Addenda to "Are the log-growth rates of city sizes distributed normally? Empirical evidence for the USA [Empir. Econ. (2017) 53:1109-1123]"

Ramos, Arturo

Universidad de Zaragoza

30 March 2019

Addenda to “Are the log-growth rates of city sizes distributed normally? Empirical evidence for the USA [Empir. Econ. (2017) 53:1109-1123]”

ARTURO RAMOS[a]

March 30, 2019

Abstract

We update the recently published paper [A. Ramos, Empir. Econ. (2017) 53:1109-1123] on the basis of another important paper [H. S. Kwong and S. Nadarajah, Physica A (2019) 513:55-62]. Specifically, we introduce the 3-normal (3N) and 3-logistic (3L) distributions and compare them with the best of our distributions in the firstly mentioned paper, namely the “double mixture exponential Generalized Beta of the second kind (dmeGB2)”. The main result is that the dmeGB2 remains to be the best model when studying log-growth rates of USA city populations to date. However, if one does not want to achieve such a high precision when describing the data, the 3L emerges to be a very good model for the same purposes.

Keywords: log-growth rates; normal distribution; logistic distribution; GB2 distribution; USA population log-growth rates **JEL:** C46, R11, R12

[a]Department of Economic Analysis, *Universidad de Zaragoza* (SPAIN) aramos@unizar.com.

1 Introduction

The paper Ramos (2017), published in the journal *Empirical Economics*, deals with the question of whether the log-growth rates of USA city sizes follow a normal distribution or not; if in the negative, we tried to provide a better alternative to which such quantities could adhere. We were successful in that task, showing that the normal distribution is strongly rejected always when considering several different types of USA log-growth of population data, and for the same samples a new distribution emerged as the selected one. This is the so called “double mixture exponential Generalized Beta 2 (dmeGB2)”.

After that, a very recently published paper (Kwong and Nadarajah, 2019) appeared in relation with the fit of the Pareto Tails and Lognormal Body (PTLN) distribution (Luckstead and Devadoss, 2017; Luckstead et al., 2017) to the samples of USA and Indian cities. The former paper proposed as alternatives the convex linear combination of three and five lognormal distributions and showed a better fit than the one in the latter papers.

The paper Kwong and Nadarajah (2019) is important and relevant for the contents of Ramos (2017) for mainly two reasons. The first is the question about whether taking a convex linear combination of three normal distributions provides a real improvement with respect to a single normal distribution, that is strongly rejected as we have mentioned already. The second is about the idea already present in Ramos (2017) regarding that the distribution of city sizes (polulations) and of log-growth rates (difference of the natural logarithms of populations) can be taken to be of the same family, the second of them being the exponentiated version of the first one. Since for the USA the paper of Kwong and Nadarajah (2019) shows a better fit of the convex linear combination of three or five lognormals, it is natural to consider the convex linear combination of three[1] normal distributions (3N) for our samples of USA population log-growth rates.

[1] It is enough to consider three, five offers a very limited improvement of the fit and the computational burden increases greatly.

We advance that taking this convex linear combination of three normals is not satisfactory enough with respect to the results already obtained in Ramos (2017) so we propose to take convex linear combinations of three instances of a similar distribution to the normal one, namely the logistic distribution (3L). This allows to obtain better results. When comparing everything with the previous dmeGB2, the last one is still generally the preferred model by statistical criteria, although the 3L emerges as a very manageable and appropriate model for the considered types of data. In this way we update the information presented in Ramos (2017) in a timely and relevant manner.

The paper is organized as follows. In Section 2 we describe the distributions to be added to the study of Ramos (2017). In Section 3 we recall the empirical datasets and the methodology employed in the current study, which is essentially the same as in the mentioned previous paper. In Section 4 we describe the new results obtained. Finally, Section 5 concludes.

2 The newly considered distributions

For a start, the best model of Ramos (2017) when studying log-growth rates of USA city populations happened to be the "double mixture exponential Generalized Beta 2 (dmeGB2)", see the Section 3.4 therein. We will use it in this Addenda in exactly the same way, so we will not reproduce here the material and make reference to the original source.

For the new distributions to be taken into account, let us consider first the probability density function of the normal distribution of parameters $\mu \in \mathbb{R}$, $\sigma > 0$ for the variable $g \in \mathbb{R}$ that stands for log-growth rates

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

and similarly, the probability density function of the logistic distribution of parameters

$\mu \in \mathbb{R}, \sigma > 0$

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

Then, the first added distribution that we will include is inspired by Kwong and Nadarajah (2019), and is the convex linear combination of three normal distributions (3N), with probability density function as follows

$$\begin{aligned} & f_{3\mathrm{N}}(g, \mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, p_1, p_2) \\ & \quad = p_1 f_{\mathrm{N}}(g, \mu_1, \sigma_1) + p_2 f_{\mathrm{N}}(g, \mu_2, \sigma_2) + (1 - p_1 - p_2) f_{\mathrm{N}}(g, \mu_3, \sigma_3) \end{aligned}$$

where $p_1, p_2, 1 - p_1 - p_2 \in [0, 1]$ are the weights of the different normal distributions in the convex linear combination.

Likewise, the probability density function of the convex linear combination of three logistic distributions (3L) is given by

$$\begin{aligned} & f_{3\mathrm{L}}(g, \mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, p_1, p_2) \\ & \quad = p_1 f_{\mathrm{L}}(g, \mu_1, \sigma_1) + p_2 f_{\mathrm{L}}(g, \mu_2, \sigma_2) + (1 - p_1 - p_2) f_{\mathrm{L}}(g, \mu_3, \sigma_3) \end{aligned}$$

with similar meanings for the weights $p_1, p_2, 1 - p_1 - p_2 \in [0, 1]$.

Thus, these distributions 3N and 3L depend on eight parameters to be estimated by Maximum Likelihood (ML) by numerical means, in a similar way as we did for the dmeGB2 in Ramos (2017) (with the command `mle` of MATLAB®).

3 Review of the datasets and the methodology

The data sets employed in this Addenda are exactly the same as in the original paper Ramos (2017). They comprise the log-growth rates of populations of the USA Incorporated Places in the period 1990-2000, an analogous thing for the set of USA All Places

(Incorporated and Unincorporated) in the period 2000-2010, the log-growth rates of the set of USA CCA clusters of Rozenfeld et al. (2011) in the period 1991-2000 for a radius definition of the clusters of 2km., and some three extra data sets taken from Sánchez-Vidal et al. (2014), namely, the log-growth rates of the USA places that are one decade old in 1910 (d1 1910), five decades old in 1950 (d5 1950) and nine decades old in 1990 (d9 1990). The descriptive statistics of all these data sets can be seen on Table 1 of Ramos (2017).

The methodology followed is described shortly in the following and adheres closely to that used in Ramos (2017) with an adding and a replacement. Firstly, we have obtained the ML estimators of the parameters of the added distributions with the command `mle` of MATLAB® and computed the Standard Errors (SE) by the indications of Efron and Hinkley (1978) and McCullough and Vinod (2003). Next, we have computed the means and the Standard Deviations (SD) of the variable g according to the empirical samples and those provided by the different distributions in each case. Also, in order to provide some measure of goodness-of-fit we have computed the Kolmogorov-Smirnov (KS), Cramér-von Mises (CM) and Anderson-Darling (AD) tests for each sample and distribution. Additionally, in order to reinforce the statistics, we have computed also the msd and pseudo-R^2 metrics inspired by Duranton (2007) in the corresponding cases, like in Ramos (2017). As a measure of the information provided by the different models, we have computed the Maximum Log-likelihoods, the Akaike Information Criterion (AIC) and Bayesian or Schwarz Information Criterion (BIC) (Burnham and Anderson, 2002, 2004), very well adapted to the ML estimation that we have performed before. Nextly, and this has not been included in Ramos (2017), we have computed the Vuong's tests (Vuong, 1989) since the considered models are non-nested, in order to assess whether they are statistically equivalent or one is significantly preferred out of the pair confronted in the test. If the statistic, that is distributed as a $N(0,1)$ random variable, is positive, then the first distribution in the test is favoured; if the statistic is negative, the second model in the pair is favoured instead. Finally, we have offered as well a graphical approximation of the fit in the different cases by means of standard qq-plots (in Ramos (2017) we have offered a different kind of graphical approach), since they may show deviations of the empirical samples from the estimated models in a very reliable way.

4 Results

We describe briefly the new results next. In Table 1 and 2 we show the ML estimators of the 3N and 3L distributions, respectively, and jointly with the Standard Errors (SE). We observe that the estimations are rather precise, and the process of the numerical estimations have been relatively fast with few iterations. Had we used (in an obvious notation) the 5N or 5L (or higher), the process of estimating would have been much longer and the convergence much slower.

In Table 3 we show in turn the empirical means and Standard Deviations and those obtained via the different distributions. We can observe that the second ones are in general close to the first ones, specially the means.

In Table 4 we show the computed `msd` and pseudo-R^2 quantities inspired by Duranton (2007) for the considered models. In it we can see that the preferred model by these means is that with lowest `msd` and highest R^2, and in this case the dmeGB2 turns out to be the selected model almost always, with the sole exception of the sample of Incorporated Places 1990-2000, for which is selected the 3L.

On its turn, Table 5 shows the results (p value, (statistics)) of the KS, CM, AD tests. The dmeGB2 is never rejected by these tests and the used samples. The 3N is rejected by the KS for the sample of All Places 2000-2010 and by the three tests for the sample of CCA 1991-2000 (2km), and the 3L is rejected only by the AD test for the lastly mentioned sample (the KS and CM tests do not reject the 3L for such a sample by a very small margin). This criterion shows a slight superiority of the dmeGB2 in the variety of samples, but the 3L goes very close to it.

The information criteria run as follows. In Table 6 we show the Maximum Log-likelihoods and the corresponding AIC and BIC. We see that the selected model for the CCA 1991-2000 (2km) sample is clearly the dmeGB2 both by AIC and BIC, and for the Incorporated Places 1990-2000 sample the AIC selects the same model, but the BIC selects the 3L. For the samples of places one, five and nine decades old in 1910, 1950 and 1990, respectively, the selected model is always the 3L by both AIC and BIC but with a very small margin.

In order to discriminate between the three models by a hypothesis test that allows for non-nested models as those used in this Addenda, we compute the corresponding Vuong's tests, and we show the results in Table 7. From it, we can see that the dmeGB2 is always preferred to the 3N (and statistically equivalent at the 5% significance level for the samples of places one, five and nine decades old in 1910, 1950, 1990, respectively). The 3L and dmeGB2 are almost always statistically equivalent except for the sample of CCA 1991-2000 (2km), where the dmeGB2 is strongly selected. Moreover, the dmeGB2 is preferred always to the 3L except for the sample of All Places 2000-2010, where the 3L is slightly preferred. Finally, the 3L is preferred to the 3N always, although the former is strongly not statistically equivalent to the latter for the samples of Incorporated Places 1990-2000 and All Places 2000-2010.

Finally in the graphs of the qq-plots we see deviations, if any, that are consistent with the previous results, specially to the KS, CM and AD tests and the preference of the models as explained before.

5 Conclusions

We have added two more distributions to the study performed in Ramos (2017), inspired by the very recent work of Kwong and Nadarajah (2019), namely, the 3N and 3L defined in Section 2. We have compared the new results with the performance of the best distribution in Ramos (2017) in order to see whether the 3N improves upon the bad performance of a single normal distribution when studying log-growth rates of USA city populations. We have introduced for motives of comparison another distribution, the 3L (logistic instead of a normal). The results are rather clear. By the previous statistical criteria, the dmeGB2 is almost always the preferred model, showing the superiority with regards the 3N and the 3L. In turn, the 3L offers a better performance than the 3N, which shows that the simple convex linear combinations of normals are not enough to ensure an excellent fit, although the improvement of the 3N is still remarkable with respect to a single normal distribution. The 3L reveals itself to be an excellent alternative to the dmeGB2: it is much more simply coded and estimated, the formulae are much more simple and even with that it is a model which is almost always non-rejected. With similar computational prerequisites, it is superior to the 3N so the 3L is clearly preferable to the former. However, if one wants to obtain the most precise results available today in a variety of instances, one may still use the dmeGB2.

Also, in Kwong and Nadarajah (2019) it is shown that the sample of the USA All Places in 2010 can be very well described by the 3LN, which is the original result that has motivated this Addenda. We can show that other USA samples can be very well described by the 3LN and also by the 3LL (lognormal replaced by the loglogistic), which is work in progress. This confirms again the idea that populations and their log-growth rates can be taken to follow distributions of the same family (exponentiated version for the log-growth rates), which holds at least for USA data sets of city populations.

It is even possible to define other distributions alternative to those three, for example (with an obvious notation) the 3GB2 or 3-versions of special cases of the GB2

(McDonald, 1984; McDonald and Xu, 1995) other than the normal and the logistic distributions. But then the number of parameters may be much higher, the numerical estimation issues may come into play and the convergence may be much slower, it could be the case that no real improvement is obtained since we already have three distributions (dmeGB2, 3N, 3L) that are non-rejected in a number of cases.

Table 1: MLE estimators and standard errors (SE) for the 3N.

	μ_1 (SE)	σ_1 (SE)	μ_2 (SE)	σ_2 (SE)
Ip 1990-2000	0.351 (0.025)	0.750 (0.018)	0.117 (0.004)	0.251 (0.003)
Ap 2000-2010	0.287 (0.035)	1.044 (0.026)	0.081 (0.003)	0.272 (0.003)
CCA 1991-2000 (2km)	0.398 (0.025)	0.581 (0.018)	0.193 (0.002)	0.145 (0.002)
d1 1910	0.371 (0.039)	0.786 (0.027)	0.215 (0.012)	0.334 (0.010)
d5 1950	0.081 (0.012)	0.307 (0.009)	0.363 (0.061)	0.775 (0.042)
d9 1990	0.081 (0.010)	0.265 (0.008)	0.331 (0.069)	0.770 (0.049)
	μ_3 (SE)	σ_3 (SE)	p_1 (SE)	p_2 (SE)
Ip 1990-2000	0.018 (0.001)	0.095 (0.001)	0.058 (0.003)	0.385 (0.006)
Ap 2000-2010	-0.009 (0.001)	0.089 (0.001)	0.041 (0.002)	0.356 (0.005)
CCA 1991-2000 (2km)	0.051 (0.001)	0.078 (0.001)	0.021 (0.001)	0.323 (0.004)
d1 1910	0.071 (0.008)	0.173 (0.007)	0.158 (0.010)	0.471 (0.021)
d5 1950	-0.013 (0.005)	0.148 (0.004)	0.365 (0.020)	0.069 (0.007)
d9 1990	0.016 (0.004)	0.110 (0.003)	0.369 (0.018)	0.051 (0.006)

Table 2: MLE estimators and standard errors (SE) for the 3L.

	μ_1 (SE)	σ_1 (SE)	μ_2 (SE)	σ_2 (SE)
Ip 1990-2000	0.120 (0.004)	0.135 (0.002)	0.357 (0.022)	0.364 (0.010)
Ap 2000-2010	0.088 (0.004)	0.156 (0.002)	-0.011 (0.001)	0.055 (0.001)
CCA 1991-2000 (2km)	0.163 (0.001)	0.067 (0.001)	0.360 (0.011)	0.196 (0.005)
d1 1910	0.176 (0.009)	0.161 (0.005)	0.010 (0.012)	0.087 (0.007)
d5 1950	0.347 (0.056)	0.394 (0.024)	0.232 (0.030)	0.172 (0.016)
d9 1990	0.306 (0.053)	0.359 (0.024)	0.074 (0.009)	0.129 (0.004)
	μ_3 (SE)	σ_3 (SE)	p_1 (SE)	p_2 (SE)
Ip 1990-2000	0.011 (0.001)	0.059 (0.001)	0.371 (0.007)	0.069 (0.003)
Ap 2000-2010	0.325 (0.035)	0.535 (0.017)	0.339 (0.002)	0.621 (0.002)
CCA 1991-2000 (2km)	0.033 (0.001)	0.040 (0.001)	0.417 (0.005)	0.048 (0.002)
d1 1910	0.363 (0.030)	0.374 (0.013)	0.569 (0.015)	0.203 (0.011)
d5 1950	-0.014 (0.004)	0.105 (0.002)	0.082 (0.009)	0.129 (0.015)
d9 1990	0.005 (0.004)	0.066 (0.002)	0.073 (0.008)	0.411 (0.024)

Table 3: Comparison of the empirical means and standard deviations with those provided by the estimated distributions.

	Empirical		dmeGB2		3N		3L	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ip 1990-2000	0.075	0.262	0.075	0.260	0.075	0.242	0.076	0.259
Ap 2000-2010	0.035	0.282	0.035	0.273	0.035	0.282	0.036	0.277
CCA 1991-2000 (2km)	0.105	0.156	0.105	0.155	0.105	0.142	0.103	0.149
d1 1910	0.186	0.415	0.186	0.415	0.186	0.395	0.185	0.415
d5 1950	0.047	0.312	0.047	0.312	0.047	0.312	0.047	0.313
d9 1990	0.056	0.261	0.056	0.261	0.056	0.244	0.056	0.259

Table 4: Computation of the msd (in units of 10^{-3}) and the pseudo-R^2 inspired by Duranton (2007) for the considered distributions. The most favoured values are marked in bold.

	dmeGB2		3N		3L	
	msd	R^2	msd	R^2	msd	R^2
Ip 1990-2000	0.64	0.9907	0.70	0.9899	**0.39**	**0.9943**
Ap 2000-2010	**1.00**	**0.9874**	1.72	0.9785	1.03	0.9870
CCA 1991-2000 (2km)	**0.09**	**0.9963**	0.44	0.9817	0.68	0.9721
d1 1910	**0.18**	**0.9990**	0.37	0.9979	0.30	0.9983
d5 1950	**0.16**	**0.9984**	0.44	0.9954	0.27	0.9972
d9 1990	**0.52**	**0.9924**	0.68	0.9900	0.57	0.9917

Table 5: Kolmogorov–Smirnov (KS), Cramér–von Mises (CM) and Anderson–Darling (AD) tests for the considered distributions. p values (statistics) in all cases. Non-rejections at the 5% level are marked in bold.

	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)**
CCA 1991-2000 (2km)	**0.798 (0.004)**	**0.886 (0.048)**	**0.927 (0.314)**
d1 1910	**0.9979 (0.007)**	**0.9997 (0.015)**	**0.9999 (0.103)**
d5 1950	**0.9307 (0.010)**	**0.9523 (0.036)**	**0.9906 (0.199)**
d9 1990	**0.9737 (0.009)**	**0.9883 (0.026)**	**0.9933 (0.188)**
	3N		
	KS	CM	AD
Ip 1990-2000	**0.146 (0.009)**	**0.217 (0.229)**	**0.156 (1.592)**
Ap 2000-2010	0.011 (0.011)	**0.071 (0.403)**	**0.062 (2.314)**
CCA 1991-2000 (2km)	0.002 (0.012)	0.005 (0.857)	0.001 (6.291)
d1 1910	**0.9272 (0.010)**	**0.9341 (0.040)**	**0.9790 (0.232)**
d5 1950	**0.9808 (0.009)**	**0.9662 (0.033)**	**0.9857 (0.215)**
d9 1990	**0.7765 (0.012)**	**0.7839 (0.0648)**	**0.8764 (0.372)**
	3L		
	KS	CM	AD
Ip 1990-2000	**0.152 (0.009)**	**0.620 (0.093)**	**0.664 (0.584)**
Ap 2000-2010	**0.778 (0.005)**	**0.654 (0.087)**	**0.720 (0.526)**
CCA 1991-2000 (2km)	**0.055 (0.009)**	**0.053 (0.450)**	0.006 (4.371)
d1 1910	**0.9940 (0.007)**	**0.9955 (0.021)**	**0.9996 (0.130)**
d5 1950	**0.8645 (0.011)**	**0.9573 (0.035)**	**0.9877 (0.2093)**
d9 1990	**0.4570 (0.016)**	**0.6671 (0.084)**	**0.7459 (0.501)**

Table 6: Maximum log-likelihoods, Akaike Information Criteria (AIC) and Bayesian or Schwartz Information Criteria (BIC) for the considered distributions. The most favoured values of AIC and BIC in each case are marked in bold.

	dmeGB2			3N		
	Log-likelihood	AIC	BIC	Log-likelihood	AIC	BIC
Ip 1990-2000	3509	**-6998**	-6920	3458	-6901	-6838
Ap 2000-2010	5625	-11,231	-11,150	5532	-11,048	-10,983
CCA 1991-2000 (2km)	19,771	**-39,521**	**-39,438**	19,586	-39,156	-39,090
d1 1910	-1388	2795	2856	-1391	2799	2848
d5 1950	-254	529	589	-256	528	576
d9 1990	446	-873	-813	440	-865	-817
	3L					
	Log-likelihood	AIC	BIC			
Ip 1990-2000	3507	-6997	**-6934**			
Ap 2000-2010	5636	**-11,256**	**-11,191**			
CCA 1991-2000 (2km)	19,615	-39,214	-39,147			
d1 1910	-1389	**2793**	**2842**			
d5 1950	-255	**526**	**574**			
d9 1990	442	**-868**	**-820**			

Table 7: p values (statistics) of Vuong's tests for the considered distributions. Non-rejections at the 5% level are marked in bold.

	3N vs dmeGB2	3L vs dmeGB2	3N vs 3L
Ip 1990-2000	0.001 (-3.360)	**0.586 (-0.544)**	0.000 (-3.753)
Ap 2000-2010	0 (-4.506)	**0.376 (0.886)**	0 (-7.402)
CCA 1991-2000 (2km)	0 (-8.760)	0 (-6.920)	**0.250 (-1.150)**
d1 1910	**0.330 (-0.975)**	**0.548 (-0.600)**	**0.408 (-0.828)**
d5 1950	**0.642 (-0.465)**	**0.831 (-0.213)**	**0.622 (-0.492)**
d9 1990	**0.149 (-1.442)**	**0.066 (-1.842)**	**0.568 (-0.572)**

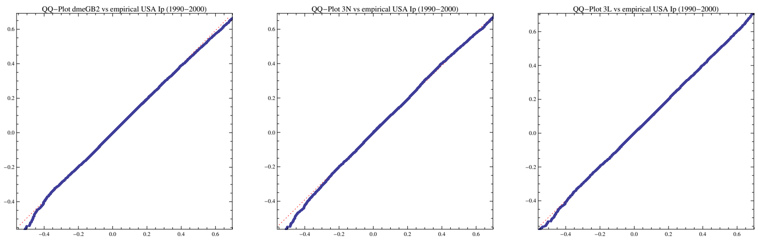

Figure 1: Qq-plots for the estimated dmeGB2, 3N and 3L distributions versus the empirical USA Ip (1990-2000) sample.

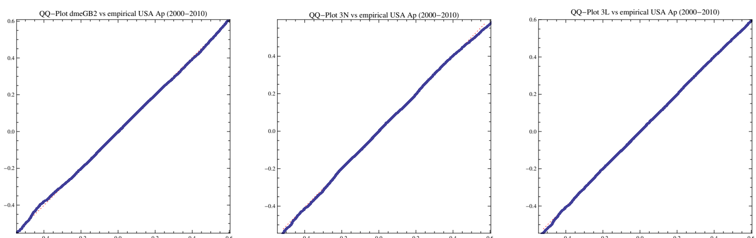

Figure 2: Qq-plots for the estimated dmeGB2, 3N and 3L distributions versus the empirical USA Ap (2000-2010) sample.

References

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.

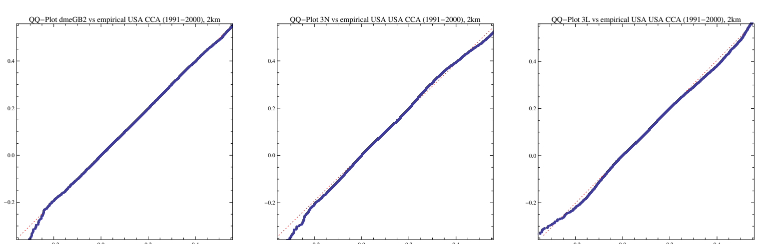

Figure 3: Qq-plots for the estimated dmeGB2, 3N and 3L distributions versus the empirical USA CCA (1991-2000), 2km sample.

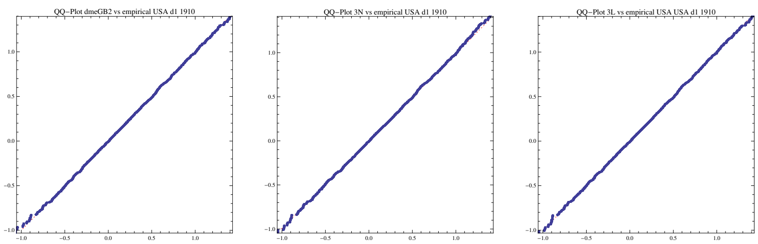

Figure 4: Qq-plots for the estimated dmeGB2, 3N and 3L distributions versus the empirical USA d1 1910 sample.

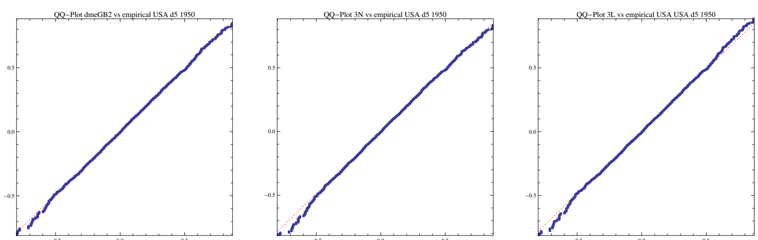

Figure 5: Qq-plots for the estimated dmeGB2, 3N and 3L distributions versus the empirical USA d5 1950 sample.

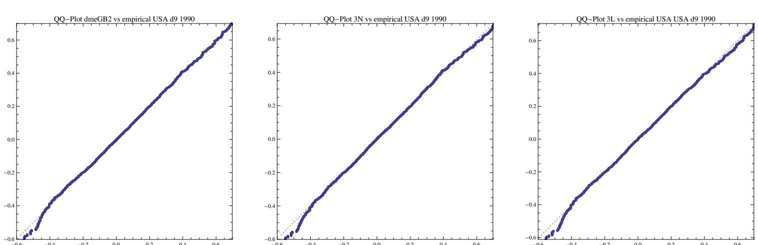

Figure 6: Qq-plots for the estimated dmeGB2, 3N and 3L distributions versus the empirical USA d9 1990 sample.

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

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.

Kwong, H. S. and Nadarajah, S. (2019). A note on “Pareto tails and lognormal body of US cities size distribution”. *Physica A: Statistical Mechanics and its Applications*, 513:55–62.

Luckstead, J. and Devadoss, S. (2017). Pareto tails and lognormal body of US cities size distribution. *Physica A: Statistical Mechanics and its Applications*, 465:573–578.

Luckstead, J., Devadoss, S., and Danforth, D. (2017). The size distribution of all Indian cities. *Physica A: Statistical Mechanics and its Applications*, 474:237–249.

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.

Ramos, A. (2017). Are the log-growth rates of city sizes distributed normally? Empirical evidence for the USA. *Empirical Economics*, 53:1109–1123.

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.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333.