THE FRENCH OVERALL CITY SIZE DISTRIBUTION

Kristian GIESEN*, Jens SÜDEKUM*

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.*

Key-words: ZIPF'S LAW, GIBRAT'S LAW, CITY SIZE DISTRIBUTIONS, DOUBLE PARETO LOGNORMAL

JEL Classification: R11, R12, O4

* Mercator School of Management, University of Duisburg-Essen, Germany;
kristian.giesen@uni-due.de ; jens.suedekum@uni-due.de.

1. 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 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 2 we introduce these data.

---

[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.

In section 3, 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, 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 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 5 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 (2012) 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

[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.

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 (2012) 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.

2. 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 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 1 shows the ten largest *communes* and their respective population sizes in 2008.

[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

Table 1. The ten largest Municipalities / Agglomerations in France, 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

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 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 3 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 (2012).

3. 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

$$P(s > S) = \left(\frac{A}{S}\right)^{\zeta}, \tag{1}$$

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 $R = N$. $P(s>S)$, so the parameters of eq. (1) can be estimated by

$$log\,(R) = K - \zeta\, log(S) \tag{2}$$

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. (2) by simple OLS. [5]

3.1. 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-size relationship is graphically illustrated in figure 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 (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 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

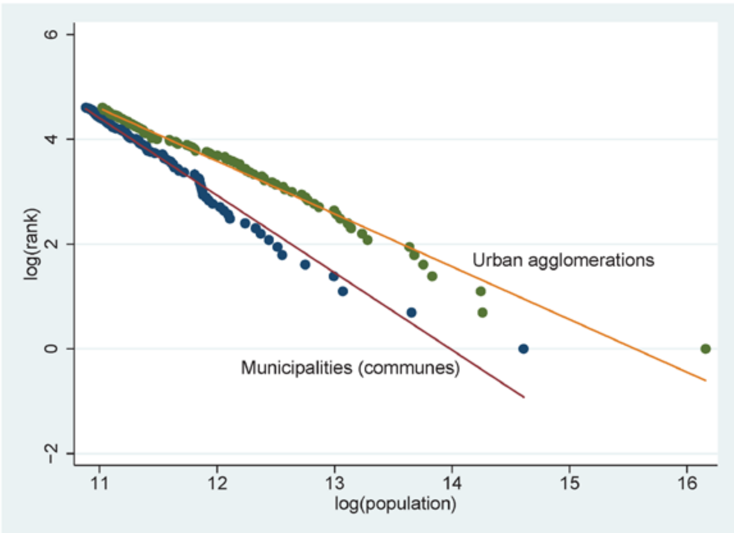
[5] There are also more sophisticated ways of estimating the Zipf coefficient, see e.g. Gabaix and Ibragimov (2011) or, for an overview, Gabaix and Ioannides (2004). However, since this is not the focus of our paper we only use the simplest and most standard rank-size regression technique.

[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.

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. (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).

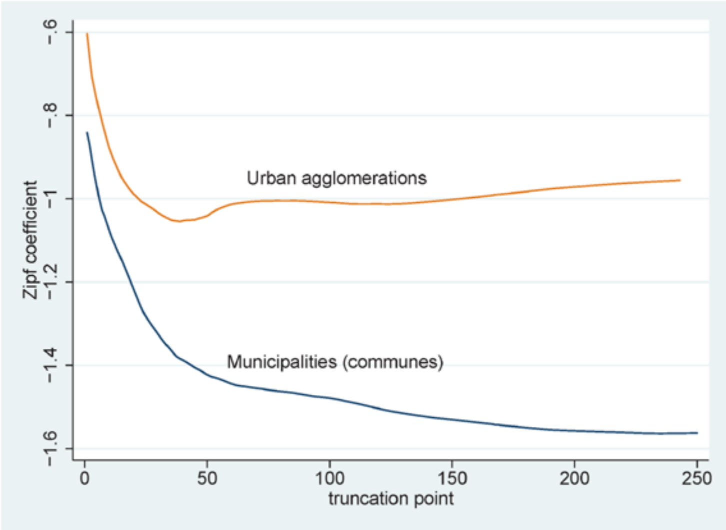
The third message is that the exact Zipf's law apparently fails to hold in the French case. The estimated slope coefficient deviates substantially from one, particularly when leaving Paris out of the regression ($\zeta = 1.576$), but also when leaving it in and using all cities ranked 1 to 100 in the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

Figure 1. Results of a Zipf regression, using the largest 100 cities



In figure 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.

Figure 2. The Zipf coefficient and the truncation point



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.

3.2. Urban areas

Before moving to this analysis of the overall CSD, we briefly consider the other data set where French cities are defined as urban agglomeration areas. In figure 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

[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.

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

4. 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 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

---

[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.

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.

4.1. 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 et al. (2010), and Figure 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.

Figure 3. Kernel density estimate of the French and the US overall CSD

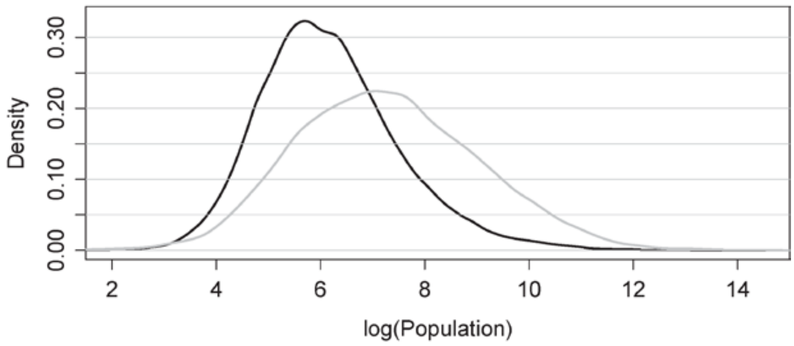
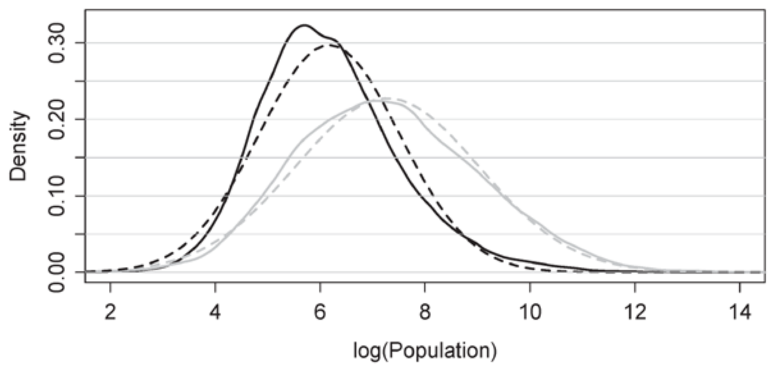


Figure 4. Kernel density estimate of the French and the US overall CSD



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 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.

4.2. 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 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 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.

[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.

5. 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, 2012) 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.

5.1. 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*:

$$f(S)=\frac{\alpha\beta}{\alpha+\beta}\left[S^{\beta-1}e^{\left(\beta\mu_0+\frac{\beta^2\sigma_0^2}{2}\right)}\Phi^c\left(\frac{log(S)-\mu_0+\beta\sigma_0^2}{\sigma_0}\right)+S^{-\alpha-1}e^{\left(\alpha\mu_0+\frac{\alpha^2\sigma_0^2}{2}\right)}\Phi\left(\frac{log(S)-\mu_0-\alpha\sigma_0^2}{\sigma_0}\right)\right]. \quad (3)$$

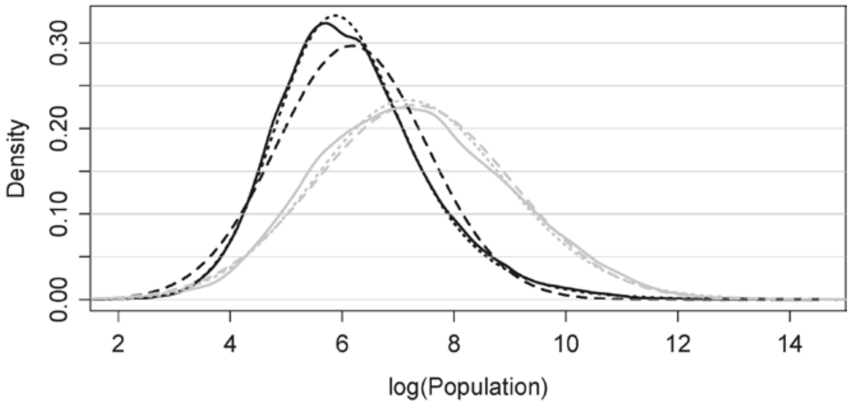
The parameters α and β are coefficients to regulate the tails, whereas μ_0 and σ_0 determine the location and the spread of the distribution. Φ represents the normal cdf and $\Phi^c = 1 - \Phi$ represents the complementary cdf. A special feature of this distribution is that if S is large, then $f(S) : S^{-\alpha-1}$ and if *S* is small, then $f(S) : S^{\beta-1}$. 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 (3) by maximum likelihood.[10] The best fit for the French data is achieved with parameters α = 1.016, β = 3.358, μ_0 = 5.588, and σ_0 = 0.882, yielding a log likelihood equal to -288,178 (also see table 2). In figure 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 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

[10] We utilize the log-likelihood function and the corresponding starting values as proposed by Reed (2002).

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.

Figure 5. Kernel density estimate and fitted LN distribution.



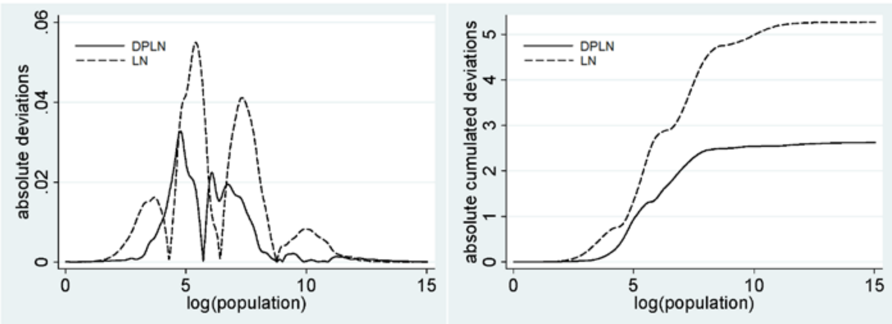
The DPLN has an advantage over the LN, because it is the more flexible functional form with four instead of two parameters. It therefore achieves a better data fit almost by definition. However, various model selection tests show that the DPLN also achieves a better *adjusted* data fit, when it is penalized for having more degrees of freedom. In particular, we use the log likelihoods of the LN and the DPLN as reported in table 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 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 $V = \frac{1}{2}\left(BIC_{DPLN} - BIC_{LN}\right)$. 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.

Table 2. Estimated parameters and formal selection tests

	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	

Figure 6. Deviations of LN and DPLN to the empirical distribution



For the US Census place data, we depict the fitted DPLN as the dotted grey line in Figure 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 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.

5.2. 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 (2012) 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

[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).

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 distribution.[12] Reed (2002) and Reed and Jorgensen (2004) have shown that the DPLN distribution as given in eq. (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 (2012) 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.

[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.

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 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 (2012) 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.

6. 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 (2012) 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 (2011) 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.

REFERENCES

Ades, A., Glaeser, E., 1995. Trade and Circuses: Explaining Urban Giants. *Quarterly Journal of Economics* 110, 195–228.

Bosker, M., Buringh, E., van Zanden, J.L., 2012. From Baghdad to London: Unravelling Urban Development in Europe and the Arab World 800-1800, forthcoming: *Review of Economics and Statistics*.

Bosker, M., Buringh, E., 2011. City Seeds: Geography and the Origins of European Cities, *CEPR Discussion Paper 8066*.

Cheshire, P., 1999. Trends in Sizes and Structures of Urban Areas. *Handbook of Regional and Urban Economics 3,* 1339-1372.

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

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

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

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.

Gabaix, X., Ioannides, Y., 2004. The Evolution of City Size Distributions. In: Henderson, V., Thisse, J. (Eds.), *Handbook of Regional and Urban economics,* Vol 4. Amsterdam: North-Holland.

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

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.

Giesen, K., Suedekum, J., 2012. The Size Distribution Across All “Cities”: A Unifying Approach. *CESifo Working Paper* 3730.

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

Henderson, J., Wang, H., 2007. Urbanization and City Growth: The Role of Institutions. *Regional Science and Urban Economics* 37, 283-313.

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

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

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

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.

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

Nitsch V., 2004. Zipf Zipped. Journal of Urban Economics 57, 86-100.

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

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

Reed, W., Jorgensen, M., 2004. The Double Pareto-Lognormal Distribution - A New Parametric Model for Size Distribution. *Communications in Statistics* 33, 1733-1753.

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.

Rozenfeld, H.D., Rybski, D., Andrade Jr., J.S., Batty, M., Stanley, H.E., Makse, H.A., 2008. Laws of Population Growth. *Proceedings of the National Academy of Sciences U.S.A.*, 105, 18702–07.

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.

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

LA DISTRIBUTION RANG-TAILLE DE TOUTES LES VILLES FRANÇAISES : UN ESSAI D'ÉVALUATION

Résumé - *Cet article propose d'étudier la distribution rang-taille de toutes les villes françaises en 2008, en incluant les plus petites localités. La distribution rang-taille des plus grandes villes françaises suit la loi de Zipf, sauf Paris qui, par son poids démographique, s'écarte de cette distribution canonique. Si l'on considère l'ensemble des villes la loi de Zipf n'est plus valide, et une distribution lognormale semble plus adaptée. Néanmoins, la meilleure adéquation est donnée par une distribution double Pareto lognormale (DPLN), ce qui correspond à des résultats obtenus sur les Etats-Unis et d'autres pays. Les implications théoriques de ce résultat sont discutées dans cet article.*

Mots-clés : DISTRIBUTION RANG-TAILLE DES VILLES, LOI DE ZIPF, LOI DE GIBRAT, DISTRIBUTION DOUBLE PARETO LOGNORMALE.