International Journal of Applied Mathematics

Volume 28 No. 6 2015, 715-725
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v28i6.6

MODELING EMPIRICAL DISTRIBUTIONS OF FIRM SIZE WITH q-DISTRIBUTIONS

Evandro Marcos Saidel Ribeiro[1], Gilberto Aparecido Prataviera[2 §]

[1,2]Department of Business Administration, FEARP
University of São Paulo
Ribeirão Preto, SP - 14040-905, BRAZIL

Abstract: We explore generalized q-distributions – q-exponential and q-Weibull – as underlying models for firm size distributions. We analyze the empirical distributions of Brazilian and USA open share companies. Total Assets and Total Revenue are used as variables for firm size. Although earlier studies have concentrated on large values of the size variable, where typical power-law behavior is observed, here we analyze the short and long range behavior. Our results indicate that q-Weibull distribution produces the best log-likelihood estimate for the entire data set. However, we have verified that for asymptotic values the typical power-law behavior is best described by q-Weibull in Brazil, and by q-exponential in USA.

AMS Subject Classification: 62-07, 62P25, 62P20
Key Words: q-distributions, firm size, power-law

1. Introduction

Several phenomena in physics, sociology, economics and finance, are described by probability distributions presenting asymptotically a power-law decay (see [17, 9, 7] and references therein). The research on power-law focuses on their observation, the power-law tail characterization, and in proposing models generating power-law distributions. A promising approach for such phenomena has been obtained by using a generalized concept of entropy [23] for nonextensive systems, and whose maximization under certain constraints results in the so-

Received: August 15, 2015

§Correspondence author

called q-distributions [25, 26, 5, 18]. The q-distributions decay with a power law for large values of its variables. Moreover, some empirical studies indicates that q-distributions, besides reproducing the power-law for extreme values, are suitable to fit empirical distributions of data in the whole range of their values [18, 24, 19, 6, 27, 20, 10, 1].

Studies on power law have a long history in economics and finance [9]. In particular, firm size distributions has been a subject of constant interest in economics and in the emerging field of econophysics [21, 15, 22, 3, 4, 13, 14, 11, 16, 2]. Firm size is important for several economic and accounting models; firm size can be a determinant factor in public policies, tax, valuation, and so on. Measures for firm sizes, as Total Assets and Total Revenue, present extremum values in their empirical observation that are troublesome in econometric models based on gaussian distribution. Earlier studies has shown that these extremum values are indeed part of data whose distributions are characterized by a power-law decay [17]. However, these studies were limited in describing or proposing models generating data for the range of values where the power-law is valid. Instead, recent works - distribution of newspapers [19], distribution of delay time in trains [6], among others [27, 20, 10, 1] - indicate that q-distributions can fit the distribution for the whole range of their values. These works suggest that q-distributions could be useful for fitting firm size. Such a study is important to get a better description of empirical data and to indicate possible underlying processes that should be considered to improve theoretical dynamical models.

In this work we explore q-distributions of probability for fitting firm size distributions. We consider firm size distributions of Brazilian and USA open share companies. We compare the q-exponential and q-Weibull distributions with empirical data containing the observations of Total Assets and Total Revenue. The parameters of q-distributions are estimated by the maximum log-likelihood method. Our results indicate that q-Weibull distribution produces the best log-likelihood estimate for the entire data set for both countries. However, we verify that for asymptotic values the typical power-law behavior is best described by q-Weibull in Brazil, and by q-exponential in USA.

The article is organized as follows: In Section 2 we present a brief review of some basics properties of q-distributions. In Section 3 we consider the empirical distributions of Total Assets and Total Revenue for Brazil and USA and their fitting using q-distributions. Finally, in Section 4 we present the conclusions.

2. q-Distributions

The q-distributions arise from a maximization of the so-called Tsallis entropy [25, 26, 5] subject to certain constraints. As we are interested in distributions of firm size, let $p(x)$ be the probability density function (PDF) for some firm size variable. The probability that some firm has size less then x is given by the cumulative distribution function (CDF),

$$P(x) = \int_0^x p(x')dx', \tag{1}$$

and the complementary cumulative distribution function (CCDF), $P^C(x)$, is obtained by

$$P^C(x) = \int_x^\infty p(x')dx' = 1 - P(x), \tag{2}$$

In the following we summarize some basic properties of the PDF and CDF functions for q-distributions, specifically, the q-exponential and q-Weibull distributions.

2.1. q-Exponential

The q-exponential is given by the following probability density function (PDF) [18]

$$p_{qe}(x) = p_0 \left[1 - (1-q)\,\frac{x}{x_0}\right]^{1/(1-q)}, \tag{3}$$

where $1 - (1-q)\,x/x_0 \geq 0$ and $p_0 = (2-q)/x_0$. In the limit of $q \to 1$ Eq. (3) recovers the usual exponential function, $p_e(x) = p_0 \exp(-x)$. If $q < 1$ Eq. (3) has a finite value for any finite real value of x, while for $q > 1$ Eq. (3) exhibits a power-law asymptotic behavior

$$p_{qe}(x) \approx x^{-1/(q-1)}. \tag{4}$$

The CCDF associated to Eq. (3) is given by

$$P^C_{qe}(x) = p'_0 \left[1 - \left(1-q'\right)\frac{x}{x'_0}\right]^{1/(1-q')}, \tag{5}$$

where, $p'_0 = p_0 x_0/(2-q)$, $q' = 1/(2-q)$ and $x'_0 = x_0/(2-q)$.

2.2. q-Weibull Distribution

The q-Weibull PDF is given by

$$p_{qw}(x) = p_0 \frac{r x^{r-1}}{x_0^r} \left[1 - (1-q) \left(\frac{x}{x_0} \right)^r \right]^{1/(1-q)}, \tag{6}$$

where $1 - (1-q)(x/x_0)^r \geq 0$ and $p_0 = (2-q)$. The CCDF associated to Eq. (6) is given by

$$P_{qw}^C(x) = p_0' \left[1 - (1-q') \left(\frac{x}{x_0'} \right)^r \right]^{1/(1-q')}, \tag{7}$$

where, $p_0' = p_0/(2-q)$, $q' = 1/(2-q)$ and $x_0' = x_0/(2-q)^{1/r}$. If $q < 1$ Eq. (3) has a finite limit [18], while for $q > 1$ Eq. (6) exhibits a power-law behavior

$$p_{qw}(x) \approx x^{-\xi}, \tag{8}$$

with $\xi = (1-r)$ for small x and $\xi = r[(2-q)/(q-1)] + 1$ for large x. In the limit of $q \to 1$ Eq. (6) recovers the Weibull PDF function

$$p_w(x) = p_0 \frac{r x^{r-1}}{x_0^r} \exp[-(x/x_0)^r]. \tag{9}$$

and its associated CCDF function

$$P_w^C(x) = p_0 \exp[-(x/x_0)^r]. \tag{10}$$

In addition, the q-Weibull distribution reproduces the q-exponential distribution in the limit of $r \to 1$.

3. Empirical Analysis

In this section we present the numerical results and a discussion on the fitting of data using q-xponential and q-Weibull. The data set consists of Total Assets and Total Revenue of Brazilian and USA companies. The data set was obtained from Economatica© database and refers to 324 Brazilian and 982 USA companies in 2010. Table 1 presents the sample data descriptive statistics. Analogously to Picoli *et al*[18], in order to compare the adjustments we also present the fitting for the Weibull distribution.

		x_m	x_M	$\bar{x}$	s	CV
Assets	USA	0.060	781.818	12.513	39.306	3.14
	Brazil	2.87E-06	198.488	4.290	15.372	3.58
Revenue	USA	1.01E-04	275.564	7.514	19.360	2.58
	Brazil	4.93E-04	104.933	2.351	7.662	3.259

Table 1: Total Assets and Total Revenue descriptive statistics: x_m is the minimum value, x_M is the maximum value, $\bar{x}$ is the mean value, s is the standard deviation, (in units of 10^9 US dolar) and $CV = s/\bar{x}$ is the coefficient of variation

		Total Assets		Revenue	
		Brazil	USA	Brazil	USA
q-Exponential	x_0	0.080	0.265	0.117	0.238
	q	1.532	1.365	1.494	1.390
q-Weibull	x_0	0.165	0.171	0.226	0.199
	q	1.292	1.574	1.257	1.471
	r	0.681	1.493	0.691	1.154
Weibull	x_0	0.449	0.704	0.517	0.685
	r	0.518	0.693	0.548	0.668

Table 2: Estimated Parameters. The parameters are in units of $\bar{x}$.

To obtain the distributions parameters from data we use the maximum log-likelihood method [8]. Among several methods of fitting, it has been addressed [7, 12] that power-law distributions, which corresponds of extremum values of q-distributions, are estimated with more accuracy and robustness using maximum likelihood estimation (MLE). The parameters are obtained by maximizing the log-likelihood function [8], defined as

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log p(x_i|\boldsymbol{\theta}), \tag{11}$$

where $\boldsymbol{\theta}$ is a vector of parameters.

The comparison between empirical and theoretical distributions is done by analyzing the cumulative distributions as defined in (2), and whose parameters are obtained from the log-likelihood. The empirical complementary cumulative distribution is obtained by

$$P_e^C(x_i) = 1 - P(x_i), \tag{12}$$

		Total Assets		Total Revenue	
		Brazil	USA	Brazil	USA
q-Exponential	$\alpha = \frac{2-q}{q-1}$	0.88	1.74	1.02	1.56
q-Weibull	$\alpha = r\frac{2-q}{q-1}$	1.65	1.11	2.00	1.29

Table 3: Power-law CCDF exponent $\alpha = \xi - 1$ at large values of variables, where ξ is the exponent for the pdf asymptotic limit.

with $P(x_i)$ being the proportion of data smaller or equal to x_i. Here, in order to obtain a smooth empirical CCDF we consider- the $P(x_i)$ defined by

$$P(x_i) = (i - 0.5)/n, \qquad \text{for } i = 1, \ldots, n. \tag{13}$$

The parameters obtained from the maximum log-likelihood are presented in Table 2. The q-Weibull distribution presents the best fitting for all variables in both countries companies.

Let us first consider the Total Asset distribution. Figs. 1a and 1b show the empirical and fitted Complementary Cumulative Distribution Function (CCDF) and Cumulative Distribution Function(CDF), respectively, for Brazilian companies in logarithmic scales. Plots of CCDF and CDF were done in order to emphasize the long- and short-range behavior. The results of Total Assets distribution for USA companies are shown in Fig. 2. Although the best fit for both countries is given by the q-Weibull distribution, by comparing Figs. 1 and 2 we observe that, while in Brazil the q-Weibull presents a good long- and short-range agreement, in USA the long range data are located between the q-Weibull and q-exponential.

For the Total Revenue distributions we can observe in Figs. 3 and 4 a similar qualitative behavior when comparing to the Total Assets distributions of Figs. 1 and 2, respectively.

The power-law exponent α, for the $P^C(x)$, obtained by considering the CCDF in the limit of large values of variables is shown in Table 3. The values in Table 3 indicate some interesting differences between country distributions. While for the USA companies we have approximately the same power-law limit exponent in the q-Weibull distribution for both variables, in Brazil we observe a significantly difference between them. Also, the power-law exponent for USA firms are closer to unit, Zipff law [3], which is consistent with previous studies. However, for Brazilian firms beside the difference in the exponents in each variable, the exponents obtained in the best fit are not consistent with Zipf law.

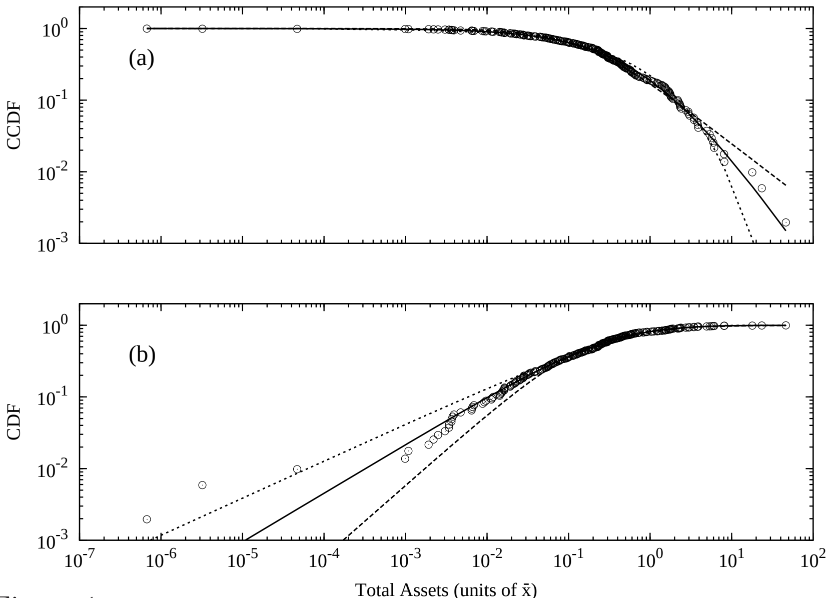

Figure 1: Distribution of Total Assets for Brazilian firms: (a) Complementary Cumulative Distribution (CCDF), (b) Cumulative Distribution (CDF). empirical data (circles); q-Weibull (solid line); q-exponential (long dashes); Weibull (short dashes). The estimated parameters are given in Table (2). The variable and the parameters are in units of $\bar{x}$.

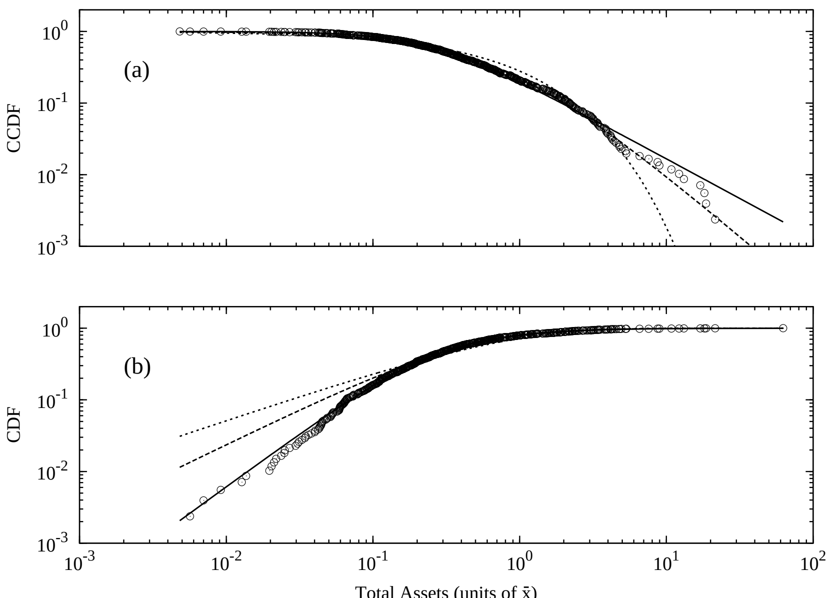

Figure 2: Same as in Fig. (1) but for USA firms.

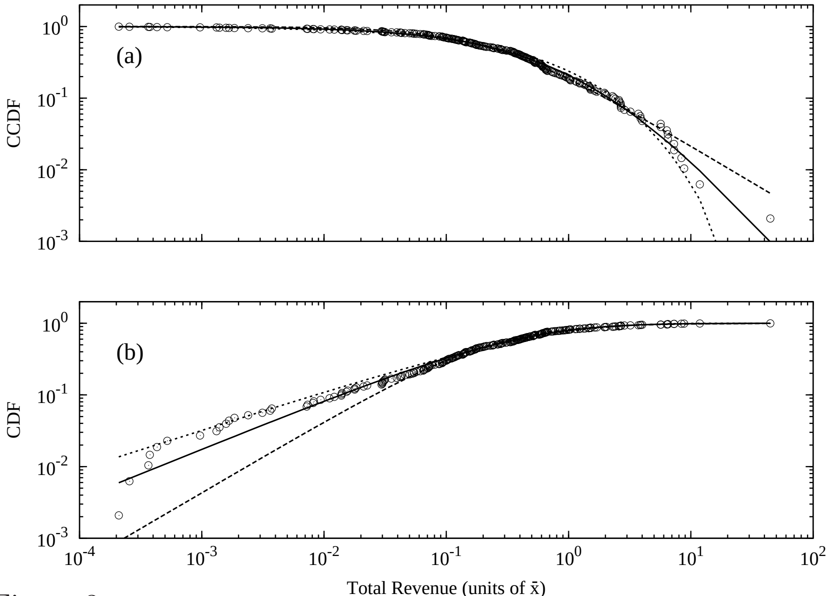

Figure 3: Distribution of Total Revenue for Brazilian firms: (a) Complementary Cumulative Distribution (CCDF), (b) Cumulative Distribution (CDF). empirical data (circles); q-Weibull (solid line); q-exponential (long dashes); Weibull (short dashes). The estimated parameters are given in Table (2)The variable and the parameters are in units of $\bar{x}$.

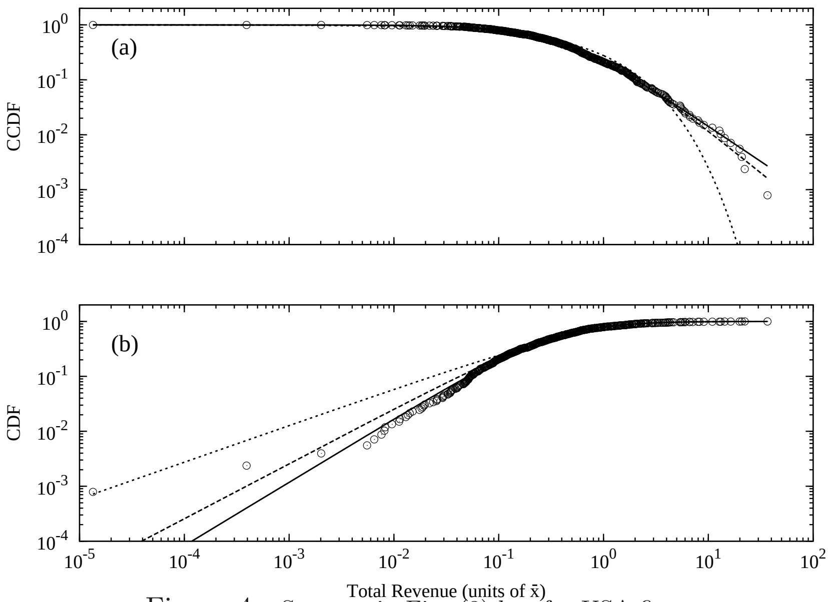

Figure 4: Same as in Fig. (3) but for USA firms.

In addition, by comparing q-distributions for Brazilian firms in Figs. 1 and 3 we observe that for small values of variables the estimative of q-Weibull is above the q-exponential one, while for large values the opposite is observed. For USA firms, in Figs 2 and 4 we observe a reverse behavior instead.

4. Conclusions

In this work we have considered generalized q-distributions to model empirical distributions of firm size. In particular, we have compared Brazilian and USA firm size distributions. The primary goal was to verify if q-distributions are suitable to fit not only the typical power-law behavior at large values, but the whole range of values of firm size. Specifically, the q-exponential and q-Weibull distributions were considered, and its parameters were estimated using the method of maximum likelihood. The best fitting to the empirical data was obtained with the q-Weibull distribution.

We verified that for asymptotic values some qualitative differences are observed between the two countries. The q-Weibull large values limit agrees with Zipff law for USA firms. For Brazilian firms the power law exponents indicate disagreement to the Zipf law. For small values, when considering Total assets, there is a good fitting to the q-weibull distribution for both countries, except for a few extreme small values of Brazilian firms. For Total Revenue a small-value disagreement is observed for USA companies.

References

[1] A.D. Anastasiadis, M.P. de Albuquerque, M.P. de Albuquerque and D.B. Mussi, Tsallis q-exponential describes the distribution of scientific citations - a new characterization of the impact, *Scientometrics*, **83** (2010), 205-218.

[2] H. Aoyama, Y. Fujiwara, Y. Ikeda, H. Iyetomi and W. Souma, *Econophysics and Companies Statistical: Life and Death in Complex Business Networks*, Cambridge University Press, New York (2010).

[3] R.L. Axtell, Zipf distribution of US firm sizes, *Science*, **293** (2001), 1818-1820.

[4] R.L. Axtell, Firm sizes: Facts, formulae, fables and fantasies, *Center on Social and Economic Dynamics Working Paper*, **44** (2006).

[5] C. Beck, Generalised information and entropy measures in physics, *Contemporary Physics*, **50** (2009), 495-510.

[6] K. Briggs and C. Beck, Modelling train delays with q-exponential functions, *Physica A*, **378** (2007), 498-504.

[7] A. Clauset, C.R. Shalizi and M.E.J. Newman, Power-Law distributions in empirical data, *SIAM Review*, **51** (2009), 661-703.

[8] G. Cowan, *Statistical Data Analysis*, Clarendon Press, Oxford (1998).

[9] X. Gabaix, Power laws in economics and finance, *Annual Review of Economics*, **1** (2009), 255-293.

[10] K. Gangopadhyay and B. Basu, City size distributions for India and China, *Physica A*, **388** (2009), 2682-2688.

[11] J. di Giovanni, A.A. Levchenkoand R. Ranciered, Power laws in firm size and openness to trade: Measurement and implications, *Journal of International Economics*, **85** (2011), 42-52.

[12] M.L. Goldstein, S.A. Morris and G.G. Yena, Problems with fitting to the power-law distribution, *Eur. Phys. J. B*, **41** (2004), 255-258.

[13] J. Growiec, F. Pammolli, M. Riccaboni and H.E. Stanley, On the size distribution of business firms, *Economic Letters*, **98** (2008), 207-212.

[14] S.H. Kang, Z. Jiang, C. Cheong and Seong-Min Yoon, Changes of firm size distribution: The case of Korea, *Physica A*, **390** (2011), 319-327.

[15] R.E. Lucas, On the size distribution of business firms, *The Bell Journal of Economics*, **9**, No 2 (1978), 508-523.

[16] R.N. Mantegna and H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, New York (2000).

[17] M.E.J. Newman, Power laws, Pareto distributions and Zipf's law, *Contemporary Physics*, **46** (2005), 323-351.

[18] S. Picoli Jr., R.S. Mendes and L.C. Malacarne, q-exponential, Weibull, and q-Weibull distributions: an empirical analysis, *Physica A*, **324** (2003), 678-688.

[19] S. Picoli Jr., R.S. Mendes and L.C. Malacarne, Statistical properties of the circulation of magazines and newspapers, *Europhys. Lett.*, **72** (2005), 865-871.

[20] S. Picoli Jr., R.S. Mendes, L.C. Malacarne and R.P.B. Santos, q-distributions in complex systems: a brief review, *Brazilian Journal of Physics*, **39**, No 2 (2009), 468-474.

[21] H.A. Simon and C.P. Bonini, The size distribution of business firms, *American Economic Review*, **48** (1958), 607-617.

[22] J. Sutton (1997), Gibrat's legacy, *Journal of Economic Literature*, **35**, No 1 (1997), 40-59.

[23] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, *J. Stat. Phys.*, **52** (1988), 479-487.

[24] C. Tsallis and M.P. de Albuquerque, Are citations of scientific papers a case of nonextensivity? *European Journal of Physics B*, **13** (2000), 777-780.

[25] C. Tsallis, R.S. Mendes and A.R. Plastino, The role of constraints within generalized nonextensive statistics, *Physica A*, **261** (1998), 534-554.

[26] E. Van der Straeten and C. Beck, Superstatistical distributions from a maximum entropy principle, *Phys. Rev. E* **78** (2008), # 051101.

[27] H.S. Yamada and K. Iguchi, q-exponential fitting for distributions of family names, *Physica A*, **387** (2008), 1628-1636.