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

Ramos, Arturo

Universidad de Zaragoza

14 July 2015

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

ARTURO RAMOS[*]

July 14, 2015

Abstract

We study the decennial 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.

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

JEL: C46, R11, R12.

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

[*]Department of Economic Analysis, Universidad de Zaragoza (SPAIN) aramos@unizar.es. I thank constructive comments from Fernando Sanz-Gracia on a previous version of the manuscript, although all remaining errors are mine. This work is supported by the project ECO2013-45969-P of the Spanish Ministry of Economy and Competitiveness and by the Aragon Government, ADETRE Consolidated Group.

1 Introduction

Several studies have dealt with the theory of the growth process of cities. However, (almost) none of the published works deal with the study of the parametric description of the distribution of city growth rates. 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 (2015) they have been 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 has also 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 of the studied cases, with the so-called "double mixture exponential Generalized Beta 2". In the estimated cases of this distribution, the variances are always finite. This new distribution will offer a performance quite better than the normal distribution.[3]

The rest of the article is organized as follows. Section 2 describes the databases

[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).

[3] It is worth mentioning the recent work of Schluter and Trede (2013) where the authors consider a model with the conclusion that the normalized growth city size distribution follows a Student$-t$.

used. Section 3 introduces the parametric distributions used in this paper. Section 4 describes the empirical results obtained. Finally, Section 5 concludes.

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 1890-2000. These include governmental units classified under state laws as cities, towns, boroughs or villages. Alaska, Hawaii and Puerto Rico have not been considered due to data limitations. The data have been collected from the original documents of the annual census published by the US Census Bureau.[4] These data sets were first introduced in González-Val (2010), see therein for details, and later used in other works like González-Val et al. (2015); Ramos and Sanz-Gracia (2015). 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.

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 (2015).

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 use their US clusters data based on the radius of 2 km. and for the years 1991 and 2000. Data sets of

[4] `http://www.census.gov/prod/www/decennial.html` Last accessed: July 7$^{\mathrm{th}}$, 2015.

this type have been used in Ioannides and Skouras (2013) and Giesen and Suedekum (2014).

[Table 1 near here]

We offer in Table 1 the descriptive statistics of the used data for the US.

3 Description of the presented distributions

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

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

where $x_{i,t}$ is the population of city 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 recall 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.

[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 distributions well-known 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)), and the (non-standardized) Student-t distribution, see, e.g., Johnson et al. (1995) and references therein. The results for the distributions not presented here are available from the author upon request.

3.2 The double mixture exponential Generalized Beta 2 (dmeGB2)

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

Then, let us consider

$$\begin{aligned} f_{\mathrm{eGB2}}(g,a,b,p,q) &= \frac{a\exp((g-b)ap)}{B(p,q)\,(1+\exp(a(g-b)))^{p+q}} \\ \mathrm{cdf}_{\mathrm{eGB2}}(g,a,b,p,q) &= \frac{1}{B(p,q)}B\left(\frac{\exp(a(g-b))}{1+\exp(a(g-b))},p,q\right) \\ u(g,\zeta) &= \exp(-\zeta g) \\ l(g,\rho) &= \exp(\rho g) \end{aligned}$$

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 four 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 like 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 introduce here, which yields the best results out of the ones we have dealt with, is based in one distribution of the work Ramos and Sanz-Gracia (2015), simply taking the exponential of the variable under study, although the meaning of the new variable will be log-growth rates. This distribution 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. They are also imposed equal weight of the distributions of the mixing at the tails, like in Ioannides and Skouras (2013), in order 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:

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

where the constants are given as follows:

$$
\begin{aligned}
d_2^{-1} &= 1-\nu+\frac{\exp(-\rho\epsilon)\,\nu\,\rho\,\mathrm{cdf}_{\mathrm{eGB2}}(\epsilon,a,b,p,q)\,l(\epsilon,\rho)}{f_{\mathrm{eGB2}}(\epsilon,a,b,p,q)} \\
e_2^{-1} &= \frac{(1-\nu)\,\exp(\epsilon\rho)}{\rho\,\mathrm{cdf}_{\mathrm{eGB2}}(\epsilon,a,b,p,q)}+\frac{\nu\, l(\epsilon,\rho)}{f_{\mathrm{eGB2}}(\epsilon,a,b,p,q)} \\
c_2^{-1} &= 1-\theta+\frac{\zeta\,\theta\,\exp(\tau\zeta)\,(1-\mathrm{cdf}_{\mathrm{eGB2}}(\tau,a,b,p,q))\,u(\tau,\zeta)}{f_{\mathrm{eGB2}}(\tau,a,b,p,q)} \\
a_2^{-1} &= \frac{(1-\theta)\,\exp(-\tau\zeta)}{\zeta\,(1-\mathrm{cdf}_{\mathrm{eGB2}}(\tau,a,b,p,q))}+\frac{\theta\, u(\tau,\zeta)}{f_{\mathrm{eGB2}}(\tau,a,b,p,q)} \\
b_2^{-1} &= e_2\,\frac{\exp(\epsilon\rho)}{\rho}+\mathrm{cdf}_{\mathrm{eGB2}}(\tau,a,b,p,q)-\mathrm{cdf}_{\mathrm{eGB2}}(\epsilon,a,b,p,q)+\frac{a_2}{\zeta\,\exp(\tau\zeta)}
\end{aligned}
$$

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 those accounted for in Ramos and Sanz-Gracia (2015); 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 recall briefly the empirical results concerning the US samples on use.

We have computed the log-growth rates between each two consecutive cross-sections of our data. In order to avoid infinite values we have removed the observations for which at least one of the population values is zero. The descriptive statistics of the data so obtained is given in Table 2.

[Table 2 near here]

After the computation of the log-growth rates we have estimated the studied distributions by the method of Maximum Likelihood (ML), using the software MATLAB and MATHEMATICA. We report on Table 3 the estimated values of the parameters for the 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 dataset, see simply Table 2. We see that the estimations are rather precise in almost all cases.

[Table 3 near here]

We have computed numerically as well the means and the standard deviations of the variable g according to the estimated dmeGB2 distributions, which are shown in Table 4. From it, we observe that the computed means and standard deviations are almost identical to those of the empirical samples, and more importantly, that all of them are finite.

[Table 4 near here]

In order to assess the goodness of fit of the two distributions explicitly shown in this paper, we use 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 5. Very briefly, the normal distribution is *strongly* rejected always by the three tests. Meanwhile, the dmeGB2 is not rejected 100% of the cases, and not by a small margin precisely. Thus, the new dmeGB2 models always, with a high degree of accuracy, the studied decennial log-growth rates of US.

[Table 5 near here]

Additionally, we compute more metrics allowing to select 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):

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

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 generate 100 random samples[6] and the results are shown in Table 6. From it, it is clear that the dmeGB2 provides a much better fit than the normal distribution.

[6] Each of these samples is of the sample size of the empirical data. The total generated observations range from about 1,900,000 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.

[Table 6 near here]

Also, we have computed the Akaike Information Criterion (AIC) and Bayesian or Schwarz Information Criterion (BIC) (Burnham and Anderson, 2002, 2004), very well adapted to the maximum likelihood estimation we have performed before. For the computed AIC and BIC see the Table 7.

By these two different types of criteria we see that the dmeGB2 greatly outperforms the normal distribution when considering the decennial log-growth rates of US city sizes, in spite of the fact that our new distribution depends on ten parameters instead of the two parameters of the normal distribution.

[Table 7 near here]

As a complement of the KS, CM, AD, msd, pseudo-R^2, AIC and BIC criteria, we show in Figure 1 an informal graphical approximation of the obtained fits for two of the used samples. We 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 $1-\mathrm{cdf}$ 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 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 3.2) 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 size.

Likewise, the normal distribution for the log-growth rates is clearly rejected empirically in all of 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 of 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 to this respect. We hope to address this issue in further work.

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.

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.

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

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

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 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.

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. and Sanz-Gracia, F. (2015). US city size distribution revisited: Theory and empirical evidence. *Working Paper, available at Munich RePec* `http://mpra.ub.uni-muenchen.de/64051/`.

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.

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.

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 data samples used

Sample	Obs.	Mean	SD	Min.	Max.
Inc. places 1990	19,120	7,978	71,874	2	7,322,564
Inc. places 2000	19,296	8,968	78,015	1	8,008,278
All places 2000	25,358	8,232	68,390	1	8,008,278
All places 2010	29,461	7,826	65,494	1	8,175,133
US CCA 1991 (2 km)	30,201	8,180	104,954	1	12,511,237
US CCA 2000 (2 km)	30,201	8,977	108,342	1	12,734,150

Table 2: Descriptive statistics of the log-growth rates for the consecutive samples used

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
US CCA 1991-2000 (2 km)	30,201	0.105	0.156	-2.398	3.773

Table 3: ML estimators and standard errors (SE) for the dmeGB2 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 2

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)	
US CCA 1991-2000 (2 km)	3.04 (0.32)	-0.121 (0.004)	0.265 (0.023)	
	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)
US CCA 1991-2000 (2 km)	20.40 (0.10)	-0.019 (0.001)	1.581 (0.012)	0.432 (0.003)
	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)	
US CCA 1991-2000 (2 km)	-0.02 (0.02)	2.15 (0.21)	0.02 (0.04)	

Table 4: Means and standard deviations (SD) according to the estimated dmeGB2 and the studied log-growth rate samples. Observe that the values are almost identical to those of Table 2.

Sample	Mean	SD
Ip 1990-2000	0.075	0.260
Ap 2000-2010	0.035	0.273
US CCA 1991-2000 (2 km)	0.105	0.155

Table 5: 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 are marked in boldface

Sample	normal		
	KS	CM	AD
Ip 1990-2000	0 (0.130)	0 (111.162)	0 (623.525)
Ap 2000-2010	0 (0.150)	0 (198.399)	0 (1112.25)
US CCA 1991-2000 (2 km)	0 (0.097)	0 (95.971)	0 (597.806)
	dmeGB2		
	KS	CM	AD
Ip 1990-2000	**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)**
US CCA 1991-2000 (2 km)	**0.798 (0.004)**	**0.886 (0.048)**	**0.927 (0.314)**

Table 6: 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 boldface.

Sample	normal		dmeGB2	
	msd	R^2	msd	R^2
Ip 1990-2000	13.12	0.8095	**0.64**	**0.9907**
Ap 2000-2010	22.66	0.7160	**1.00**	**0.9874**
US CCA 1991-2000 (2 km)	4.44	0.8168	**0.09**	**0.9963**

Table 7: 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		
	log-likelihood	AIC	BIC
Ip 1990-2000	-1,548	3,100	3,116
Ap 2000-2010	-3,817	7,638	7,655
US CCA 1991-2000 (2 km)	13,302	-26,600	-26,584
	dmeGB2		
	log-likelihood	AIC	BIC
Ip 1990-2000	3,509	**-6,998**	**-6,920**
Ap 2000-2010	5,625	**-11,231**	**-11,150**
US CCA 1991-2000 (2 km)	19,771	**-39,521**	**-39,438**

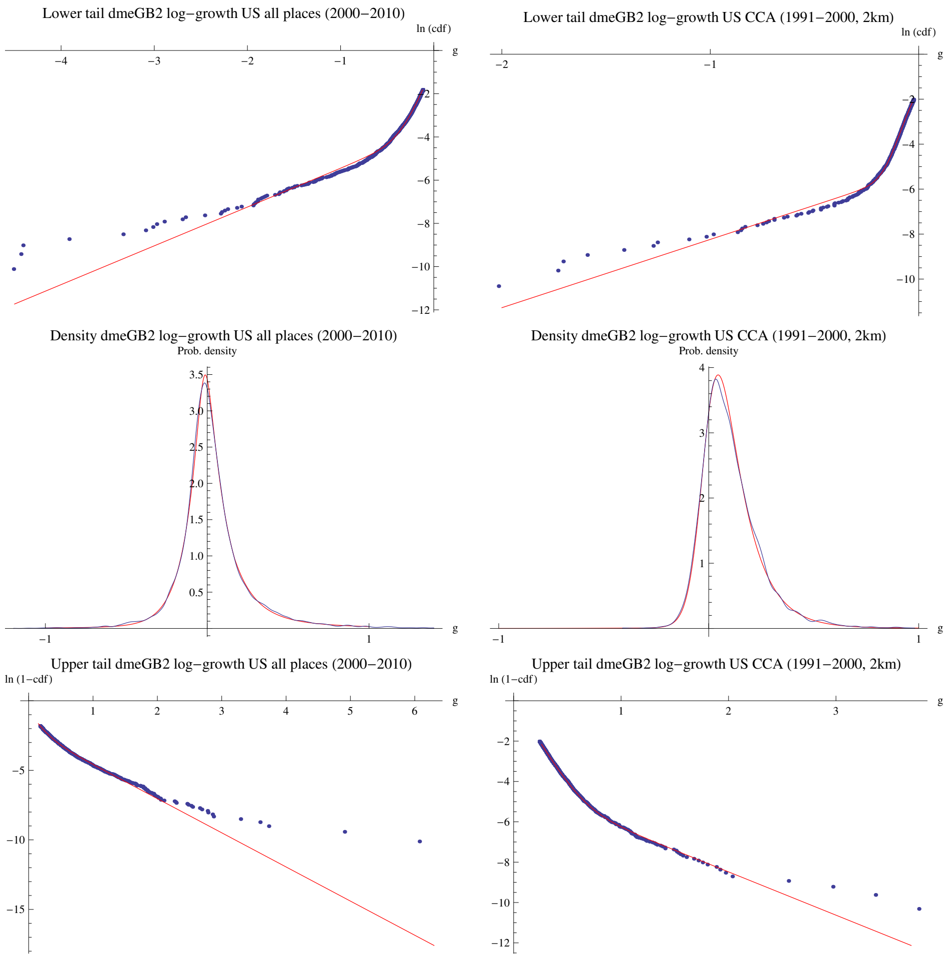

Figure 1: First row: empirical and estimated dmeGB2 $\ln(\mathrm{cdf})$ for the lower tail. Second row: empirical (Gaussian kernel density, bandwidth=0.02) and estimated dmeGB2 density functions. Third row: empirical and estimated dmeGB2 $\ln(1-\mathrm{cdf})$ for the upper tail. Left-hand column: log-growth rates of all US places 2000-2010 and dmeGB2. Right-hand column: log-growth rates of US CCA clusters 1991-2000, 2 km and dmeGB2. Empirical in blue, estimated in red in all cases.