Physica A 461 (2016) 756–766

Contents lists available at ScienceDirect

Physica A

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

Modelling income data using two extensions of the exponential distribution

Enrique Calderín-Ojeda [a,*], Francisco Azpitarte [b,c], Emilio Gómez-Déniz [d,e]

[a] Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Australia
[b] Melbourne Institute of Applied Economics and Social Research, The University of Melbourne, Australia
[c] Brotherhood of St Laurence, Australia
[d] Department of Quantitative Methods in Economics and TiDES Institute, Spain
[e] University of Las Palmas de Gran Canaria, Spain

HIGHLIGHTS

- Two extensions of the Exponential model to describe income distributions with zero incomes are presented.
- The Exponential ArcTan (EAT) distribution and the composite EAT–Lognormal model.
- They provide a better characterization of income distributions than the Gamma model.
- The capacity of these models is shown using income data for Australia for the period 2001–2012.

ARTICLE INFO

Article history:
Received 15 February 2016
Received in revised form 13 May 2016
Available online 15 June 2016

Keywords:
Income distribution
Australia
Mixture model
Exponential distribution
Lognormal distribution
Zero Income

ABSTRACT

In this paper we propose two extensions of the Exponential model to describe income distributions. The Exponential ArcTan (EAT) and the composite EAT–Lognormal models discussed in this paper preserve key properties of the Exponential model including its capacity to model distributions with zero incomes. This is an important feature as the presence of zeros conditions the modelling of income distributions as it rules out the possibility of using many parametric models commonly used in the literature. Many researchers opt for excluding the zeros from the analysis, however, this may not be a sensible approach especially when the number of zeros is large or if one is interested in accurately describing the lower part of the distribution. We apply the EAT and the EAT–Lognormal models to study the distribution of incomes in Australia for the period 2001–2012. We find that these models in general outperform the Gamma and Exponential models while preserving the capacity of the latter to model zeros.

1. Introduction

The parametric analysis of income distributions has received considerable attention in the economics and econophysics literature. Following the pioneering work of Vilfredo Pareto [1], many functional forms have been proposed in the literature to study income distributions.[1] The statistical performance of these models will depend on the features of the data and the

* Corresponding author.
E-mail address: ecalderin@unimelb.edu.au (E. Calderín-Ojeda).

1 This includes the Lognormal distribution, the Exponential law, as well as more complex models with more parameters such as the Singh–Maddala, the Gamma, and the Generalized Beta of the Second Kind. For a detailed discussion of these models and its application to the analysis of income distributions see Ref. [2].

http://dx.doi.org/10.1016/j.physa.2016.06.047

capacity of the model to capture those characteristics. In particular, the choice of the parametric model is highly influenced by the presence of observations with zero incomes. Largely overlooked in the income distribution literature, the presence of zeros rules out the possibility of using models which have been proven to give a good fit to income data like the Lognormal, Gamma or GB2 as these models do not include the zero in their domains. Many researchers overcome this problem by excluding the observations with zero incomes from the analysis and assuming that the density at zero equals zero. This approach, however, is likely to be invalid, especially when the number of zeros is large and the analyst is interested in describing the bottom part of the income distribution for the study of poverty, inequality or the polarization of incomes.

An alternative approach is to analyse income distributions using parametric models that can accommodate the zeros when fitting the model to the empirical data. This a more sensible approach as the analyst ought to make the best use of the information from the data including those observations with zero income. The Exponential model is particularly suitable option as it has positive density at zero. In fact, as Banerjee et al. [3] show using Australian data, the Exponential distribution gives a good description of most of the income distribution although it fails to capture some features of the upper part of the distribution.

This paper contributes to the existing literature by proposing two simple extensions of the Exponential model to describe income distributions: the Exponential ArcTan (EAT) distribution which is achieved by using the methodology derived in Ref. [4] or in Ref. [5] and the composite EAT–Lognormal model which is derived following the procedure given in Ref. [6]. Thus this paper adds to the limited research on modelling distributions with null or negative values. This includes the so-called Dagum Type-II distribution proposed by Dagum [7] which is a four-parameter model with positive density at zero. Clementi et al. [8] propose the κ-generalized statistical distribution, a three-parameter model with positive density at zero, to analyse the income distribution in the US. They found that this model in general outperforms models like the Singh–Maddala and Dagum type I.[2] We illustrate the suitability of the new models using income data for Australia for the period from 2001 to 2012. We fit the models to the distributions of household disposable income which include a non-trivial number of zeros. Our empirical results show that the EAT and EAT–Lognormal provide in general a better fit to the data than the Gamma and the Exponential models. Importantly, this is achieved without significantly increasing the number of parameters of the model which makes these models particularly attractive to model income distributions in the presence of zeros.

The rest of the paper is organized as follows. In Section 2 we present the new models and their most relevant properties. Section 3 discusses the application of the new models to study changes in the distribution of household disposable income in Australia for the period 2001–2012. In the first part of this section we describe the data sources used for the analysis. We then present the main results derived from the empirical application and we discuss the advantages of the models introduced in this paper with respect to other parametric models widely used for the analysis of income distributions. Finally, Section 4 includes the conclusions and some issues for further research.

2. Parametric models

2.1. The exponential ArcTan distribution

Let us initially consider the half-Cauchy distribution (see Ref. [11]) truncated at $\alpha > 0$ with probability density function (pdf) given by

$$f(y) = \frac{1}{\tan^{-1}\alpha}\frac{1}{1+y^2}, \quad 0 < y < \alpha. \tag{1}$$

Now, let $\bar{F}(x)$ be the survival function of a random variable X with support in $[a, b]$, where a and b can be finite or non-finite and consider also the transformation $y = \alpha\bar{F}(x)$. Then, the corresponding pdf of the random variable X obtained from (1) results

$$f(x; \alpha) = \frac{1}{\tan^{-1}\alpha}\frac{\alpha f(x)}{1+[\alpha\bar{F}(x)]^2}, \tag{2}$$

for $a \le x \le b$ and $\alpha > 0$. The survival function of X, derived from (2) by integration, is provided by

$$\bar{F}(x; \alpha) = \frac{\tan^{-1}(\alpha\bar{F}(x))}{\tan^{-1}\alpha}. \tag{3}$$

Besides, (2) and (3) are appropriate density and survival functions, respectively when the support of the parameter α is extended to $(-\infty, \infty) - \{0\}$, satisfying that $\bar{F}(x; \alpha) = \bar{F}(x; -\alpha)$. Additionally, by taking in (3) limit when α approaches to zero and applying L'Hospital's rule, it is simple to derive that the parent survival function, $\bar{F}(x)$, is obtained as a limiting case. In particular, when $\bar{F}(x)$ is replaced by the survival function of the exponential distribution, the Exponential ArcTan (EAT) distribution is obtained. The family of survival functions in (3) has been recently applied to the classical Pareto distribution

[2] Clementi et al. [9] use the same model to study the distribution of wealth in the US. For a review of the parametric models that have been proposed for the analysis of wealth distributions see Ref. [10, Ch. 4].