Physica A 442 (2016) 193–196

Contents lists available at ScienceDirect

Physica A

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

The power law distribution for lower tail cities in India

Stephen Devadoss [a], Jeff Luckstead [b,*], Diana Danforth [b], Sherzod Akhundjanov [c]

[a] *University of Idaho, United States*
[b] *University of Arkansas, United States*
[c] *Washington State University, United States*

HIGHLIGHTS

- India is a predominantly rural country with numerous villages.
- Power-law behavior of small Indian cities.
- Lower-tail Indian cities follow the reverse-Pareto distribution.

ARTICLE INFO

Article history:
Received 26 February 2015
Received in revised form 12 May 2015
Available online 16 September 2015

Keywords:
India
Lower tail cities
Size distribution
Power law

ABSTRACT

The city size distribution for lower tail cities has received scant attention because a small portion of the population lives in rural villages, particularly in developed countries, and data are not readily available for small cities. However, in developing countries much of the population inhabits rural areas. The purpose of this study is to test whether power law holds for small cities in India by using the most recent and comprehensive Indian census data for the year 2011. Our results show that lower tail cities for India do exhibit a power law.

1. Introduction

The city size distribution literature has extensively studied upper tail cities because much of the population lives in urban areas and data for large cities are readily available. Many studies have shown that Zipf's law for upper tail cities is a regularly observed phenomenon. For example, Krugman [1], Gabaix [2], and Ioannides and Overman [3] conclude that the distribution of large cities in the United States is stable and Zipf's law fits upper tail cities with statistical regularity.[1]

In contrast, the city size distribution for lower-tail cities has received scant attention because a small portion of the population lives in rural villages, particularly in developed countries, and data are not readily available for small cities.[2] However, in developing countries much of the population resides in rural areas. For example, in India about 69 percent of the population lives in rural areas [8]. Recent studies by Gangopadhyay and Basu [9] and Luckstead and Devadoss [6] found power-law behavior of upper-tail Indian cities. However, because India is predominantly a rural country with numerous

* Corresponding author.
E-mail addresses: devadoss@uidaho.edu (S. Devadoss), jluckste@uark.edu (J. Luckstead), ddanfort@uark.edu (D. Danforth), sherzod.akhundjanov@email.wsu.edu (S. Akhundjanov).

1 However, several studies have shown Zipf's law fails to hold for upper tail cities particularly for several developing countries [4–6].

2 One exception is Reed [7] who studied the lower tail distribution for various states in the United States and Spain and showed that the degree of linearity in the lower tail is stronger than that of the upper tail.

http://dx.doi.org/10.1016/j.physa.2015.09.016
0378-4371/

small and medium villages, it is worth examining the size distribution of the lower tail cities for developing countries such as India. Thus, the purpose of this study is to investigate whether power law holds for small cities in India by using the most recent and comprehensive Indian census data for the year 2011.

2. Methodology

Consider m cities in the lower tail and rank them in ascending order such that x_1 is the smallest city and x_m is the largest city. Denote $F(x)$ as the CDF of x, then $\frac{i}{m} \approx F(x_i)$.[3] Based on Reed [11] (see the Appendix for more detail), the PDF and CDF of a reverse Pareto random variable x are

$$f(x) = \beta \frac{x^{\beta-1}}{x_m^{\beta}} \tag{1}$$

$$F(x) = \left(\frac{x}{x_m}\right)^{\beta}, \tag{2}$$

where the location parameter is $x_m = \max(x)$ such that $x_m \geq x_i > 0$ and the shape parameter is $\beta > 0$. For a reverse Pareto law of small cities, $\log(\Pr(X < x))$ is linearly related to $\log(x)$ with a positive slope for β [7]. Substitution of the CDF in (2) into $\frac{i}{m} \approx F(x_i)$ results in $\frac{i}{m} \approx \left(\frac{x_i}{x_m}\right)^{\beta}$. If $\beta = 1$, then $\frac{i}{m} \approx \frac{x_i}{x_m}$, and the rank of a city is proportional to its size.

For an independently and identically distributed sample of m lower tail cities, the log likelihood function of the reverse Pareto distribution is

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

Maximization of this function with respect to β yields the first-order condition

$$\frac{\partial \mathcal{L}}{\partial \beta} = \frac{(m-1)}{\beta} - (m-1)\log(x_m) + \sum_{i=1}^{m-1}\log(x_i) = 0. \tag{4}$$

The above equation can be solved to obtain

$$\hat{\beta} = \frac{(m-1)}{(m-1)\log(x_m) - \sum_{i=1}^{m-1}\log(x_i)}. \tag{5}$$

In the next section, we estimate the parameter β for lower tail Indian cities.

3. Data and results

The data for Indian cities for the 2011 census were collected from the Census of India [8]. According to this census, cities are classified as villages based on the following criteria.[4] All places with (1) no municipality, corporation, and cantonment board, (2) a population of less than 5000, (3) at least 25% of the male workers employed in agricultural occupations, and (4) a population density of less than 400 per square km are classified as villages.

To determine the cutoff of the lower tail cities, we generate the histogram of the log of city sizes of the full sample (Fig. 1). Based on the inflection point of the lower tail of this figure, we can ascertain that the log of x_m, i.e., the largest city in the lower tail, is about 4.5, which translates into a population of about 90. For this truncation point, the sample size is 37,153 with a mean city size of 44.99 and standard deviation of 26.33. In addition, to provide robust estimates, we consider different truncation points for $\log(x_m)$ ranging from 4.09 to 4.79, which translate into a population of 60–120. The corresponding sample size for this range of $\log(x_m)$ is 24,849–49,826.

Table 1 presents the estimated values of β and their standard errors for various sample sizes of lower tail cities. The estimated values of β generally show a slight positive trend as the sample size increases. For the truncation point $x_m = 110$, the estimated values of β is exactly equal to 1. Similar observations can be also made for $x_m = 100$. These results show βs are very close to 1, indicating lower-tail power law behavior. Our results are robust in that the estimated values of β are close to 1 for various sample sizes corresponding to different truncation points of x_m. Since the standard errors are very small, the beta estimates for various truncation points are highly significant.

[3] Also see Stanley et al. [10].

[4] While our paper considers cities based on the administrative definition, Rozenfeld et al. [12] find that cities based on a geographical notion can generate Zipf's law results for a larger sample with a cutoff city size of 12,000.