Are the log-growth rates of city sizes distributed normally? Empirical evidence for the US

Arturo Ramos

Abstract We have studied the log-growth population rate distributions of the US incorporated places (resp., all places) for the period 1990-2000 (resp. 2000-2010) and the recently constructed US City Clustering Algorithm (CCA) population data in the period 1991-2000. Also, we have considered the samples of US incorporated places that are one decade old in 1910, five decades old in 1950 and nine decades old in 1990.

An excellent parametric description of these log-growth rates is obtained by means of a newly introduced distribution called "double mixture exponential Generalized Beta 2 (dmeGB2)". The normal distribution is not the one empirically observed for the same datasets.

Keywords urban log-growth rates distribution · exponential distribution · exponential Generalized Beta 2 distribution · US population log-growth rates

  JEL: C46, R11, R12.

1 Introduction

Several studies have dealt with the theory of the growth process of cities, see, e.g., Duranton and Puga (2014). Also, several articles have dealt with the temporal dimension of city growth, sometimes finding deviations from the usual Gibrat's Law (Cuberes, 2011; Sánchez-Vidal et al., 2014; Giesen and Suedekum, 2014; Desmet and Rappaport,

A. Ramos
Dept. of Economic Analysis
Universidad de Zaragoza
Gran Vía 2, 50005-Zaragoza (SPAIN)
Tel.: +34 876 55 4784
Fax: +34 976 76 1996
E-mail: aramos@unizar.es

2015). In addition, the classical article of Ioannides and Overman (2003) studies growth rates but only in a non-parametric way, using 3D stochastic kernels. However, almost none of the published works deal with the study of the parametric description of the distribution of city growth rates, one exception being Schluter and Trede (2013). This is possibly due to the lack of good data sets in order to carry on the study until very recent times. In Ramos and Sanz-Gracia (2016) we have used some examples of this kind of data to study the city size distribution of the US, with remarkable success. Using these datasets, the computation of the log-growth rates is relatively easy so the study of their distribution is a natural subsequent task.

This research also has theoretical implications, since Gibrat's process, as it is described in Sutton (1997) and references therein, Eeckhout (2004) and Delli Gatti et al. (2005), takes the log-growth rates to be normally distributed. For another overview of Gibrat's Law see, e.g., González-Val et al. (2014). If, empirically, the former assumption happens not to hold, and moreover an alternative description for the log-growth rates is found with associated finite variances,[1] then one of the usual assumptions of Gibrat's process would deserve a reconsideration.[2]

In this article we have succeeded in parameterizing the distribution of log-growth rates with a newly introduced functional form in all the cases studied, with the so-called "double mixture exponential Generalized Beta 2 (dmeGB2)". In the estimated cases of this distribution, the variances are always finite, and almost equal to the empirical variances. This new distribution will offer a quite better performance than the normal distribution. However, in order to assess the robustness of the results with regards to other alternatives, we will analyze the Student-t and asymmetric Laplace normal distributions as well.[3]

The rest of the article is organized as follows. Section 2 describes the databases used. Section 3 introduces the parametric distributions used in this paper. Section 4 describes the empirical results obtained. Finally, in Section 5 we have offered a discussion and some conclusions.

2 The databases

We have used in this article data about US urban centers from three sources. The first is the decennial data of the US Census Bureau of "incorporated places" without any size restriction, for the period 1900-2000. These include governmental units classified under state laws such as cities, towns, boroughs or villages. Alaska, Hawaii and Puerto Rico have not been considered due to data limitations. The data has been collected from the original documents of the annual census published by the US

---

[1] The assumption of the finite variances for the log-growth rates is essential for the application of the standard Central Limit Theorem, rather than the assumption that the log-growth rates are normal. For alternative Central Limit Theorems when studying city size, see, e.g., Lee and Li (2013).

[2] In this article we are not testing whether the city size distribution is lognormal, something implied if Gibrat's Law is fulfilled (Eeckhout, 2004). That is investigated in other articles, like for example Giesen et al. (2010); González-Val et al. (2015); Ramos and Sanz-Gracia (2016).

[3] It is worth recalling that Schluter and Trede (2013) considered a model with the conclusion that the normalized growth city size distribution of German cities follows a Student-t.

Census Bureau.[4] These data sets were first introduced in González-Val (2010), see therein for details, and later used in other works like Sánchez-Vidal et al. (2014); González-Val et al. (2015); Ramos and Sanz-Gracia (2016). For the sake of brevity in this paper, we will consider the necessary data for constructing the 1990-2000 log-growth rates of incorporated places. In addition, we will take directly the one decade of age in 1910 (d1 1910), five decades of age in 1950 (d5 1950) and nine decades of age in 1990 (d9 1990) samples of the database constructed for the article Sánchez-Vidal et al. (2014) in order to perform a robustness check of our approach with regards to the age of cities in the case of the US incorporated places during the 20th century. We have taken three representative samples with as high as possible number of observations according to the whole database.

The second source consists of all US urban places, unincorporated and incorporated, and without size restrictions, also provided by the US Census Bureau for the years 2000 and 2010. The data for the year 2000 was first used in Eeckhout (2004) and later in Levy (2009), Eeckhout (2009), Giesen et al. (2010), Ioannides and Skouras (2013) and Giesen and Suedekum (2014). The two samples were also used in González-Val et al. (2015); Ramos and Sanz-Gracia (2016).

The third comes from a different and recent approach to defining city centers, described in detail in Rozenfeld et al. (2008, 2011). They use a so called "City Clustering Algorithm" (CCA) to get "an automated and systematic way of building population clusters based on the geographical location of people." (*op. cit.*) We have used their US clusters data based on the radius of 2 km. and for the years 1991 and 2000. Data sets of this type have been used in Ioannides and Skouras (2013) and Giesen and Suedekum (2014).

[Table 1 near here]

We have offered in Table 1 the descriptive statistics of the used log-growth data for the US.

3 Description of the distributions presented

In this section we will introduce the distributions used along the paper[5] for the (two consecutive periods) log-growth rates, denoted by

$$g_{i,t} = \log x_{i,t} - \log x_{i,t-1} \in (-\infty, \infty)$$

---

[4] http://www.census.gov/prod/www/decennial.html Last accessed: June 24th, 2016.

[5] From a practical point of view, it is our interest in this paper to obtain a very good parametric fit of the log-growth rate distributions. For that, we have first tried several well-known distributions in the economics literature: the normal, the asymmetric exponential power (AEP) of Bottazzi and Secchi (2011), which generalizes the Laplace distribution of, e.g., Johnson et al. (1995), Stanley et al. (1996) and references therein, the α-stable distribution, see, e.g., Zolotarev (1986); Uchaikin and Zolotarev (1999) and references therein (the calculations for the α-stable distribution have been performed using the STABLE software of Robust Analysis Inc., see http://www.robustanalysis.com/) the generalized hyperbolic distribution (Barndorff-Nielsen, 1977; Barndorff-Nielsen and Halgreen, 1977; Barndorff-Nielsen and Stelzer, 2005), the (non-standardized) Student-t distribution, see, e.g., Johnson et al. (1995) and references therein, and the asymmetric double Laplace normal (Reed, 2002, 2003; Reed and Jorgensen, 2004; Manas, 2009). The results for the distributions not presented here are available from the author upon request.

where $x_{i,t}$ is the city population i at time t. When a fixed t is taken we will simply write $g \in (-\infty, \infty)$ for the variable of all log-growth rates of the cross-sections taken.

3.1 Normal distribution

Firstly, we have recalled the normal distribution for the log-growth rates g. We thus set

$$f_{\mathrm{n}}(g, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(g-\mu)^2}{2\sigma^2}\right)$$

where μ is real and $\sigma > 0$ are, respectively, the mean and the standard deviation of the variable g according to this distribution.

3.2 The non-standarized Student-t distribution

The non-standardized Student-t distribution for the log-growth rates g, see, e.g., Johnson et al. (1995) and references therein, is given by the following probability density function

$$f_{\mathrm{Stu}}(g, \mu, \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma} \left(1 + \frac{1}{\nu}\left(\frac{g-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$$

where $\mu \in \mathbb{R}$ (location parameter), $\sigma > 0$ (scale parameter), $\nu > 0$ is the number of degrees of freedom, and $\Gamma(\cdot)$ denotes the Gamma function. Particular cases of this distribution are the Cauchy distribution ($\nu = 1$) and the normal distribution ($\nu = \infty$). If $1 < \nu \leq 2$ the variance of the distribution becomes infinite. This distribution has been used to study city size log-growth rates of Germany by Schluter and Trede (2013).

3.3 The asymmetric double Laplace normal (adLn)

The third distribution in our study will be the asymmetric double Laplace normal distribution (adLn), introduced by (Reed, 2002, 2003; Reed and Jorgensen, 2004) and later used, e.g., by Manas (2009):

$$
\begin{aligned}
&f_{\mathrm{adLn}}(g, \alpha, \beta, \mu, \sigma) \\
&= \frac{\alpha\beta}{2(\alpha+\beta)} \exp\left(\alpha\mu + \frac{\alpha^2\sigma^2}{2}\right) \mathrm{e}^{-\alpha g}\left(1 + \mathrm{erf}\left(\frac{g-\mu-\alpha\sigma^2}{\sqrt{2}\sigma}\right)\right) \\
&\quad - \frac{\alpha\beta}{2(\alpha+\beta)} \exp\left(-\beta\mu + \frac{\beta^2\sigma^2}{2}\right) \mathrm{e}^{\beta g}\left(\mathrm{erf}\left(\frac{g-\mu+\beta\sigma^2}{\sqrt{2}\sigma}\right) - 1\right)
\end{aligned}
$$

where erf is the error function associated to the normal distribution and $\mu \in \mathbb{R}$, $\alpha, \beta, \sigma > 0$ are the four parameters of the distribution. It has the property that it approximates different exponential laws in each of its two tails: $f_{\mathrm{adLn}}(g) \approx \mathrm{e}^{-\alpha g}$

when $g \to \infty$ and $f_{\mathrm{adLn}}(g) \approx \mathrm{e}^{\beta g}$ when $g \to -\infty$. The body is approximately normal, although it is not possible to exactly delineate the switch between the normal and the exponential behaviors since the adLn distribution is the convolution of an asymmetric double Laplace with a normal distribution.

3.4 The double mixture exponential Generalized Beta 2 (dmeGB2)

For our new distribution "double mixture exponential Generalized Beta 2" we have first defined some basic functions which will be employed by the former.

Then, let us consider

$$f_{\mathrm{eGB2}}(g, a, b, p, q) = \frac{a \exp((g - b)ap)}{B(p, q)\left(1 + \exp(a(g - b))\right)^{p+q}}$$

$$\mathrm{cdf}_{\mathrm{eGB2}}(g, a, b, p, q) = \frac{1}{B(p, q)} B\left(\frac{\exp(a(g - b))}{1 + \exp(a(g - b))}, p, q\right)$$

$$u(g, \zeta) = \exp(-\zeta g)$$

$$l(g, \rho) = \exp(\rho g)$$

The f_{eGB2} ($\mathrm{cdf}_{\mathrm{eGB2}}$) is the exponential version of the Generalized Beta of the second kind density (resp., cumulative distribution function, cdf) (McDonald, 1984; McDonald and Xu, 1995; Kleiber and Kotz, 2003),

$$B(z, p, q) = \int_0^z t^{p-1}(1 - t)^{q-1}\, dt\,, \quad z \in [0, 1]$$

is the incomplete Beta function and $B(p, q) = B(1, p, q)$ is the Beta function. The three parameters a, p, q are positive shape parameters and $b \in \mathbb{R}$ is a location parameter. The function $u(g, \zeta)$ will model the decreasing exponential part of the upper tail of our new distribution, where $\zeta > 0$, and $l(g, \rho)$ corresponds to the increasing exponential lower tail, with $\rho > 0$. The functions u, l are not normalized at this stage as in Ioannides and Skouras (2013). Note that if the variable x follows a Pareto distribution and $y = \ln x$, then y follows an exponential distribution.

The new distribution we have introduced here, which yields the best results out of the ones we have dealt with, has two tails which are exponential with a mixture of exponential Generalized Beta 2, and body of this last type. The switch between the tails and the body occurs at two exact thresholds ϵ (lower tail-body) and $\tau > \epsilon$ (body-upper tail). For the lower tail, the combining coefficient will be denoted by $\nu \in (0, 1)$, and by $\theta \in (0, 1)$ for the upper tail. We require continuity of the density function at the threshold points and overall normalization to one. Equal weight of the distributions of the mixing at the tails are also imposed, as in Ioannides and Skouras (2013), so that the parameters ν, θ control the proportion of each component of the combination in the lower (resp. upper) tail.

The resulting composite density is given by:

$$f_{\mathrm{dmeGB2}}(g, \rho, \epsilon, \nu, a, b, p, q, \tau, \zeta, \theta)$$
$$= \begin{cases} b_2[(1-\nu)\, d_2\, f_{\mathrm{eGB2}}(g, a, b, p, q) + \nu\, e_2\, l(g, \rho)] & g < \epsilon \\ b_2\, f_{\mathrm{eGB2}}(g, a, b, p, q) & \epsilon \le g \le \tau \\ b_2[(1-\theta)\, c_2\, f_{\mathrm{eGB2}}(g, a, b, p, q) + \theta\, a_2\, u(g, \zeta)] & \tau < g \end{cases}$$

where the constants (i.e., quantities that do not depend on the variable g) are given as follows:

$$d_2^{-1} = 1 - \nu + \frac{\exp(-\rho\epsilon)\,\nu\,\rho\,\mathrm{cdf}_{\mathrm{eGB2}}(\epsilon, a, b, p, q)\,l(\epsilon, \rho)}{f_{\mathrm{eGB2}}(\epsilon, a, b, p, q)}$$

$$e_2^{-1} = \frac{(1-\nu)\,\exp(\epsilon\rho)}{\rho\,\mathrm{cdf}_{\mathrm{eGB2}}(\epsilon, a, b, p, q)} + \frac{\nu\,l(\epsilon, \rho)}{f_{\mathrm{eGB2}}(\epsilon, a, b, p, q)}$$

$$c_2^{-1} = 1 - \theta + \frac{\zeta\,\theta\,\exp(\tau\zeta)\,(1 - \mathrm{cdf}_{\mathrm{eGB2}}(\tau, a, b, p, q))\,u(\tau, \zeta)}{f_{\mathrm{eGB2}}(\tau, a, b, p, q)}$$

$$a_2^{-1} = \frac{(1-\theta)\,\exp(-\tau\zeta)}{\zeta\,(1 - \mathrm{cdf}_{\mathrm{eGB2}}(\tau, a, b, p, q))} + \frac{\theta\,u(\tau, \zeta)}{f_{\mathrm{eGB2}}(\tau, a, b, p, q)}$$

$$b_2^{-1} = e_2\,\frac{\exp(\epsilon\rho)}{\rho} + \mathrm{cdf}_{\mathrm{eGB2}}(\tau, a, b, p, q) - \mathrm{cdf}_{\mathrm{eGB2}}(\epsilon, a, b, p, q) + \frac{a_2}{\zeta\,\exp(\tau\zeta)}$$

This distribution depends on ten parameters $(\rho, \epsilon, \nu, a, b, p, q, \tau, \zeta, \theta)$ to be estimated below by Maximum Likelihood (ML). Also, this distribution can be obtained in an exact way from an optimization model similar to that of Ramos and Sanz-Gracia (2016); we enclose a MATHEMATICA® notebook with the main optimization equations as supplementary material. The model is based heavily on a previous model by Parker (1999).

4 Results

In this Section we have recalled briefly the empirical results concerning the US samples on use.

We have estimated the studied distributions by the method of Maximum Likelihood (ML), using the software MATLAB® and MATHEMATICA®. We have reported on Tables 2, 3 and 4 the estimated values of the parameters for the Student-t, the adLn and dmeGB2 and the corresponding standard errors (SE) computed according to Efron and Hinkley (1978) and McCullough and Vinod (2003). The ML estimators for the parameters of the normal distribution are exact, being the mean and standard deviation of each empirical data sample, see simply Table 1. We can see that the estimations are rather precise in almost all cases, possibly except the estimations of ϵ for the dmeGB2 in the samples of d1 1910, d5 1950 and d9 1990. Also, we have not been able to estimate the adLn for the sample of all US places (2000-2010).

[Table 2 near here]

[Table 3 near here]

[Table 4 near here]

We have computed numerically as well the means and the standard deviations (SD) of the variable g according to the estimated Student-t, adLn and dmeGB2 distributions, which are shown in Table 5. From it, we can observe that the computed means and standard deviations for the adLn are very similar to the empirical ones, and those for the dmeGB2 are almost identical. More importantly, the computed standard deviations according to the dmeGB2 are always finite. In contrast, the estimated Student-t for the sample of all US places (2000-2010) has infinite standard deviation.

[Table 5 near here]

In order to assess the goodness of fit of the four distributions explicitly shown in this paper, we have used three standard statistical tests: the Kolmogorov–Smirnov (KS) test, the Crámer–von Mises (CM) test and the Anderson–Darling (AD) test. These test are very powerful when the sample size is as high as in the cases of this article (Razali and Wah, 2011) and the last one is particularly useful when one wants to see the adequacy of the distribution at the tails, see, e.g., Cirillo (2013). The results are shown on Table 6. Very briefly, the normal and Student-t distributions are *strongly* rejected always by the three tests. The adLn is not rejected by the three tests for the samples d1 1910, d5 1950 but otherwise always rejected. Meanwhile, the dmeGB2 is not rejected in 100% of the cases, and not by a small margin precisely.

[Table 6 near here]

Additionally, we have computed more metrics allowing to select the best one amongst the hypothesized distributions, namely the msd and the pseudo R^2 quantities adapted from Duranton (2007) to this particular case (we simply replace the log-variable by the variable under study):

$$\text{msd} = \frac{1}{m} \sum_{j=1}^{m} [\text{Actual log growth rate}(j)$$
$$-\text{Mean Simulated log growth rate}(j)]^2 \qquad (1)$$
$$R^2 = 1 - \frac{\text{msd}}{\text{var}} \qquad (2)$$

where var is the empirical variance for log-growth rates and m is the number of observations in the empirical sample.

For the msd and R^2 quantities, we have generated 100 random samples[6] and the results are shown in Table 7. From it, it is clear that the dmeGB2 provides a much better fit than the other distributions.[7]

[Table 7 near here]

Also, we have computed the Akaike Information Criterion (AIC) and Bayesian or Schwarz Information Criterion (BIC) (Burnham and Anderson, 2002, 2004), very

---

[6] Each of these samples is of the sample size of the empirical data. The total generated observations range from about 298,000 to 3,020,000 depending on the case under study and we hope the results to be statistically significant. We have chosen a number of generated samples reasonably high enough while maintaining computational feasibility.

[7] The pseudo-R^2 becomes negative for the Student-t and the sample of all US places (2000-2010). This is because the estimated Student-t in this case has infinite standard deviation and the generated samples according to the former induce a huge msd quantity.

well adapted to the maximum likelihood estimation we have performed before. For the computed AIC and BIC see the Table 8.

By these two different types of criteria we can see that the dmeGB2 greatly outperforms the other distributions when considering the log-growth rates of US city sizes for all the used samples, in spite of the fact that our new distribution depends on ten parameters instead of less parameters of the other distributions. In particular, the dmeGB2 is much better than the normal distribution for describing these log-growth rates.

[Table 8 near here]

As a complement of the KS, CM, AD, msd, pseudo-R^2, AIC and BIC criteria, we have shown in Figure 1 an informal graphical approximation of the obtained fits for three of the used samples. We can observe excellent fits with small deviations, if any, at the tails (the deviations at the tails are subject to an amplification effect, see, e.g., González-Val et al. (2013)). However, the overall fit of the densities is visually excellent. Let us remark that on the plots of the tails the cdf for the lower tail or $1 - \text{cdf}$ for the upper tail are nearly exponential, and therefore the graphs are almost linear, in agreement with previous knowledge (Johnson et al., 1995; Stanley et al., 1996; Bottazzi and Secchi, 2011).

[Figure 1 near here]

5 Discussion and conclusions

In the preceding Section we have seen that a very appropriate parametric model for the log-growth rate distribution of the city size of the US is the newly introduced (in Subsection 3.4) dmeGB2.

In our opinion, the excellent parametric fit of this distribution is by itself a significant advance of the theory of the growth of city sizes.

Likewise, the normal distribution for the log-growth rates is clearly rejected empirically in all our samples, so one of the assumptions of the Gibrat's process (see, e.g., Sutton (1997) and references therein, Eeckhout (2004) and Delli Gatti et al. (2005)) may not hold, and it deserves a reconsideration.

The variances given by the dmeGB2 in all our cases of study are finite, so we have found an example of distribution for the log-growth rates of city size for the US, always not rejected empirically and with finite variances. This is an alternative to the normal distribution.

This does not mean that other assumptions of Gibrat's process do not hold in principle. On the contrary, more research can be done in this respect, see the recent article Ramos (2015).

However, one might wonder how the city size distribution would be under the log-growth rates behaving in the described way. We should recall first that regarding the US city size distribution the recent work of Ramos and Sanz-Gracia (2016) has appeared. In it, an empirically observed distribution ("threshold double Pareto Generalized Beta 2 (tdPGB2)" for places is derived in an exact way from a theoretical model of maximization of the net output of the system of cities in the US. In this

model, the resulting city size distribution has parameters that depend on the elasticities of the production function with respect to the population, to the number of cities, and of the congestion costs with respect to population. The threshold parameters separating the lower tail, the body and the upper tail are endogenously determined by the ML estimation procedure. In this way an economic model generates exactly the city size distribution. Time changes in the mentioned elasticities induce time changes in the city size distribution, and therefore log-growth rates are generated. It is striking that the observed US city log-size distributions and log-growth rates have distributions of the *same* family dmeGB2, according to Ramos and Sanz-Gracia (2016) and this paper.

Additionally, the usual Gibrat's process takes the log-growth rates to be normal, and then, the resulting log-sizes are obtained adding up the initial log-sizes and the consecutive log-growth rates. If the log-growth rates are identically and independently distributed (i.i.d.) as normal distributions, the resulting distribution for log-sizes will be normal as well, because the *convolution* of normal distributions is again a normal distribution, in an *exact* way.

The class of distributions which are *exactly* closed under convolution is known to consist of the α-stable ones, see, e.g., Zolotarev (1986); Uchaikin and Zolotarev (1999) and references therein. Except in the case of $\alpha = 2$, which corresponds to the normal distribution, these distributions have the problem of possessing infinite variances.

In this article we have proved empirically that the log-growth rates of US city sizes do not follow a normal distribution and a better model than all of the previously used distributions has been proposed successfully, namely the dmeGB2. This distribution has finite variances for the cases studied (and almost equal to the empirical ones). Since the observed distributions for US city log-sizes and log-growth rates are so similar in practical terms, this leads us to think that the new family of distributions dmeGB2 may be *approximately* closed under convolution in a sense that has to be made precise in statistical terms.[8] If this property holds, it might be a link between the observed log-growth rates and the resulting observed city log-size distribution, with a similar rationale of the standard process of generating a lognormal distribution for city sizes out of normal distributions for log-growth rates.

All these implications are to be compared with current theories of urban growth, see, e.g., Duranton and Puga (2014); the review of this work from this new perspective will probably shed new light into the determinants of city size and city growth.

In conclusion, our research complements that of other authors and possibly opens a new avenue for further investigations.

---

[8] In another paper or papers as it is out of the scope of the present article.

References

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 353(1674):401–419.

Barndorff-Nielsen, O. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 38:309–311.

Barndorff-Nielsen, O. and Stelzer, R. (2005). Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian Lévy processes. *Scandinavian Journal of Statistics*, 32:617–637.

Bottazzi, G. and Secchi, A. (2011). A new class of asymmetric exponential power densities with applications to economics and finance. *Industrial and Corporate Change*, 20(4):991–1030.

Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer-Verlag.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33:261–304.

Cirillo, P. (2013). Are your data really Pareto distributed? *Physica A*, 392:5947–5962.

Cuberes, D. (2011). Sequential city growth: Empirical evidence. *Journal of Urban Economics*, 69:229–239.

Delli Gatti, D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M., and Palestrini, A. (2005). A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic Behavior & Organization*, 56:489–512.

Desmet, K. and Rappaport, J. (2015). The settlement of the United States, 1800 to 2000: The long transition towards Gibrat's Law. *Journal of Urban Economics. Forthcoming.* http://dx.doi.org/10.1016/j.jue.2015.03.004.

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

Duranton, G. and Puga, D. (2014). The growth of cities. In Durlauf, S. N. and Aghion, P., editors, *Handbook of Economic Growth*. North-Holland.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451.

Eeckhout, J. (2009). Gibrat's law for (all) cities: Reply. *American Economic Review*, 99:1676–1683.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65(3):457–482.

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

Giesen, K., Zimmermann, A., and Suedekum, J. (2010). The size distribution across all cities-double Pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137.

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.

González-Val, R., Lanaspa, L., and Sanz-Gracia, F. (2014). New evidence on Gibrat's Law for cities. *Urban Studies*, 51(1):93–115.

González-Val, R., Ramos, A., and Sanz-Gracia, F. (2013). The accuracy of graphs to describe size distributions. *Applied Economics Letters*, 20(17):1580–1585.

González-Val, R., Ramos, A., Sanz-Gracia, F., and Vera-Cabello, M. (2015). Size distribution for all cities: Which one is best? *Papers in Regional Science*, 94(1):177–197.

Ioannides, Y. M. and Overman, H. G. (2003). Zipf's law for cities: An empirical examination. *Regional Science and Urban Economics*, 33(2):127–137.

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

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions. Volume 2*. John Wiley & Sons.

Kleiber, C. and Kotz, S. (2003). *Statistical size distributions in Economics and actuarial sciences*. Wiley-Interscience.

Lee, S. and Li, Q. (2013). Uneven landscapes and city size distributions. *Journal of Urban Economics*, 78:19–29.

Levy, M. (2009). Gibrat's law for (all) cities: Comment. *American Economic Review*, 99:1672–1675.

Manas, A. (2009). French butchers don't do quantum physics. *Economics Letters*, 103:101–106.

McCullough, B. D. and Vinod, H. D. (2003). Verifying the solution from a nonlinear solver: A case study. *American Economic Review*, 93(3):873–892.

McDonald, J. B. (1984). Generalized functions for the size distribution of income. *Econometrica*, 52(3):647–665.

McDonald, J. B. and Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, 66:133–152.

Parker, S. C. (1999). The generalised beta as a model for the distribution of earnings. *Economics Letters*, 62:197–200.

Ramos, A. (2015). Log-growth distributions of US city sizes and non-Lévy processes. *Working Paper, available at Munich RePec* `http://mpra.ub.uni-muenchen.de/66561/`.

Ramos, A. and Sanz-Gracia, F. (2016). US city size distribution revisited: Theory and empirical evidence. *Working Paper, available at Munich RePec* `http://mpra.ub.uni-muenchen.de/71928/`.

Razali, N. M. and Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2:21–33.

Reed, W. J. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science*, 42:1–17.

Reed, W. J. (2003). The Pareto law of incomes–an explanation and an extension. *Physica A*, 319:469–486.

Reed, W. J. and Jorgensen, M. (2004). The double Pareto-lognormal distribution–a new parametric model for size distributions. *Communications in Statistics-Theory and Methods*, 33(8):1733–1753.

Rozenfeld, H. D., Rybski, D., Andrade, J. S., Batty, M., Stanley, H. E., and Makse, H. A. (2008). Laws of population growth. *Proceedings of the National Academy of Sciences*, 105(48):18702–18707.

Rozenfeld, H. D., Rybski, D., Gabaix, X., and Makse, H. A. (2011). The area and population of cities: new insights from a different perspective on cities. *American Economic Review*, 101:2205–2225.

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

Schluter, C. and Trede, M. (2013). Gibrat, Zipf, Fisher and Tippet: City size and growth distributions reconsidered. *Working Paper 27/2013 Center for Quantitative Economics WWU Münster*.

Stanley, M. H. R., Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A., and Stanley, H. E. (1996). Scaling behaviour in the growth of companies. *Nature*, 379:804–806.

Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature*, 35:40–59.

Uchaikin, V. V. and Zolotarev, V. M. (1999). *Chance and stability. Stable distributions and their applications*. VSP.

Zolotarev, V. M. (1986). *One-dimensional stable distributions*. American Mathematical Society.

Table 1: Descriptive statistics of the log-growth rates for the used samples

Sample	Obs	Mean	SD	Min	Max
Ip 1990-2000	19,048	0.075	0.262	-4.467	3.581
Ap 2000-2010	24,685	0.035	0.282	-5.278	6.075
CCA 1991-2000 (2 km)	30,201	0.105	0.156	-2.398	3.773
d1 1910	3,291	0.186	0.415	-1.914	3.723
d5 1950	3,088	0.047	0.312	-2.398	2.705
d9 1990	2,987	0.056	0.261	-1.580	3.581

Table 2: ML estimators and standard errors (SE) for the Student-t and the studied log-growth rate samples. The estimators for the normal distribution are the mean and standard deviation of the log-growth data, see Table 1

Sample	Student-t		
	μ (SE)	σ (SE)	ν (SE)
Ip 1990-2000	0.040 (0.001)	0.124 (0.001)	2.123 (0.033)
Ap 2000-2010	0.004 (0.001)	0.107 (0.001)	1.804 (0.022)
CCA 1991-2000 (2 km)	0.083 (0.001)	0.097 (0.001)	3.426 (0.053)
d1 1910	0.148 (0.006)	0.263 (0.005)	3.001 (0.135)
d5 1950	0.014 (0.004)	0.179 (0.003)	2.680 (0.115)
d9 1990	0.031 (0.003)	0.135 (0.003)	2.394 (0.098)

Table 3: ML estimators and standard errors (SE) for the asymmetric double Laplace normal (adLn) and the studied log-growth rate samples.

Sample	adLn			
	a (SE)	b (SE)	μ (SE)	σ (SE)
Ip 1990-2000	5.114 (0.040)	8.519 (0.079)	-0.003 (0.001)	0.009 (0.006)
Ap 2000-2010	–	–	–	–
CCA 1991-2000 (2 km)	7.755 (0.048)	19.465 (0.180)	0.027 (0.001)	0.038 (0.001)
d1 1910	3.005 (0.058)	4.679 (0.106)	0.067 (0.005)	0.074 (0.012)
d5 1950	4.197 (0.084)	6.537 (0.154)	-0.039 (0.004)	0.055 (0.008)
d9 1990	5.230 (0.105)	7.851 (0.182)	-0.008 (0.003)	0.029 (0.007)

Table 4: ML estimators and standard errors (SE) for the dmeGB2 and the studied log-growth rate samples.

Sample	dmeGB2			
	ρ (SE)	ϵ (SE)	ν (SE)	
Ip 1990-2000	2.69 (0.13)	-0.000 (0.014)	0.082 (0.006)	
Ap 2000-2010	1.80 (0.08)	-0.000 (0.012)	0.054 (0.003)	
CCA 1991-2000 (2 km)	3.04 (0.32)	-0.121 (0.004)	0.265 (0.023)	
d1 1910	3.27 (0.15)	0.000 (0.026)	0.490 (0.034)	
d5 1950	1.79 (0.31)	0.000 (0.080)	0.032 (0.010)	
d9 1990	2.89 (0.33)	0.000 (0.027)	0.085 (0.017)	
	a (SE)	b (SE)	p (SE)	q (SE)
Ip 1990-2000	34.79 (0.27)	-0.006 (0.001)	0.327 (0.003)	0.193 (0.002)
Ap 2000-2010	54.61 (0.38)	-0.017 (0.001)	0.192 (0.002)	0.150 (0.001)
CCA 1991-2000 (2 km)	20.40 (0.10)	-0.019 (0.001)	1.581 (0.012)	0.432 (0.003)
d1 1910	11.92 (0.31)	-0.067 (0.006)	0.911 (0.038)	0.199 (0.006)
d5 1950	22.17 (0.55)	-0.080 (0.005)	0.353 (0.009)	0.122 (0.005)
d9 1990	63.83 (1.91)	-0.041 (0.004)	0.155 (0.005)	0.038 (0.002)
	τ (SE)	ζ (SE)	θ (SE)	
Ip 1990-2000	0.31 (0.01)	2.57 (0.08)	0.51 (0.02)	
Ap 2000-2010	0.17 (0.01)	2.45 (0.06)	0.46 (0.01)	
CCA 1991-2000 (2 km)	-0.02 (0.02)	2.15 (0.21)	0.02 (0.04)	
d1 1910	0.36 (0.03)	5.09 (0.30)	0.38 (0.04)	
d5 1950	0.16 (0.01)	7.94 (0.37)	0.49 (0.03)	
d9 1990	0.11 (0.01)	7.91 (0.26)	0.74 (0.02)	

Table 5: Means and standard deviations (SD) according to the estimated distributions and the studied log-growth rate samples. Those corresponding to the normal distribution are exactly the mean and standard deviations of the empirical samples. Compare with the values in Table 1.

Sample	Student-t		adLn		dmeGB2	
	Mean	SD	Mean	SD	Mean	SD
Ip 1990-2000	0.040	0.514	0.075	0.228	0.075	0.260
Ap 2000-2010	0.004	∞	–	–	0.035	0.273
CCA 1991-2000 (2 km)	0.083	0.150	0.105	0.144	0.105	0.155
d1 1910	0.148	0.456	0.186	0.402	0.186	0.415
d5 1950	0.014	0.355	0.047	0.288	0.047	0.312
d9 1990	0.031	0.332	0.056	0.232	0.056	0.261

Table 6: p-values (statistics) of the Kolmogorov–Smirnov (KS), Cramér–Von Mises (CM) and Anderson–Darling (AD) tests for the used samples and density functions. Non-rejections at the 5% significance level are marked in bold

Sample	normal			Student-t		
	KS	CM	AD	KS	CM	AD
Ip 1990-2000	0 (0.129)	0 (109.391)	0 (614.161)	0 (0.042)	0 (9.401)	0 (90.109)
Ap 2000-2010	0 (0.154)	0 (204.089)	0 (1132.13)	0 (0.037)	0 (9.897)	0 (87.714)
CCA 1991-2000 (2 km)	0 (0.097)	0 (97.453)	0 (602.344)	0 (0.054)	0 (16.381)	0 (185.934)
d1 1910	0 (0.086)	0 (9.103)	0 (52.647)	0.002 (0.033)	0.004 (0.933)	0 (9.300)
d5 1950	0 (0.101)	0 (11.783)	0 (68.490)	0.010 (0.030)	0.010 (0.758)	0 (8.568)
d9 1990	0 (0.120)	0 (16.284)	0 (91.614)	0.003 (0.033)	0.010 (0.764)	0 (7.566)
	adLn			dmeGB2		
	KS	CM	AD	KS	CM	AD
Ip 1990-2000	0 (0.029)	0 (3.450)	0 (24.177)	**0.770 (0.005)**	**0.588 (0.099)**	**0.248 (1.253)**
Ap 2000-2010	–	–	–	**0.689 (0.005)**	**0.734 (0.073)**	**0.678 (0.569)**
CCA 1991-2000 (2 km)	0.017 (0.010)	0.008 (0.779)	0 (8.315)	**0.798 (0.004)**	**0.886 (0.048)**	**0.927 (0.314)**
d1 1910	**0.543 (0.014)**	**0.620 (0.093)**	**0.405 (0.914)**	**0.9979 (0.007)**	**0.9997 (0.015)**	**0.9999 (0.103)**
d5 1950	**0.119 (0.022)**	**0.218 (0.229)**	**0.086 (2.053)**	**0.9307 (0.010)**	**0.9523 (0.036)**	**0.9906 (0.199)**
d9 1990	0.013 (0.030)	0.046 (0.474)	0.012 (3.727)	**0.9737 (0.009)**	**0.9883 (0.026)**	**0.9933 (0.188)**

Table 7: Values of the msd (in units of 10^{-3}) and of the pseudo R^2 inspired by Duranton (2007) for the used samples and distributions. The most favoured values are marked in bold. From the definition of the pseudo R^2 in Eq. (2) it may happen that it becomes negative if msd is huge.

Sample	normal		Student-t	
	msd	R^2	msd	R^2
Ip 1990-2000	13.12	0.8095	25.36	0.6319
Ap 2000-2010	22.66	0.7160	130.22	-0.6325
CCA 1991-2000 (2 km)	4.44	0.8168	2.15	0.9115
d1 1910	13.11	0.9239	11.33	0.9342
d5 1950	12.03	0.8763	8.46	0.9130
d9 1990	11.60	0.8298	8.12	0.8810
Sample	adLn		dmeGB2	
	msd	R^2	msd	R^2
Ip 1990-2000	4.46	0.9352	**0.64**	**0.9907**
Ap 2000-2010	–	–	**1.00**	**0.9874**
CCA 1991-2000 (2 km)	1.57	0.9352	**0.09**	**0.9963**
d1 1910	1.11	0.9935	**0.18**	**0.9990**
d5 1950	2.89	0.9703	**0.16**	**0.9984**
d9 1990	3.87	0.9432	**0.52**	**0.9924**

Table 8: Maximum log-likelihoods, AIC and BIC for the used distributions and log-growth rates samples. The lowest values of AIC and BIC for each sample are marked in boldface

| Sample	normal			Student-t		
	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
Ip 1990-2000	-1,548	3,100	3,116	3,067	-6,129	-6,105
Ap 2000-2010	-3,817	7,638	7,655	5,244	-10,482	-10,458
CCA 1991-2000 (2 km)	13,302	-26,600	-26,584	18,284	-36,561	-36,536
d1 1910	-1,775	3,554	3,567	-1,443	2,893	2,911
d5 1950	-784	1,572	1,584	-301	608	626
d9 1990	-228	459	471	410	-813	-795
	adLn			dmeGB2		
	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
Ip 1990-2000	3,053	-6,098	-6067	3,509	**-6,998**	**-6,920**
Ap 2000-2010	–	–	–	5,625	**-11,231**	**-11,150**
CCA 1991-2000 (2 km)	19,251	-38,493	-38,460	19,771	**-39,521**	**-39,438**
d1 1910	-1,405	2,818	**2,842**	-1,388	**2,795**	2,856
d5 1950	-295	598	622	-254	**529**	**589**
d9 1990	384	-760	-736	446	**-873**	**-813**
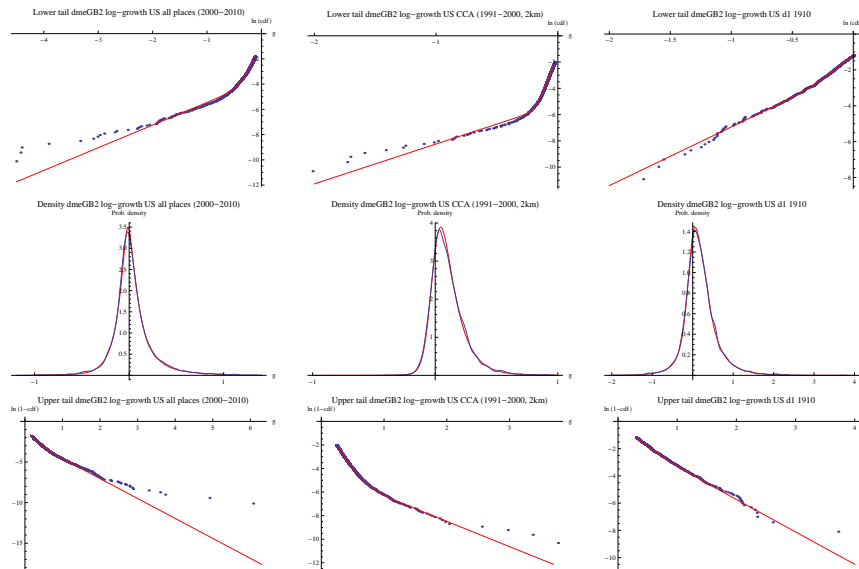
Fig. 1: First row: empirical and estimated dmeGB2 $\ln(\text{cdf})$ for the lower tail. Second row: empirical (Gaussian kernel density, bandwidth=0.02) and estimated dmeGB2 density functions. Third row: empirical and estimated dmeGB2 $\ln(1 - \text{cdf})$ for the upper tail. Left-hand column: log-growth rates of all US places (2000-2010) and dmeGB2. Middle column: log-growth rates of US CCA clusters (1991-2000), 2 km and dmeGB2. Right-hand column: log-growth rates of US d1 1910 and dmeGB2. Empirical in blue, estimated in red in all cases.