Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

The distribution of all French communes: A composite parametric approach

Enrique Calderín-Ojeda

Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Australia

HIGHLIGHTS

- The distribution of the size of all French settlements from 1962 to 2012 is examined.
- The analysis is completed by using the composite Lognormal–Pareto distribution.
- Lognormal upper tail Pareto distribution nests the composite Lognormal–Pareto model.
- The upper quartile is closely described by a Pareto distribution for early periods.
- For later periods most of the city size dynamics is explained by a Log normal model.

ARTICLE INFO

Article history:
Received 15 September 2015
Received in revised form 7 December 2015
Available online xxxx

Keywords:
City size
Composite models
France
Lognormal
Pareto

ABSTRACT

The distribution of the size of all French settlements (communes) from 1962 to 2012 is examined by means of a three-parameter composite Lognormal–Pareto distribution. This model is based on a Lognormal density up to an unknown threshold value and a Pareto density thereafter. Recent findings have shown that the untruncated settlement size data is in excellent agreement with the Lognormal distribution in the lower and central parts of the empirical distribution, but it follows a power law in the upper tail. For that reason, this probabilistic family, that nests both models, seems appropriate to describe urban agglomeration in France. The outcomes of this paper reveal that for the early periods (1962–1975) the upper quartile of the commune size data adheres closely to a power law distribution, whereas for later periods (2006–2012) most of the city size dynamics is explained by a Lognormal model.

1. Introduction

Traditionally, two important principles have been considered to analyse the empirical distribution of the city size. The Zipf's law [1] and the Gibrat's law of proportionate growth of cities [2]. On the one hand, the Zipf's law in the city distribution system indicates that the second largest city is half the size of the largest, the third largest city a third the size of the largest and the nth is the nth size of the largest one. This implies that the city sizes follow a power law distribution. For example, Moura and Ribeiro [3] studied the Zipf's law for Brazilian cities and Gangopadhyay and Basu [4] analysed the size distributions of urban agglomerations for India and China by estimating the scaling exponent for Zipf's law. Moreover, other different approaches based on generalization of the Pareto distribution have been suggested in the literature. In this regard, Sarabia and Prieto [5] developed a model to describe Spanish city size data by means of the Pareto Positive Stable distribution. Similarly, Gómez-Déniz and Calderín-Ojeda [6] examined the arrangement of urban agglomerations in Australia and New Zealand by using the Pareto ArcTan distribution. On the other hand, the Gibrat's law asserts that the Lognormal distribution

E-mail address: enrique.calderin@unimelb.edu.au.

http://dx.doi.org/10.1016/j.physa.2016.01.018
0378-4371/

Please cite this article in press as: E. Calderín-Ojeda, The distribution of all French communes: A composite parametric approach, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.01.018

emerges when the size of the cities grows randomly but proportionately. In this sense, many papers have argued that the effect of fitting city size data by means of the Pareto distribution vanishes when the whole population range is included in the sample, without excluding medium and small cities. For instance, Anderson and Ge [7] determined that the Lognormal model is preferable to the Pareto distribution by using size distribution of Chinese cities. Similarly, Eeckhout [8] showed that the Lognormal distribution provides a good fit to the size of all cities in the US employing data from 2000 census. In addition, examination of the transition from Lognormal to Pareto has been considered in the urban economics literature. On this subject, Luckstead and Devadoss [9] studied the city size distribution of China and India for seven decades; in their paper they concluded that the Chinese city distribution is explained by a Lognormal model between 1950–1990 and by a Pareto distribution in 2010. In contrast, the Indian cities fluctuate from Lognormal in the earlier periods to Zipf in the most recent periods.

Nevertheless, the assertion that the Lognormal distribution provides a better fit to city size data was disputed by Levy [10] who claimed that in the top range of the largest cities, the size distribution diverges dramatically from the Lognormal distribution and it is in excellent agreement with a straight line. In this sense, the distribution of the settlement size can be divided into two regions: the bottom and middle ranges where the empirical data are explained by the Lognormal distribution, and the top range where the empirical distribution fits a power law distribution. Relying on this idea, Giesen et al. [11] used the four-parameter Double Pareto Lognormal distribution, a distribution that is Pareto in the upper and lower tails and Lognormal in between, to explain the distribution of all cities by using untruncated city size data from eight countries. Additionally, González-Val et al. [12] showed that Double Pareto Lognormal distribution provides the best fit to describe city size data in the US, Spain and Italy. However, this model is unable to estimate the thresholds where the lower tail ends and the upper tail begins. Recently, Ioannides and Skouras [13] suggested the Lognormal upper tail Pareto distribution, a four-parameter continuous spliced model that combines the Lognormal distribution in the lower tail and middle part of the distribution and the Pareto distribution in the upper tail. In this model the break point between the two components is endogenously estimated from the data. Puente-Ajovín and Ramos [14] use the seven-parameter threshold double Pareto Singh–Maddala distribution to describe the French, German, Italian and Spanish city size data. This is a distribution with Pareto behaviour in the lower and upper tails and Singh–Maddala body.

In the last few years, composite distributions have been used in actuarial statistics to model loss data when the claims faced by insurers consist of a mixture of moderate and large claims. In this paper a composite Lognormal–Pareto distribution with unrestricted mixing weights is proposed to describe the distribution of the population size of all settlements (*communes*) in France for different years in the period between 1962 and 2012. This composite model, that was firstly introduced by Scollnik [15], uses a Lognormal distribution up to an unknown threshold value, endogenously estimated from the data, and a two-parameter Pareto density thereafter. Next, continuity and differentiability conditions are imposed at the threshold to yield a smooth density function and to reduce from four to three the number of parameters to be estimated. The resulting model is similar in shape to the Lognormal distribution but with a thicker tail. This model is nested in the Lognormal upper tail Pareto distribution and it includes as particular cases the Lognormal, Pareto and Zipf distributions. Numerical results show that the upper quartile of the commune size distribution is close to Pareto for the earlier years, and the empirical data are better explained by the Lognormal distribution for the most recent years.

The reminder of the paper is organized as follows. In Section 2, the methodology used in this work is described by providing a short review of the Lognormal upper tail Pareto and composite Lognormal–Pareto models with unrestricted mixing weights together with brief comments on parameter estimation. Next, in Section 3 the data, numerical illustrations and graphical methods of model assessment based on Zipf's plots and log–log plots are presented. Finally, the paper ends with a concluding section.

2. Methodology

Ioannides and Skouras [13] proposed a Lognormal upper tail Pareto (LUTP) spliced model to describe the US city size distribution. This model combines a Lognormal distribution in the lower tail and main bulk of the distribution and a Pareto distribution in the upper tail. The probability density function (pdf) of the LUTP distribution is given by

$$f(x) = \begin{cases} r\,\dfrac{f_1(x)}{F_1(\theta)}, & 0 < x \le \theta \\ (1-r)\,f_2(x), & \theta \le x < \infty. \end{cases} \quad (1)$$

In (1)

$$f_1(x) = \frac{1}{\sqrt{2\pi}\,x\,\sigma}\,\exp\left(-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right), \quad x > 0 \quad (2)$$

is the pdf of the Lognormal distribution, where $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter and

$$F_1(\theta) = \Phi\left(\frac{\ln\theta-\mu}{\sigma}\right) \quad (3)$$