Physica A 444 (2016) 158–162

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Size distribution of U.S. lower tail cities

Stephen Devadoss [a], Jeff Luckstead [b,*]

[a] University of Idaho, United States
[b] University of Arkansas, United States

HIGHLIGHTS

- A large number of small cities are in the lower tail.
- Application of reverse Pareto and reverse general Pareto to U.S. lower-tail cities.
- U.S. lower-tail cities follow reverse-Pareto distribution.

ARTICLE INFO

Article history:
Received 13 May 2015
Received in revised form 9 September 2015
Available online 19 October 2015

Keywords:
Lower tail cities
Reverse Pareto
Size distribution
United States

ABSTRACT

Studies that analyzed the size distribution of U.S. cities have mainly focused on the upper tail and showed that these cities adhere to Zipf's law. However, even though a large number of cities are in the lower tail, very few studies have examined the distribution of these small cities because of data limitations. We apply reverse Pareto and reverse general Pareto distributions to analyze U.S. lower tail cities. Our results show the power law behavior of lower tail U.S. cities is accurately represented by both the reverse Pareto and general Pareto.

1. Introduction

Studies that analyzed the size distribution of U.S. cities have mainly focused on the upper tail (135 largest metropolitan areas) and showed that these cities adhere to Zipf's law (see for example Refs. [1–4]).[1,2] Almost all past studies have analyzed the distribution of large cities because of the high concentration of population in these cities. However, even though a large number of cities are in the lower tail, very few studies have examined the distribution of these small cities because of data limitations. But, starting in 2000, the U.S. Census has provided a substantially expanded data set that includes all locations called "places" which cover all cities, towns, and villages. This new and expanded data has paved the way for several studies to analyze the size distribution of all U.S. cities [9–11]. However, these studies clearly identified the difficulty in examining both the upper and lower tail simultaneously, particularly using the standard rank-size plots on a log–log scale due to heavy distortion of upper tail cities for descending rank and lower tail cities for ascending rank. Reed [12] emphasized the need to analyze these tails separately because lower tail cities follow reverse Pareto (defined below), whereas upper tail cities exhibit Pareto. Then, Reed applied reverse Pareto to study lower tail cities in California and West Virginia in the United States and Cantabria and Barcelona in Spain. A recent study by Devadoss et al. [13] found that lower tail cities in India do exhibit

* Corresponding author.
E-mail addresses: devadoss@uidaho.edu (S. Devadoss), jluckste@uark.edu (J. Luckstead).

[1] Studies have also shown that the Zipf exponent depends on the sample size [5,6]. Furthermore, Rozenfeld et al. [7] presented evidence that Zipf's law is supported by geographic rather than administrative boundaries.

[2] Stanley et al. [8] have also analyzed firm size distributions.

http://dx.doi.org/10.1016/j.physa.2015.09.077
0378-4371/

power law behavior.[3,4] Because a large portion of cities are in the lower tail and these cities have received scant attention in the literature, the purpose of this paper is to analyze the size distribution of U.S. lower tail cities.

2. Methodology

The size of lower tail cities, $x_1 \leq x_2 \leq, \ldots, \leq x_{n-1} \leq x_n$ with ranking 1 to n, can be represented by the reverse Pareto distribution, with PDF and CDF

$$g(x|\alpha, \mu) = \alpha \frac{x^{\alpha-1}}{\mu^{\alpha}} \tag{1}$$

$$G(x|\alpha, \mu) = \left(\frac{x}{\mu}\right)^{\alpha}, \tag{2}$$

where $\mu = \max(x)$ is the location parameter and α is the shape parameter.

We apply the maximum likelihood method to estimate α. For a sample of n independently and identically distributed (i.i.d.) lower tail cities, the log-likelihood function is

$$\mathcal{L}(\alpha, \mu|x_1, \ldots, x_{n-1}) = (n-1)\log(\alpha) - \alpha(n-1)\log(\mu) + (\alpha-1)\sum_{i=1}^{n-1}\log(x_i). \tag{3}$$

Maximize the above function with respect to α and solve the first-order condition to obtain

$$\hat{\alpha} = \frac{(n-1)}{(n-1)\log(\mu) - \sum_{i=1}^{n-1}\log(x_i)}. \tag{4}$$

To predict city sizes $(\hat{x}_p)$, we substitute $\hat{\alpha}$ into the reverse Pareto CDF Eq. (2) and solve for $\hat{x}_p = \mu G(\cdot)^{1/\hat{\alpha}}$. The log of actual and predicted values of x can be plotted against the log rank to obtain the rank-size plot.

We also estimate the lower tail distribution using the more flexible reverse-general Pareto (GP).[5] To our knowledge, our study is the first to apply reverse-general Pareto to estimate lower tail size distribution. The PDF and CDF for GP are

$$f(x|\beta, \mu, \sigma) = \frac{\beta}{\sigma}\left(1 + \frac{x-\mu}{\sigma}\right)^{\beta-1} \tag{5}$$

$$F(x|\beta, \mu, \sigma) = \left(\frac{x-\mu+\sigma}{\sigma}\right)^{\beta}, \tag{6}$$

where $\mu = \max(x)$ is the location parameter, σ is the scale parameter, and β is the shape parameter. With $\mu = \sigma$, reverse general Pareto becomes reverse Pareto; thus the former nests the latter.

For n i.i.d. samples of lower tail cities, the log-likelihood function of reverse general Pareto is

$$\mathcal{L}(\beta, \mu, \sigma|x_1, \ldots, x_{n-1}) = (n-1)\log(\beta) - (n-1)\log(\sigma) + (\beta-1)\sum_{i=1}^{n-1}\log\left(1 + \frac{x_i-\mu}{\sigma}\right).$$

Since the maximization of this function does not yield closed form solutions for β and σ, we use nonlinear optimization to obtain $\hat{\beta}$ and $\hat{\sigma}$. Based on the restriction $\beta = \alpha$ and $\mu = \sigma = \max(x)$, we can test whether the reverse general Pareto nests reverse Pareto using the likelihood ratio test by calculating the values of the log-likelihood function for the unrestricted $(\mathcal{L}^u)$ and restricted $(\mathcal{L}^r)$ models and using the likelihood ratio (LR) test

$$LR = 2\left(\log \mathcal{L}^u - \log \mathcal{L}^r\right) \overset{a}{\sim} \chi^2(2).$$

Substituting $\hat{\beta}$ and $\hat{\sigma}$ into the reverse general Pareto CDF (6), we can solve for the predicted values: $\hat{x}_{gp} = \hat{\sigma}F(\cdot)^{1/\hat{\beta}} + \hat{\mu} - \hat{\sigma}$. We can graph the log of actual and predicted city sizes against the log of the rank to generate the rank-size plot.

We employ rank-size plots and the Kolmogorov–Smirnov (KS) test to analyze the goodness of fit for both the reverse Pareto and general Pareto distributions for U.S. small cities.

[3] Luckstead and Devadoss [14] studied the growth process of Indian large cities and concluded that Gibrat's law holds for these cities.

[4] See Devadoss and Luckstead [15] for the growth process of lower tail cities in the United States.

[5] Urzúa [5] and Luckstead and Devadoss [16] have applied general Pareto to upper tail city distributions.