ARTICLE IN PRESS

Physica A ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Physica A

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

New composite distributions for modeling industrial income and wealth per employee

Martin Wiegand, Saralees Nadarajah *

School of Mathematics, University of Manchester, Manchester M13 9PL, UK

HIGHLIGHTS

- A three part composite distribution is proposed for modeling income and wealth.
- The proposed distribution gives 60 percent reduction in squared error compared to Soriano-Hernández et al. (2017).
- The proposed distribution gives 50 percent reduction in absolute error compared to Soriano-Hernández et al. (2017).

ARTICLE INFO

Article history:
Received 10 June 2017
Received in revised form 2 October 2017
Available online xxxx

Keywords:
Composite distribution
Forbes Global 2000
Income
Wealth

ABSTRACT

Forbes Magazine offers an annual list of the 2000 largest publicly traded companies, shedding light on four different measurements: Sales, profits, market value and assets held. Soriano-Hernández et al. (2017) modeled these wealth metrics using composite distributions made up of two parts. In this note, we introduce different composite distributions to more accurately describe the spread of these wealth metrics.

1. Introduction

In this note, we investigate the Global 2000 data set, compiled by Forbes Magazine. It features the 2000 largest companies, with their most important financial indicators, namely annual profits, sales, market value and assets along with employee count and a ranking system based on a combined value.

An analysis was published by Soriano-Hernández et al. [1], where two part models were introduced along with a number of different distribution combinations to predict the percentage of companies below a certain wealth threshold. The tail distribution remained a Pareto distribution of type I for all estimations, combined with either a log normal, gamma or exponential distribution modeling the body part of the sample. These distributions were chosen on the grounds of previous successful modeling in finance [2–4] or modeling of gas propagation in physics [5].

The basic principle was to divide the data into two parts, and introduce a partial distribution for each part. Both distributions would then be connected by a hard cut-off, leading to a non-continuous, abrupt transition. Formally, the probability density function (PDF) and the cumulative distribution function (CDF) of the composite model can be specified by

$$f(x) = \begin{cases} f_1(x), & \text{if } x \leq \theta, \\ [1 - F_1(\theta)] f_2(x), & \text{if } \theta < x, \end{cases}$$

* Corresponding author.
E-mail address: Saralees.Nadarajah@manchester.ac.uk (S. Nadarajah).

https://doi.org/10.1016/j.physa.2017.11.106

and

$$F(x)=\begin{cases}F_1(x), & if x\leq\theta,\\ F_2(x)+F_1(\theta)-F_1(\theta)F_2(x), & if \theta<x,\end{cases}$$

respectively.

Rather than modeling the values provided directly, it was decided that a quotient of a metric and the employee count would be more expressive, giving insight into how much returns the employees generate. This led to the assumption that all observed businesses can be divided into two categories, depending on the number workforce employed. The first category would be companies in retails like Wal-Mart, which have to rely on large numbers of employees for their services. On the other hand we have companies like Apple, which due to the nature of their products and production processes can perform with comparatively low employee number in relation to their revenue.

The focus on the application of this model is less in fitting a PDF or CDF to the data, but in a decreasing percentage function $100\left[1-\widehat{F}(x)\right]$ to describe the amount of companies below a certain wealth metric. In this note, we introduce a new approach which suggests a third subpopulation. We then show that it provides a considerable improvement in fits over the old approach. We verify our results with a number of error measures reflecting the goodness of fit and comparative plots, visualizing the areas of greatest improvement.

2. Composite models

Contrary to the previously proposed model, we argue that a distinct third population group is present, leading us to the construction of a three part composite model. The best fitting two part models in Fig. 1 do not entirely capture the proper curvature of the percentage function. More strikingly, they also differ considerably in the tails. While the parts closer to the catenation point are still somewhat adequately captured, higher values stray further away. This is especially evident in the profit and sales plots, where we hypothesize that a third section at around 0.1 billion and 0.05 billion could diminish this deviation.

Furthermore, we like to introduce a smoother variant of the composite model, which in its two part form has been introduced by Bakar and Nadarajah [6]. The PDF and CDF are provided below for arbitrary distributions with PDFs f_1, f_2 and CDFs F_1, F_2 merged at point $\theta\in\mathbb{R}$ with weight $\Psi\in\mathbb{R}^+$:

$$f(x)=\begin{cases}\dfrac{1}{1+\Phi}\dfrac{f_1(x)}{F_1(\theta)}, & if x\leq\theta,\\ \dfrac{\Phi}{1+\Phi}\dfrac{f_2(x)}{1-F_2(\theta)}, & if \theta<x,\end{cases}$$

and

$$F(x)=\begin{cases}\dfrac{1}{1+\Phi}\dfrac{F_1(x)}{F_1(\theta)}, & if x\leq\theta,\\ \dfrac{\Phi}{1+\Phi}\left[1+\Phi\dfrac{F_2(x)-F_2(\theta)}{1-F_2(\theta)}\right], & if \theta<x.\end{cases}$$

We now expand this model by a third partial distribution with PDF f_3 and CDF F_3. Additionally we now mark $\theta_1<\theta_2$ as the catenation points between the composites and Ψ, Θ as the weights. This yields

$$f(x)=\zeta\begin{cases}\dfrac{f_1(x)}{F_2(\theta_1)}, & if x\leq\theta_1,\\ \Phi\dfrac{f_2(x)}{F_2(\theta_2)-F_2(\theta_1)}, & if \theta_1<x\leq\theta_2,\\ \Psi\dfrac{f_3(x)}{1-F_3(\theta_2)}, & if \theta_2<x\end{cases}\tag{1}$$

and

$$F(x)=\zeta\begin{cases}\dfrac{F_1(x)}{F_1(\theta_1)}, & if x\leq\theta_1,\\ 1+\Phi\dfrac{F_2(x)-F_2(\theta_1)}{F_2(\theta_2)-F_2(\theta_1)}, & if \theta_1<x\leq\theta_2,\\ 1+\Phi+\Psi\dfrac{F_3(x)-F_3(\theta_2)}{1-F_3(\theta_2)}, & if \theta_2<x.\end{cases}$$

For convenience, we introduce $\zeta=\frac{1}{1+\Phi+\Psi}$ as scaling parameter. Soriano-Hernández et al. [1] have proposed most prominently the gamma and log normal distributions for the body (f_1, F_1) and a Pareto type I distribution for the tail (f_2, F_2). We mostly agree with the choice of distributions being used for the body component, albeit introducing a beta Weibull

Please cite this article in press as: M. Wiegand, S. Nadarajah, New composite distributions for modeling industrial income and wealth per employee, Physica A (2017), https://doi.org/10.1016/j.physa.2017.11.106.