Q-LOG EXPONENTIAL DISTRIBUTION IN URBAN AGGLOMERATION

Aurelia CĂȘARU*, Irina BĂNCESCU, Vasile PREDA****

*PhD School of Mathematics, University of Bucharest, Romania
(aurelia.casaru@gmail.com)
**National Institute for Economic Research, Bucharest, Romania

DOI: 10.19062/1842-9238.2018.16.2.9

Abstract: *A purpose of urban theory is to describe how cities develop based on the number of inhabitants. Many statistical models and laws of growing of cities have been suggested, such as Zipf's, Mandelbrot-Zipf's and Gibrat's laws. This paper studies the urban agglomeration of Romania for the years between 2007-2017, using the q-log exponential distribution.*

Keywords: *q-log exponential distribution, Romania's cities distribution, Tsallis statistics.*

1. INTRODUCTION

Cities develop in different ways all around the world, depending on many social-economic factors such as: economic growth of the area, economic activity, ethnic factors, infrastructure, and not only. Statisticians have studied the cities distribution for a long time, considering big cities, small cities or all together. Some laws of probability have Pareto tails for the lower and upper tails and different bodies: log-normal or Singh-Maddala [1,2]**.** Other statistical models used in urban theory are the q-exponential, Pareto, log-normal [3,4], and more recently, the q-log distribution family [5].

This paper studies the cities and municipalities agglomeration of Romania from 2007 to 2017 using the q-log exponential distribution. We apply the Kolmogorov-Smirnov test and graphically show how well this probability law models the data.

The paper is organized as follows. In Section 2, we present the q-log exponential distribution. Empirical analysis of Romania's cities population is performed in Section 3, while Section 4 concludes the paper.

2. METHODOLOGY

The q-log-location-scale exponential model was first introduced in 2018, as a submodel of q-log-location-scale distributions [5]. This class of distributions has been obtained by applying a q-logarithm Tsallis transformation to a baseline location-scale distribution. The q-logarithm and q-exponential functions are defined by

$\log_q^T(x) = \log(x), if x > 0, q = 1$ and $\log_q^T(x) = \dfrac{x^{q-1}-1}{q-1}, if x > 0, q \neq 1$

while $e_q(x) = \exp(x), if q = 1$, $e_q(x) = [1+(q-1)x]^{1/(q-1)}, if q \neq 1, 1+(q-1)x > 0$, and $e_q(x) = 0^{1/(q-1)}, if q \neq 1, 1+(q-1)x \leq 0$ where q is a real parameter.

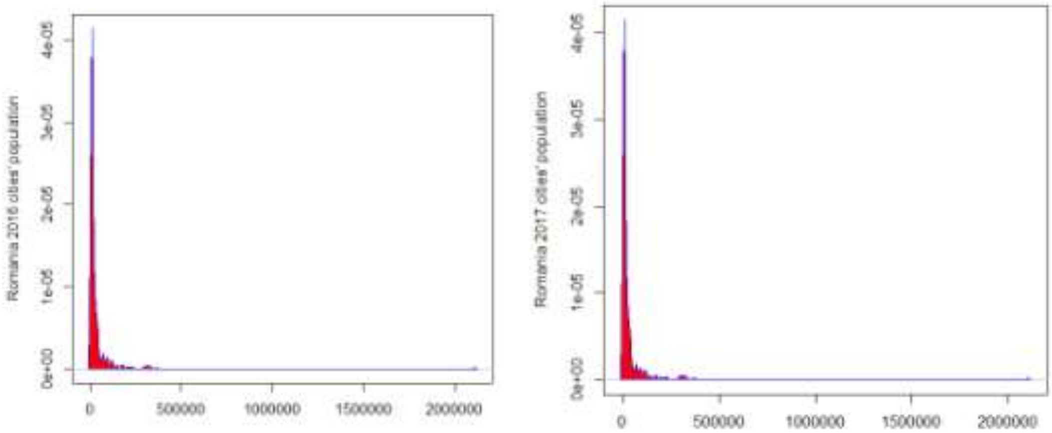

FIG. 1. Empirical density of city size "Romania2016" and "Romania2017" data

The q-log exponential distribution of parameters ρ > 0, θ > 0, $q_1 \in (0,1)$, and $q_2 > 1$ is obtained considering as baseline distribution the q_1-exponential model. These two statistical models are presented next.

The q_1-exponential distribution is defined by the following cumulative distribution and density functions, respectively

$$G(x) = 1 - \left[e_{q_1}(-\rho x)\right]^{q_1},\ g(x) = q_1 \rho e_{q_1}(-\rho x), q_1 > 0, \rho > 0, 0 < x < C_{q_1\rho} \quad (1)$$

Where

$$C_{q_1\rho}(1) = \infty, if q_1 \in (0,1) \text{and} C_{q_1\rho} = \frac{1}{\rho(q_1 - 1)}, if q_1 > 1.$$

The q-log exponential distribution (qLE) of parameters q_1, q_2, ρ, and θ is defined by the following distribution and density functions, respectively

$$F(x) = 1 - \left[e_{q_1}\left(-\rho \log_{q_2}^{T}\left(\frac{x}{\theta}\right)\right)\right]^{q_1},\ f(x) = \frac{q_1\rho}{\theta}\left(\frac{x}{\theta}\right)^{q_2 - 2} e_{q_1}\left(-\rho \log_{q_2}^{T}\left(\frac{x}{\theta}\right)\right), x > \theta$$

where $q_1 \in (0,1), q_2 > 1, \rho > 0$, while θ>0 is chosen as the minimum value of each dataset considered.

3. EMPIRICAL ANALYSIS

In this section, we discuss the analysis of cities' size distribution in Romania between 2007 and 2010. We perform the Kolmorogov-Smirnov test based on maximum likelihood estimation of parameters.

3.1 Data

Considering demographic data provided by INS we analyze the cities' population data using q-log exponential distribution. Some characteristics of the datasets considered such as maximum and minimum values, number of observations, measures of skewness and kurtosis, standard deviation, and mean are displayed in Table 1. It can be observed that for each datasets, the measure of kurtosis is extremely high, suggesting a heavy-tail distribution. Also, the skewness is high for these datasets.

Because all datasets have almost the same values for these measures, we have only displayed the empirical densities for years 2016 and 2017. The empirical densities of datasets ”Romania2016”, and ”Romania2017” are displayed in Fig. 1.

Table 1. Descriptive statistics of Romania cities population

Year	Nr. of obs.	Mean	SD	Min	Max	Skewness	Kurtosis
2007	319	40,161.72	131,960.5	1,811	2,156,978	13.20	205.09
2008	319	40,051.41	131,970.5	1,784	2,158,816	13.23	205.78
2009	319	40,011.10	132,033.1	1,750	2,160,627	13.24	206.11
2010	319	39,960.93	132,077.1	1,732	2,162,037	13.26	206.41
2011	319	39,816.94	131,755.7	1,710	2,157,282	13.27	206.63
2012	319	39,671.51	131,410.2	1,704	2,151,758	13.27	206.70
2013	319	39,589.31	130,890.8	1,695	2,140,816	13.23	205.72
2014	319	39,428.26	129,371.9	1,674	2,110,752	13.14	203.55
2015	319	39,317.36	128,847.3	1,677	2,100,519	13.11	202.87
2016	320	39,243.34	129,026.9	1,684	2,107,399	13.15	203.86
2017	320	39,136.24	128,783.3	1,663	2,103,251	13.14	203.81

3.2 Parameter estimation and discussion

In order to assess if the q-log exponential model is appropriate to model the datasets considered, we utilize the Kolmogorov-Smirnov (KS) test. Also, we discuss the maximum likelihood estimation of the parameters. Distributions having few parameters are well fitted to data by maximum likelihood method. Hence, the maximum likelihood method applied for qLE model is described next.

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from $qLE(\rho, q_1, q_2, \theta)$ distribution of parameters ρ > 0, θ > 0, $q_1 \in (0,1)$ and $q_2 > 1$. The log-likelihood function for the vector of parameters $\delta = (\rho, q_1, q_2, \theta)^T$ can be expressed as

$$l(\delta) = n\log(q_1) + n\log(\rho) - nlog(\theta) + \sum_{i=1}^{n}(q_2 - 2)[\log(x_i) - \log(\theta)] + \sum_{i-1}^{n}\log e_{q_1}\left(-\rho\log_{q_2}^{T}\left(\frac{x_i}{\theta}\right)\right)$$

The log-likelihood can be maximized by solving the nonlinear likelihood equations obtained by differentiating the equation above. However, the maximum likelihood estimator of θ is very simple

$$\theta = \min x_i$$

In other words, we choose the parameter θ to be equal to the smallest value of the dataset considered for analysis. The components of the score vector U(δ) are

$$U_\rho(\delta) = \frac{\partial l}{\partial \rho} = \frac{n}{\rho} - \sum_{i=1}^{n}\log_{q_2}^{T}\left(\frac{x_i}{\theta}\right)\left[e_{q_1}\left(-\rho\log_{q_2}^{T}\left(\frac{x_i}{\theta}\right)\right)\right]^{1-q_1}$$

$$U_{q_1}(\delta) = \frac{\partial l}{\partial q_1} = \frac{n}{q_1} - \sum_{i=1}^{n}\left\{\frac{\log\left[1 + (q_1 - 1)\left(-\rho\log_{q_2}^{T}\left(\frac{x_i}{\theta}\right)\right)\right]}{(q_1 - 1)^2} + \frac{\rho\log_{q_2}^{T}\left(\frac{x_i}{\theta}\right)}{(q_1 - 1)\left[1 + (q_1 - 1)\left(-\rho\log_{q_2}^{T}\right)\left(\frac{x_i}{\theta}\right)\right]}\right\}$$

$$U_{q_2}(\delta)=\frac{\partial l}{\partial q_2}=\sum_{i=1}^{n}\log\left(\frac{x_i}{\theta}\right)-\sum_{i=1}^{n}\left[e_{q_1}\left(-\rho\log_{q_2}^{T}\left(\frac{x_i}{\theta}\right)\right)\right]^{1-q_1}\frac{\rho}{(q_2-1)^2}\left\{\left(\frac{x_i}{\theta}\right)^{q_2-1}\left[\log\left(\frac{x_i}{\theta}\right)(q_2-1)+1\right]+1\right\}$$

Solving the nonlinear likelihood equations above requires the use of numerical Methods that admit restrictions, such as an extended Nelder-Mead method [6].

Table 2 displays the maximum likelihood estimates of Romania's cities population. The standard errors were calculated by considering 500 bootstrapped samples, while the software used is R. All parameter estimates are highly significant as indicated by the low standard errors. The MLE of $\hat{\rho}$ ranges from 0.0284 to 0.0320, of $\hat{q}_1$ ranges from 0.20417 to 0.21728, while of $\hat{q}_2$ ranges from 4.58516 to 4.82080. The smallest city in Romania for year 2007 had a population of 1,811 inhabitants, while in 2017 this value decreased to 1,663 inhabitants. The estimate of parameter θ is taken as the minimum value of each dataset.

Table 2. Parameter estimates of q-log exponential distribution of Romania's cities population

Year	Parameter estimators (standard errors)			
	ρ	$\hat{q}_1$	$\hat{q}_2$	θ
2007	0.0320 (0.6836)	0.20417 (0.02952)	4.82080 (0.54629)	1,811
2008	0.0308 (0.65560)	0.20610 (0.03013)	4.79259(0.54740)	1,784
2009	0.0291 (0.68319)	0.20745 (0.03194)	4.76868 (0.57177)	1,750
2010	0.0288 (0.69591)	0.20941 (0.03049)	4.73258 (0.52923)	1732
2011	0.0284 (0.69357)	0.21164 (0.03059)	4.69677 (0.52232)	1,710
2012	0.0288 (0.68885)	0.21293 (0.03277)	4.67398 (0.54680)	1,704
2013	0.0292 (0.69707)	0.21453 (0.03077)	4.64519 (0.51047)	1,695
2014	0.0292 (0.67057)	0.21650 (0.03288)	4.60674 (0.51194)	1,674
2015	0.0304 (0.69024)	0.21687 (0.03106)	4.59454 (0.49957)	1,677
2016	0.0310 (0.67105)	0.21648 (0.03209)	4.60079 (0.54209)	1,684
2017	0.0304 (0.65839)	0.21728 (0.03227)	4.58516 (0.54138)	1,663

To predict city sizes $\hat{x}$, we substitute θ, and $\hat{q}_2$ into the q-log exponential CDF and solve for

$$\hat{x}=\hat{\theta}e_{\hat{q}_2}\left[\frac{-1}{\hat{\rho}}log_{\hat{q}_1}^{T}\left[(1-F(.))^{1/\hat{q}_1}\right]\right]$$

The log of actual and predicted values of x can be plotted against the log rank to obtain the rank-size plot.

3.3 Graphical analysis

In this section, we graphically analyze the modelling of Romania's cities population. We perform the Kolmogorov- Smirnov test and display the rank-size plots of both data and predicted values. The Kolmogorov-Smirnov (KS) test considers the goodness-of-fit by analyzing the supremum of the difference between the theoretical and empirical CDF. Table 3 reports the KS test values of q-log exponential distribution based on the selected data. A p-value of the KS test close to 1 indicates extreme evidence for the data to have come from the distribution fitted.

Table 3. Kolmogorov-Smirnov test results

Year	KS	p-value
2007	0.02612	0.981
2008	0.02646	0.978
2009	0.02631	0.979
2010	0.02653	0.978
2011	0.02692	0.974
2012	0.02694	0.974
2013	0.02719	0.972
2014	0.02727	0.971
2015	0.02676	0.976
2016	0.02622	0.980
2017	0.02698	0.973

CONCLUSIONS

Romania's cities population can be very well modelled by means of q-log exponential distributions for each year. Since a large portion of the population of Romania (56.4%) is living in cities, the economic activities in this part of the country are important to the national economic growth. At present, in 2017, the capital, Bucharest, is the most developed city in the country, having more than 2 millions inhabitants, while the second largest city Iasi, has 368,866 inhabitants. This fact suggests a large gape in development between the capital and the rest of the cities.

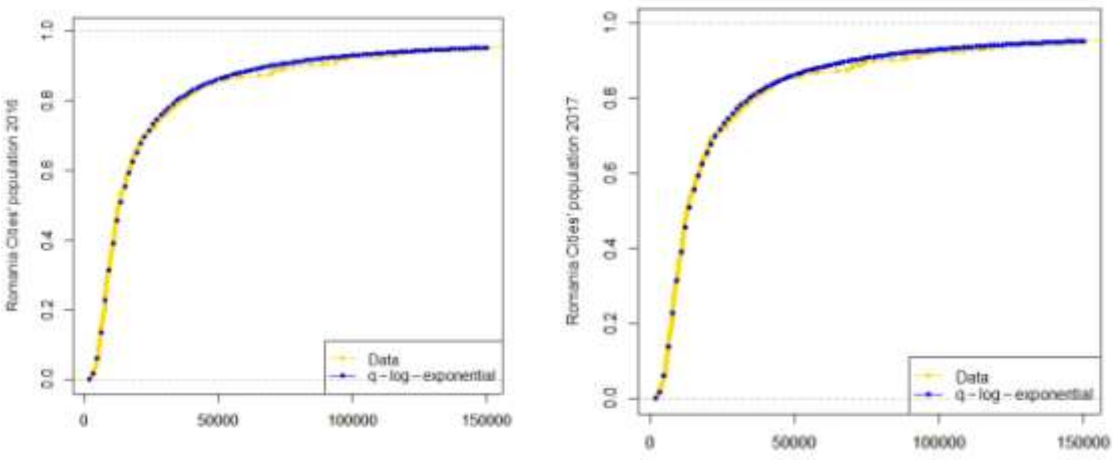

FIG. 2. Empirical and theoretical cumulative distribution of city size "Romania2016" and "Romania2017" data

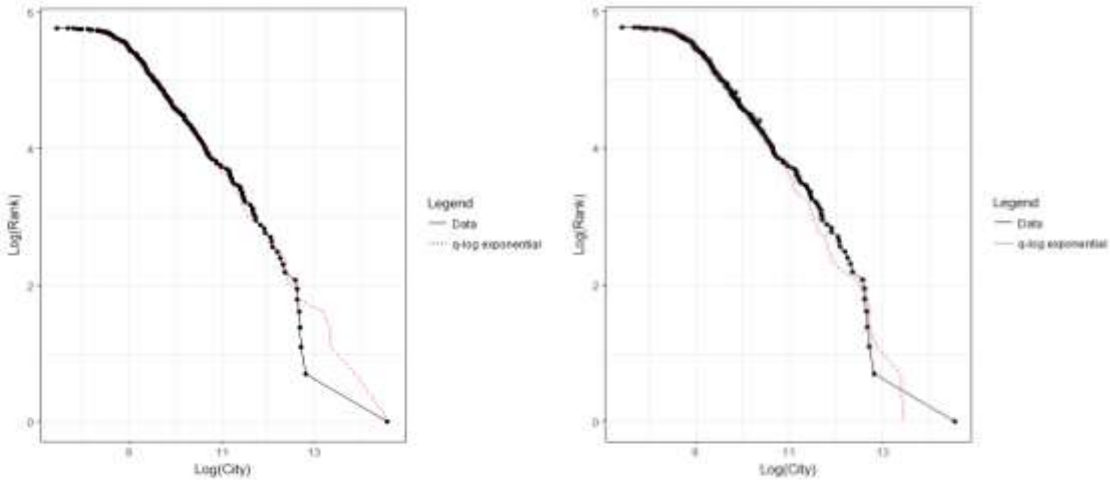

FIG. 3. Rank-size plots for 2016 and 2017

REFERENCES

[1] J. Luckstead, S. Devadoss, *Pareto tails and lognormal body of US cities size distribution*, *573-578*, Physica A: Statistical Mechanics and its Applications 465, 2017;

[2] M. Puente-Ajovın, A. Ramos, *On the parametric description of the French, German, Italian and Spanish city size distributions*, *489-509*, The Annals of Regional Science, 54(2), 2015;

[3] L.C. Malacarne, R.S. Mendes, E.K. Lenzi, *q-Exponential distribution in urban agglomeration*, Physical Review E 65, (1), 2001;

[4] P. Cirillo, *Are your data really Pareto distributed? 5947–5962,* Physica A: Statistical Mechanics and its Applications 392 (2013);

[5] I. Bancescu, *q-log-distributions. Log-concavity and Log-convexity, 1-16*, The European Physical Journal Plus, 133, 2018;

[6] Subrahmanyam M.B. *An Extension of the simplex method to constrained nonlinear optimization*, 311-319, Journal of Optimization Theory and Applications, 62(2), 1989.