DISCUSSION PAPER SERIES

IZA DP No. 3928

Zipf's Law for Cities in the Regions and the Country

Kristian Giesen
Jens Suedekum

January 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Zipf's Law for Cities in the Regions and the Country

Kristian Giesen
University of Duisburg-Essen

Jens Suedekum
University of Duisburg-Essen and IZA

Discussion Paper No. 3928
January 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 3928
January 2009

ABSTRACT

Zipf's Law for Cities in the Regions and the Country*

The salient rank-size rule known as *Zipf's law* is not only satisfied for Germany's national urban hierarchy, but also for the city size distributions in single German regions. To analyze this phenomenon, we build on the insights by Gabaix (1999) that Zipf's law follows from a stochastic growth process. In particular, Gabaix shows that if the regions follow Gibrat's law, we should observe Zipf at both the regional and the national level. This theory has never been addressed empirically. Using non-parametric techniques we find that Gibrat's law holds in *each* German region, irrespective of how "regions" are defined. In other words, Gibrat's law and therefore Zipf's law tend to hold *everywhere* in space.

JEL Classification: R11, O4

Keywords: city size distributions, city growth, Zipf's law, Gibrat's law, rank-size rule

Corresponding author:

Jens Suedekum
Mercator School of Management
University of Duisburg-Essen
Lotharstrasse 65
47057 Duisburg
Germany
E-mail: jens.suedekum@uni-due.de

* We thank Marcus Berliant, Maarten Bosker, Steven Brakman, Xavier Gabaix, Harry Garretsen, Wen-Tai Hsu, Stefan Hupfeld, Volker Nitsch, Henry Overman, Kurt Schmidheiny, Hiroki Watanabe, and seminar participants at the 2008 Annual Meeting of the North American Regional Science Council (NARSC) in New York City for helpful comments and suggestions. All errors and shortcomings are solely our responsibility.

1 Introduction

Zipf's law is one of the most striking empirical laws in economics, describing a remarkably stable regularity in the spatial structure of market economies (Krugman, 1996). As is well known, it states that the upper tail of the city size distribution within an area (say, the US) and at any point in time can be approximated 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 twice as large as the second-largest city (Los Angeles), three times as large as the third-largest city (Chicago), and so on. Following the seminal contributions by Auerbach (1913) and Zipf (1949) this alleged empirical rule has been addressed dozens of times.[1] Virtually all of these 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 also tends to hold in single regions of a country.

FIGURE 1 HERE

Figure 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. When running a standard Zipf regression of the type $\log(Rank) = \log(a) - \zeta \log(Size)$ we estimate $\zeta = 1.027$ with std.error 0.325 and $R^2 = 0.99$. As suggested by Zipf's law, 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. In other German regions (including Germany as a whole) this picture looks similar, even though the slope coefficients are not as close to $\zeta = 1$ as in Hessen (see below).

The main aim of this paper is to shed light on *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), then subject to frictions a steady state city size distribution is implied that obeys Zipf's law.[3] In other words, Gabaix establishes Gibrat's law as a statistical explanation for Zipf's

[1] Comprehensive studies on Zipf's law are provided by Rosen and Resnick (1980) and Soo (2005). Nitsch (2005) conducts a meta-study, and Gabaix and Ioannides (2004) present a survey of the literature.

[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 that systematically addresses intra-national city size distributions. Furthermore, we focus on the relationship between Gibrat's law and Zipf's law from a regional perspective. This aspect has not been studied empirically so far.

[3] Eeckhout (2004) shows that the pure form of Gibrat's law will generate a log-normal city size distribution, not the Pareto. Eeckhout (2004) furthermore shows that a log-normal distribution fits the overall size distribution of *all* places (cities and small towns) better than the Pareto. Zipf's law is a feature pertaining only to the upper tail of the size distribution. In this paper we concentrate on this upper tail in order to address the salient Zipf's law from a regional perspective.

law, thereby opening new avenues for theoretical and empirical research on the rank-size rule. Theoretical papers on the economic microfoundations of Zipf's law typically 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 is the basis for studies such as Ioannides and Overman (2003), who test Zipf's law indirectly by investigating if Gibrat's law holds at the national level in the US.

Yet 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 region, then Zipf's law will be satisfied for all regional and also for the national city size distribution.[4] To the best of our knowledge this theoretical insight has never been addressed empirically. In this paper we provide a first test of Gibrat's law on a regional level, contemplating different concepts of a "region". Most naturally we analyze the German Länder, which can be thought of as politically well-defined clubs of spatially adjacent cities. Moreover, since Gabaix's theory is not confined to regions with historical or administrative content, we also consider two other concepts. Firstly, we construct *random* regions, i.e., random draws from the population of large German cities that need not be adjacent to one another. Secondly, we build spatial clubs of large cities that are geographically adjacent, but that need not belong to the same Federal State.

It turns out that Gibrat's law holds not only at the national level in Germany, but in *each* region regardless of which type. According to Gabaix (1999) Zipf's law should then also be valid in each region and in the country - and in fact, this appears to be true. These findings 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 national level. Moreover, our findings suggest that the regular pattern for urban growth that arises at the national level holds more generally. Urban growth is scale independent not only in the aggregate across all cities (or in random draws from this aggregate), but also in non-random samples. Gibrat's law actually seems to hold *everywhere* in space.

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

A casual illustration of Gibrat's rule of proportionate growth is to say that the size of a city contains no information about its future growth rate. More precisely, cities grow randomly with the same expected growth rate and same variance. Gabaix (1999) shows that under certain conditions this growth process will converge to a Pareto distribution with exponent equal to -1: Zipf's law. This, however, requires some frictions in the growth process since the pure form of Gibrat's law generates a log-normal city size distribution, not the Pareto (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

[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.

$$\frac{dS_{it}}{S_{it}} = \mu dt + \sigma dB_{it}. \tag{1}$$

S_{it} reflects the size of city i at time period t normalized by the overall size of the economy. 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 (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 will converge 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

$$\zeta = \frac{1}{1 - S_{min}/\bar{S}}, \tag{2}$$

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 $S_{min} > 0$ this process also predicts a power law distribution, yet with exponent $\zeta > 1$. This case would not be Zipf's law exactly, but a closely related form of a power law (see Brakman et al. 1999 on this distinction).

Moreover, consider a country that is composed of R regions. Gibrat's law holds for each region, thus $G(S) = a_r/S^{\zeta}$ describes the regional city size distribution, where $\zeta \sim 1$ as S_{min} goes to zero. The national city size distribution is then characterized by:

$$G(S) = a/S^{\zeta} \quad \text{where } a = \sum_{r=1}^{R} \lambda_r a_r, \;\; \sum_{r=1}^{R} \lambda_r = 1 \tag{3}$$

and again $\zeta \sim 1$. In words, if Gibrat's law and hence Zipf's law holds within each region, then Zipf's law holds for the country. This is even true if regions differ in their average growth rate, as long as urban growth is scale independent within each region. This theoretical result, proposition 2 in Gabaix (1999), is the basis for our analysis in section 5.

3 Data

The data set for this study is provided by the German Federal Statistical Office (Statistisches Bundesamt). It contains area size (in square kilometers) and the number of inhabitants

[5]This result has recently been extended by Cordóba (2008a,b), who shows that Gibrat's law is not only a sufficient but also a necessary condition for Zipf's law: if Zipf's law holds at the national level, Gibrat's law must hold at the national level. This does *not* imply, however, that Gibrat's law must necessarily hold in every subset of cities within the nation, i.e., in every region.

for 2143 German cities, covering the time period from 1975 to 1997. This data set 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 2004). 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 in our data set is classified by the administratively defined boundaries, i.e., our data follows the "city proper" concept. The alternative would be "urban agglomeration" data, which aggregates main cities and suburban areas that often form own administrative units into metropolitan areas (MAs). Unfortunately such data is only available for a single year (1997) from an official source, which forces us to mainly use the city proper data.

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 primate 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.

4 The national level: Germany

In this section we first address Gibrat's law and Zipf's law at the national level, following the typical approach in the literature.

4.1 Gibrat's law

For the analysis of Gibrat's law we use *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.

[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.

Under the null 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 we take a first look at this issue 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 on size, and yields a first impression that urban growth appears to be very similarly distributed across different city size classes.

FIGURES 2, 3 HERE

In Figures 3a and 3b we provide non-parametric estimates for the conditional means and variances of city growth rates. The estimation was performed using the Nadaraya-Watson technique (see Nadaraya 1964; Watson 1964; Härdle 1992) which estimates the expectation of growth conditional on size. The underlying regression equation is

$$g_i = m(S_i) + \epsilon_i \quad \text{with} \quad m(S_i) = E[g|S_i], \tag{4}$$

where m, the unknown regression function, is calculated according to

$$\hat{m}_h(s) = \frac{n^{-1} \sum_{i=1}^{n} K_h(s - S_i) g_i}{n^{-1} \sum_{i=1}^{n} K_h(s - S_i)}. \tag{5}$$

In this calculation m is a locally weighted average, where the kernel K (in our case the Epanechnikov kernel) is the weighting function with bandwidth h. We use bandwidth $h = 0.5$ as our benchmark.[7]

We find that city growth rates are indeed independent of initial city size. 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. In other words, we cannot formally reject Gibrat's law for Germany.

These findings should be set into perspective to Bosker et al. (2008), who address the stability of the German city size distribution over a longer time period (1925-1999). They argue that Gibrat's law held perfectly in Germany prior to WWII, but not in the post-war time period 1945-1999 where they found that small cities gained population relative to large ones. This implies that the German city size distribution was permanently affected by the WWII-shock.[8] Our findings indicate that the transition from the pre- to the post-war distribution has been completed until 1975, the starting point of our data set. That is, if Gibrat's law did not hold in Germany in the first decades after the war during the transition towards a new distribution, it seems to hold again in the more recent period.

[7]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.

[8]In this respect Germany would be a special case due to the massive shocks of WWII, 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) or Eaton and Eckstein (1997).

4.2 Zipf's law

As Gibrat's law holds at the national level in Germany, how about Zipf's law? In Figure 4a we plot the log of the city's rank in Germany's overall urban hierarchy (#1 for Berlin, #2 for Hamburg, #3 for München, and so on) against log 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. By inspection the relationship is linear with only small outliers, and in fact, linear regressions yield R^2-levels beyond 0.98.[9] For the estimation we have considered two alternative methods that both draw on the recent contribution by Gabaix and Ibragimov (2008) who show that standard OLS yields biased results in rank-size regressions: (i) a normal log rank-log size regression with corrected std.errors (see legend of Figure 1 for details), (ii) a refined OLS regression which uses the log of (rank-1/2) as the LHS variable (also with corrected std.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.

FIGURE 4 HERE

Recall from section 2 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 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 we use that type of data.

This is shown in Figure 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) appear to be outliers as they are a bit "too small".[10] 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.

[9]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.

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

5 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.

5.1 Random and non-random regions

We contemplate three different concepts of a "region". Firstly, we build *random* regions which are constructed by randomly drawing 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, which we normalize with the mean and standard deviation of growth rates in the random region. These 20x22 normalized growth rates are used to estimate conditional mean and variance with the same methodology (the Nadaraya-Watson approach) as before.

Secondly, we 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). The Länder are *non-random* draws from the population of large German cities. Unlike the random regions they form spatially adjacent clubs. 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 state.

Finally, we construct clubs of cities that are spatially adjacent but that do not necessarily belong to the same Federal State. 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.[11] Notice that the number of members differs across clubs, that one city can belong to more than one club, and that the respective central city need not be the largest member in its club. We also consider two other definitions (see below). The resulting spatial clubs are also non-random draws from the population of large German cities, yet without any political or administrative content.

5.2 Results

Figure 5 summarizes the results for 8 representative examples of **random regions**, where $N = 20$ is the number of cities for each draw. Across these examples we consistently cannot reject the null hypotheses of proportionate growth, for virtually no range of city sizes. For other examples of random draws this is similar. These findings are in accordance with our expectations: If the properties of scale invariance of mean and variance of normalized growth rates are satisfied in the total population of large cities, they should also be satisfied (except for white noise) in random draws from this population.

The interesting question is whether the validity of Gibrat's law indicates the validity of Zipf's law in these random groups of cities, as posited in Gabaix's theory. Our findings

[11] We have constructed a 71x71 distance matrix for this exercise using standard route planning software.

in fact support this theory. In the Zipf plots that we provide for the random draws (see the left column of Figure 5) we find the well-known linear rank-size relationships. The estimated slopes $-\zeta$ range from -0.87 to -1.22 for these examples, using from now on only the normal rank-size regression with corrected std. errors.

To address Zipf's law for the random regions more systematically, we summarize in Figure 6a the distribution of the OLS estimates of $-\zeta$ 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 6b we show the distribution of $-\zeta$ across 500 draws of size $N = 100$.[12] 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 towns from the lower tail of the size distribution where Zipf's law is known to perform worse (Eeckhout, 2004). The power law shape of the distribution is, however, robust across all random regions.

FIGURES 5-7 AND TABLE 1 HERE

Turning now to the **German Länder**, notice that the scale independence of urban growth 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. 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.[13]

Empirically, however, this is not the case in Germany. We rather find that Gibrat's law holds in *each* Federal State. For no State and over the entire range of city sizes we cannot reject the null of zero normalized growth rates and constant variance equal to one (see the right two columns in Figure 7), the only slight exception being the small cities in Rheinland-Pfalz. In the left column of Figure 7 we show the Zipf plots and in Table 1 we summarize our estimation results for the Länder. The Zipf coefficients range from $\zeta = 0.93$ in Bayern to $\zeta = 1.37$ in Nordrhein-Westfalen (all highly significant), with 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 4.[14] Hence, we find that city size distributions follow on average a similar power law pattern in the Länder as in the national aggregate, that is for Western Germany as a whole.

[12]For the construction of these regions we did not stick to the population of the 71 largest cities but we have included also smaller towns (with minimum population size 6,000) in the population of cities from which to draw.

[13]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.

[14]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.

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 Nadaraya-Watson estimates for conditional city growth for 8 of the 71 spatial clubs in Figure 8, where one club consist of all cities with distance below $d = 200$ kilometers around the respective central member. The picture that emerges from Figure 8 is that urban growth is scale independent also 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 $M = 20$ 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 largest German cities and constructed a club as the collection of the $M = 20$ largest cities inside this circle. We have dropped those clubs with less than 20 members, which led to a total of 68 clubs.

FIGURES 8 AND 9 HERE

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 practically never be rejected in any range of city sizes.[15] Looking at 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 9a-9c summarize the distribution of $-\zeta$ 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.

Summing up, it seems to be the case that the property of scale independent urban growth in Western Germany is not confined to the Länder; Gibrat's law holds also in areas that extend across State boundaries. Furthermore, we again find support for the claim that the validity of Gibrat's law in an area comes along with Pareto-distributed city sizes in that area.

6 Conclusions

In this paper we have shown that mean and variance of annual urban growth rates in Western Germany are scale independent. Gibrat's law is not only satisfied at the national level, or in random draws from the aggregate population of (large) German cities. It continues to hold in non-random samples, namely in the German Länder and in various types of cross-state clubs of spatially adjacent cities.

According to Gabaix (1999) we would then expect Zipf's law for city sizes to be satisfied as well, both on a national and on a regional level, and in fact we do find strikingly

[15]The Nadaraya-Watson plots for the two alternative definitions of a spatial club are omitted here for brevity. All plots for all types of regions can be made available upon request.

linear rank-size relationships in the country and in each of the various types of regions. These results support Gabaix's (1999) theoretical insight that stochastic city growth at the regional level implies a power law shape of the distribution of city sizes at the regional and national level. There are some deviations of the estimated shape parameters of this power law from $\zeta = 1$, but this is not inconsistent with the theory, since the exact Zipf's law is only achieved in the limit.

The other main message of this paper is that Gibrat's law appears to hold *everywhere* in space, that is for *each* region regardless of how it is defined. In principle it could be the case that urban growth is not scale independent in single non-random subsets of cities, even if Gibrat's law holds in the aggregate. However, this is not what is happening in Germany. Rather it seems to be the case that Gibrat's law is satisfied in each of the meaningful subsets of large German cities, not only in the aggregate.

These empirical findings are consistent with theories of proportional urban growth (such as Eaton and Eckstein 1997), where all cities in a country (regardless of initial size) grow with the same rate. Our results may inspire future theoretical work, as they show that urban growth is proportional even on a lower geographical scale. According to our results cities appear to grow with the same normalized rate regardless of initial size and regardless of location inside the country. One can thus think of the regions as miniatures of the country when it comes to urban growth structures and, consequently, when it comes to city size distributions.

References

[1] Auerbach F., (1913), "Das Gesetz der Bevölkerungskonzentration", Petermanns Geographische Mitteilungen 59, 74-76.

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

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

[4] 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.

[5] 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.

[6] Cheshire P., (1999), "Trends in Sizes and Structures of Urban Areas", Handbook of Urban and Regional Economics, vol. 3, North-Holland, Amsterdam.

[7] Cordóba J.-C., (2008a), "A Generalized Gibrat's Law", International Economic Review 49, 1463-1468.

[8] Cordóba J.-C., (2008b), "On the Distribution of City Sizes", Journal of Urban Economics 63, 177-197.

[9] Eaton J., Eckstein Z., (1997), "Cities and Growth: Theory and Evidence from France and Japan", Regional Science and Urban Economics 27, 443-74.

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

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

[12] Gabaix X., Ibragimov R., (2008), "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents", mimeo, Stern School of Business, New York University.

[13] Gabaix X., Ioannides Y., (2004), "The Evolution of City Size Distributions", Handbook of Urban and Regional Economics, Vol 4. Amsterdam: Elsevier Science, North-Holland.

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

[15] Härdle W., (1992), "Applied Nonparametric Regression", Cambridge University Press.

[16] Ioannides Y., Overman H., (2003), "Zipf's Law for Cities: An Empirical Examination", Regional Science and Urban Economics 33, 127-37.

[17] Krugman P., (1996), "The Self-Organizing Economy", Cambridge: Blackwell

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

[19] Nitsch V., (2005), "Zipf Zipped", Journal of Urban Economics 57, 86-100.

[20] 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.

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

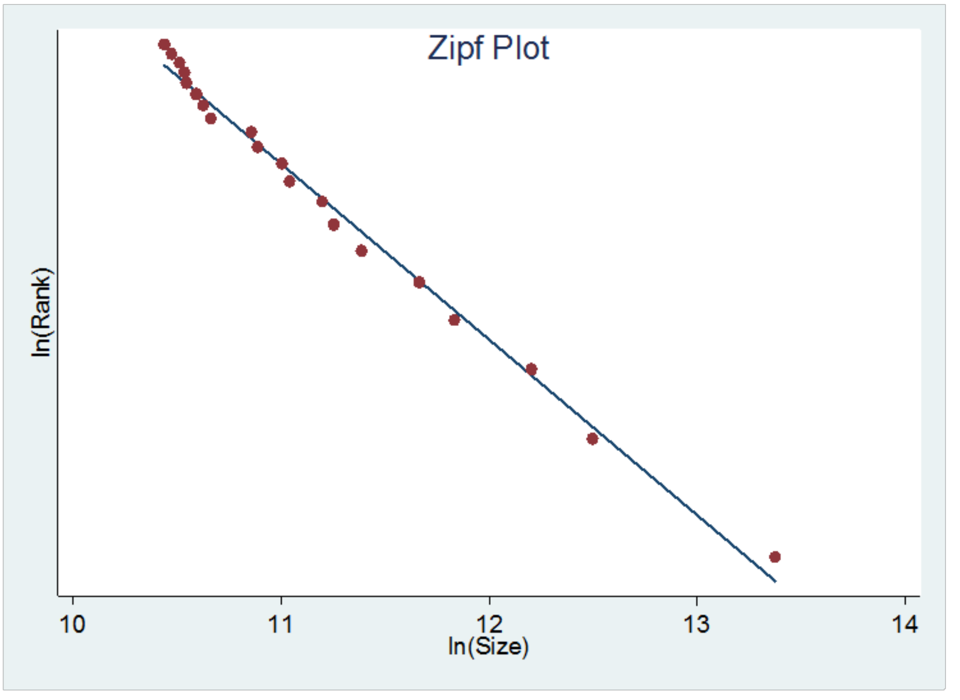
[22] Silverman B., (1986), "Density Estimation for Statistics and Data Analysis", New York: Chapman and Hall.

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

[24] Watson G., (1964), "Smooth Regression Analysis", Sankhya Series A 26, 359-372

[25] Zipf G., (1949), "Human Behavior and the Principle of Least Effort", Cambridge: Addison-Wesley.

Figure 1) Zipf Regression for Hessen



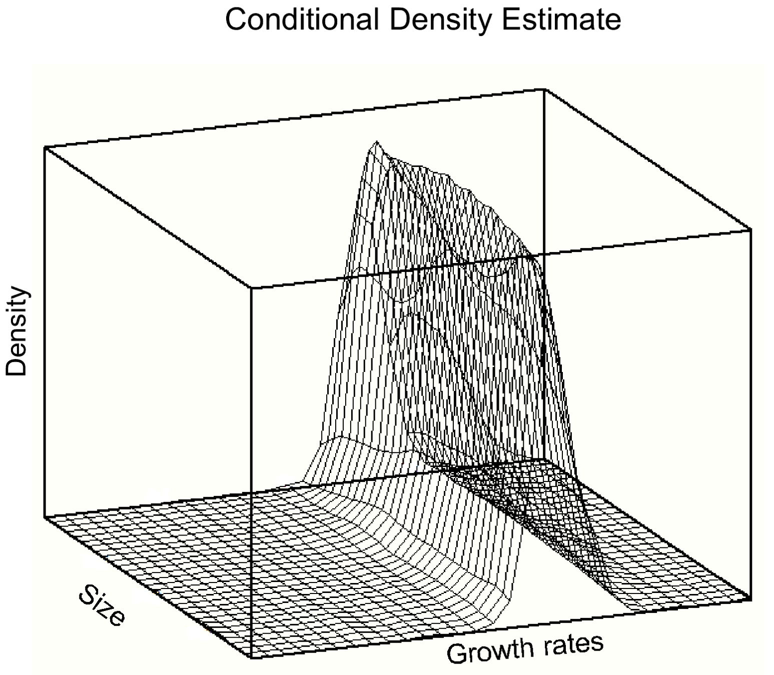
ln (Rank)= 13.60 - 1.027 ln (Size)

R^2=0.9920 (GI-corr. std. error: 0.325, p-value < 0.01)

Obs: 20

Legend: The std. errors are corrected according to the procedure suggested in Gabaix and Ibragimov (2008). GI-corr. std. errors are calculated according to $\sqrt{\frac{2}{n}}\hat{\zeta}$. Gabaix and Ibragimov (2008) show that simple OLS std. errors are biased. The simple OLS std. error for this regression would be 0.026.

Figure 2) Conditional Stochastic Kernel for Germany



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.

Figure 3a) Nadaraya-Watson Estimator: Mean Normalized Growth Rate

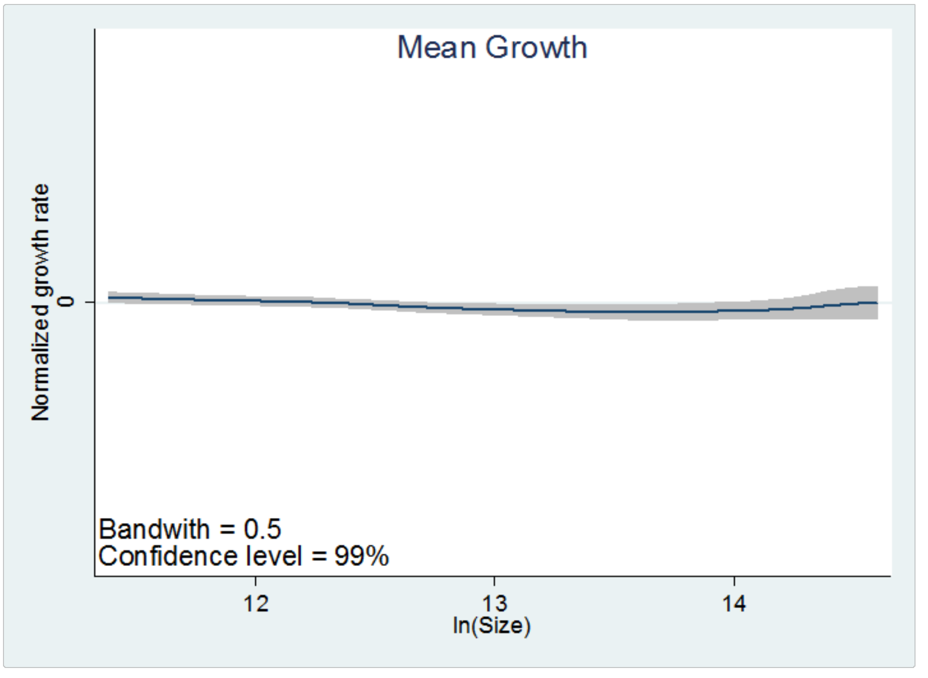


Figure 3b) Nadaraya-Watson Estimator: Normalized Variance of Growth Rate

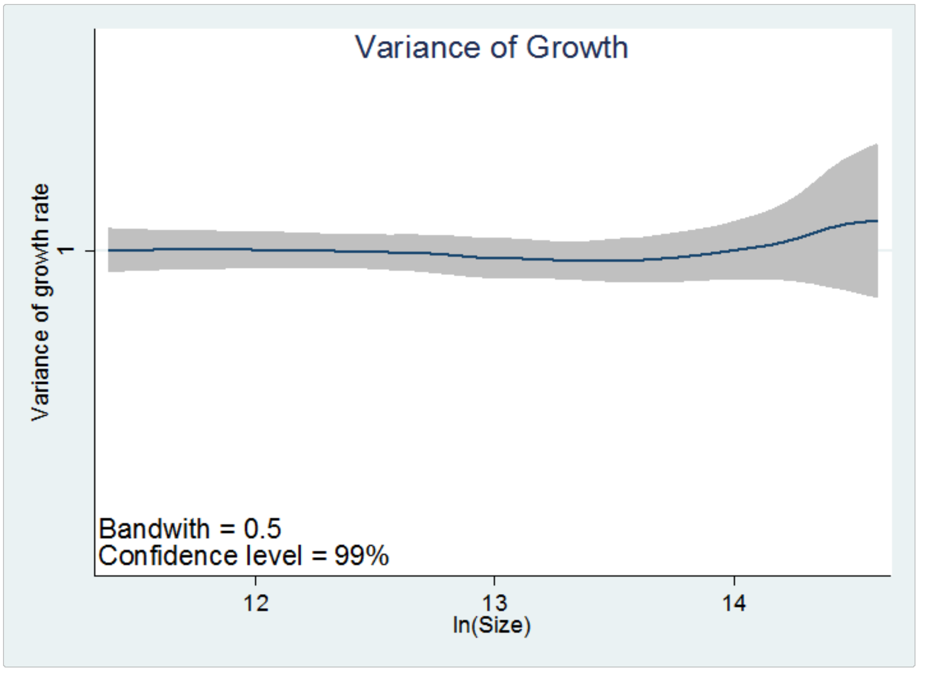
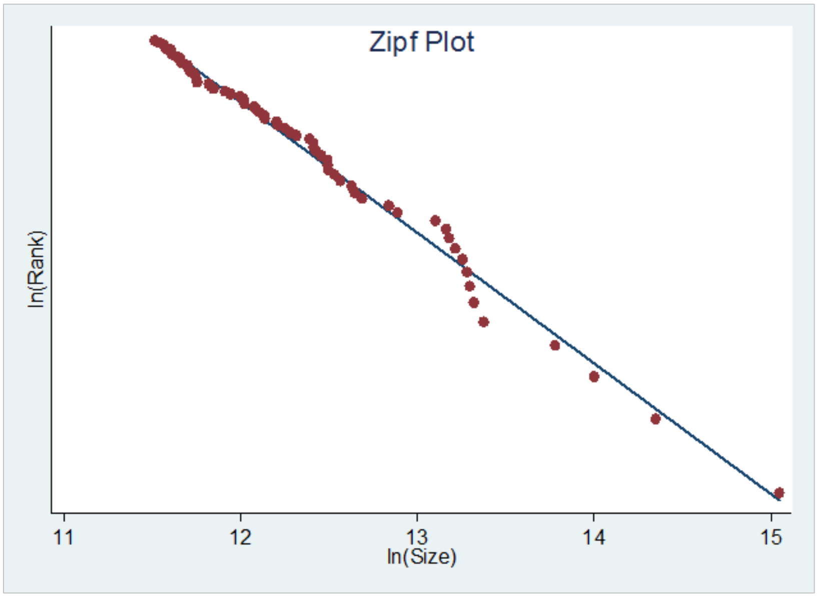


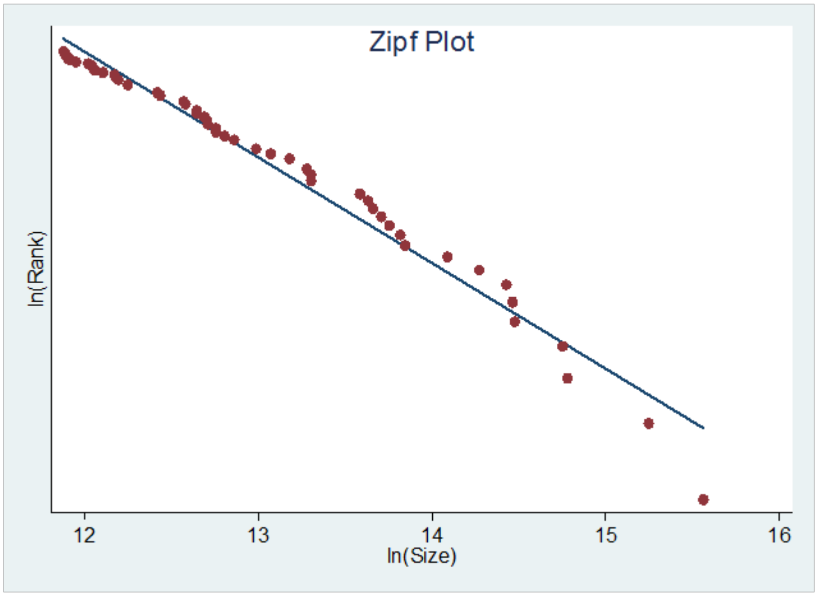
Figure 4a) Zipf Regression for Western Germany (City Proper Data)



OLS regression
ln (Rank)= 18.45 - 1.23 ln (Size)
R^2=0.9899 (GI.-corr. std. error: 0.206)
Obs.: 71

Gabaix Ibragimov regression
ln (Rank-1/2)= 19.53 – 1.32 ln (Size)
R^2=0.9863 (GI.-corr. std. error: 0.222)
Obs.: 71

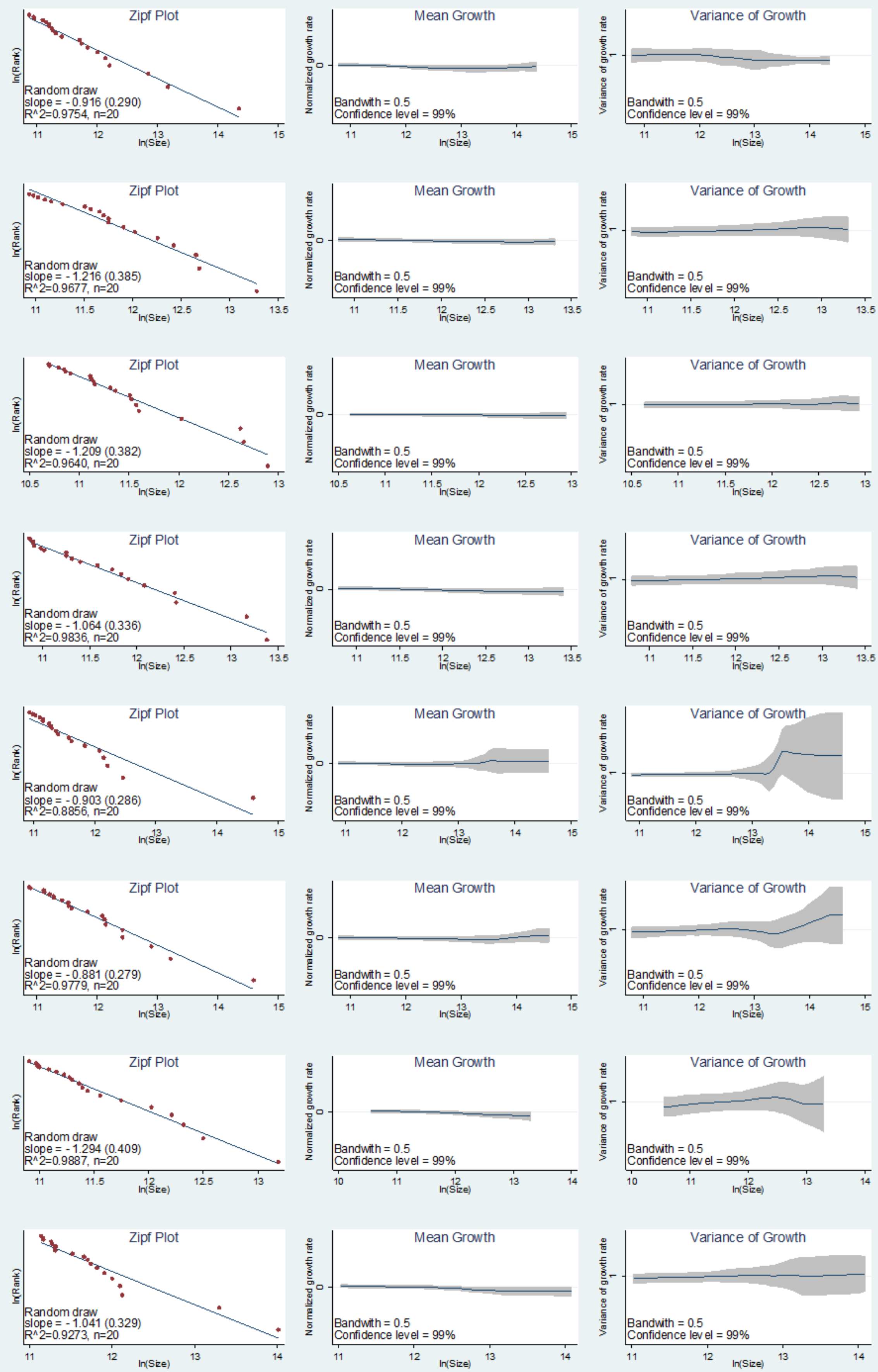
Figure 4b) Zipf Regression for Western Germany (Urban Agglomeration Data)



OLS regression
ln (Rank)= 15.56 – 0.96 ln (Size)
R^2=0.9739 (GI.-corr. std. error: 0.192)
Obs.: 50

Gabaix Ibragimov regression
ln (Rank-1/2)= 16.46 – 1.03 ln (Size)
R^2=0.9532 (GI.-corr. std. error: 0.206)
Obs.: 50

Figure 5) Random draw simulations: Zipf and Nadaraya-Watson Regression



The Gabaix-Ibragimov corrected std. err. is presented in parentheses.

Figure 6a) Distribution of Zipf Coefficients from Random draw Simulation (N=20)

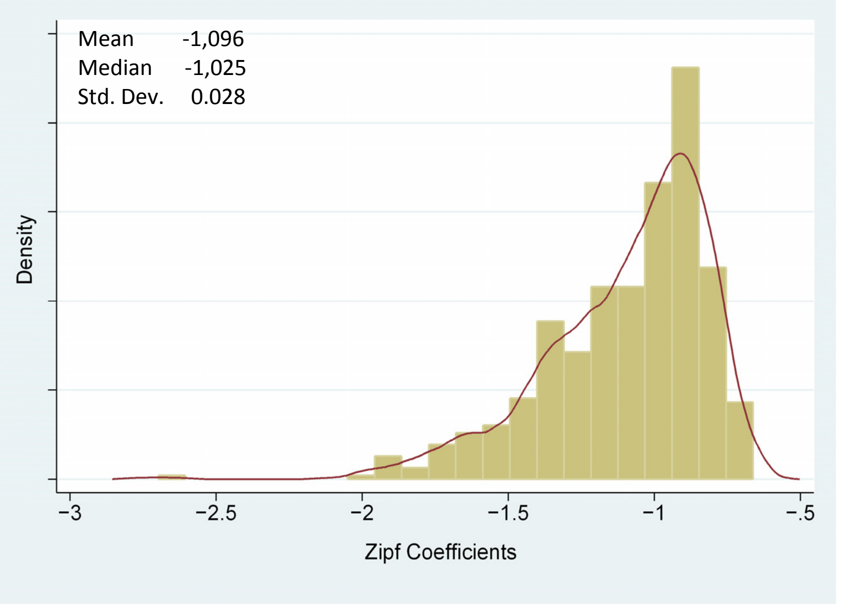
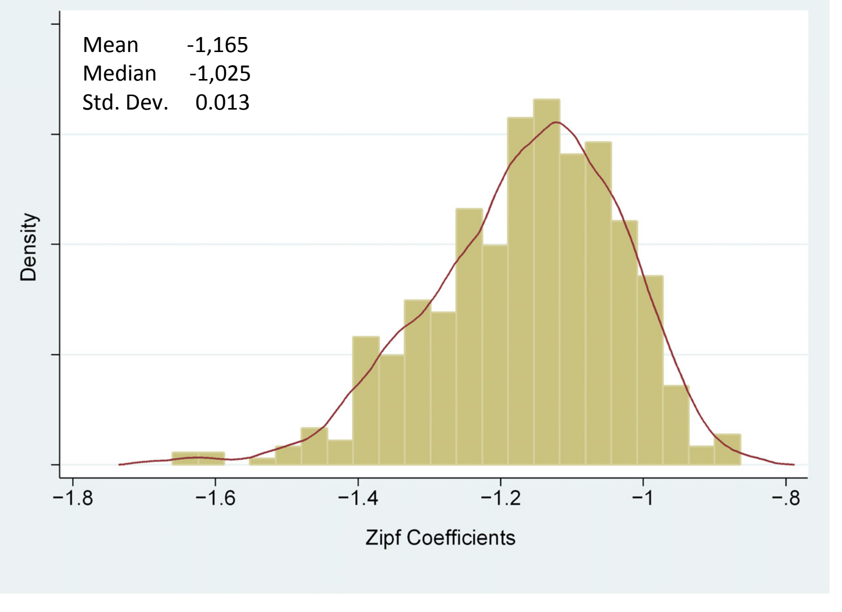
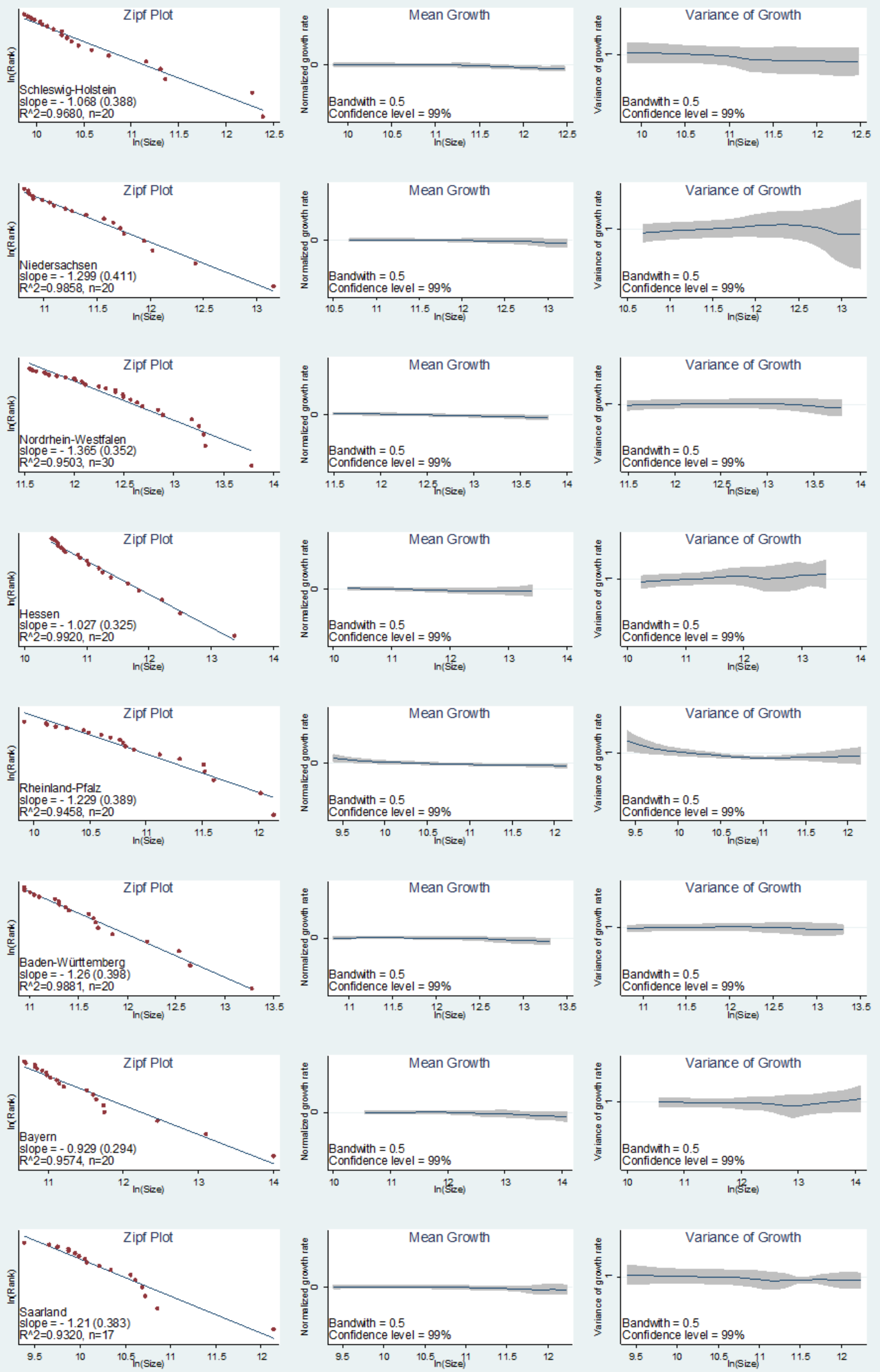


Figure 6b) Distribution of Zipf Coefficients from Random draw Simulation (N=100)



Legend: Based on 500 simulations of size N=20 (fig. 6a) and, respectively, N=100 (fig.6b). For each simulation a standard Zipf regression was conducted. Figures 6a and 6b report the distribution of the slope coefficients -ζ across the 500 simulations. All estimated slope coefficients were highly statistically significant using GI-corr. std. errors.

Figure 7) The Federal German States: Zipf and Nadaraya-Watson Regression



The Gabaix-Ibragimov corrected std. errors are presented in parentheses.

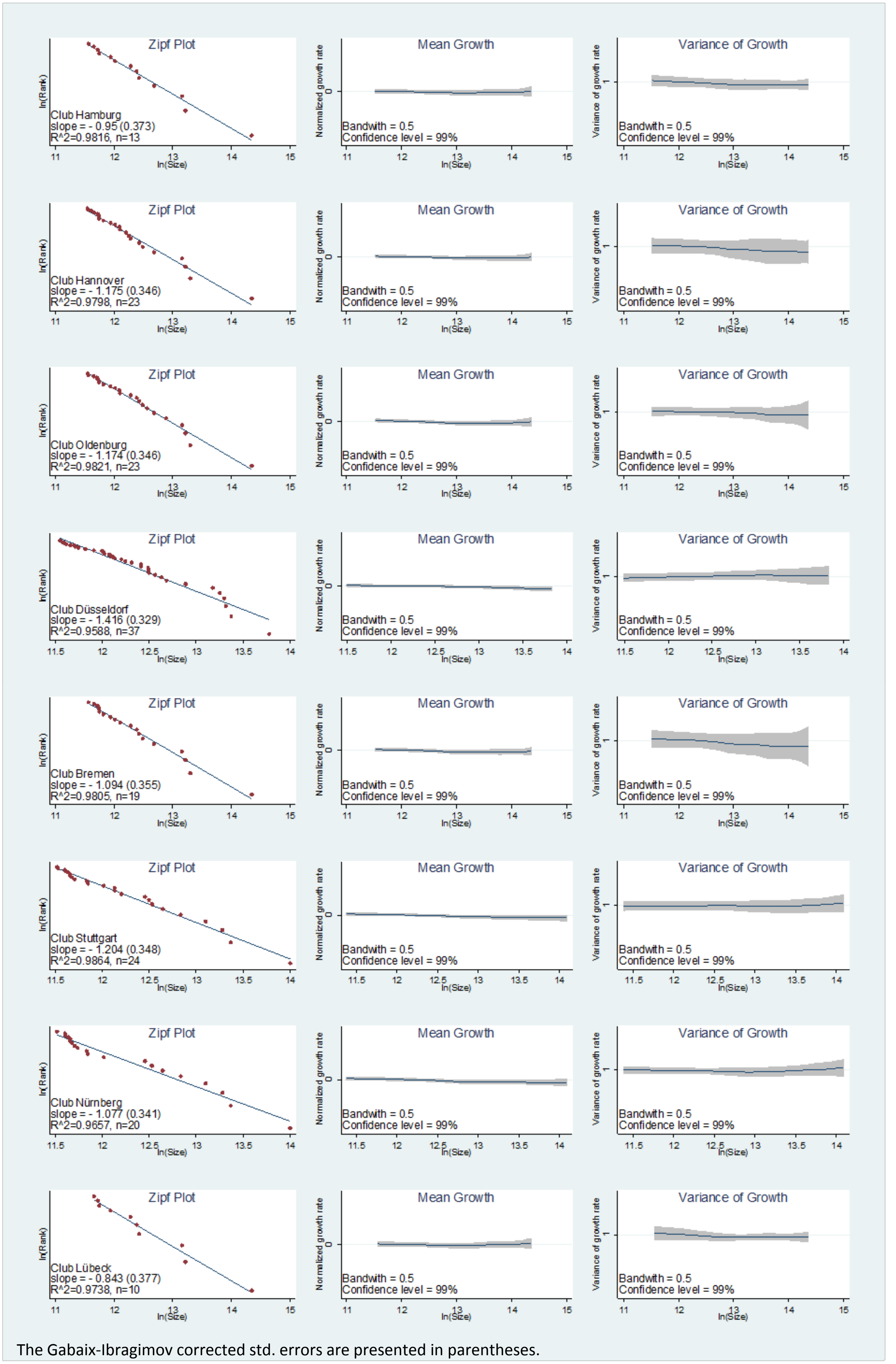
Table 1) The Federal States of Western Germany

	Number of Cities	Minimum City Size	Maximum City Size	**Zipf Coefficient**	R^2
Schleswig-Holstein	20	19 940	240 516	**1.068 (.388)**	**0.9680**
Niedersachsen	20	49 814	520 670	**1.299 (.411)**	**0.9858**
Nordrhein-Westfalen	30	103 872	964 311	**1.365 (.352)**	**0.9503**
Hessen	20	34 128	643 469	**1.027 (.325)**	**0.9920**
Rheinland- Pfalz	20	20 224	186 136	**1.229 (.389)**	**0.9458**
Baden Württemberg	20	56 781	585 274	**1.26 (.398)**	**0.9881**
Bayern	20	43 707	1 205 923	**0.929 (.294)**	**0.9574**
Saarland	17	11 946	186 402	**1.21 (.383)**	**0.9320**

Mean Zipf Coefficient across Federal States: 1.173
Median Zipf Coefficient across Federal States: 1.220

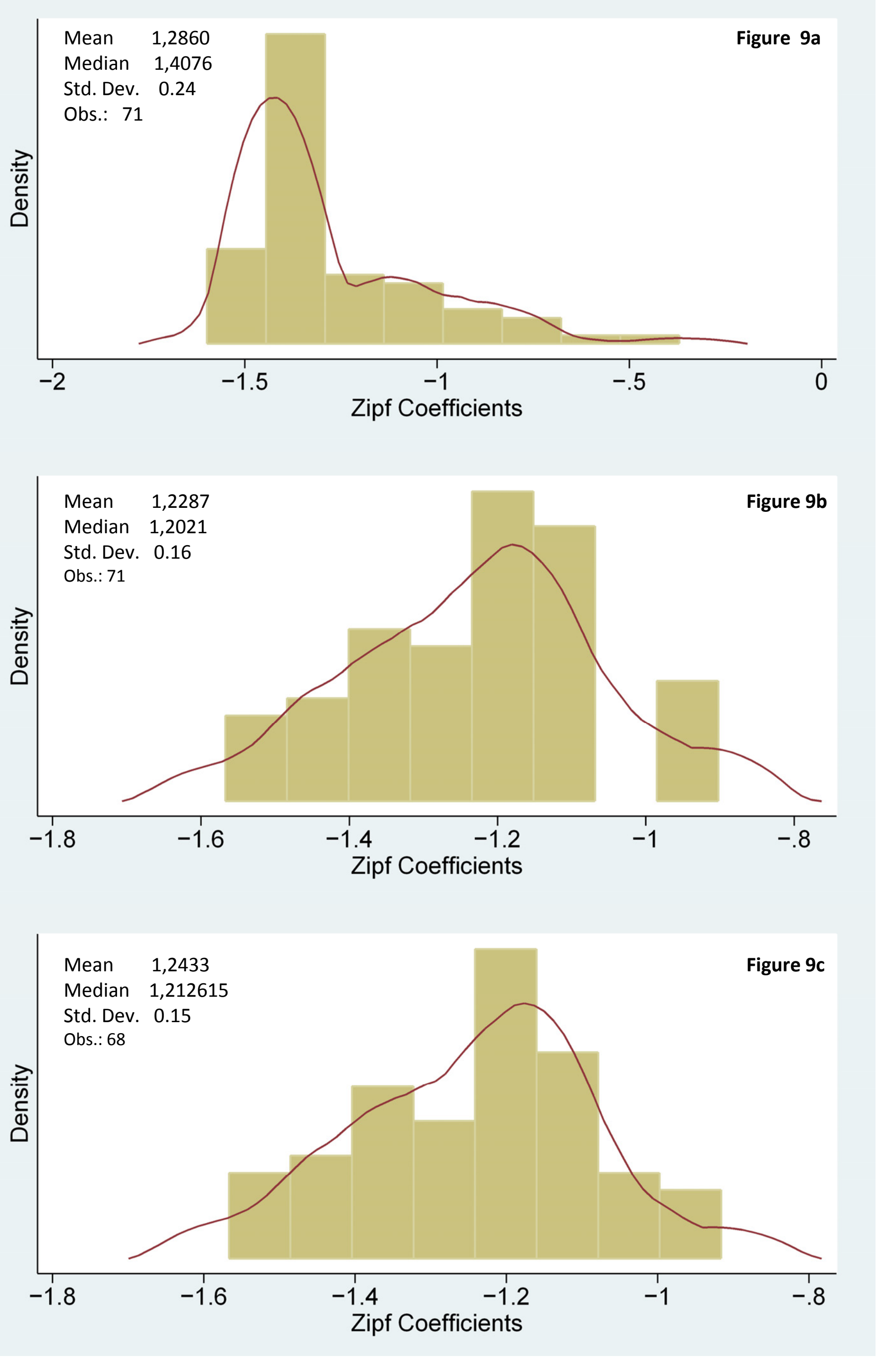
GI-corr. std. errors of the Zipf coefficients are shown in parentheses.

Figure 8) Non-Random Clubs of Cities: Zipf and Nadaraya-Watson Regression



The Gabaix-Ibragimov corrected std. errors are presented in parentheses.

Figure 9) Distribution of Zipf Coefficients from Non-Random 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 200 km (fig. 9a)

Rule 2: Select a city and include the 20 cities with the smallest distances to the selected one (fig. 9b)

Rule 3: Select a city and include the largest 20 cities within a radius of 300 km (fig. 9c)

Figures 9a-9c report the distribution of the slope coefficients -ζ across the spatial clubs. All estimated slope coefficients were highly statistically significant using GI-corr. std. errors.