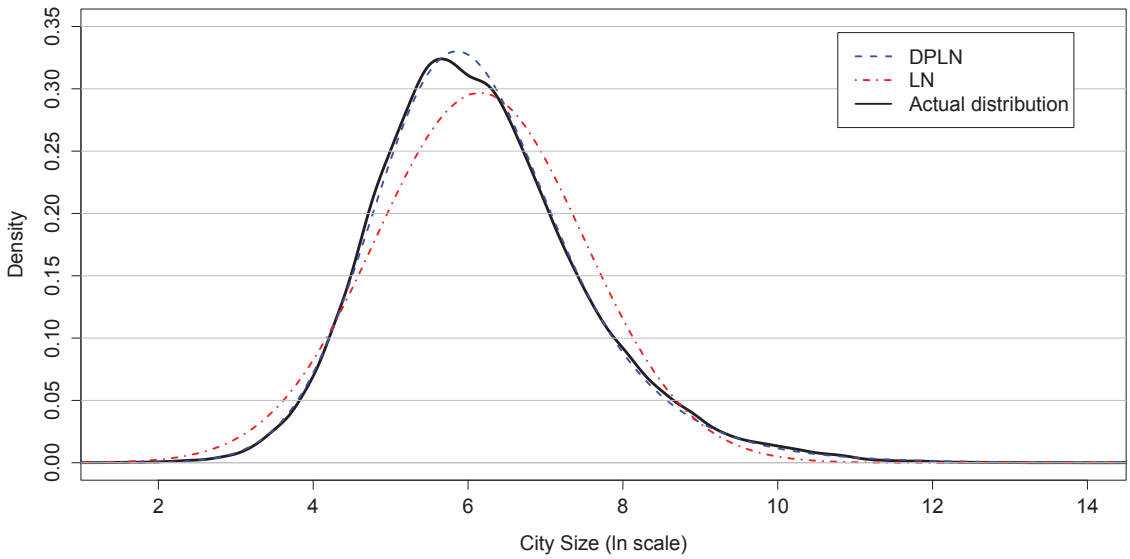

Figure 3.3: City size distribution, France 2006.

distributions. In common with Eeckhout (2004, 2009) we find that the LN does a good job in fitting the data. Yet, the empirical distributions in all countries exhibit a distinctive power law pattern in the tails, as also noted by Levy (2009) for the large cities in the US. Though the LN can be consistent with such a power law pattern under certain conditions (see Mitzenmacher 2003), this feature of the data is more precisely captured by the DPLN. In other words, the DPLN is *even better* compatible with Zipf's law among large cities while following a lognormal shape in other ranges. Our results may therefore bring Eeckhout and the older Zipf literature even closer together. In contrast to that literature, which has mostly drawn conclusions from truncated samples, in our case the Zipfian power law emerges as an upper tail feature of the un-truncated distribution.

3.3 Urban growth processes and steady-state city size distributions

In the model by Eeckhout (2004) an economy consists of a fixed number of locations across which workers are freely mobile. The spatial equilibrium results from a trade-off between positive and negative size externalities that accrue within but do not spill over across locations. In every time period, each location is hit by an idiosyncratic and random productivity shock. Cities eventually grow according to the pure form of Gibrat's law, which can be described as $dPop_{it}/Pop_{it} = \mu dt + \sigma dB_{it}$, where $dPop_{it}/Pop_{it}$ is the percentage change of population in city i at time t. The parameter μ is trend growth, and B_{it} is an independent shock with mean zero and variance σ^2. As already anticipated by Gibrat (1931), such a stochastic proportionate growth process with additive random shocks asymptotically leads to a lognormal (LN) distribution.[3]

Reed (2002) develops a model that is more statistical in nature. Cities grow stochastically as under Gibrat's law, but in every time interval dt there is the probability λdt that a new city emerges as a satellite of an existing one.[4] The initial size of the new city is drawn from a LN distribution with mean μ_0 and variance σ_0^2. These new cities then also exhibit proportionate growth. At time t there are $e^{\lambda t}$ cities in total, some of which are older than others. Reed (2002) proves that this growth process, which resembles the Yule-process first described in biology (see Yule 1925), asymptotically leads to a "double Pareto lognormal" (DPLN) distribution, with density

[3]In another influential paper, Gabaix (1999) has shown that Zipf's law follows as the limiting distribution of an augmented version of Gibrat's law that includes a lower bound for city sizes; also see Gabaix and Ioannides (2004).

[4]In the pure form of Gibrat's law there is no creation of new cities.

$$f(x) = \frac{\alpha\beta}{\alpha+\beta}\left[x^{\beta-1}e^{\left(\beta\mu_0+\frac{\beta^2\sigma_0^2}{2}\right)}\Phi^c\left(\frac{log(x)-\mu_0+\beta\sigma_0^2}{\sigma_0}\right)\right.$$
$$\left.+x^{-\alpha-1}e^{\left(\alpha\mu_0+\frac{\alpha^2\sigma_0^2}{2}\right)}\Phi\left(\frac{log(x)-\mu_0-\alpha\sigma_0^2}{\sigma_0}\right)\right]. \tag{3.1}$$

α and β are the Pareto coefficients for the upper and the lower tail, respectively, and μ_0 and σ_0 are the lognormal body parameters. Φ represents the normal cumulative density function (cdf) and $\Phi^c = 1 - \Phi$ represents the complementary cdf.

Details about the properties of this distribution can be found in Reed and Jorgensen (2005). It is shown there, for example, that a DPLN distributed random variable X can be represented as UV_1/V_2, where U, V_1 and V_2 are independent and U is a LN distribution with parameters μ_0 and σ_0 and V_1 and V_2 are Pareto distributions with shape parameters α and β, respectively. The DPLN is unimodal if $\beta > 1$ and can be written as a mixture of a right-handed and a left-handed Pareto-lognormal limiting distribution which, respectively, arises if $\alpha \to \infty$ or $\beta \to \infty$. It is not possible to exactly delineate the lognormal body part and the Pareto-distributed tails. That is, we cannot pin down parametrically at which city size the upper tail of the DPLN starts (or where the lower tail ends), although informal approximations are of course possible. Last, the simple LN distribution is nested in the DPLN if $\alpha \to \infty$ and $\beta \to \infty$.

Our paper can be seen as the first attempt in the literature to discriminate between the two theories of urban growth, the pure Gibrat's law and the generalized version by Reed (2002). We do so by comparing which of the theoretical steady-state distributions, LN or DPLN, is the preferred model for empirical city size distributions. So far, the fit of these distributions has only been addressed separately. For the DPLN this has been done by Reed (2002), but only for four regions (two US states and two Spanish provinces) and not in comparison to the LN.

3.4 The overall city size distribution: LN versus DPLN

3.4.1 Data

The basic data problem for our study is that un-truncated settlement size data, which are needed to fit an entire distribution, are not yet easily available for many countries. What is available are truncated samples of large cities with population size above some threshold level. Such data sets are, for example, used in the cross-country investigation of Zipf's law by Soo (2005) and exist for virtually all countries in the world. It is nevertheless possible to obtain un-truncated settlement size data at least for some countries, and in this paper we present evidence for eight cases. For brevity we mainly focus on Germany,

the US, and France, but we additionally include Brazil, Czech Republic, Hungary, Italy and Switzerland in the analysis.

The data for the US is the same that has been used by Eeckhout (2004, 2009) and Levy (2009). It is provided by the US census and includes population sizes for 25,359 settlements ("places") in the year 2000, ranging from 1 to roughly 8m inhabitants in New York City. This data set has two main limitations. First, a "place" is not defined according to economic criteria but follows an administrative definition that, moreover, varies considerably across US states.[5] An alternative geographical unit are metropolitan statistical areas (MSAs), which are defined in a more meaningful way but are subject to a minimum population size.[6] Second, the census places, although not subject to a minimum size, do not comprehensively represent the entire US population but only about 74% of it. For Germany the data is provided by the federal statistical office (*Statistisches Bundesamt*). The so-called "DESTATIS" database includes population sizes for 2,075 cities in the year 2006. This data set has comparable problems. A German city is also defined according to administrative boundaries. In addition, the historical awarding of "city rights" is decisive as to whether a settlement is counted as a city or not. The smallest city (Arnis) has 309 inhabitants. Overall, the German data set covers about 72% of the total population in the year 2006. The French data set as provided by the national statistical office (*INSEE*) includes the sizes of 36,674 French administratively defined settlements (communes) in the year 2006. It provides the best coverage as it basically represents the entire French population.

As for the other countries, we consulted the web pages of various national statistical offices (see http://www.bls.gov/bls/other.htm) to check for un-truncated settlement size data. In most cases such data are not freely provided, but for the five additional countries mentioned above they are publicly available. Further details about these data can be found in table 3.1, where we report the number of settlements, the percentage of the overall population that is covered, as well as the minimum and the maximum city size for each country.

3.4.2 Maximum-Likelihood estimation of the LN and the DPLN distributions

The first step in the analysis is to fit both the LN and the DPLN parameterizations to the data by using the maximum likelihood (ML) method. Reed and Jorgenson (2005: eq. 28) explicitly derive the log-likelihood function of the DPLN distribution; for the LN this is a standard exercise.

[5] The precise definition of places is explained in the *Geographic Areas Reference Manual* available online under http://www.census.gov/geo/www/garm.html.

[6] Also see Cuberes (2011) on the pros and cons of administrative versus economic definitions of cities.

	Germany (2006)		**United States (2000)**		**France (2006)**		**Brazil (2007)**	
N	2,075		25,358		36,674		5,564	
Coverage	0.72		0.74		0.97		0.99	
Min	309		1		1		804	
Max	3,404,037		8,008,278		2,181,371		10,886,518	
	LN	**DPLN**	**LN**	**DPLN**	**LN**	**DPLN**	**LN**	**DPLN**
$\hat{\mu}$	9.425	-	7.277	-	6.152	-	9.384	-
$\hat{\sigma}$	1.114	-	1.753	-	1.344	-	1.137	-
$\hat{\alpha}$	-	1.442	-	1.225	-	1.028	-	1.093
$\hat{\beta}$	-	4.638	-	3.131	-	5.355	-	7.247
$\hat{\mu_0}$	-	8.947	-	6.777	-	5.367	-	8.607
$\hat{\sigma_0}$	-	0.841	-	1.526	-	0.919	-	0.685
$ln(L_j^i)$	-22,726	-22,687	-234,773	-234,710	-288,484	-287,387	-60,823	-60,536
	Czech (2009)		**Hungary (2009)**		**Italy (2009)**		**Switzerland (2008)**	
N	6,249		3,152		8,101		2,706	
Coverage	0.99		0.99		0.99		0.99	
Min	3		12		33		19	
Max	1,233,211		1,712,210		2,724,347		358,540	
	LN	**DPLN**	**LN**	**DPLN**	**LN**	**DPLN**	**LN**	**DPLN**
$\hat{\mu}$	6.132	-	6.778	-	7.845	-	6.953	-
$\hat{\sigma}$	1.228	-	1.336	-	1.332	-	1.339	-
$\hat{\alpha}$	-	1.066		1.196	-	1.559	-	2.139
$\hat{\beta}$	-	4.818		2.070	-	3.691	-	3.287
$\hat{\mu_0}$	-	5.401		6.425	-	7.474	-	6.790
$\hat{\sigma_0}$	-	0.766		0.922	-	1.135	-	1.215
$ln(L_j^i)$	-48,467	-48,164	-26,749	-26,702	-77,370	-77,336	-23,442	-23,439

Table 3.1: Data summary and estimated parameters of LN and DPLN distribution.

Legend: Un-truncated settlement size data are publicly available from the websites of the respective national statistical offices. See http://www.bls.gov/bls/other.htm for a comprehensive list. N is the number of data points (cities) in country i. Coverage is the percentage of the total population in country i and the respective year that is represented by the data set. Min and Max are the population size of the smallest and the largest settlement in the data set. Settlements are classified according to administrative boundaries. See the websites of the national statistical offices for details. Parameters are estimated with the maximum likelihood method. $ln(L_j^i)$ is the log-likelihood of distribution $j = LN, DPLN$ in the respective country i.

Table 3.1 summarizes the estimated parameters and the corresponding log-likelihoods of the two parameterizations for the eight countries. Notice that the estimated upper tail parameters of the DPLN distribution ($\hat{\alpha}$) are in some cases (France, Czech Republic) very close to unity. This corresponds to an exact validity of Zipf's law. One has to be careful, however, comparing these estimates with Zipf coefficients from the literature, i.e., with shape parameters of a fitted Pareto distribution. This is because the "Zipf coefficients" are highly sensitive to the chosen threshold city size. For example, when running a standard rank-size regression of the type $\log(\text{Rank}) = \log(C) - \zeta \cdot \log(\text{Size})$ for Germany, we estimate $\hat{\zeta} = 1.27$ when including only cities with more than 100,000 inhabitants in the regression, $\hat{\zeta} = 1.34$ with a threshold of 200,000, $\hat{\zeta} = 1.23$ with a threshold of 50,000, and so forth. In other words, a Zipf coefficient exactly equal to unity arises, if at all, only under special assumptions on the minimum city size when fitting a Pareto distribution. It should therefore come as no surprise that the estimates for $\hat{\alpha}$ also deviate from unity. As for the lower tail parameter, there is no focal point to compare to. As can be seen, $\hat{\beta}$ is consistently far greater than unity, but there is no theory saying that the power law in the lower tail should be such that the second-smallest settlement within a country is twice as large as the smallest, or the like. Even when running a "naive" rank-size regression for the lower tail in Germany, $\log(\text{Rank}) = \log(C) + \xi \cdot \log(\text{Size})$, we obtain values that are not even close to unity ($\hat{\xi}$=2.48 for cities smaller than 2,000, $\hat{\xi}$=2.09 for cities smaller than 6,000, and so forth). Generally speaking, an advantage of using un-truncated data is that one does not have to make such arbitrary choices about size thresholds.

We now turn to several informal (visual) and formal tests of the performance of the fitted DPLN versus the simpler but more rigid LN distribution.

3.4.3 Informal tests

The solid lines in figures 3.1-3.3 show non-parametric kernel density estimations (KDE) of the actual city size distributions in Germany, the US, and France in logarithmic scale using Silverman's optimal bandwidth. The dot-dashed lines in these figures represent the fitted LN, and the dashed lines the fitted DPLN distributions with parameters given above. Upon inspection both parameterizations decently fit the actual distribution in all countries.

In order to address their relative performance, we plot the pointwise vertical differences between the empirical and the two competing theoretical cumulative density functions (cdfs) in the left panels of figures 3.4-3.6. The right panels show the cumulated deviations at different city sizes. As can be seen, the pointwise differences are larger for the LN than for the DPLN in almost all ranges. A standard Kolmogorov-Smirnov (KS-) test looks at the supremum of these pointwise differences across the entire distribution. This

supremum is clearly larger for the LN than for the DPLN in all three countries. Hence, the KS-test would reject the former parameterization earlier than the latter. The cumulated deviations of the DPLN consistently remain below those of the LN, especially in France where we actually have the best relative performance of the DPLN.

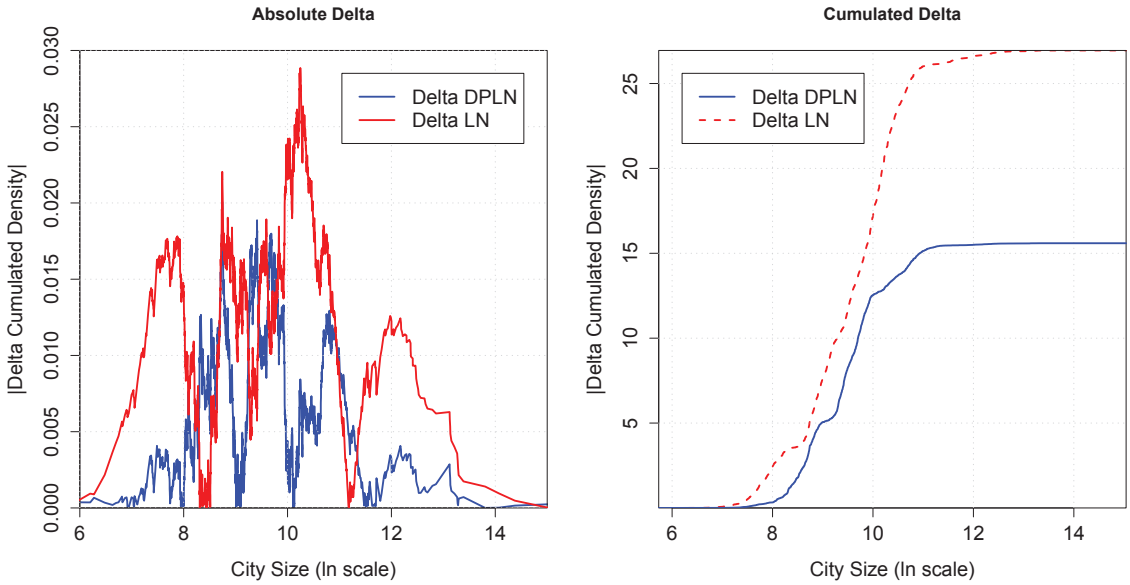

Figure 3.4: Deviations of LN and DPLN to the empirical distribution - Germany 2006.

In figures 3.7-3.9 we plot the empirical cdfs for the three countries with respective confidence bands, which are constructed by using approximations for the critical levels of the 95% KS-test statistics (see Bickel and Doksum, 2001).[7] The panels on the left refer to the overall cdf and the panels on the right zoom onto the upper tail. For the German case, both theoretical distributions consistently fall inside the 95% confidence band. In other words, *statistically* both distributions cannot be rejected at the 5%-level. For the case of the US, both distributions are sometimes located outside the band in the bottom and medium range, which can be detected in the left panel of figure 3.8. However, it can be shown that the LN tends to fall outside that band more often and more clearly than the DPLN. Focussing only on the upper tail (right panel of figure 3.8), both the DPLN and the LN are located inside the 95% confidence band throughout. Hence, both parameterizations cannot be rejected in that range of city sizes roughly exceeding exp(10)≈ 22,000 inhabitants. For the case of France the graphical analysis reveals particularly clearly that the DPLN delivers a better fit than the LN, as the latter distribution is actually rejected for a quite wide range of city sizes.

Analogous figures can be provided for the five additional countries for which we have

[7] A similar technique has been used by Eeckhout (2009), who constructs a confidence band for the theoretical (LN) distribution and analyzes if the actual distribution falls inside that band. Our approach of constructing a confidence band for the empirical cdf is useful, because we jointly consider the performance of two theoretical distributions.

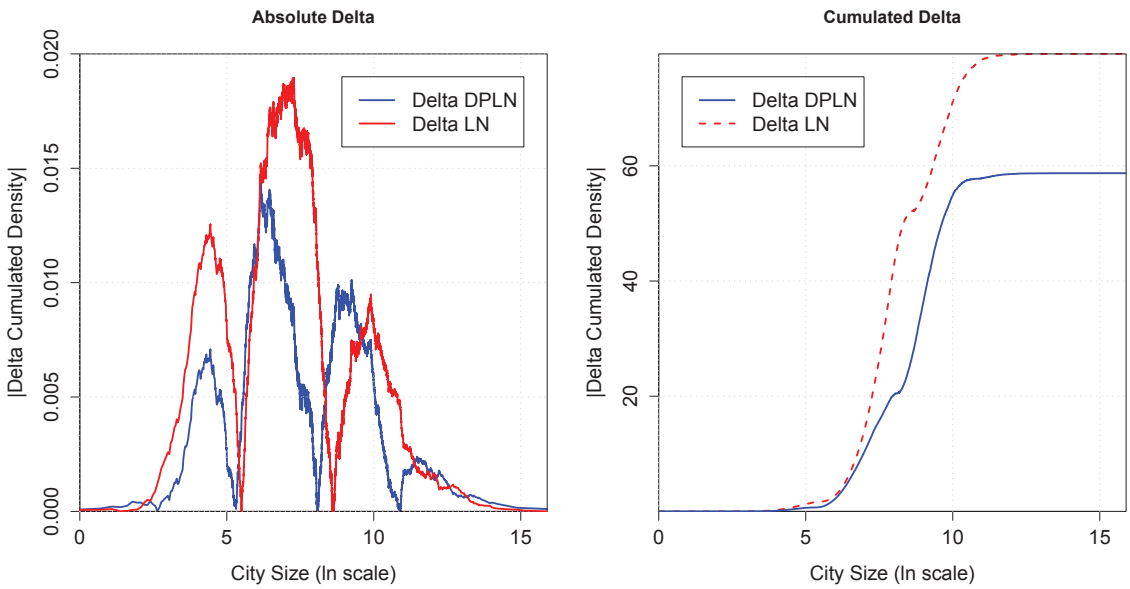

Figure 3.5: Deviations of LN and DPLN to the empirical distribution - USA 2000.

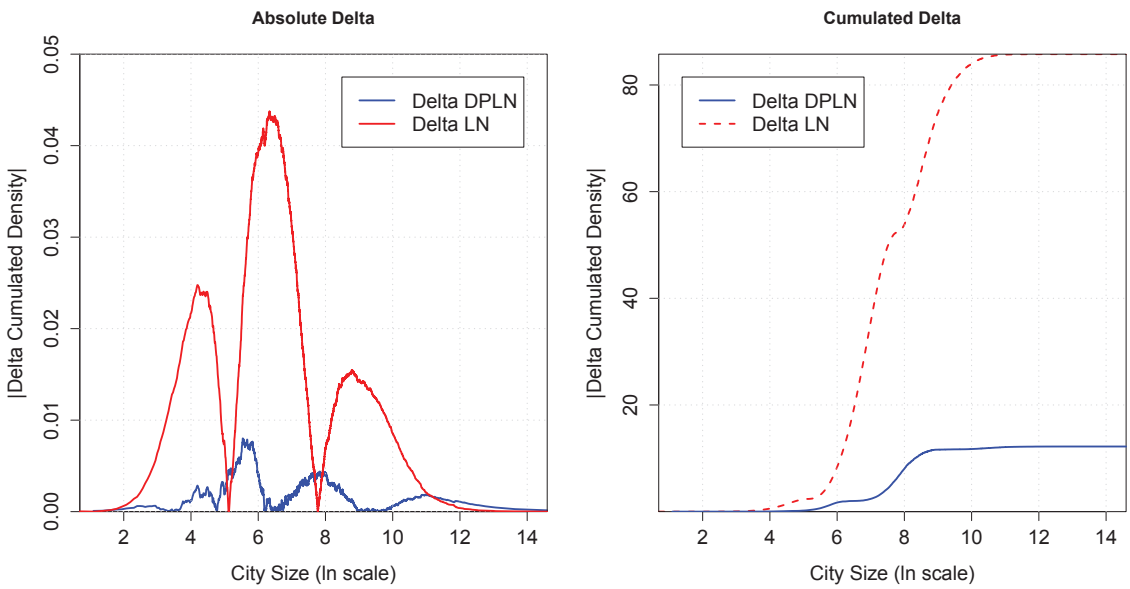

Figure 3.6: Deviations of LN and DPLN to the empirical distribution - France 2006.

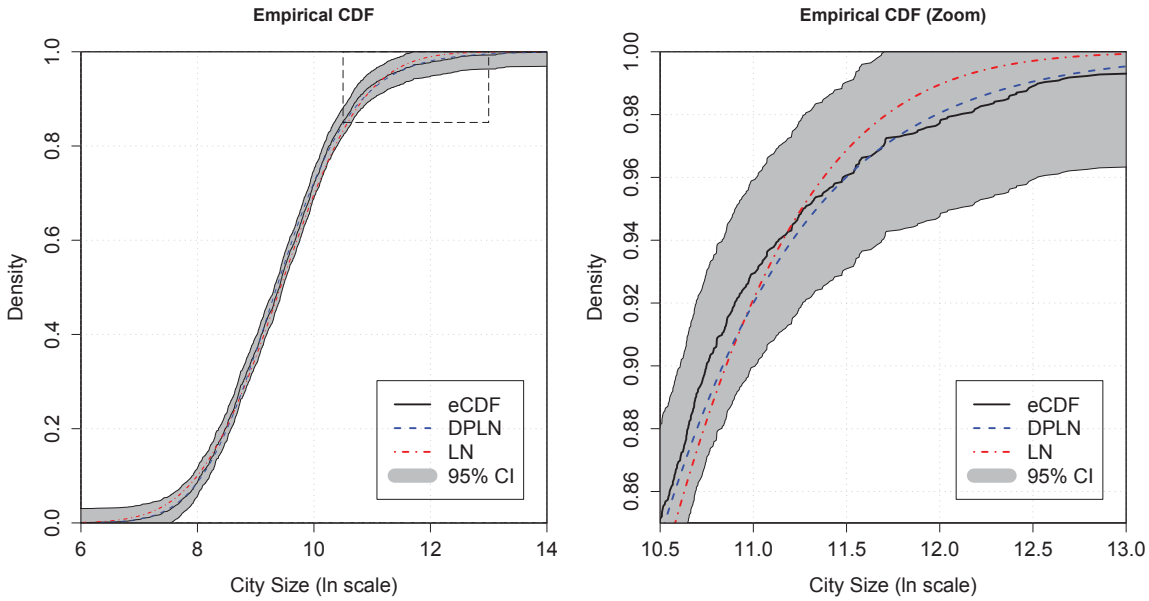

Figure 3.7: Empirical city size distribution with 95 %-confidence interval - Germany 2006.

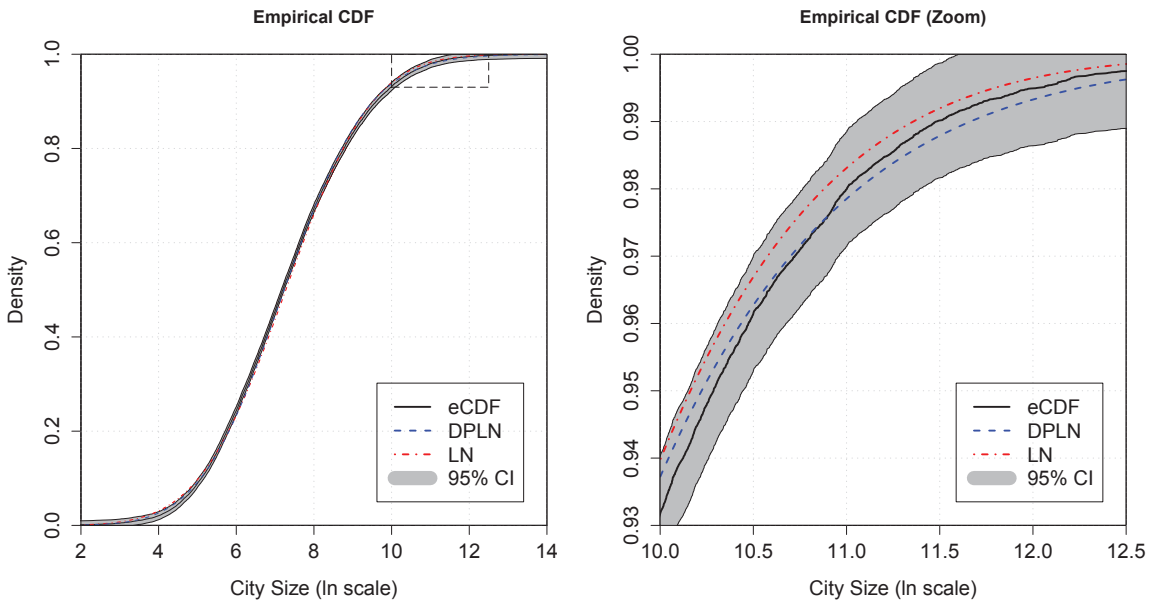

Figure 3.8: Empirical city size distribution with 95 %-confidence interval - USA 2000.

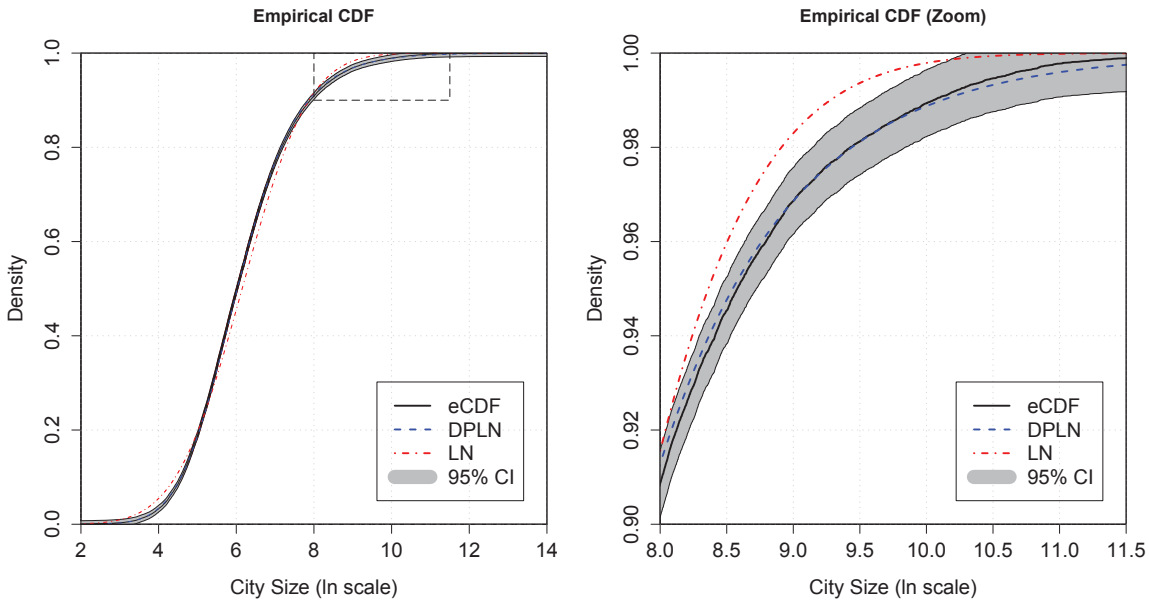

Figure 3.9: Empirical city size distribution with 95 %-confidence interval - France 2006.

sufficient data, but they are omitted for brevity. They reveal a qualitatively similar picture: The LN fits the data well, and most of the time it cannot be rejected statistically. However, the DPLN delivers a better performance both in the body and in the tails of the distribution.

3.4.4 Formal tests

Turning now to more formal tests, table 3.2 condenses the information from figures 3.1-3.3 by integrating up the pointwise vertical differences of the respective theoretical from the empirical distribution. In Germany, for example, the deviations sum up to 15.59 for the DPLN and to 26.94 for the LN (also see right panel of figure 3.10), which implies that the LN has 72% higher cumulated deviations. Results look similar for the other countries, i.e., the LN leads to a larger sum of deviations everywhere. The performance difference is particularly strong in France and in the Czech Republic, and smallest in Switzerland and in the US.

From a statistical point of view, the DPLN has a natural advantage as it is the more flexible functional form. We therefore use the log-likelihoods reported in table 3.1 to compute Akaike's information criterion (AIC) and the related Schwarz criterion (also called "Bayesian information criterion", BIC). Both are model selection criteria that trade-off the precision of a hypothesized distribution and the number of parameters that need to be estimated. Table 3.2 reports the results. By construction, the distribution with the lower numerical value of the AIC (BIC) is favored. Looking first at the AIC, we find that the values for the DPLN are consistently lower than for the LN distribution in all countries. Turning to the BIC, we obtain a consistent result for seven cases, but for

	Germany (2006)		**US (2000)**		**France (2006)**		**Brazil (2007)**	
	LN	**DPLN**	**LN**	**DPLN**	**LN**	**DPLN**	**LN**	**DPLN**
Cum. Diff.	26.95	15.59	79.43	58.72	85.78	12.21	140.39	55.38
AIC	45,457	45,382	469,550	469,428	576,971	574,781	121,650	121,079
BIC	45,468	45,404	469,566	469,461	576,988	574,815	121,664	121,106
LR (p-Value)	78.24 (0.01)		126 (0.01)		2194.2 (0.01)		575.2 (0.01)	
Bayes Factor	< 0.0001		< 0.0001		< 0.0001		< 0.0001	
Jeffrey's Scale	Strong for DPLN		Strong for DPLN		Strong for DPLN		Strong for DPLN	
	Czech (2009)		**Hungary (2009)**		**Italy (2009)**		**Switzerland (2008)**	
	LN	**DPLN**	**LN**	**DPLN**	**LN**	**DPLN**	**LN**	**DPLN**
Cum. Diff.	53.06	6.12	29.19	9.13	38.99	8.34	5.19	3.50
AIC	96,938	96,335	53,501	53,412	154,744	154,680	46,888	46,886
BIC	96,951	96,362	53,514	53,436	154,758	154,708	46,900	46,910
LR (p-Value)	606.5 (0.01)		93.7 (0.01)		67.6 (0.01)		5.92 (0.052)	
Bayes Factor	< 0.0001		< 0.0001		< 0.0001		140.5	
Jeffrey's Scale	Strong for DPLN		Strong for DPLN		Strong for DPLN		Strong for LN	

Table 3.2: Model comparison LN versus DPLN.

Legend: The values in the first row report the cumulated vertical deviations between the respective theoretical distribution and the empirical distribution. The Akaike information criterion for country i and distribution j is computed as $AIC_j^i = 2 \cdot k_j - 2 \cdot ln(L_j^i)$ and the Schwarz criterion as $BIC_j^i = k_j \cdot ln(N^i) - 2 \cdot ln(L_j^i)$, with k_j denoting the number of free parameters of distribution j, N^i the number of data points (cities) in country i, and $ln(L_j^i)$ the log-likelihood as reported in table 3.1. Both model selection criteria favor the distribution j that yields the lower numerical value. The likelihood-ratio test statistic is calculated according to $LR^i = 2 \cdot (ln(L_{DPLN}^i) - ln(L_{LN}^i))$ and follows the $\chi^2(2)$-distribution. The critical value for a hypothesis test at the 5%-level is equal to 5.99. The Bayes factor for country i is obtained by $B^i \approx exp(S^i)$, where $S^i = \frac{1}{2}(BIC_{DPLN}^i - BIC_{LN}^i)$. The value of B^i can be interpreted by using Jeffrey's scale (see Kass and Raferty 1995), which implies strong evidence in favor of DPLN if $B^i < 1/10$, moderate evidence if $1/10 < B^i < 1/3$, and weak evidence if $1/3 < B^i < 1$. Values of B^i larger than one indicate evidence in favor of the LN distribution.

Switzerland the BIC is now in favor of the LN distribution. In the Swiss case the DPLN only leads to a marginally better fit than the LN (the log-likelihoods are almost the same). Since the BIC penalizes the use of additional parameters stronger than the AIC does, the former criterion thus indicates that the simpler model (LN) is sufficient while the latter criterion is still in favor of the richer model (DPLN). For the other seven countries both statistical selection criteria agree that the DPLN is the better suited parameterization.

Given the nested structure of LN and DPLN, we can also compare model performance by a standard likelihood-ratio test. The log-likelihoods are, respectively, denoted by $ln(L^i_{LN})$ and $ln(L^i_{DPLN})$ for country i, and the test statistic $LR^i = 2 \cdot (ln(L^i_{DPLN}) - ln(L^i_{LN}))$ follows the $\chi^2(2)$-distribution as the DPLN has two parameters more than the LN. As can be seen in table 3.2, the null hypothesis that the DPLN leads to no significant improvement can be rejected at a very high confidence level (P-value below 1%). The only exception is Switzerland, where we cannot reject the null at the 5%-level. Finally, another approach to model comparison are Bayes factors. This technique is a flexible Bayesian analogue to the likelihood-ratio test, and does not even require one model to be nested in the other. As shown in Kass and Raferty (1995), Bayes factors can be easily approximated by using the Schwarz criterion (BIC). Specifically, to compare the LN and the DPLN distribution we can calculate the Bayes factor for country i as $B^i \approx exp(S^i)$, where $S^i = \frac{1}{2}(BIC^i_{DPLN} - BIC^i_{LN})$. The value of B^i can be interpreted by using Jeffrey's scale, and the results in table 3.2 indicate that there is *strong* evidence in favor of the DPLN. Consistent with our previous results, we find that Switzerland is an exception as the LN is the strongly preferred model for that country.

Summing up, with the exception of Switzerland all model selection criteria clearly show that the DPLN is the better suited model for the true city size distribution, even after being penalized for having more parameters than the LN.

3.4.5 Rank-size plots for the upper tail

Last, reminiscent of the debate between Levy (2009) and Eeckhout (2009), we analyze the top range in greater detail by using rank-size plots which are a standard tool in the Zipf literature. This final part focuses on the German case for brevity. The dots in figure 3.10 refer to the actual city sizes of the 100 largest cities (accounting for roughly 27m people or 33 % of the German population) and their respective rank in the national urban hierarchy. The dot-dashed line represents a random sample of the fitted LN distribution, where we rely on 500 iterations. This line indicates how the rank-size plot would look like if the underlying city size distribution were a LN with parameters given above. Similarly, the dashed line represents the sample of the fitted DPLN. The plot on the left is in logarithmic scale. It reveals that the DPLN fits the data very well in the upper tail, which is consistent with the argument by Levy (2009) that the sizes of the largest

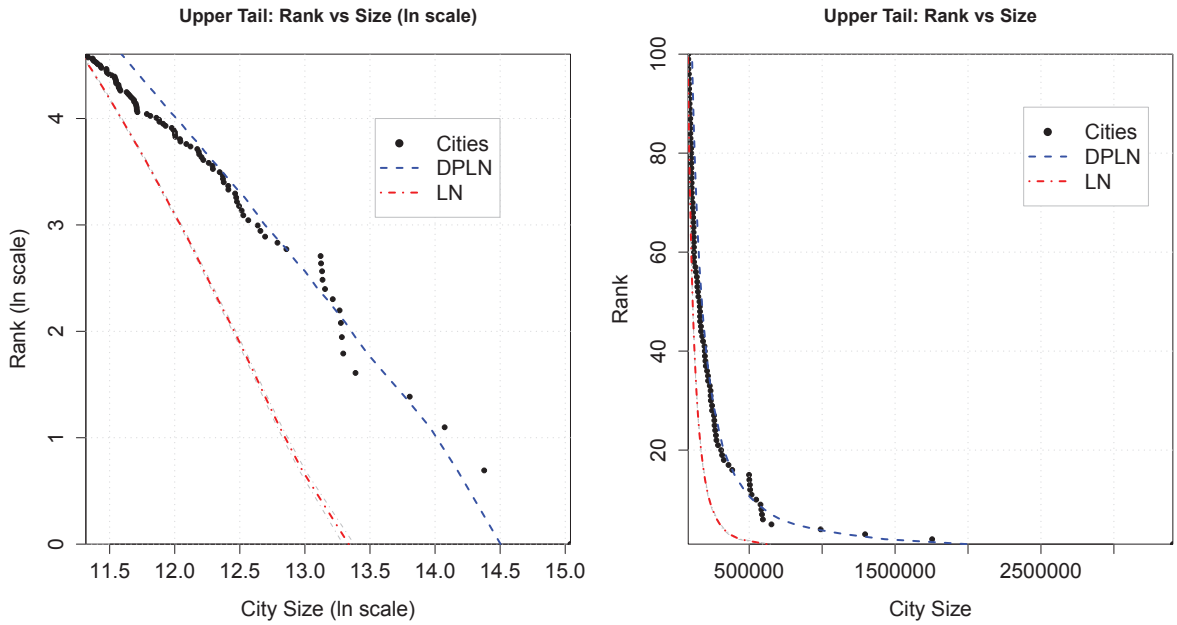

Figure 3.10: Rank-Size plot for the 100 largest cities - Germany 2006.

cities follow a power law. To address the issue raised by Eeckhout (2009), that the low rank logarithm observations lead to a bias in log-log-plots, we provide the same chart in standard scale on the right. This figure leads to essentially the same insight, however.[8]

It should be noted that these rank-size plots for the upper tail are just one possible method of addressing the goodness of fit of a theoretical distribution. In this paper we have considered several alternative approaches that typically contemplate the *overall* size distribution and not only the upper tail. We thereby followed the notion by Eeckhout (2009) who argues that the focus on the large cities is problematic as the definition of the truncation point is mostly arbitrary.

3.5 Conclusion

The various methods that have been used in this paper lead to a consistent picture: Although the lognormal (LN) does a good job in fitting the empirical city size distribution across all settlements of a country, the "double Pareto lognormal" (DPLN) distribution does a better job - even after taking into consideration that there are more parameters to be estimated.

Our findings have two main implications. First, they suggest that urban growth across all cities may be better described by the generalized Gibrat process developed in

[8] A similar plot can also be produced for the lower tail of the distribution. Among the 100 smallest cities we also find a distinctive power law pattern that is precisely in line with the predictions of the DPLN distribution. Furthermore we have also conducted an analysis by using log-density-plots, similar as in Eeckhout (2009). That approach also corroborates our findings of the better performance of the DPLN.

Reed (2002), rather than by the pure form of Gibrat's law. Even though our evidence is indirect, as we do not compare the growth processes directly but the theoretical steady-state distributions, it is consistent with some recent work which also points out that the pure Gibrat's law does not perform well when taking into account all types of settlements (see Michaels et al., 2008). Second, our findings may reconcile the recent debate about city size distributions between Eeckhout (2004, 2009) and Levy (2009) and thereby also build a bridge to the older Zipf literature. The DPLN parameterization implies that city size distributions have a lognormal shape over a wide range, but feature a distinct power law pattern in the tails. These features, in particular the mixture of lognormal with Pareto behavior among large cities, are nicely consistent with the empirical findings by Levy (2009) which have been recently confirmed by Ioannides and Skouras (2009). The urban growth process formalized in Reed (2002) and the resulting asymptotic DPLN distribution may therefore theoretically rationalize those empirical observations.

An issue that is not covered in this paper are the economic microfoundations of urban growth processes. For the pure form of Gibrat's law there already exist economic theories that clarify the foundations for scale-independent urban growth (most notably Eeckhout 2004). The theory by Reed (2002) is still more statistical in nature. It would be interesting to explore which economic forces can give rise to the mechanism of random city formation that is crucial for the Reed-Yule-process. One could, for example, try to extend the Eeckhout-model to allow for an endogenous number of locations by incorporating city birth and death in the style of Henderson (1974). Some recent papers have started, though in a somewhat different context, to explore such questions (e.g., Rossi-Hansberg and Wright 2007), but certainly more work is needed in this area.

Chapter 4

The overall French city size distribution

4.1 Abstract

We analyze the overall size distribution across all French settlements in the year 2008. The sizes of the largest French cities follow the famous Zipf's law fairly closely, with Paris being a notable outlier. However, for the overall city size distribution (CSD), Zipf's law is not a useful approximation. We show that the lognormal (LN) distribution does a reasonable job in fitting the overall French CSD. Yet, it is clearly outperformed by a different parameterization – the double Pareto lognormal (DPLN) distribution. This is consistent with our previous findings for city sizes in the US and other countries. We discuss the implications of these results for urban growth theory.

4.2 Introduction

The famous Zipf's law is probably the most extensively studied empirical regularity in urban economics. It states that the largest cities within a country approximately follow a Pareto distribution with shape parameter equal to minus one. This law is frequently expressed in an equivalent form as the rank-size rule for city sizes, where it states that the country's largest city is roughly twice as large as the second-largest, three times as large as the third-largest city, and so on.[1]

A major drawback of this traditional literature on Zipf's law, however, is its focus on the upper tail of the city size distribution (CSD). In former times, researchers interested in the CSD of some country were forced to focus only on the largest cities within that country, simply because reliable data about population sizes were only available for them but not for smaller cities, towns, villages, etc. As data availability improved, it became increasingly clear that Zipf's law is not a useful description for the *overall* CSD, but that it pertains – if at all – only in the upper tail. This, however, raises several questions: Where does the upper tail start, i.e., what is a "large" city? As we show below, this issue is actually crucial because the empirical performance of Zipf's law depends systematically on the number of cities included in the analysis. Even more fundamentally, the question arises why one should truncate the sample of settlements in the first place if data for the overall population distribution across space is available. Why should we focus only on the top of the urban hierarchy and forget about the rest, if data does not force us to do so?

The recent urban literature has therefore shifted its attention away from the upper tail and towards the *overall* size distribution across *all* "cities" of the country.[2] In that literature, which has been initiated by Eeckhout (2004) in his seminal article, at least

[1] Comprehensive studies found that city sizes in most countries indeed closely follow a Pareto distribution, but that the Zipf coefficients often deviate from unity. See Rosen and Resnick (1980), Soo (2005) or Nitsch (2005).

[2] From now on, we use the term "city" synonymously also for small towns and villages.

three main issues came up that are intensively debated ever since: First, what is the most appropriate parameterization for the overall CSD? Second, what is the relationship of this CSD with the traditional Zipf's law, i.e., has the new evidence basically invalidated decades of research on the rank-size rule? Third, and maybe most importantly, where do these parameterizations come from and what can we learn from them about the engines of urban growth?

In this paper, we focus on the case of France and reconsider some of the recent issues and controversies about overall CSDs. Most work in that area, including our own, has been done for the US urban system. Focussing on a leading European country is thus of interest in its own sake. Furthermore, the available data for settlement sizes in France (the *communes*) are outstandingly good by international standards, whereas the comparable US data are plagued by many more concerns regarding their comprehensiveness and accuracy. In section 4.3 we introduce these data.

In section 4.4, we start along traditional lines and focus only on the upper tail. We show that the largest French cities fairly closely follow a Pareto distribution, even though Paris is much larger than it accordingly "should be". Yet, whether we generally find evidence for or against the exact Zipf's law crucially depends on the definition of the upper tail, i.e., where we truncate the sample of cities. In section 4.5 , we then move to the overall French CSD. Eeckhout (2004) has provided a theory according to which the overall CSD should converge to a lognormal (LN) distribution, and he showed that the LN indeed fits the size distribution across US "cities" (defined as Census places) quite well. This is bad news for the traditional Zipf literature. If the "true" distribution is LN, there is no Pareto distribution among large cities. Why have so many papers then found evidence for Zipf's law? The answer according to Eeckhout is that these studies may have simply misperceived the LN for the Pareto by looking only at a sample of large cities, because the two distributions have similar properties in the upper tail. In short, Zipf's law is just an illusion!

Several authors, most notably Levy (2009), Ioannides and Skouras (2009) and Malevergne et al. (2011), have contested this conclusion and argued that the LN has serious deficits in matching the US places data. In particular, they argue that the LN may fit well for small and medium-sized places, but that the sizes of the large cities are distinctively closer to a Pareto than implied by the LN distribution. They hence argue that the "true" parameterization for the overall CSD should consist of a LN which then switches to Pareto behaviour beyond a certain threshold city size. They do, however, not provide a theory why such a functional form for the overall CSD should emerge endogenously in an urban system. For the French case, we find that the LN distribution does at best a reasonable job in matching the city size data, comparatively much worse than in the US. Interestingly, the deficits of the LN arise over the entire range of city sizes and not just in the upper tail, as can be seen in Figure 4.4 below. This suggests that an ad-hoc mixture

model for the overall CSD that mechanically switches from LN to Pareto at some point may also have a hard time matching the French data.

In section 4.6 we then provide a resolution to this puzzle and suggest a parameterization that fits the French overall CSD extremely closely: the *double Pareto lognormal* (DPLN) distribution. In previous research, see Giesen and Suedekum (2012a) and Giesen et al. (2010), we have shown that this flexible distribution closely fits the overall CSD in the US and in other countries.[3] The first bottom-line message of this paper is, therefore, that the French overall CSD can be approximated by the same functional form that also performs very well elsewhere. This robust evidence in favour of the DPLN is good news for the older Zipf literature. In contrast to the LN, the DPLN is fully consistent with a Zipfian power law pattern that emerges as an upper tail feature of an overall functional form. When the underlying "true" distribution is DPLN, claiming that the sizes of large cities follow a power law is no systematic mistake. Zipf's law is, hence, not an illusion!

Even more importantly, in Giesen and Suedekum (2012a) we develop a micro-founded economic model of an urban system where city sizes endogenously converge to a DPLN distribution. In other words, the DPLN is not an ad-hoc functional form that is chosen purely on the basis of data fit. It has an explicit theoretical foundation and can be rationalized by an economic model that combines scale-independent urban growth with age heterogeneity across cities. The second bottom-line message is, hence, that the French case analyzed in this paper yields further corroborating evidence for our urban growth model which apparently matches cross-sectional CSDs in many countries very successfully.

4.3 Data

The main data set that we use in this paper comes from the French National Institute of Statistics and Economic Studies (INSEE). It contains the population sizes of 36,682 French municipalities *(communes)* in the year 2008 (including the overseas departments), in total accounting for 63,961,859 people.[4] The *communes* are administrative units, so their boundaries are legally and not economically defined. In that sense, they correspond to the US Census places that have been used in most of the recent urban literature, including Eeckhout (2004). However, the French administrative settlement size data is more comprehensive and subject to much less concern than its US counterpart.

The key issue here is that the Census places only represent about 74 % of the total US population in the year 2000. The remaining 26 % live in settlements that are neither

[3]For the US, this is true both when using administratively defined Census places as the unit of analysis, but also when using the recently developed area clusters by Rozenfeld et al. (2011) which are constructed from the "bottom-up" by using high resolution data on population density in the US.

[4]Many more details about this data as well as a historical excursion when and why it was first collected can be found under http://www.insee.fr/fr/methodes/nomenclatures/cog/documentation.asp (last accessed on April 9, 2012).

counted as "incorporated" nor as "Census designated" places. Whether a settlement is an official Census place or not, is not primarily selected based on its population size. There are Census places with only one or two inhabitants. However, especially settlements in the rural parts of relatively large metropolitan areas are often not considered as "places" and are thus ignored in the data set. What is more, the definition of Census places also varies quite substantially across the US Federal States. These problems raise the concern that the Census places may be a selective or biased representation of the overall US CSD, since it is unknown how the remaining 26 % of the US population not captured by the data spreads across space.

The French administrative data set does not face such issues. It basically represents the entire French population in 2008 and thus gives an comprehensive portray of the overall (untruncated) French CSD, ranging from the administrative entity of Paris with 2,211,297 inhabitants down to the commune of Rochefourchat with exactly one inhabitant. Table 4.1 shows the ten largest *communes* and their respective population sizes in 2008.

Municipalities		Agglomerations	
Name	Size	Name	Size
Paris	2,211,297	Paris	10,413,386
Marseille	851,420	Marseille	1,557,950
Lyon	474,946	Lyon	1,536,974
Toulouse	439,553	Lille	1,015,744
Nice	344,875	Nice	941,490
Nantes	283,288	Toulouse	871,961
Strasbourg	272,116	Bordeaux	836,162
Montpellier	252,998	Nantes	586,078
Bordeaux	235,891	Toulon	559,246

Table 4.1: The ten largest Municipalities/Agglomerations in France, 2008.

Still, there is the concern that the single units are defined according to administrative boundaries which can be quite arbitrary. Because of this, *communes* are often treated as separate units/cities even though they are essentially part of the same city. A principal alternative is to abandon administrative data and to use urban agglomerations data instead. We also consider such data in this paper, more specifically the population sizes of the major 247 French urban areas in 2008, as also provided by the INSEE. Here, the Paris agglomeration is on top of the urban hierarchy with a population of more than 10 million people (also see table 4.1 for the ten largest French agglomerations). However, this data in total only represents around 37 million people, i.e., less than 60 % of the total French population. It is also selective in the sense that it is truncated from below (the smallest urban area is Bar-le-Duc with 19,321 inhabitants), and that it does not include the rural population outside the big cities. For our analysis of the overall French CSD

these data are thus less useful, although we may still use it in section 4.4 where we focus only on the upper tail.

A novel and very interesting approach of defining cities has recently been developed by Rozenfeld et al. (2008, 2011). Here, cities are defined from the "bottom-up" by using an algorithm on high resolution data on population densities in a country. The advantage of this approach of defining "cities" (also called "area clusters") is that it comprehensively portrays the overall distribution of the entire population across space. It completely ignores artificial administrative boundaries, but it is not limited to metro areas beyond a certain threshold size. Unfortunately, such area clusters data – which would be ideally suited for our type of analysis – does not yet exist for France to the best of our knowledge. So far, Rozenfeld et al. (2008) have only provided it for the US and Great Britain, and we have analyzed that data in our previous research, see Giesen and Suedekum (2012a).

4.4 Large cities in France: Zipf's law?

Zipf's law is, strictly speaking, about two different statements. The first statement claims that city sizes follow a Pareto distribution. The second statement is that the slope coefficient of the Pareto is equal to minus one. Under a Pareto, cities are thus distributed according to

$$P(s > S) = \left(\frac{A}{S}\right)^{\zeta}, \tag{4.1}$$

where ζ denotes the shape parameter of the Pareto distribution, also known as the Zipf coefficient. The rank of a city in the urban hierarchy is given by $R = N \cdot P(s > S)$, so the parameters of eq. (4.1) can be estimated by

$$log(R) = K - \zeta log(S). \tag{4.2}$$

where $K = \zeta log(A) + log(N)$. If Zipf's law holds exactly, we have $\zeta = 1$.

We arrange the data so that cities are ordered by their size and labeled with their respective rank; Paris has rank 1, Marseille has rank 2, Lyon has rank 3, and so on. We then run the standard rank-size regression as stated in eq. (4.2) by simple OLS.[5]

4.4.1 The communes

We start off with the municipalities data and focus on the 100 largest French *communes*. This truncated sample of cities, where the threshold rank $\bar{R}$ is basically chosen arbitrarily, represents 13,895,689 people, i.e., around 22% of the total French population. The rank-

[5]There are also more sophisticated ways of estimating the Zipf coefficient, see e.g. Gabaix and Ibragimov (2011) or, for an overview, Gabaix and Ioannides (2004). However, since this is not the focus of our paper we only use the simplest and most standard rank-size regression technique.

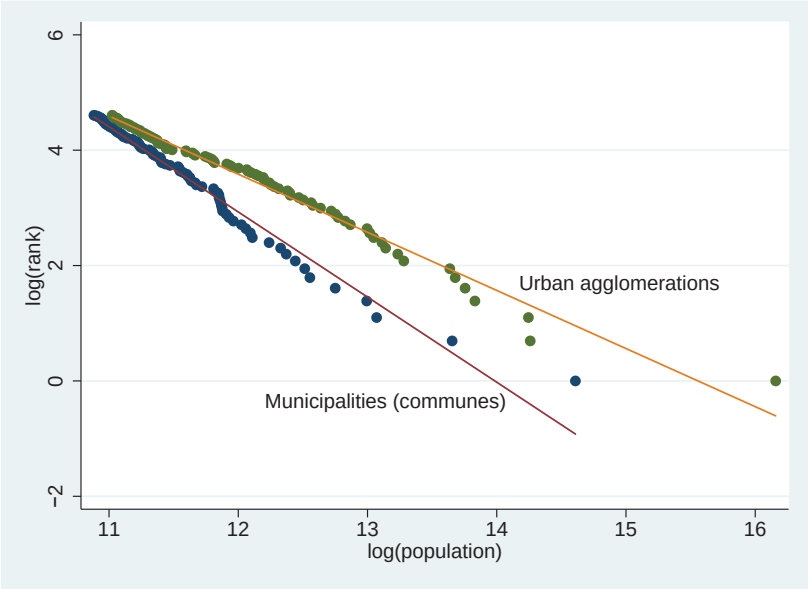

Figure 4.1: Zipf regression for the 100 largest municipalities/agglomerations.

size relationship is graphically illustrated in figure 4.1, where we depict the log population size of the cities on the horizontal and their log rank in the urban hierarchy on the vertical axis. When estimating the rank-size regression (4.2) for these 100 cities, we obtain a slope coefficient of $\zeta = 1.476$ with a standard error of $\sigma = 0.022$ and a R^2 of 0.98.

This regression and the corresponding scatter plot in figure 4.1 convey three main messages: First, the graphical rank-size relationship looks almost linear, which is equivalent to saying that the city sizes of the largest French communes tend to follow a Pareto distribution fairly closely. This statement is supported by the overwhelmingly high R^2 level of the linear regression.[6] Second, there is one clear outlier: Paris. The capital city of France is much larger than it "should be" according to a power law for city sizes, also when focussing on administrative city definitions. This is a quite typical pattern discussed in detail by Ades and Glaeser (1995) who show that particular political forces often cause the capital city to be unusually large in the urban hierarchy. If we leave Paris out of the picture, the rank-size relationship would appear even more linear, and in fact, when estimating eq. (4.2) only for the cities ranked 2-100, we obtain an even higher $R^2 = 0.991$ and a slope coefficient of $\zeta = 1.576$ (std.err. 0.015).

The third message is that the exact Zipf's law apparently fails to hold in the French case. The estimated slope coefficient deviates substantially from one, particularly when leaving Paris out of the regression ($\zeta = 1.576$), but also when leaving it in and using all cities ranked 1 to 100 in the French urban hierarchy ($\zeta = 1.476$). This evidence *against*

[6] It is well known that rank-size regressions automatically yield high R^2 levels, simply because of the ordering of cities by rank. Monte Carlo simulations show that, even if city sizes hypothetically followed a uniform distribution, such regressions would still deliver an R^2 around 0.8 (also see Gan et al., 2006). However, R^2 levels exceeding 0.98 cannot be regarded as artificial evidence for a Pareto distribution, but those levels can only be obtained if there is actually a power law relationship in the sizes of the cities.

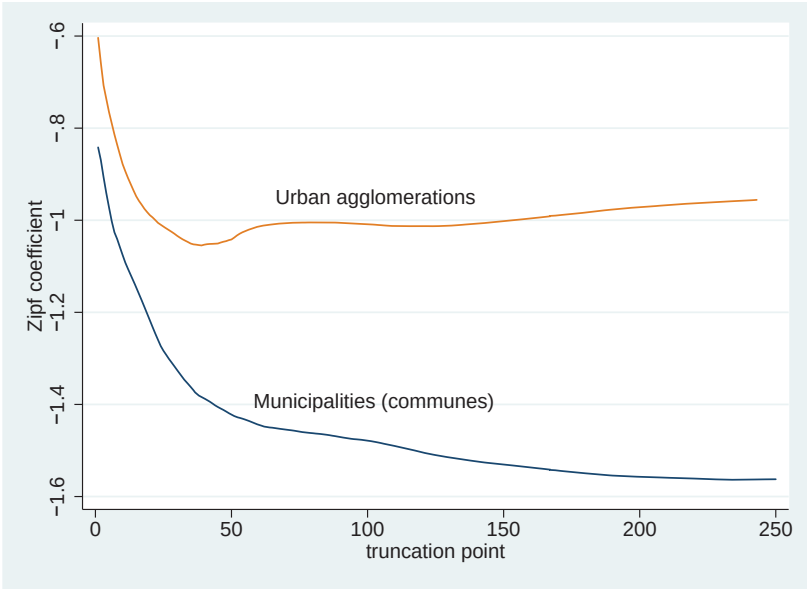

Figure 4.2: The Zipf coefficient and the truncation point.

the exact Zipf's law is, however, very sensitive to the arbitrary truncation point $\bar{R}$ where the the CSD is cropped. Suppose we set $\bar{R} = 10$, i.e., we focus only on the ten largest communes. In that case, we get $\zeta = 1.001$ (std.err. 0.071, R^2=0.96) and would have to conclude that Zipf's law holds exactly. If we only take the largest five cities, we have $\zeta = 0.841$ (std.err. 0.079, R^2=0.97), and so on.

In figure 4.2 we show how the estimated slope coefficient ζ varies with the choice of the truncation point $\bar{R}$. The figure shows that a wider definition of the upper tail (a higher $\bar{R}$) tends to increase the Zipf coefficient in absolute terms. More generally, the figure shows that Eeckhout's (2004) important insight about the US urban system also applies for France: By the choice of the truncation point, researchers can manipulate whether they obtain evidence in favour of or against the exact Zipf's law. A power law shape for city sizes seems to prevail almost regardless of how the truncation point is set (as long as $\bar{R}$ is not too large) , but whether the slope coefficient is close to the magical $\zeta = 1$ depends very much on the definition of the "upper tail" of the CSD.

There are no generally accepted rules how this truncation point should be chosen, and if there are rules, they tend to lack economic foundations (see Chesire, 1999).[7] More fundamentally, even if one could agree on an appropriate definition of $\bar{R}$, the question remains why we should truncate the sample of cities in the first place. Why should the cities below this threshold be disregarded, even though we do have detailed knowledge about their population sizes? Because of issues like this, researchers have gradually departed from analyses focussed only on the upper tail, and towards inquiries about the *overall* size distribution across *all* settlements of a country.

[7]This general point also implies that a cross-country comparison of ζ is difficult, because one has to make sure that comparable rules for the choice of the truncation points are applied in all countries.

4.4.2 Urban areas

Before moving to the analysis of the overall CSD, we briefly consider the other data set where French cities are defined as urban agglomeration areas. In figure 4.1, we illustrate the rank-size relationship when we analogously focus on the 100 largest urban areas, together representing 32,433,021 people, or 51.9% of the total French population. Again we find that the Paris area is "too large" given the benchmark of a perfect power law. The second-largest agglomeration, Marseille, is "too small" given this benchmark. However, by and large, that rank-size relationship still looks almost linear, and when estimating the standard Zipf regression for these 100 cities we obtain a highly significant slope coefficient equal to $\zeta = 1.0075$ (standard error 0.131) and a R^2 of 0.983. In other words, across the largest 100 French urban areas, Zipf's law holds exactly. Figure 4.2 suggests that the slope coefficient ζ is also much less sensitive to the truncation point for the urban agglomeration data. Even when including all 247 urban areas, we get $\zeta = 0.955$ (std.err. 0.005, R^2=0.99) which is not much different from the coefficient estimated before.[8]

The main advantage of the urban agglomeration data is that it ignores arbitrary administrative boundaries in the definition of cities. In that sense, it is preferable to the *communes*. The evidence in figures 4.1 and 4.2 suggests that, across those sensibly defined cities, Zipf's law seems to be quite stable – maybe except in the very upper tail. However, recall that the urban areas together capture only 60 % of the French population, so it is unclear if Zipf's law continues to hold so well if we included also the remaining 40 %. A step ahead would be to develop a concept of cities that does not proceed along administrative boundaries, but that still captures the entire population living in the country. The "bottom-up" approach by Rozenfeld et al. (2008, 2011), who define area clusters for the US and Great Britain, seems highly promising in that respect. However, as said before, such data does not yet exist for France. For the US, Rozenfeld et al. (2011) found that Zipf's law very well describes the size distribution across all area clusters larger than 13,000 inhabitants. Similarly, in Great Britain, Zipf performed well for clusters larger than 5,000 people. But outside that upper tail, the law again breaks down and cities no longer obey to a Pareto distribution.

Generalizing those results to France, we can speculate that the power law shape probably continues to hold even a bit further down the urban hierarchy, if more comprehensive data about area clusters below the smallest recorded urban area (Bar-le-Duc with 19,321 inhabitants) were available. However, eventually Zipf's law would very likely break down as well, once we have moved down the hierarchy far enough. In other words, also with urban agglomeration data, Zipf's law is not a useful description for the overall CSD. We have to think about different parameterizations, while bearing in mind that a Zipfian

[8] There are some deviations when we focus only on the very largest urban areas. For example, with $\bar{R}$=10 we get $\zeta = 0.7894$ (std.err. 0.099, R^2=0.88), and so clear evidence against Zipf's law and even against a power law shape.

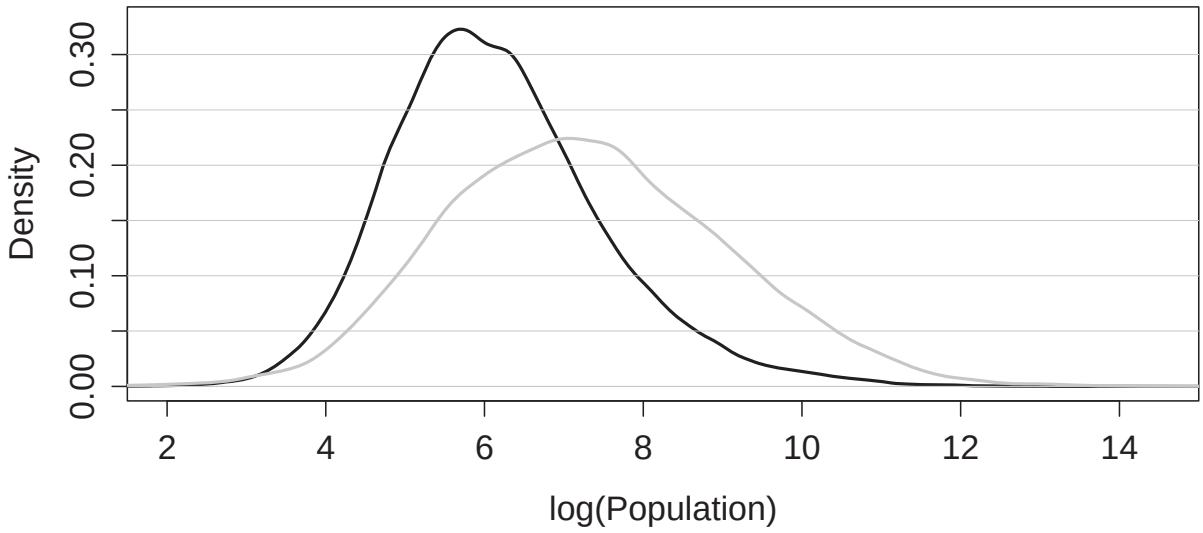

Figure 4.3: Kernel density estimate of the French and the US overall CSD.

power law seems to be really pervasive in the upper tail.

4.5 The overall city size distribution

From now on, we concentrate on the *overall* French CSD and thus on the administratively defined *communes* as the unit of analysis. In figure 4.3 we depict a kernel density estimation of the size distribution across all 36,682 municipalities where population sizes are in logarithmic scales, see the solid black curve. For the purpose of comparison, we also provide the comparable overall CSD for the US in that figure, more specifically the empirical log size distribution across 25,359 US Census places in the year 2000 (see the solid grey curve). It becomes very clear that a Pareto parameterization cannot possibly fit the overall CSDs, neither in France nor in the US. The log settlement sizes rather appear to be close, at least visually, to a normal distribution, though with different variance across countries.

4.5.1 Preliminaries: Random urban growth and the LN distribution

Eeckhout (2004) provides a theory according to which the overall CSD of a country should converge to a lognormal (LN) distribution. That theory is based on a random urban growth process where cities grow according to the pure Gibrat's law. More details about Eeckhout's model follow below. Applying the LN parameterization to the US data, Eeckhout (2004) indeed finds that it does a good job in matching the size distribution across Census Places. We have verified this result in our previous research, see Giesen

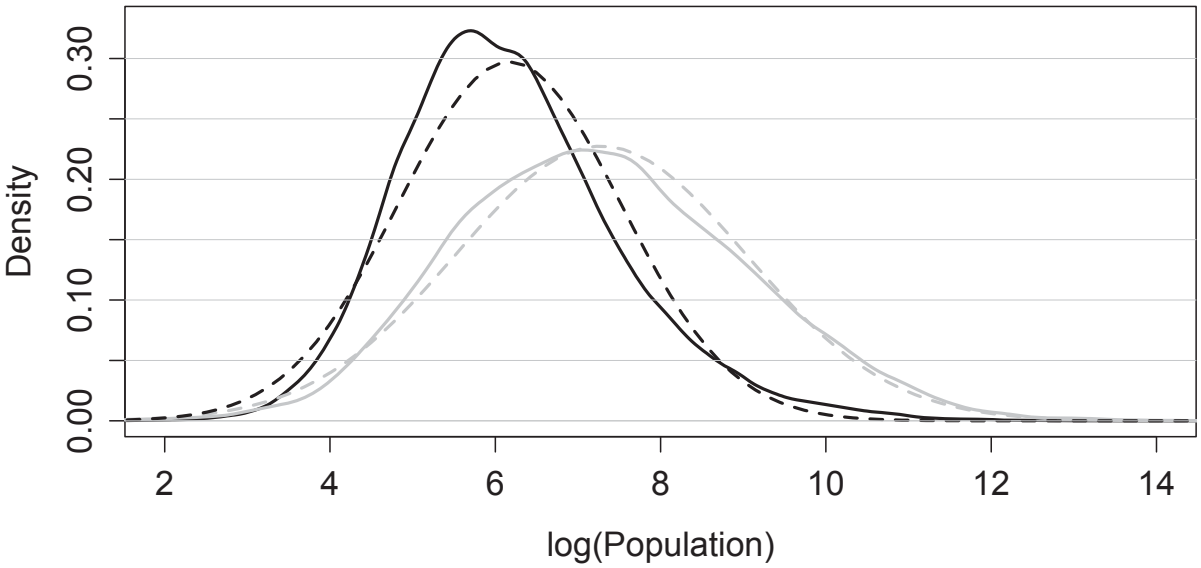

Figure 4.4: Kernel density estimate of the French and the US overall CSD.

et al. (2010), and Figure 4.4 illustrates this. As can be seen from the broken grey line, which represents the fitted LN distribution, it certainly does not deliver a perfect but still a decent fit.

This evidence for the overall CSD thus lends empirical support to Eeckhout's (2004) urban growth model. Yet, it has quite delicate implications for the traditional literature on Zipf's law. As a matter of fact, the LN does not feature a power law in the upper tail and, hence, it is strictly speaking not compatible with Pareto and Zipf. Why have so many previous studies then provided evidence for a Zipfian power law among large cities? The reason according to Eeckhout (2004, 2009) is that the LN and the Pareto distribution have similar properties in the upper tail and can become virtually indistinguishable. In other words, Zipf can be observed among large cities in practice, because the Pareto closely resembles the true size distribution (the LN) in the top range.[9] The definition of a "large city" also matters in this respect. As we have shown above, the estimated Pareto slope coefficients depend crucially on the truncation point within the sample of cities. Eeckhout (2004) proves that, if the underlying "true" distribution is LN, the coefficient estimate ζ is decreasing in $\bar{R}$ – a pattern that we have actually found for France in figure 4.2 and that can also be observed for the US data. Summing up, when the overall CSD is actually a LN, previous studies on Zipf's law may have fallen for an illusion.

[9]Also see Mitzenmacher (2004), who shows that the density or the countercumulative distribution function of the LN generate a "nearly straight" line in logarithmic plots when the variance is large. A power law (Pareto) would generate exactly a straight line in such plots.

4.5.2 Does the LN fit the the French data?

In this subsection we investigate whether the suggested LN parameterization fits the French city size data. Using maximum likelihood estimation, we find that the best fit of a LN parameterization to the empirical size distribution for French communes is achieved with parameters $\mu = 6.173$ and $\sigma = 1.343$, delivering a value of the log likelihood equal to -289,238.5. In figure 4.4 we depict the best fitting LN distribution as the broken black line.

Judged by pure visual inspection, it can be seen that the overall fit of the LN to the French data is fair at best. There are notable deviations, which occur over the entire range of city sizes. One issue is that the empirical CSD seems to have a fatter upper tail than the LN. In the lower tail, it is the other way around: The LN has more mass in the range of very small settlement sizes than the empirical distribution. More generally speaking, the actual French CSD exhibits a slight skew to the left, a distributional feature that by construction cannot be replicated by the LN which is symmetrical in logarithmic scales.

Comparing the data fit of the LN between France and the US, figure 4.4 shows that the LN fits much better to the US Census places than it does to the French communes. This conclusion can also be supported by a more formal statistical approach. We ran Kolmogorov-Smirnoff tests and compared the p-values for the null that the data follow a LN. We find that this hypothesis is rejected much earlier for France than for the US. For the case of France we thus obtain much weaker support for the theoretical framework by Eeckhout (2004) which predicts an asymptotic LN shape for the overall CSD. We leave the full discussion for later, but already preview our argument, which is that this difference may be caused by the stronger age heterogeneity of French cities as compared to American cities.

4.6 The DPLN distribution

In this section we suggest an alternative parameterization for the French overall CSD, the so-called "Double Pareto Lognormal" (DPLN) distribution. We then briefly outline the genesis of the DPLN and describe our urban growth model (see Giesen and Suedekum, 2012a) that endogenously leads to DPLN distributed city sizes. The model by Eeckhout (2004), which leads to a LN distribution, can be seen as a special case of our more general framework, and we discuss the origin of the differences below.

	French municipalities (2008)	
N	36,674	
Min	1	
Max	2,211,297	
	DPLN	LN
α	1.016	-
β	3.358	-
μ	5.588	6.173
σ	0.882	1.343
AIC	576,348	578,473
BIC	576,314	578,456
$ln(L_j)$	-288,178.0	-289,238.5
LR (p-value)	2121 (0.01)	
Bayes Factor	<0.0001	
Jeffrey's Scale	Strong for DPLN	

Table 4.2: Estimated parameters and formal selection tests.

4.6.1 Parameterization and data fit

The DPLN distribution was initially developed by the Canadian statistician and economist William J. Reed (2002). It has the following density for city sizes S:

$$f(S) = \frac{\alpha\beta}{\alpha+\beta}\left[S^{\beta-1}e^{\left(\beta\mu_0+\frac{\beta^2\sigma_0^2}{2}\right)}\Phi^c\left(\frac{log(S)-\mu_0+\beta\sigma_0^2}{\sigma_0}\right)\right.$$
$$\left.+S^{-\alpha-1}e^{\left(\alpha\mu_0+\frac{\alpha^2\sigma_0^2}{2}\right)}\Phi\left(\frac{log(S)-\mu_0-\alpha\sigma_0^2}{\sigma_0}\right)\right]. \tag{4.3}$$

The parameters α and β are coefficients to regulate the tails, whereas μ_0 and σ_0 determine the location and the spread of the distribution. Φ represents the normal cdf and $\Phi^c = 1-\Phi$ represents the complementary cdf. A special feature of this distribution is that if S is large, then $f(S) \sim S^{-\alpha-1}$ and if S is small, then $f(S) \sim S^{\beta-1}$. The DPLN therefore incorporates a Pareto distribution in the upper and a reverse Pareto distribution in the lower tail. Another special feature is that it nests the LN as a limiting case when $\{\alpha, \beta\} \to \infty$. For other values the body of the distribution is also close to a lognormal shape. However, the DPLN should not be though of as a rigid mixture of LN and two Paretos. It is rather a flexible parameterization that has several distributional features which the LN or the mixture model of LN and Pareto cannot capture. In particular, the DPLN can be skewed in log scale and its kurtosis can have positive or negative excess, i.e., it can be more peaked (leptokurtic) or more flat (platykurtic) than the LN.

It is straightforward to estimate the parameters of the DPLN as given in (4.3) by maximum likelihood.[10] The best fit for the French data is achieved with parameters

[10] We utilize the log-likelihood function and the corresponding starting values as proposed by Reed

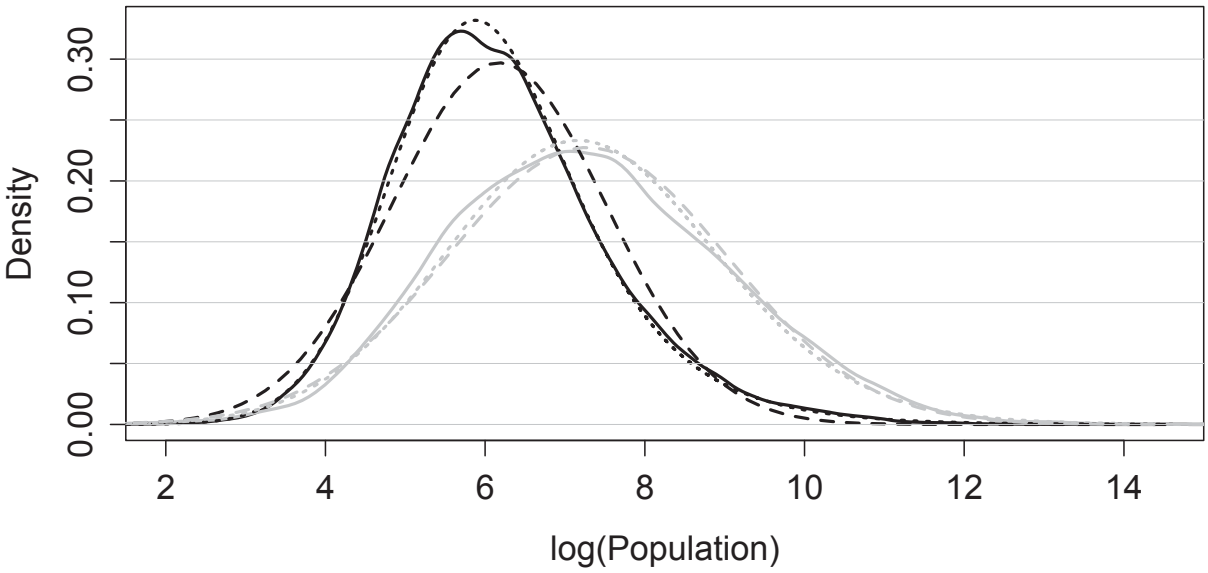

Figure 4.5: Kernel density estimate and best fit LN distribution.

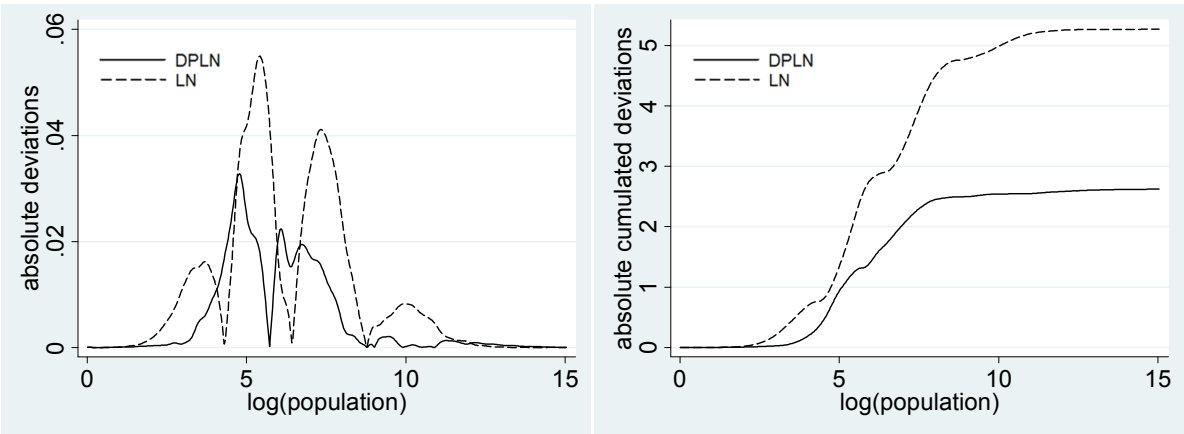

Figure 4.6: Deviations of LN and DPLN to the empirical distribution.

$\alpha = 1.016, \beta = 3.358, \mu_0 = 5.588$, and $\sigma_0 = 0.882$, yielding a log likelihood equal to $-288,178$ (also see table 4.2). In figure 4.5, the dotted black line represents the fitted DPLN distribution for France. Already visually it is clear that the DPLN fits the French city size data much better than the LN. Except for the small bump that occurs at log city sizes around 6, the DPLN is almost everywhere closely in line with the empirical CSD, while this is certainly not the case for the LN. The better fit is confirmed in figure 4.6, where we show the vertical deviations of both hypothesized parameterizations from the empirical CSD. The left panel depicts the pointwise, and the right panel the cumulated deviations. As can be seen, the DPLN fits the data better than the LN almost throughout the entire range of the distribution, and it has much lower overall deviations.

The DPLN has an advantage over the LN, because it is the more flexible functional

(2002).

form with four instead of two parameters. It therefore achieves a better data fit almost by definition. However, various model selection tests show that the DPLN also achieves a better *adjusted* data fit, when it is penalized for having more degrees of freedom. In particular, we use the log likelihoods of the LN and the DPLN as reported in table 4.2 to compute Akaike's information criterion (AIC) and the related Schwarz criterion (also called "Bayesian information criterion", BIC). Both criteria trade off the precision of a hypothesized distribution and the number of parameters. Table 4.2 reports the results. By construction, the distribution with the lower numerical value of the AIC (BIC) is favored. As can be seen, for both criteria we find that the values for the DPLN are lower than for the LN, thus implying that the DPLN is the better model from a statistical point of view.

Given the nested structure of LN and DPLN, we can also compare model performance by a standard likelihood-ratio test. The test statistic $LR = 2 \cdot (ln(L_{DPLN}) - ln(L_{LN}))$ follows the $\chi^2(2)$-distribution as the DPLN has two parameters more than the LN. It can be shown that the null hypothesis that the DPLN leads to no significant improvement compared to the LN can be rejected at a very high confidence level (P-value below 1%). Finally, another approach for model comparison are Bayes factors. This technique is a flexible Bayesian analogue to the likelihood-ratio test, and does not even require one model to be nested in the other. As shown in Kass and Raferty (1995), Bayes factors can be easily approximated by using the Schwarz criterion (BIC). Specifically, to compare the LN and the DPLN distribution we can calculate the Bayes factor as $B \approx exp(V)$, where $V = \frac{1}{2}(BIC_{DPLN} - BIC_{LN})$. The value of B can be interpreted by using Jeffrey's scale, and the results indicate that there is *strong* evidence in favor of the DPLN.

For the US Census place data, we depict the fitted DPLN as the dotted grey line in Figure 4.5. The performance difference between LN and DPLN is much less pronounced than in the French case. All model selection criteria would still favor the DPLN as the more appropriate functional form (also see Giesen et al., 2010), but the margin of improvement is lower. For example, when calculating the AIC for the US data, we obtain AIC(DPLN)=469,428 and AIC(LN)=469,550. That is, the AIC of the DPLN is only 0.026 % below the LN's AIC. For the BIC we have BIC(DPLN)=469,461 and BIC(LN)=469,566 in the US case, i.e., a value around 0.022 % lower. In the French case, the performance difference is around 16 to 18 times more pronounced, corroborating the visual impression that is delivered by figure 4.5.

Summing up, all model selection criteria clearly show that the DPLN is a very well suited functional form for the French empirical CSD, much better (even in adjusted terms) than the LN. In that respect, the French case is in line with the evidence that we have established in our previous research, where we show that the DPLN matches empirical CSDs both across countries and for different ways of defining "cities" very well.

4.6.2 Genesis of the DPLN

The DPLN is not an ad-hoc parameterization that is chosen purely to achieve a good data fit. In Giesen and Suedekum (2012a) we show that it actually emerges endogenously from a dynamic economic model of an urban system that combines scale-independent urban growth (Gibrat's law) as in Eeckhout (2004) with endogenous city creation and age heterogeneity across cities.

In Eeckhout's (2004) model, there is an economy with a fixed population and a given number of locations across which workers are freely mobile. The locations differ by their exogenous total factor productivities, and in every time period each location is hit by an idiosyncratic productivity shock that is drawn from a probability distribution with mean $\mu = 0$ and variance $\sigma^2 > 0$. At the city level, there is a trade-off between positive and negative size externalities that accrue within but do not spill over across locations. In a spatial equilibrium utility is equalized across locations, since workers are perfectly mobile across space. If a city experiences a positive productivity shock, this attracts people into the respective location. The negative externalities dominate at the city level, however, and this prevents a degenerate CSD where the entire population wants to concentrate in a single location. At the aggregate level there is no productivity growth, i.e., the single locations' productivities (and ultimately population sizes) evolve randomly without an aggregate trend.

From the perspective of a single city at some point in time t_0, this growth process (Gibrat's law) directly implies that its expected log population size T years ahead will follow a normal distribution. This is essentially a manifestation of the central limit theorem, as cities face random productivity shocks and their sizes thus also evolve randomly over time.[11] The overall CSD of the country in a given point in time aggregates the sizes of all cities that exist at that time. As long as all cities start from the same initial conditions and are subject to the same growth process for the same amount of time, which is the case in Eeckhout's (2004) model, this aggregation problem is easy: All cities have the same LN size probability distribution, which in turn is then also equivalent to the country's overall CSD. Things are more complicated, however, if cities are heterogenous.

Suppose cities are created at different points in time, so that there is age heterogeneity across cities. This is a highly realistic assumption, both for France and for other countries: Some cities are older than others. Furthermore, suppose there is aggregate productivity growth in the country, i.e., the distribution from which cities receive their *i.i.d.* shocks has a positive mean. Then, older cities are – in expectation – larger than younger cities, simply because they had longer time to grow. To obtain the overall CSD in that case, one needs to aggregate the city-specific size probability distributions according to the city age

[11] In another influential paper, Gabaix (1999) has shown that Zipf's law follows as the limiting distribution of an augmented version of Gibrat's law that includes a lower bound for city sizes; also see Gabaix and Ioannides (2004).

distribution.[12] Reed (2002) and Reed and Jorgensen (2005) have shown that the DPLN distribution as given in eq. (4.3) is the closed-form solution for the mixture of many LN distributions where the mixing parameter is exponentially distributed. In our context, this means that if age is exponentially distributed across cities, while all cities simply grow according to Gibrat's law (with positive drift) and thus have LN size probability distributions, this will asymptotically lead to DPLN distributed city sizes.

The framework by Eeckhout (2004) corresponds to the simple mixture case: there is a fixed number of cities without systematic differences in initial sizes or city ages. In that case, the country's overall CSD actually follows a LN distribution. One way to generate DPLN instead of LN distributed city sizes is to simply assume that cities differ by age, such that the age distribution is exponential. The aggregation of the city-specific size probability distributions would then do the job: Older cities have conditional CSDs with higher means, since they are around for a longer time, and with an exponentially distributed mixing parameter (city age) the country's overall CSD would become a DPLN.

In Giesen and Suedekum (2012a) we do not rely on such an exogenous age heterogeneity, but we consider an extension of the Eeckhout framework where an exponential age distribution across cities results endogenously. First of all, we allow for positive growth in the economy's overall population. For a given number of cities, this would imply decreasing welfare levels of time, ceteris paribus. Since negative size externalities prevail at the city level, having to fit more people into a given number of locations means that people would be worse off. We therefore consider a social planner who can create new cities, subject to a fixed resource cost per city (for housing, infrastructure, etc.). We show that the planner would create cities at a constant rate. More specifically, the optimal rate of city creation is equivalent to the population growth rate, which in turn smoothes welfare over time. With this time path for city creation, the city age profile endogenously converges to an exponential distribution. Since existing cities grow according to Gibrat's law, due to the random productivity shocks and perfect mobility of workers across cities, this in turn implies that city sizes asymptotically follow the DPLN distribution.

Constant growth in the number of cities is a natural outcome within our modelling framework, given that population grows at a constant rate as well. Still, it may be a delicate empirical issue because we typically do not observe persistent exponential growth in the number of cities within a country. However, recall that the crucial driver behind the exact functional form of the DPLN is the exponential city age distribution, which per se seems to be empirically much less implausible. That age distribution may also prevail if the number of cities does *not* grow at a constant rate over time, at least not persistently. In particular, suppose that city creation takes place only in an early phase

[12] Put differently, the *conditional* CSD, given the city's age, is a LN distribution, since size probability distributions are identical for all cities that have the same age. The *unconditional* CSD is a mixture of many LN distributions with parameters dependent on the cities ages.

of history where new settlements are developed. Say, in this early phase, the rate of city creation and the population growth rate are both constant. Then, at some point in time, say $\bar{t}$, population growth and city creation stop as the economy has now matured. At time $\bar{t}$, the city age profile is exponential and the oldest cities are, in expectation, the largest ones. Projected into the future, the city age distribution will remain a (shifted) exponential as cities get older in parallel. Also the differences in city sizes that exist in $\bar{t}$ will be projected into the future. In expectation, the largest cities in $\bar{t}$ will also be the largest one in $\bar{t}+1$, and so on. The overall CSD is thus still a mixture of heterogeneous city-specific size probability distributions, reflecting the size differences at $\bar{t}$, and will thus continue to follow a DPLN shape, though an increasingly fuzzy one given the variance of the idiosyncratic shocks to city productivity and size.

Summing up, in Giesen and Suedekum (2012a) we have extended Eeckhout's (2004) urban growth framework and considered several realistic features that were missing in the baseline model: aggregate productivity growth, aggregate population growth, and most importantly, age heterogeneity across cities. The overall CSD implied by our more general model – the DPLN – is much closer to the data (in France and in other countries) than the theoretically implied CSD of the baseline version, the LN. In our model, the crucial element of age heterogeneity arises endogenously from constant growth in the number of cites. However, there are also other ways of getting at an exponential city age distribution.

4.7 Conclusions and discussion

In this paper we have shown that the DPLN distribution provides an excellent fit to the French overall city size distribution, consistent with previous research for the urban systems in the US and other countries. Our research in this area can, in our view, potentially settle several controversies in the literature on urban growth and city size distributions.

There is still a lively debate how to parameterize overall CSDs, and especially about the relationship of this parameterization with the older literature on Zipf's law. If the "true" model of the CSD is a LN distribution, this would be bad news for the old Zipf literature. It would mean that researchers have made a systematic mistake for decades, by thinking that they have detected a power law for large cities, whereas in fact it was something else that only looks similar like a Zipfian power law. When the "true" model is the DPLN, there is no discrepancy between the old and the new literature on CSDs. The DPLN distribution actually features a power law in the upper tail, so previous research did not succumb to an illusion.

Other researchers have suggested alternative ways of bridging those literatures. In particular, Levy (2009), Ioannides and Skouras (2009) and Malevergne et al. (2011) have

all suggested that an appropriate parameterization for the overall CSD should involve some combination of LN in the body and Pareto in the (upper) tail of the distribution. None of these authors have developed a theory-based distribution, however, that can be rationalized by an underlying urban growth model. This is the particular benefit of the DPLN distribution. We can make explicit not only the stochastic foundations of the DPLN, but even provide a fully micro-founded economic model in which city sizes endogenously converge to this overall CSD. The distributional properties of the DPLN are similar in spirit to the ad-hoc functional forms advocated by the other authors, but even slightly more flexible than a rigid convex combination of LN and Pareto.

Another key advantage of the DPLN is that it delivers a very good fit for many different data sets. In Giesen and Suedekum (2012a) we show that the the LN parameterization may be well suited to match the US Census places data, but it fails miserably to match the overall CSD when using the recently developed "area clusters" data by Rozenfeld et al. (2008, 2011) where cities are economically and not administratively defined. The DPLN, however, fits the overall CSD for both definitions very closely. In this paper, we have focussed on the French case, and showed that France is no exception in this respect.

In fact, the data fit of the DPLN is actually much better for France than for the US administrative city units, the Census places. Having described the underlying model(s) of the LN and the DPLN, we can even hypothesize why this is the case. According to our theoretical framework, the country's overall CSD should have a more distinctive DPLN pattern the stronger is the age heterogeneity across cities within that country. If all cities were equally old, our model would predict that the CSD becomes again a LN. If some cities are much older than others within the country, however, there is a distinctive power law pattern in the upper tail and the cities located in the upper tail should – on average – also be much older than the cities in the bottom range of the size distribution.

Systematic empirical research on the age profile of cities within and across countries is still a largely neglected topic in urban economics, probably because reliable data on city creation dates are difficult to obtain. There are some marvellous recent attempts in this direction, e.g. the works by Bosker and Buringh (2010) and Bosker et al. (2012) that should be pushed further much more. Also there is little empirical work on the evolution of the number of cities in a country, particularly when small settlements ought to be included in the analysis. Notable exceptions include Henderson and Wang (2007) or Gonzáles-Val (2010). However, even if a full empirical analysis is beyond the scope of this paper, comparing France and the US in terms of the age heterogeneity of their cities is a relatively easy exercise. The oldest American city is probably Jamestown, VA, which was founded in 1607. The French urban system is much older, so in short, age heterogeneity across cities is much stronger in France than in the US. Consistently, we find that the DPLN outperforms the LN by a higher margin in France.

Chapter 5

Random urban growth and endogenous city formation

5.1 Abstract

This paper builds a dynamic general equilibrium model of random urban growth with endogenous city formation. The model features positive and negative local population externalities along with city specific productivity enhancing amenities. Cities are heterogeneous in their age and grow according to Gibrat's law. Migration is induced by intercity differences in wages, rental prices and commuting costs. In the spatial equilibrium, city sizes follow the *double Pareto lognormal distribution* and the empirical lognormal-shaped overall city size distribution and the well known upper-tail Zipfian power law are features of this model.

5.2 Introduction

Age heterogeneity across cities is a largely neglected element in the literature on urban growth. The central point of this paper is to highlight the role of city formation and therefore the meaning of age heterogeneity for explaining city size distributions.

The paper begins by building an economic model of random urban growth. Within the model, city sizes result from counteracting agglomeration and dispersion forces. Cities grow according to Gibrat's law, whilst a rising population distributes endogenously over an endogenously rising number of cities.[1] Each city has an idiosyncratic productivity parameter, which evolves randomly over time with a positive drift. As a consequence, older cities are on average more productive and therefore more attractive to citizens. City age heterogeneity is then accompanied by city size heterogeneity, and older cities host a larger share of the overall population. The spatial equilibrium of the model is characterized by the feature that city sizes follow the DPLN distribution.

The characteristic features of the model are Gibrat's law and age heterogeneity among cities. Many studies have confirmed that urban growth follows Gibrat's law, at least approximately, for example the study of Ioannides and Overman (2003) and Eeckhout (2004) for the US, Giesen and Suedekum (2010) for Germany and Eaton and Eckstein (1997) for France and Japan. On the other hand, there are some studies that provide evidence for deviations from Gibrat's law, for example the study of Rozenfeld et al. (2008). The existence of a stable city size distribution over time, however, seems to confirm that there is no stable pattern of growth with respect to size. Neither do small cities grow faster than larger cities, in which case all cities would have the same size, nor do large cities grow faster than small cities, which would cause the disappearance of small cities.

[1] Gibrat's law is fulfilled, if cities of all sizes have the same expected growth rate and the same expected variance of growth rates. In other words, Gibrat's law demands that there is no pattern of growth regarding to city size.

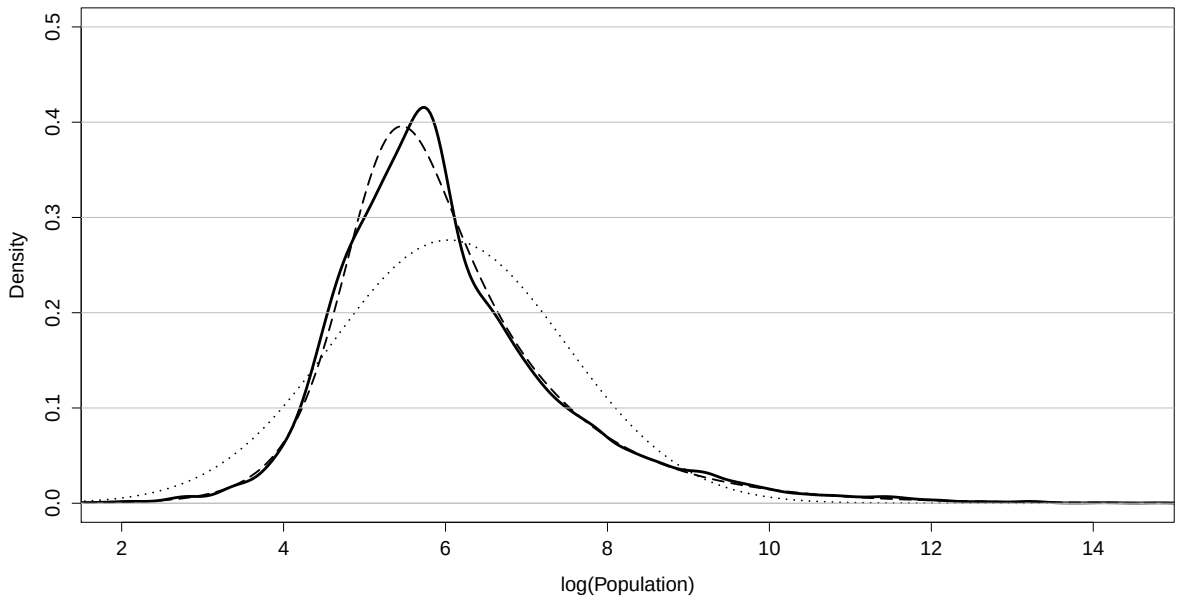

Figure 5.1: Kernel density estimate of the British city size distribution and best fit LN and DPLN.

The empirical data on city creation, and therefore on age heterogeneity, is very limited, but the importance is supported by several studies. Henderson and Wang (2007) report that in the time interval 1960-2000, the worldwide number of cities with population size larger than 100,000 increased from 1220 to 2684. Dobkins and Ioannides (2001) show that the number of urban territories in the US increased from 1 in 1760 to 555 in 1990 and Bairoch (1988) finds that the number of cities in Europe larger than 20,000 citizens increased from 39 in the year 1000 to 130 in the year 1760.

To convince the reader of the flexibility and the impressive fit of the DPLN, figure 5.1 and 5.2 show the French and British overall city size distribution. The French data is typical administrative data on city sizes, whereas the British data is on agglomerations, constructed by Rozenfeld et al. (2011). In both pictures, the solid black line represents a kernel density estimate for the overall city size distribution. In addition, the dotted line represents the best fit LN and the dashed line the best fit DPLN, each estimated by maximum-likelihood. In both cases, the DPLN has a better fit than the LN. For the French data, the LN has a descent fit and is able to replicate the data well, though worse than the DPLN. For the British CCA data, the superior fit of the DPLN is more obvious. The LN has severe deficiencies in describing the data, whereas the DPLN has a very nice fit[2]. A detailed study on the superior fit of the DPLN compared to the LN is provided

[2]For both countries, the better fit of the DPLN, compared to the LN is based on several model selection criteria, as the Akaike and Bayesian information criterion, the log-likelihood ratio test, Goodness-of-fit statistics and qq-Plots, which are not shown and explained here for brevity but are available upon request from the author.

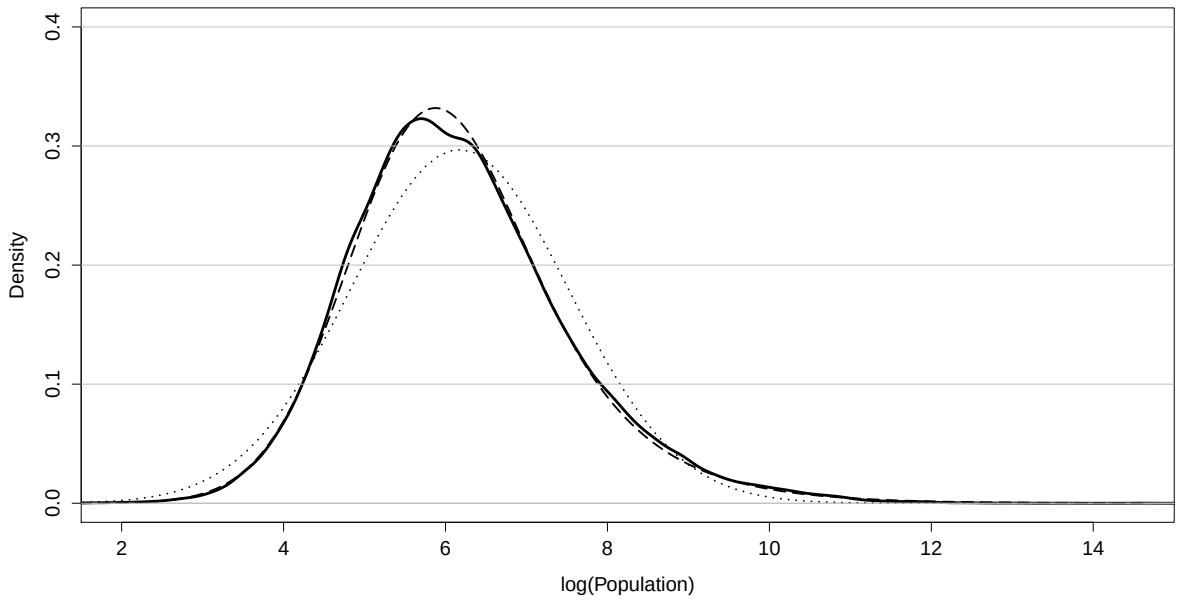

Figure 5.2: Kernel density estimate of the French city size distribution and best fit LN and DPLN.

by Giesen, Zimmermann and Suedekum (2010). There we discriminate between the LN and the DPLN distribution by analyzing un-truncated settlement size data from eight countries using various model selection criteria. We there show that the LN fits the data very well, but the DPLN to be the preferred model even after having been penalized for having more parameters than the LN.

In Urban Economics and Regional Science there is a long tradition to propose and estimate city size distributions in order to capture and document regularities across countries. There are three main unsolved issues within that literature. The first issue is on the correct parameterization of the underlying distribution. One the one side, there is a mature literature, arguing that the large cities in almost all countries follow a Pareto distribution, or even adhere exactly to the well known rank-size rule known as Zipf's law, for which the Pareto slope parameter needs to be unity. Various studies, including the international study by Soo (2005) or the meta-analysis by Nitsch (2005), show the Pareto distribution to be a good description for almost all countries and the slope parameter to be close to unity. On the other side, recent developments focus on the overall distribution instead of only the upper tail. Eeckhout (2004) analyzes the overall city size distribution of the US, using Census 2000 data, including even small villages of less than 100 inhabitants. His finding is that the overall distribution is lognormal (LN). This stands in stark contrast to the literature that focuses only on the large cities, as an overall LN city size distribution is not compatible with a Pareto distribution and Zipf's law. Eeckhouts argument is that earlier studies confused the LN with the Pareto, as the

latter is virtually indistinguishable from the upper tail of the LN. This finding raised an interesting discussion, on whether the right tail of the distribution is Pareto or LN, see Eeckhout (2004, 2009), Levy (2009) and Malevergne et al. (2011).

The second issue is on the correct definition of a city. It has long been argued that administratively defined cities are inadequate for analyzing city size distributions. The drawback of this data is that cities are defined by legal boundaries and are thereby not able to capture genuine agglomerations, formed by economic forces. The remedy to this data problem is *constructed* data on metropolitan areas, which are unfortunately only available for a small fraction of a countries distribution, if at all. Hence, Rozenfeld et al. (2011) construct data sets for the overall distribution, where each city is defined by an algorithm that identifies agglomerations.[3] Surprisingly, their findings in analyzing the US and GB cluster size distribution contradicts Eeckhout's claim. They show that the LN fails and that a very large range is best described by Zipf's law.

The third issue is on the underlying forces, responsible for shaping similar city size distributions across countries. Gabaix (1999) argues that a modified version of Gibrat's law is responsible for the Zipfian power law. The resulting city size distribution of this modified version of Gibrat's law however, cannot account for the lognormal body. Eeckhout (2004) in contrast argues that Gibrat's law in its simplest form, with a homogeneous and fixed set of cities shapes the lognormal body. This resulting city size distribution of this simplistic version of Gibrat's law however, cannot account for the upper tail Pareto distribution.

The DPLN is able to clarify all three problems. As shown below, the DPLN is able to resolve the first issue, as it nests both, the lognormal as well as the Pareto. In Giesen and Suedekum (2012a) we show that the DPLN also solves the second problem, as it is able to describe both kind of data much better than the lognormal or Pareto distribution. The model presented in this paper explains the third issue, as it shows that the DPLN is the natural outcome of an economy, where cities grow according to Gibrat's law, and where cities are of different ages.

Parallel to the empirical findings, various studies aim at explaining urban growth, the existence of cities and their sizes. Those models can be divided into two categories. The first category is concerned with deterministic models with predictable outcomes. The deterministic models aim especially at identifying and explaining the determinants of urban growth and the reasons for the existence of cities. This literature, beginning with Henderson (1974) mostly relies on the concept of an optimal or efficient city size, as utility or welfare within a city is modeled to be a hump-shaped function in population size. Some well known papers are Eaton and Eckstein (1997), Black and Henderson (1999)

[3]Their approach is to use a city clustering algorithm (CCA) on high resolution data. This algorithm expands a city to its boundaries and stops to expand the city, if there is no more dense enough populated space to expand on.

or Henderson and Venables (2009). An important challenge of urban growth models is to be in accordance with the empirical findings from the city size distribution literature. As is highlighted by Duranton (2011), the classical deterministic urban growth models lack on this important feature, as they fail to generate distributions close to actual city size distributions.

The second category is concerned with random models. The common feature of those models is to incorporate a stochastic process for some kind of city or industry specific variable. In contrast to the deterministic models, the random models are able to replicate features of empirical city size distributions, like Gabaix (1999). He presents one of the most important models among the random urban growth models, which is the widely accepted explanation for Zipf's law. In his setup cities face temporary amenity shocks and cities grow according to Gibrat's law. He shows that a small modification of Gibrat's law, by introducing a lower bound on city sizes, leads to Zipf's law. Another important contribution is by Duranton (2007), who proposes a model in which random innovations cause industry reallocations. This induces shocks to city population and the model is able to replicate a convex upper tail city size distribution that is shown to be even closer to the data than Zipf's law. In Rossi-Hansberg and Wright (2007) the key mechanism is the formation of new cities. There, migration is induced by industry specific shocks and the model is able to replicate Gibrat's law as well as Zipf's law. In contrast to those contributions, which aim explicitly at explaining the size distribution of large cities, the model of Eeckhout (2004) aims at explaining the overall city size distribution. He builds a model of local externalities, where migration occurs due to city specific productivity shocks. In the setup, cities grow according to a simple version of Gibrat's law with equally old cities and result in a lognormal distribution, which is shown to have a decent fit to empirical city data. Unfortunately, the Eeckhout (2004) model relies on three restrictive assumptions. First, there is no population growth. Second, there is no creation of cities. Third, there is no aggregate growth in productivity or technological progress. Therefore, the resulting lognormal distribution lacks in flexibility, fails to match the constructed city size data by Rozenfeld et al. (2011) and cannot account for the Pareto distribution in the upper tail. The model presented here is a generalization of the Eeckhout-model and abolishes the three restrictive assumptions. As a consequence, the resulting city size distribution is not LN but DPLN.

The objective to random urban growth models is that the forces of urban growth are not further analyzed but simply assumed to be stochastic. This paper argues this to be a legitimate way; the attractiveness of a city, and the resulting population, is determined by such a huge amount of factors that no model can possibly account for all of them. The task of economic models should be to identify and capture the essential forces, while the remainder, which is not modeled or unknown needs to be counted as unpredictable. But, despite the unpredictability of the remaining forces, they exist and

lead to intercity differences. This motivates the random urban growth models to summarize this unpredictability in terms of a random process.[4] It is important to highlight that random models are not theory free. In fact, they still rely on microeconomic foundations, on agglomeration and dispersion forces. The model presented in this paper, for example, accounts for agglomeration forces, such as an urban wage premium and dispersion forces as congestion costs and higher housing prices. Other unpredictable factors, may they be economic, geographic, political or demographic are not modeled but captured in terms of a Brownian motion.

The structure of this paper is as follows. Section 5.3 builds a model of city creation and growth, in which city sizes follow the DPLN. Section 5.4 analyzes the features of the model and the resulting city size distribution while Section 5.5 concludes.

5.3 A model of city growth with local externalities and city creation

This section builds an economic model of urban growth with endogenous city creation. The approach is to build on and modify a dynamic and continuous time version of the framework by Eeckhout (2004). The Eeckhout model replicates a random urban growth process, where a fixed set of equally old cities follow Gibrat's law and results in a log-normal city size distribution. The model below aims at incorporating population growth, technological progress and endogenous city creation, three features, not captured by the Eeckhout model. By incorporating those straightforward features, the resulting city size distribution will be the DPLN instead of the LN.

5.3.1 The model

Within the model, a country is populated by S_t inhabitants, living in N_t cities. Population growth is $g_s > 0$ and $S_t = S_0 e^{g_s t}$. Cities grow at the, yet unspecified, endogenous rate λ_t and $N_t = N_0 e^{\int_{t=0}^{t} \lambda_t}$, where S_0 and N_0 are the starting sizes. Each city i has an idiosyncratic productivity parameter $A_{i,t}$.[5] This city specific productivity parameter evolves over time via a geometric Brownian motion with a positive drift g_A and variability ς_A according to the following equation:

$$dA_{i,t}/A_{i,t} = \epsilon^A_{i,t} \qquad \text{where} \qquad \epsilon^A_{i,t} = g_A \cdot dt + \varsigma_A \cdot \mathrm{d}B_{i,t}. \tag{5.1}$$

[4] This procedure is well accepted in financial economics, in the modeling of stock market prices. Stock market prices are influenced by such a huge amount of factors that the literature models them via Brownian motions.

[5] With the assumption of a city specific productivity parameter, the Spatial Impossibility Theorem of Starrett (1978) does not hold and a spatial equilibrium with population concentration is possible.

The Brownian motion delivers the property that a city might end up with a very high or a very low productivity parameter with a small probability, whereas the typical city will encounter growth according to the drift. The positive drift is to capture technological progress of the existing cities. A cities productivity parameter, $A_{i,t}$ therefore reflects the evolution of a cities technology. With a positive drift, older cities have higher productivity parameters on average, and as will be shown below, this translates into a higher population. The size of a city is affected by two opponent forces, both depending, on the respective city's population size. On the one hand, there is an agglomeration force: workers and firms profit from the spillovers of locating close to each other. This idea is incorporated by the factor $a_+(S_{i,t}) = S_{i,t}^{\theta}$, a local positive externality, specific to a city of size $S_{i,t}$. This factor multiplicatively enhances the productivity of each worker in this city, as the marginal product of labor is determined by:

$$y_{i,t} = A_{i,t} S_{i,t}^{\theta} = w_{i,t}. \tag{5.2}$$

Due to competitive labor markets, the marginal product of labor equals wages. In accordance with the urban wage premium literature, larger cities pay higher wages. On the other hand, there is a dispersion force: the bigger the city, the higher the degree of congestion problems. The idea of this local negative externality is incorporated by the factor $a_-(S_{i,t}) = S^{-\gamma} \in]0,1[$, which multiplicatively reduces the effective working time and can be understood as time lost in commuting. Workers have one unit of working time and they have to choose how much they spend on working $l_{i,t}$ and how much on leisure $(1 - l_{i,t})$. With the negative externality, the effective time devoted to labor is given by:

$$L_{i,t} = a_-(S_{i,t}) l_{i,t}. \tag{5.3}$$

Quite naturally, the larger the city, the more effort has to be devoted to commuting. The preference structure for residents of a specific city is such that they receive utility out of a numéraire consumption good $c_{i,t}$, leisure time $(1 - l_{i,t})$ and out of land property $h_{i,t}$ from a Cobb-Douglas utility function. Land is rented at the price $p_{i,t}$ from an absentee landlord, with H denoting the total land size, constant over time and the same to each city. Residents therefore face the following utility maximization problem:

$$u(c_{i,t}, h_{i,t}, l_{i,t}) = c_{i,t}^{\alpha} h_{i,t}^{\beta} (1 - l_{i,t})^{1-\alpha-\beta} \quad s.t. \quad c_{i,t} + p_{i,t} h_{i,t} \leq w_{i,t} L_{i,t}. \tag{5.4}$$

The solution for the prices and allocations are shown by Eeckhout (2004) to be:

$$p^*_{i,t} = \frac{\beta A_{i,t} \cdot S_{i,t}^{\theta-\gamma+1}}{H} \tag{5.5}$$

$$w^*_{i,t} = A_{i,t} \cdot S_{i,t}^{\theta} \tag{5.6}$$

$$c^*_{i,t} = \alpha \cdot A_{i,t} \cdot S_{i,t}^{\theta-\gamma} \tag{5.7}$$

$$h^*_{i,t} = \frac{H}{S_{i,t}} \tag{5.8}$$

$$l^*_{i,t} = \alpha + \beta. \tag{5.9}$$

Workers are perfectly mobile and choose their location (potentially every time period again) without frictions. In each period of time, citizens observe the vector of technological shocks, and therefore the vector of productivity parameters. This information, along with the vector of population sizes enables them to calculate indirect utility of a citizen in city i according to:

$$u_{i,t} = \Phi \left(A_{i,t} S_{i,t}^{-\Theta} \right)^{\alpha} \tag{5.10}$$

$$\begin{aligned} \text{with} \quad & \Theta = \gamma + \frac{\beta}{\alpha} - \theta \\ \text{and} \quad & \Phi = \alpha^{\alpha} H^{\beta} (1-\alpha-\beta)^{(1-\alpha-\beta)}. \end{aligned} \tag{5.11}$$

By assumption, $\Theta = \gamma + \beta/\alpha - \theta > 0$ and indirect utility within cities is decreasing in their size. [6] The city specific productivity parameters and theirs shocks directly affects wages, rents and working time and thereby induce migration. Since workers base their location decision in an inter-city utility comparison and are perfectly mobile, utility has to be equalized across cities, so that there are no incentives to move, and therefore $u_{i,t} = u_t$ and

$$A_{i,t}^{-(1/\Theta)} S_{i,t} = A_{j,t}^{-(1/\Theta)} S_{j,t} = k_t \quad \forall j \neq i \qquad \text{where} \quad k_t = \left(\frac{\Phi}{u_t} \right)^{1/(\alpha\Theta)}. \tag{5.12}$$

This spans a system of N_t equations with the unknown population sizes $S_{i,t}$ under the constraint $S_t = \int_{i=0}^{N_t} S_{i,t}$. The continuous time solution, is the vector of city sizes, with the magnitude N_t, where each vector component is given by

$$S_{i,t} = \frac{A_{i,t}^{\frac{1}{\Theta}}}{\int_{j=0}^{N_t} A_{j,t}^{1/\Theta} \mathrm{d}j} \cdot S_t. \tag{5.13}$$

[6] The traditional literature, like Henderson (1974), works with the concept of a utility function which is hump-shaped in population size. The representation of this paper allows only for a monotonic course and $\Theta < 0$ is excluded by assumption. In this case utility would be increasing in the population size, leading to only one city, with the whole country population living in that city. The assumption of $\Theta < 0$ is the analogon to a city that is on the right, decreasing part of the hump-shaped utility curve from the traditional literature.

The interpretation of this equation is straightforward, as it shows the size of a city to be a fraction of the overall population, where this fraction depends on a cities productivity parameter relative to the mass of other cities productivities. Young cities, with low $A_{i,t}$ have on average a smaller population and therefore pay lower wages, but have less congestion and lower housing prices.[7] The respective population growth rate of a city i (potentially negative) is then determined by its productivity draw, relative to the average, as well as overall population growth g_s

$$g_{i,s} = \frac{\left(1+\epsilon_{i,t}^{A}\right)^{\frac{1}{\Theta}}\left(1+g_S\right)}{\left(1+g_A\right)^{\frac{1}{\Theta}}} - 1. \tag{5.14}$$

The resulting indirect utility level in the spatial equilibrium, equal to all citizens irrespective of their location can be calculated by using the normed productivity parameter as

$$V_t^* = \Phi\left(A_t \cdot S_t^{-\Theta}\right)^{\alpha}, \qquad \text{where} \quad A_t = \left(\int_{i=0}^{N_t} A_{i,t}^{1/\Theta}\mathrm{d}i\right)^{\Theta}. \tag{5.15}$$

The national welfare function is then easily derived by the product of per capita indirect utility times population size $S_t V_t^*$ or $e^{g_s t} V_t^*$, respectively. Over time, indirect utility, as given by equation (5.15) is subject to two opponent forces. First there is technological progress, which has a positive effect on utility, since it increases the marginal product of labor. Second there is population growth, which has a negative effect on utility since it brings higher rents and more expensive commuting. This exogenous movement in indirect utility can be modeled via

$$V_{t+dt}^* = \Phi\left(\frac{(1+g_A)^{1/\Theta}}{(1+g_S)} \cdot \Omega_t\right)^{\alpha\Theta} \qquad \text{where} \quad \Omega_t = \frac{\int_{i=0}^{N_t} A_{i,t}^{1/\Theta}\mathrm{d}i}{S_t}. \tag{5.16}$$

The strength of the two opposing effects determines whether indirect utility rises or falls over time. Under the inequality $(1+g_A)^{1/\Theta} < (1+g_S)$ population growth is strong enough to outperform the positive effect of technological progress, in the absence of city creation.

5.3.2 City creation

In the model economy, cities are created by a social planner, which aims at maximizing national welfare. The creation of a city involves fixed costs for setting up technical and social infrastructure, which are paid by the existing population. Along with its creation, a new born city draws an initial productivity parameter out of a LN distribution with mean A_0. This initial productivity draw, along with zero population attracts new residents,

[7] There is no optimum city size in this setup, as in classical urban growth models like Henderson (1974), Black and Henderson (1999) or Henderson and Venables (2009). Due to city specific technology cities naturally differ in size.

which relocate from the other cities into this new city, until overall utility is equalized again. The creation of a city is therefore a public good and the impact on utility of its new residents is the same as to the rest of the population. We consider a pure birth process, so that cities can be become infinitesimally small, but cannot formally exit the process. The expected initial size of a new born city becomes

$$E[S_0] = \left(\frac{A_0}{A_t}\right)^{1/\Theta} S_t, \qquad \text{where} \quad A_t = \left(\int_{i=0}^{N_{t-dt}} A_{i,t}^{1/\Theta} \mathrm{d}i + x_t A_0^{1/\Theta}\right)^{\Theta}. \tag{5.17}$$

Analogue to equation (5.13), initial size of a new born city is related to overall productivity, which includes the productivity of new born cities, and population size. After their creation, along with the initial productivity draw, their subsequent technological progress follows the same Brownian motion as existing cities do. The social planners task is to choose the optimal number of cities to be created in each period.[8] The optimal sequence of cities is characterized by the property that it maximizes the stream of each periods national welfare, discounted with the rate ρ. Building a new city creates a trade-off between the new living environment, which has a positive effect on national wide indirect utility, and the incurrence of fixed costs of F. The objective function is the national welfare function, which needs to be reduced by the costs of creating x_t cities at time t at the price of F, measured in indirect utility χ, which is assumed to be constant and equal to unity. The planners infinite horizon problem becomes:

$$max \int_0^{\infty} e^{(g_s-\rho)t} \left(\Phi \Omega_t^{\Theta\alpha} - x_t \frac{F}{S_t}\right) \qquad s.t. \;\; \dot{\Omega} = g_\Omega \Omega_t + x_t \frac{A_0}{S_t} \tag{5.18}$$

subject to the initial condition Ω_0, the feasible set of the control variable $x_t \geq 0$ and the transition equation $\dot{\Omega}$. The transition equation describes the movement of Ω over time and captures the influence of city creation but also by exogenous technological progress and population growth, reflected by $g_\Omega = \left(\frac{(1+g_a)^{1/\Theta}}{(1+g_s)} - 1\right)$. The respective present value Hamiltonian becomes

$$H_t = e^{(g_s-\rho)t} \left(\Phi \Omega_t^{\Theta\alpha} - x_t \frac{F}{S_t}\right) + \mu \left(g_{\Omega_t} \Omega_t + x_t \frac{A_0}{S_t}\right) \tag{5.19}$$

where μ_t, the so called co-state variable, stands for the shadow value of an additional city, created in t. Using Pontryagy's principle of optimality, the static efficiency condition demands the Hamiltonian to be maximized in each period so that:

$$\frac{\delta H_t}{x_t} = 0 \;\; \rightarrow e^{(g_s-\rho)t} \mu_t A_0^{1/\Theta} = F. \tag{5.20}$$

[8] Note that there is no need to influence the distribution of citizens across cities, as this allocation is efficient, due to perfect mobility.

The first optimality condition therefore requires that at any point in time, the gain from a new city must equal its costs. Notice that due to the linearity in the choice variable x_t, this is no argument in this first order condition. The dynamic efficiency condition demands the path of the co-state variable to be

$$\frac{\delta H_t}{\delta \Omega_t} = -\dot{\lambda} \quad \rightarrow \dot{\lambda}_t = -g_\Omega \lambda_t - e^{(g_s-\rho)t} \alpha \Theta \Phi \Omega_t^{\alpha\Theta-1}. \tag{5.21}$$

The economic interpretation of this equation is that of a no-arbitrage condition and demands that in the optimum, it must not be possible to increase utility by delaying or accelerating city creation over time. The third condition to be fulfilled is the transition equation to hold

$$\frac{\delta H_t}{\mu_t} = \dot{\Omega} \quad \rightarrow \dot{\Omega} = g_\Omega \cdot \Omega_t + x_t \frac{A_0}{S_t}. \tag{5.22}$$

Furthermore, the constraints must be satisfied and binding, and the transversality condition of complementary slackness needs to be fulfilled:

$$\lim_{t\to\infty} \mu_t \geq 0 \qquad \text{and} \qquad \lim_{t\to\infty} \mu_t \cdot \Omega_t = 0. \tag{5.23}$$

In addition, the discount rate must be high enough for the integral to be bounded from above which translates into $\rho > g_S > 0$. From the optimality conditions, given by the set of differential equations, (5.20) - (5.22), we can solve for the optimal city creation rate. From equation (5.20), the linearity in the control variable indicates a boundary solution. In fact, without population growth, the solution of this problem would be a typical bang-bang solution, in the sense that as long as the shadow price of another city exceeds its costs, it would be optimal to create cities at the maximum rate, until the shadow price and costs are equalized. With population growth, however, the optimality condition in equation (5.20) is itself dynamic, so that we have a singular solution.

The time derivate of equation (5.20) postulates the optimal movement of the costate variable to be $\dot{\mu} = (g_s - \rho)e^{(g_s-\rho)t} A_0^{-1/\Theta} F$. Equating this with (5.21) and using the property that within the singular solution $\mu_t = e^{t(g_s-\rho)} A_0^{-1/\Theta} F$, the optimum state variable is given by

$$\Omega^* = \left(\frac{\alpha\Theta\Phi \cdot A_0^{1/\Theta}}{\chi F} \cdot \frac{1+g_S}{(1+\rho-g_S)(1+g_S) - (1+g_A)^{1/\Theta}} \right)^{\frac{1}{1-\alpha\Theta}}, \tag{5.24}$$

which is time constant and therefore $\dot{\Omega} = 0$. This indicates that there is an optimal utility level, which needs to be maintained and that city creation is necessary to outweigh the negative impact of population growth, net of the positive effect from technological progress. Using the property that $\dot{\Omega} = 0$ along with the transition equation, the singular

solution for the optimal control becomes

$$x_t^* = e^{g_S \cdot t} \cdot \left[\left(1 - \frac{(1+g_A)^{1/\Theta}}{1+g_S} \right) \cdot \frac{S_0}{A_0^{1/\Theta}} \right] \cdot \Omega^*. \tag{5.25}$$

The condition $x_t \geq 0$ requires that $(1+g_A)^{1/\Theta} < (1+g_S)$. The solution of the above stated social planners optimal control problem is therefore the triplet Ω_t^*, x_t^* and μ_t^* with an optimal time path, characterized by the conditions $\hat{\Omega} = 0$, $\hat{x} = g_s$ and $\hat{\mu} = g_s - \rho$.

5.3.3 Characterization of the model

We begin by focusing on city growth at the intensive margin. In the above established model, cities experience random fluctuations in their population. We notice that a cities growth rate, shown in equation (5.14), does not depend on a cities size but solely on the city specific random shock, countrywide population growth and technology progress, where the latter two are the same to each city. Since the growth rate of a city is determined by the city specific random shock, which is provided for each city and in each period from the same geometric Brownian motion, cities of all sizes grow proportional and according to the same process. This leads us to pose the following proposition on urban growth:

Proposition 1 *Urban growth is characterized by Gibrat's law.*

We continue to analyze city growth at the extensive margin. The number of cities N_t is increasing in each period with the increment x_t. One result of the above model is that the intermittent increment x_t is increasing over time along the population growth rate g_S, so that $\hat{x} = g_S$. The growth rate of cities λ_t, then translates into $\lambda_t = \hat{N} = \dot{N}_t/N_t = x_t/N_t$, which becomes $\lambda_t = \frac{e^{g_S t}}{e^{g_S t}-1} \cdot g_S$ and converges to the time constant $\lambda_t = g_S = \lambda$. This information brings us to pose the second proposition:

Proposition 2 *The age distribution of cities follows an exponential distribution.*
Proof. With a time constant city creation rate λ the number of cities, existing at t is $N_0 e^{\lambda t}$. Using this information, the cumulative distribution of the year of birth τ, is calculated by using the probability concept of Laplace to be $P(T \leq \tau) = \frac{N_\tau}{N_t} = e^{\lambda\tau - \lambda t}$. This distribution is dynamic and therefore depends on the time of observation t. The age κ of a city is then defined as $\kappa = t - \tau$. Using $K = t - T$, we see that the age distribution is characterized by $P(K > \kappa) = e^{-\lambda\kappa}$ and therefore by $P(K \leq \kappa) = 1 - e^{-\lambda(\kappa)}$. This is the cumulative distribution of an exponentially distributed random variabel with density $f(\kappa) = \lambda e^{-\lambda\kappa}$.■

After being created, a city lacks an initial productivity parameter. As stated above, the initial productivity parameter is assigned at birth, by drawing A_0 out of a lognormal distribution with parameters μ_{A_0} and σ_{A_0}. This initial productivity draw, along with zero

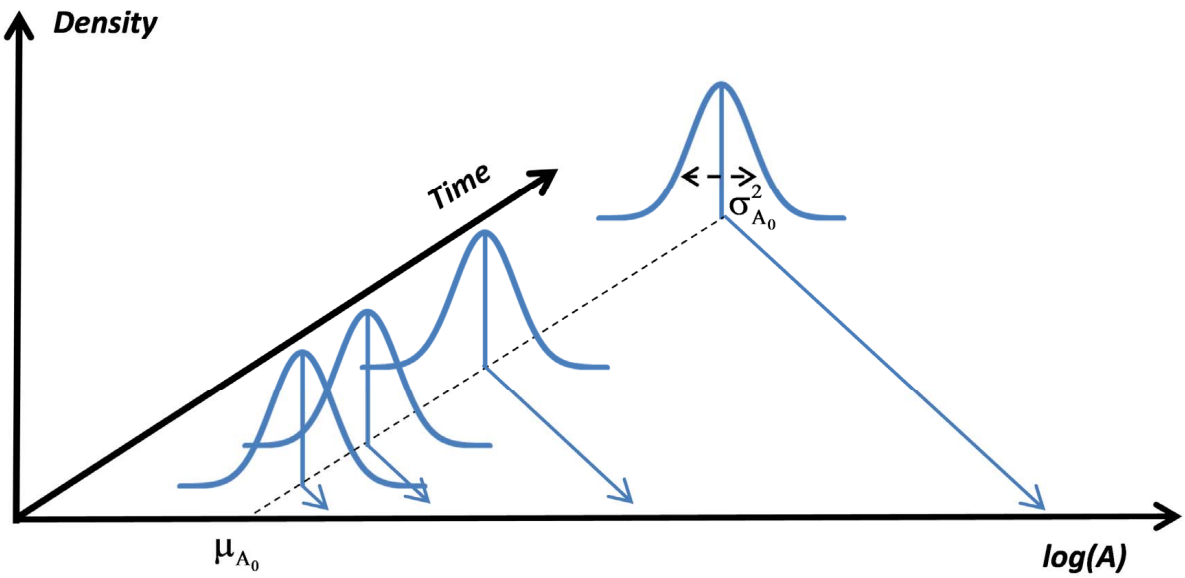

Figure 5.3: The mechanics of the random urban growth model with endogenous city formation.

population attracts residents from other cities and thereby determines initial population size.

Proposition 3 *Initial sizes of new born cities are $LN(\mu_{S_0}, \sigma_{S_0})$ distributed.*
Proof. In the spatial equilibrium with city formation according to equation (5.25), the utility of an individual is time invariant, so that $k_t = \bar{k}$. From equation (5.12) the size of a new born city is related to its initial productivity draw according to the linear transformation $log(S_0) = log(\bar{k}) + (1/\Theta) \cdot log(A_0)$. Using the transformation rules of the Normal distribution, it is clear that if $A_0 \sim LN(\mu_{A_0}, \sigma_{A_0})$ then $S_0 \sim LN(\mu_{S_0}, \sigma_{S_0})$ with $\mu_{S_0} = (1/\Theta)\mu_{A_0} + log(\bar{k})$ and $\sigma_0 = (1/\Theta)^2 \sigma_{A_0}$. ■

We can conclude that a city with initial productivity parameter A_0, and age κ has been following the Brownian motion for technological progress, given by equation (5.1) for κ periods. It will therefore have a productivity parameter of

$$logA_{i,\kappa} = logA_0 + \int_{t=0}^{\kappa} \epsilon^A_{i,t} dt. \tag{5.26}$$

We know from the central limit theorem and standard Itô calculus that this specific city will have an age dependent productivity parameter distribution, which follows a lognormal distribution according to

$$logA_{i,\kappa} \sim N\left(\mu_{A_0} + \mu_\kappa, \sigma^2_{A_0} + \sigma_\kappa\right) \tag{5.27}$$

where μ_{A_0} and σ_{A_0} determine the expectation and variability of the initial A_0-draw, whereas

$$\mu_\kappa = \left(g_A - \frac{\varsigma_A^2}{2}\right) \cdot \kappa \qquad \text{and} \qquad \sigma_\kappa = \varsigma_A^2 \cdot \kappa \tag{5.28}$$

show the evolution over time. We assume that $g_a > \frac{\varsigma^2}{2}$, so that older cities have higher productivity parameters on average. The intuition is that older cities have a longer time to develop, or in technical terms to follow the Brownian motion and to profit from the positive drift. In addition, we can already conclude, along with equation (5.13) that due to the higher productivity parameters the older cities will host a higher share of the overall population. With the information on the productivity parameter distribution of a single city at hand, it is straightforward to derive the countrywide overall distribution of productivity parameters. This is a composition of many cities of different ages and therefore different productivity parameters. Technically speaking, the overall distribution is a mixture distribution. The mixture components are the different lognormal distributions, which differ in their parameters, which in turn depend on how long the specific city has been following that process; city age as shown in equation (5.27). City age is assigned via the mixture weight, which is shown by Proposition 3 to be the exponential distribution. The density of the resulting mixture distribution is therefore the Riemann-Stieltjes integral

$$f(A) = \int LN\left(A; \mu_\kappa, \sigma_\kappa^2\right) d\,Exp(\kappa; \lambda),$$

for which Reed (2002) derives a closed form solution according to

$$\begin{aligned} f(A) = \frac{ab}{a+b}\Bigg[& A^{b-1} e^{\left(b\mu_{A_0} + \frac{b^2\sigma_{A_0}^2}{2}\right)} \Phi^c\left(\frac{log(A) - \mu_{A_0} + b\sigma_{A_0}^2}{\sigma_{A_0}}\right) \\ & + A^{-a-1} e^{\left(a\mu_{A_0} + \frac{a^2\sigma_{A_0}^2}{2}\right)} \Phi\left(\frac{log(A) - \mu_{A_0} - a\sigma_{A_0}^2}{\sigma_{A_0}}\right)\Bigg]. \end{aligned} \tag{5.29}$$

This density is called the DPLN.[9] In this formulation, Φ denotes the standard normal cdf and Φ^c the respective ccdf. The parameters $a(g_A, \varsigma_A, \lambda)$ and $b(g_A, \varsigma_A, \lambda)$ capture the information on the time specific parameters and will be analyzed below. The above enables us to make the fourth proposition:

Proposition 4 *Productivity is distributed according to the* $DPLN(a, b, \mu_{A_0}, \sigma_{A_0})$.

Figure 5.3.3 illustrates and summarizes the mechanics of Proposition 4. Cities are

[9] Reed (2002) shows that the DPLN is also compatible with a scenario in which the city creation mechanism is stopped in any point in time $\bar{t}$. In this case the mixture weight is a shifted exponential distribution and city sizes still are DPLN distributed.

born at different points in time and when born, they draw an initial productivity parameter out of a lognormal distribution. The distribution of initial sizes is time invariant, reflected by the dashed line at μ_{A_0} and the constant variance $\sigma^2_{A_0}$. Over time, the cities productivity parameters grow or shrink, however with a positive drift, indicated by the arrows. Therefore, the older the city, the higher the expected productivity parameter and the higher the variance of this parameter. At any point in time, the overall distribution is the mixture of all cities, each providing a productivity parameter, each drawn from a time specific lognormal probability distribution, depending on the cities age. Age is assigned via the exponential distribution, so that the time interval between city creation is decreasing and the number of new born cities is higher than the number of old cities.

More important than the distribution of productivity parameters, in this context is the distribution of population. In any time period, the existing population distributes over the set of cities N_t, according to equation (5.13). Reed and Jorgensen (2004) show that the family of DPLN distributions has the property of closure under power-law transformation. More specifically, using their equation (27) along with the above equation (5.12) we can determine the parameters of the DPLN for the population distribution and provide Proposition 5:

Proposition 5 *If city specific productivity is $DPLN(a, b, \mu_{A_0}, \sigma_{A_0})$ distributed then population is $DPLN(\alpha, \beta, \mu_{S_0}, \sigma_{S_0})$ distributed with $\alpha = a\Theta$, $\beta = b\Theta$, $\mu_{S_0} = (1/\Theta)\mu_{A_0} + log(\bar{k})$ and $\sigma_0 = (1/\Theta)^2\sigma_{A_0}$.*

Compare those findings to Eeckhout (2004). In Eeckhout, the productivity parameters do not have a positive drift, instead their shock is drawn in each period from a distribution with zero mean. Older cities therefore do not have a productivity advantage and there is no difference in the expected productivity parameters across different ages. The overall productivity parameter distribution is therefore a mixture of many identical lognormal distribution but with the same parameters. A mixture of equal sub-distributions takes the form of the sub-distribution itself. In Eeckhout as well, population is a closed power transformation of productivity and population is therefore lognormal distributed.

5.4 Features of the model and the DPLN

5.4.1 General features of the DPLN

The DPLN is a four parameter distribution. Besides its location parameter μ_{S_0} and the scale parameter σ_{S_0}, the DPLN is characterized by α and β, the slope parameters of the

Pareto tails[10].

The mechanism behind the shape of the DPLN density is the following: For the focus on large cities (S is large), the right summand of the distribution in parentheses vanishes, as the value of Φ^c goes to zero leaving $f(S) \sim S^{-\alpha-1}$. Large cities are therefore described by a Pareto distribution, capable of being in accord to the existing Zipf's law literature. By focusing on small cities, (S is small), the left term of the summand in parentheses vanishes, as the value of Φ goes to zero leaving $f(S) \sim S^{\beta-1}$, a Pareto distribution to depict the distribution of small cities[11]. For intermediate values of S, the right and left part of the term in parentheses interact to display the body of a lognormal distribution.

The Pareto distributions in the upper and lower tail are main features of the DPLN and give the DPLN several advantages over the LN or other *rivalry* distributions. First, with Pareto distributions, the DPLN has the possibility to produce the typical straight line behavior in log-log plots, which has been found by the Zipf's law literature in the upper tail. The second, related advantage are the respective shape parameters α and β. With those, the DPLN can adapt each tail separately to the form of the empirical distribution. It is therefore better suited to describe the behavior of very large and very small cities than the LN, as has been critizised most notably by Levy (2009). Contrary, the LN distribution is able to appear like a Pareto distribution but cannot produce the straight line behavior exactly, see Mitzenmacher (2004). In addition, the tails of the LN are just the extensions of its body and cannot be adapted separately. Third, comparing both distributions in logarithmic scales, the DPLN can be skewed and have kurtosis. It therefore does not need to be symmetric and can be more peaked (leptokurtic) or more flat (platykurtic) than a normal distribution in log scales.

A further property of the DPLN is, as analyzed below, that it nests the LN in the limit of $\{\alpha, \beta\} \to \infty$. This feature, along with the feature of its tails, allow the DPLN to be LN, as well as Pareto. This unique characteristic, infers in the recent debate by Eeckhout (2004, 2009) and Levy (2009) led in the *American Economic Review*, on whether the lognormal or Pareto distribution is appropriate for describing city sizes. Under the DPLN, both claims are correct. The lognormal is best for describing the body of the distribution, while the upper tail is a Pareto distribution, as has been formally confirmed by Malevergne et al. (2011). With the upper tail Pareto distribution, the DPLN does not call the existing literature on Zipf's law as mistaken, but only as "incomplete" in the sense that the distributional properties of small- and medium-sized places have been left aside. Furthermore, the DPLN does not call the LN to be wrong, but itself to be a more

[10]The literature about Zipf's law is especially interested in the coefficient α, which determines the slope of the upper tail Pareto distribution. It is important to keep in mind that in the context of the DPLN, α must be interpreted with caution; α (as well as β) also affect the location and scale of the distribution. The maximum likelihood procedure therefore uses the shape parameters to optimizes the tails, as well as the body over them. Therefore α cannot be interpreted as Zipf coefficient.

[11]The Pareto distribution in the lower tail is a further property of the DPLN that is in accordance with city size data, but not further analyzed here.

detailed description of the overall distribution. More detailed in a sense that its genesis takes the heterogeneity in the initial birth conditions into account.

5.4.2 Some comparative statics

In this section, we focus on how exogenous parameter changes pass through the model and affect the resulting city size distribution. The above presented model is mainly characterized by six exogenous parameters: g_A, ς_A, that characterize the stochastic process of technological progress, μ_{A_0}, σ_{A_0} that determine the distribution of the initial productivity draw, population growth g_S and the parameter Θ, which captures the information on the interaction between the agglomeration and dispersion forces. While the impact on the endogenous variables within the model is easily seen from the equilibrium values, given by equation (5.24) and (5.25), the DPLN is most easily analyzed over its moments.[12] The approach here is to focus on the DPLN distribution of the productivity parameters, the effects on the resulting city size distribution are analogue, as argued in Proposition 5. The moment generating function (mgf) of the DPLN in log scales ($a = logA$) is given by

$$M_a(\theta) = \frac{exp\left(\mu_{A_0} + \frac{\sigma_{A_0}^2 \theta^2}{2}\right)}{\lambda^{-1}\left(\lambda - \left(g_A - \frac{\varsigma_A^2}{2}\right)\theta - \frac{\varsigma_A^2}{2}\theta^2\right)}, \tag{5.30}$$

from which the first (mean) and second (variance) moment of the DPLN in log scales are easily derived.[13] The formula for the mean of the DPLN is calculated as

$$E[a] = \mu_{A_0} + \frac{g_A - \frac{\varsigma_A^2}{2}}{\lambda}. \tag{5.31}$$

The intuition of the mean is straightforward. The formula shows that the mean of the DPLN is the expected initial size μ_{A_0} plus a term that denotes the average growth of a city. Under the exponential age distribution, $1/\lambda$ is the average age of a city and with the Brownian motion $g_A - \frac{\varsigma_A^2}{2}$ is the expected growth of a city per period. We see that the mean is increasing in the initial productivity draw and the drift of the Brownian motion. The variance of the Brownian motion enters negative, a standard result of Itô calculus. The city creation rate also has a negative impact on the mean, because with a higher λ, more cities are relatively young and therefore small, thereby decreasing the mean. The variance of the DPLN is the following:

$$Var[a] = \frac{4g_A^2 - 2\lambda(\mu_{A_0} - 1)\varsigma_A^2 + \varsigma_A^4 + 4g_A\left(\lambda\mu_{A_0} - \varsigma_A^2\right) + 2\lambda^2\left(\mu_{A_0}^2 + \sigma_{A_0}^2\right)}{2\lambda^2}. \tag{5.32}$$

[12] A detailed derivation of the moment generating function of the DPLN and its properties are provided by Reed (2002) and Reed and Jorgensen (2004).

[13] Higher order moments, like third (Skewness) and fourth (Kurtosis) are available but are not shown for brevity.

The comparative statics for the variance are as well straightforward. The variance is increasing in the positive drift g_A of the Brownian motion, since older cities and young cities are more apart. The variance is also increasing in the variance σ_{A_0} and the mean μ_{A_0} of the initial productivity draw. The positive reaction to a higher mean arises because the Brownian motion is a multiplicative growth process. Higher initial sizes therefore increase the scale of the DPLN. The variance is decreasing in the city creation rate λ, because a higher city creation rate provides more young cities, clustering at low sizes. The reaction of the variance to the variability of the Brownian motion is ambiguous and described by

$$\frac{\delta V[a]}{\delta \varsigma} \begin{cases} < 0 & \text{if} \quad \lambda < 2g_A - \varsigma^2 \\ > 0 & \text{if} \quad \lambda > 2g_A - \varsigma^2 \text{ and } \mu_{A_0} > \frac{\lambda - 2g_A + \varsigma^2}{\lambda} \\ < 0 & \text{if} \quad \lambda < 2g_A - \varsigma^2 \text{ and } \mu_{A_0} > \frac{\lambda - 2g_A + \varsigma^2}{\lambda} \end{cases}. \tag{5.33}$$

The reason for this ambiguity is a trade off. On the one hand, ς^2 increases the variability of the Brownian motion, thereby increasing the variance of the DPLN. On the other hand, ς^2 occurs in the Itô-term $g_A - \frac{\varsigma^2}{2}$ and therefore decreases the average growth of cities. With a lower average growth, older cities are closer to the young cities, thereby decreasing the variance of the DPLN.

5.4.3 The meaning of the city creation rate

The above presented economic model is characterized by a stable growth path, under which Ω_t, x_t and N_t grow by λ. A further analysis of the moment generating function reveals the special meaning of the parameter λ, especially for the coherence between the DPLN and the LN. Using a partial decomposition, along the lines of Reed (2002), equation (5.30) can be rewritten as

$$M_a(\theta) = exp\left(\mu_{A_0} + \frac{\sigma_{A_0}^2 \theta^2}{2}\right) \cdot \frac{\alpha\beta}{(\alpha - \theta)(\beta + \theta)}, \tag{5.34}$$

where $\alpha(g_A, \varsigma_A, \lambda)$ and $\beta(g_A, \varsigma_A, \lambda)$ are the roots of the characteristic function, given by

$$\alpha = \frac{-2g_A + \varsigma_A^2 - \sqrt{4g_A^2 - 4g_A\varsigma_A^2 + \varsigma_A^4 + 8\varsigma_A^2 \cdot \lambda}}{2\varsigma_A^2} \tag{5.35}$$

$$\beta = -\frac{-2g_A + \varsigma_A^2 + \sqrt{4g_A^2 - 4\mu\varsigma_A^2 + \varsigma_A^4 + 8\varsigma_A^2 \cdot \lambda}}{2\varsigma_A^2}. \tag{5.36}$$

Now recall the mgf of the Normal distribution and the asymmetric Laplace distribution (ALP):

$$f_X \sim N(\mu, \sigma) \quad \rightarrow \quad M_X(\theta) = exp\left(\mu + \frac{\sigma^2\theta^2}{2}\right) \tag{5.37}$$

$$f_X \sim ALP(\alpha, \beta) \quad \rightarrow \quad M_X(\theta) = \frac{\alpha\beta}{(\alpha - \theta)(\beta + \theta)}. \tag{5.38}$$

We can now conclude that the mgf of the DPLN (in log scales) in equation (5.30) is the product of equation (5.37) and equation (5.38). Going over to normal scales, this means that the distribution of the DPLN can also be obtained from the convolution of a LN and a double Pareto distribution (DP), which is the logarithmic counterpart of the ALP, with a density of

$$f(v) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} v^{\beta-1} & \text{if} \quad v \leq 0 \\ \frac{\alpha\beta}{\alpha+\beta} v^{-\alpha-1} & \text{if} \quad v > 1 \end{cases}. \tag{5.39}$$

We are now in a position to distinguish two extreme cases. First, in the case of $\lambda = 0$, either $\alpha = 0$ or $\beta = 0$ (depending on whether $-2_A > \varsigma^2$) and the exponential distribution collapses and along with it the DPLN. The reason is that under $\lambda = 0$, there has never been the creation of any city and there is no city size distribution. In the second extreme case, we consider $\lambda \rightarrow \infty$. Using the first derivate of (5.35) and (5.36), it is easy to show that $\frac{\delta\alpha}{\delta\lambda} > 0, \frac{\delta\beta}{\delta\lambda} > 0$ and $\lambda \rightarrow \infty$ leads to $\{\alpha, \beta\} \rightarrow \infty$. From (5.39) we see that the DP distribution has a mean at zero. As α and β become larger, the tails of the distribution become steeper, the variance becomes smaller and the distributions mass of probability becomes centered around zero. Therefore the DPLN is the sum of a Normal distribution with a distribution that has zero probability, leaving the DPLN as a Normal distribution in log scales without Pareto-tails.

For values of λ between those extreme cases, we can conclude from the above analysis that a higher λ implies the DPLN to look more like a LN distribution. The intuition is straightforward; recall that the DPLN is the mixture of many lognormal distributions with different parameters which depend only on their age. From the theory on mixture distributions, we further know that the shape of the mixture is closer to the shape of the sub-distributions, the closer the sub-distributions are. With a higher λ, the steeper the exponential age distribution and the lower the variance of age and the sub-distributions are closer.

5.5 Conclusions

The recently made access to overall data on city sizes has triggered a wave of studies. In stark contrast to the mature literature on Zipf's law, those studies find a lognormal shaped city size distribution. The paper at hand proposes their co-existence as the natural outcome of an economic model of urban growth. The distinctive feature of this model is age heterogeneity accompanied with city evolution.

Within the model cities grow at the intensive and at the extensive margin and a growing population distributes endogenously over a rising number of cities. While the

intensive margin is characterized by Gibrat's law, the extensive margin is characterized by a constant city growth rate. Older cities have a longer time to experience economic and population growth and are therefore larger on average. The model features a DPLN city size distribution, which has been shown by Giesen, Suedekum and Zimmermann (2010) to have a very well data fit and to outperform the LN. The attractiveness of the DPLN, besides its superior fit, is the comprehension of the well known Pareto distribution along the lognormal body. This unique feature reconciles the findings of Eeckhout (2004) with the well-known Zipfian power law.

The pioneering work of Eeckhout (2004) and the proposal of a LN city size distribution has become the benchmark for explaining cities in their sizes and numbers. While this paper advances the DPLN, the both should not be viewed as rivalry distributions but as complements; the DPLN is just a more detailed description. In the future, the literature will eventually be enhanced by further proposals on the correct city size distribution. While those distributions will differ in their functional form and their number of parameters no serious study should propose a distribution ad-hoc, in a theory free manner.

Chapter 6

Conclusions

Conclusions

Within the last decade, the literature on city size distributions has been thrilling. Recent findings about the overall city size distribution are in stark contrast to the predictions of the mature literature that focused only on the largest cities. The contribution of this doctoral thesis is to unify between those two opposing views and to contribute to the literature in many respects both theoretically and empirically.

The bottom line is that Zipf's law is not necessarily an illusion, as proposed by Eeckhout (2004). Instead, Zipf's law is a possible feature of an upper-tier truth: the double Pareto lognormal distribution (DPLN). As shown in this work, the DPLN provides an excellent data fit, better than the famous lognormal (LN) distribution, which is proposed by the seminal study of Eeckhout (2004). The thesis therefore settles several controversies in the literature on city sizes and urban growth. Most important, it settles the recent dispute between Eeckhout (2004, 2009) and Levy (2004), conducted in the *American Economic Review*. Their dispute is concerned with the question on whether the LN or the Pareto distribution are the correct parameterization for city size distributions, especially in the upper tail. This thesis advances the DPLN under which Eeckhout's findings are true for the body of the distribution, while Levy's claim of a Pareto distribution are correct for the upper tail.

While there are many economic models that try to explain residential location choice and city formation, none of those models is in accordance with empirical data on city size distributions; they cannot explain the body or the upper tail of the distribution. One major contribution of this thesis is to offer a micro-founded economic model of random urban growth, in which a rising population distributes over a rising number of cities. The resulting city size distribution is the DPLN, and the model is therefore, to the best of my knowledge, the only existing model to replicate empirical overall city size distributions.

The DPLN and the LN are not rivalry but result from the same mechanism: Gibrat's law. The mere difference is the DPLN to be a more detailed description. In the future there will be further studies on the correct parameterization for empirical city size distributions across countries. Each study should keep in mind that the appearance and functional form of city size distributions is not of immediate interest for economists per se. It is the emergence of the empirical patterns that yield implications for economic theories like city formation, urban growth, location choice, and many more.

Chapter 7

References

References

(1) Ades, A., Glaeser, E., 1995. Trade and Circuses: Explaining Urban Giants. Quarterly Journal of Economics 110, 195-228.

(2) Auerbach, F., 1913. Das Gesetz der Bevölkerungskonzentration. Petermanns Geographische Mitteilungen 59, 74-76.

(3) Bairoch, P., 1988. Cities and Economic Development: From Dawn of History to the Present. University of Chicago Press, Chicago.

(4) Bickel, P., Doksum, K., 2001. Mathematical Statistics Volume 1: Basic Ideas and Selected Topics. Prentice Hall, New Jersey.

(5) Black, D., Henderson, V., 1999. A Theory of Urban Growth. Journal of Political Economy 107, 252-284.

(6) Black, D., Henderson, V., 2003. Urban Evolution in the USA. Journal of Economic Geography 3, 343-372.

(7) Blank, A., Solomon, S., 2000. Power Laws in Cities Populations, Financial Markets and Internet Sites (Scaling in Systems with a Variable Number of Components). Physica A 287, 279-288.

(8) Bosker, E., Brakman, S., Garretsen, H., Schramm, M., 2008. A Century of Shocks: The Evolution of the German City Size Distribution. Regional Science and Urban Economics 38, 330-347.

(9) Bosker, M., Buringh, E., 2010. City Seeds: Geography and the Origins of European Cities. CEPR Discussion Paper 8066.

(10) Bosker, M., Buringh, E., van Zanden, J.L., 2012. From Baghdad to London: Unraveling Urban Development in Europe and the Arab World 800-1800. Forthcoming: Review of Economics and Statistics.

(11) Brakman, S., Garretsen, H., van Marrewijk, C., van den Berg, M., 1999. The Return of Zipf: Towards a Further Understanding of the Rank-Size Distribution. Journal of Regional Science 39, 182-213.

(12) Champernowne, D.G., 1953. A Model of Income Distribution. Economic Journal 63, 318–351.

(13) Cheshire, P., 1999. Trends in Sizes and Structures of Urban Areas. Handbook of Urban and Regional Economics Volume 3: Applied Urban Economics, North-Holland, Amsterdam.

(14) Christaller, W., 1933. Die Zentralen Orte in Süddeutschland. Gustav Fischer, Jena.

(15) Cuberes, D., 2011. Sequential City Growth: Empirical Evidence. Journal of Urban Economics 69, 229-239.

(16) Dobkins, L.H., Ioannides, Y., 2001. Spatial Interactions Among U.S. Cities: 1900-1990. Regional Science and Urban Economics 31, 701-731.

(17) Duranton, G., Puga, D., 2004. Micro-Foundations of Urban Agglomeration Economies. Handbook of Urban and Regional Economics Volume 4: Cities and Geography, North-Holland, Amsterdam.

(18) Duranton, G., 2007. Urban Evolutions: The Fast, the Slow, and the Still. American Economic Review 97, 197-221.

(19) Duranton, G., 2011. Innovation in Cities: Classical and Random Urban Growth Models. Handbook of Research on Innovation and Entrepreneurship, Edward Elgar, Northampton.

(20) Eaton, J., Eckstein, Z., 1997. Cities and Growth: Theory and Evidence from France and Japan. Regional Science and Urban Economics 27, 443-474.

(21) Eeckhout, J., 2004. Gibrat's Law for (All) Cities. American Economic Review 94, 1429-1451.

(22) Eeckhout, J., 2009. Gibrat's Law for (All) Cities: Reply. American Economic Review 99, 1676-1683.

(23) Findeisen, S., Suedekum, J., 2008. Industry Churning and the Evolution of Cities: Evidence for Germany. Journal of Urban Economics 64, 326-339.

(24) Fujita, M., Krugman, P.R., Venables, A.J., 1999. The Spatial Economy: Cities, Regions and International Trade. MIT Press, Cambridge.

(25) Gabaix, X., 1999. Zipf's Law for Cities: An Explanation. Quarterly Journal of Economics 114, 739-767.

(26) Gabaix, X., Ioannides, Y., 2004. The Evolution of City Size Distributions. Handbook of Urban and Regional Economics Volume 4: Cities and Geography, North-Holland, Amsterdam.

(27) Gabaix, X., Ibragimov, R., 2011. Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents. Journal of Business Economics and Statistics 29, 24-39.

(28) Gan, L., Li, D., Song, S., 2006. Is the Zipf Law Spurious in Explaining City Size Distributions. Economics Letters 92, 256-262.

(29) Garmestani, A., Allen, C., Gallagher, C., Mittelstaedt, J., 2007. Departures from Gibrat's Law, Discontinuities and City Size Distributions. Urban Studies 44, 1997-2007.

(30) Gibrat, R., 1931. Les Inégalités Économiques. Librairie du Recueil Sirey, Paris.

(31) Giesen, K., Suedekum, J., Zimmermann, A., 2010. The Size Distribution Across all Cities - Double Pareto Lognormal Strikes. Journal of Urban Economics 68, 129-137.

(32) Giesen, K., Suedekum, J., 2011. Zipf's Law for Cities in the Regions and the Country. Journal of Economic Geography 11, 667-686.

(33) Giesen, K., Suedekum, J., 2012a. The Size Distribution Across All "Cities": A Unifying Approach. CESifo Working Paper 3730.

(34) Giesen, K., Suedekum, J., 2012b. The Overall French City Size Distribution. Under Revision for Region et Developpement.

(35) Glaeser, E.L., 1999. Learning in Cities. Journal of Urban Economics 46, 254–277.

(36) González-Val, R., 2010. The Evolution of US City Size Distribution From a Long Term Perspective (1900-2000). Journal of Regional Science 50, 952-972.

(37) Haerdle, W., 1992. Applied Nonparametric Regression. Cambridge University Press, Cambridge.

(38) Henderson, V., 1974. The Sizes and Types of Cities. American Economic Review 64, 640-656.

(39) Henderson, V., 1995. Innovation and Agglomeration: Two Parables Suggested by City-Size Distributions: Comment. Japan and the World Economcy 7, 391-393.

(40) Henderson, V., Wang, H., 2007. Urbanization and City Growth: The Role of Institutions. Regional Science and Urban Economics 37, 283-313.

(41) Henderson, V., Venables, A., 2009. Dynamics of City Formation. Review of Economic Dynamics 12, 233-254.

(42) Ioannides, Y., Overman, H., 2003. Zipf's Law for Cities: An Empirical Investigation. Regional Science and Urban Economics 33, 127-137.

(43) Ioannides, Y., Skouras, S., 2009. Gibrat's Law for (All) Cities: A Rejoinder. Tufts University, Economics Department Discussion Paper.

(44) Kass, R., Raferty, A., 1995. Bayes Factors. Journal of the American Statistical Association 90, 773-795.

(45) Krugman, P.R., 1996. The Self-Organizing Economy. Blackwell Publishers, Cambridge.

(46) Levy, M., Solomon, S., 1996. Dynamical Explanation for the Emergence of Power Law in a Stock Market Model. International Journal of Modern Physics C 37, 65-72.

(47) Levy, M., 2009. Gibrat's Law for (All) Cities: Comment. American Economic Review 99, 1672-1675.

(48) Lösch, A., 1940. Die Räumliche Ordnung der Wirtschaft: Eine Untersuchung über Standort, Wirtschaftsgebiete und Internationalen Handel. Gustav Fischer, Jena.

(49) Malevergne, Y., Pisarenko, V., Sornette, D., 2011. Testing the Pareto Against the Lognormal Distributions with the Uniformly Most Powerful Unbiased Test Applied to the Distribution of Cities. Physical Review E 83, 036111.

(50) Michaels, G., Rauch, F., Redding, S., 2012. Urbanization and Structural Transformation. Forthcoming: Quarterly Journal of Economics.

(51) Mitzenmacher, M., 2004. A Brief History of Generative Models for Power Law and Lognormal Distributions. Internet Mathematics 1, 226-251.

(52) Nadaraya, E., 1964. On Estimating Regression. Theory of Probability and its Applications 9, 141-142.

(53) Nitsch, V., 2005. Zipf Zipped. Journal of Urban Economics 57, 86-100.

(54) Reed, W., 2001. The Pareto, Zipf and Other Power Laws. Economic Letters 74, 15-19.

(55) Reed, W., 2002. On the Rank-Size Distribution for Human Settlements. Journal of Regional Science 42, 1-17.

(56) Reed, W., 2003. The Pareto Law of Incomes - An Explanation and an Extension. Physica A 319, 469-486.

(57) Reed, W., Jorgensen, M., 2005. The Double Pareto-Lognormal Distribution - A New Parametric Model for Size Distribution. Communications in Statistics 34, 1733-1753.

(58) Rossi-Hansberg, E., Wright, M., 2007. Urban Structure and Growth. Review of Economic Studies 74, 597-624.

(59) Rosen, K., Resnick, M., 1980. The Size Distribution of Cities: An Examination of the Pareto Law and Primacy. Journal of Urban Economics 8, 165-186.

(60) Rozenfeld, H.D., Rybski, D., Andrade, J., Batty, M., Stanley, H.E., Makse, H.A., 2008. Laws of Population Growth. Proceedings of the National Academy of Sciences U.S.A. 105, 18702-18707.

(61) Rozenfeld, H.D., Rybski, D., Makse, H.A., Gabaix, X., 2011. The Area and Population of Cities: New Insights From a Different Perspective on Cities. American Economic Review 101, 2205-2225.

(62) Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, New York.

(63) Soo, K., 2005. Zipf's Law for Cities: A Cross-Country Investigation. Regional Science and Urban Economics 35, 239-263.

(64) Starrett, D., 1978. Market Allocations of Location Choice in a Model with Free Mobility. Journal of Economic Theory 17, 21-37.

(65) Watson, G., 1964. Smooth Regression Analysis. Sankhya Series A 26, 359-372.

(66) Yule, U., 1925. A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. Philosophical Transactions of the Royal Society of London B 213, 21-87.

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