Physica A 390 (2011) 689–698

Contents lists available at ScienceDirect

Physica A

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

The Gompertz–Pareto income distribution

F. Chami Figueira [a], N.J. Moura Jr. [b], M.B. Ribeiro [c,*]

[a] *Chemical School, Federal University of Rio de Janeiro-UFRJ, Rio de Janeiro, Brazil*

[b] *Brazilian Institute for Geography and Statistics-IBGE, Rio de Janeiro, Brazil*

[c] *Physics Institute, Federal University of Rio de Janeiro-UFRJ, CxP 68532, Rio de Janeiro, RJ 21941-972, Brazil*

ARTICLE INFO

Article history:
Received 7 May 2010
Received in revised form 22 July 2010
Available online 20 October 2010

Keywords:
Income distribution
Pareto power law
Gompertz curve
Brazil's income data
Fractals

ABSTRACT

This work analyzes the Gompertz–Pareto distribution (GPD) of personal income, formed by the combination of the Gompertz curve, representing the overwhelming majority of the economically less favorable part of the population of a country, and the Pareto power law, which describes its tiny richest part. Equations for the Lorenz curve, Gini coefficient and the percentage share of the Gompertzian part relative to the total income are all written in this distribution. We show that only three parameters, determined by linear data fitting, are required for its complete characterization. Consistency checks are carried out using income data of Brazil from 1981 to 2007 and they lead to the conclusion that the GPD is consistent and provides a coherent and simple analytical tool to describe personal income distribution data.

1. Introduction

The mathematical characterization of income distribution is an old problem in economics. Vilfredo Pareto [1] was the first economist to discuss it in quantitative terms and it bears his name the law he found empirically in which the tail of the cumulative income distribution, formed by the richest part of the population of a country, follows a power law pattern. Since then, the *Pareto power law* for income distribution has been verified to hold universally, for various countries and epochs [2]. Despite the empirical success of this law, the characterization of the lower income region, representing the overwhelming majority of the population in any country, remained an open problem. Various functions with an increasing number of parameters were proposed by economists to represent the lower part, or the whole, of the income distribution [3]. However, no consensus emerged on what would be the most suitable way of representing the whole income distribution of countries.

In the middle 1990s physicists became interested in problems which until then were considered the exclusive realm of economists. Econophysicists approached these problems in a data driven mode [4–7], that is, with none, or little, consideration to the typical neoclassical economics mindframe in which axiomatic, some would say ideological [8,9], considerations take precedence over real data [6,10]. Ignoring this empirically flawed mindset [11–18], efforts have been made by econophysicists, helped later by a few non-representative economists, to carefully study real data of economic nature. This gave a new impetus to the income distribution problem due to an emerging body of fresh results, as well as hints from statistical physics on how it could be dynamically modeled [19].

Drăgulescu and Yakovenko [20], Christian Silva [21], Yakovenko and Rosser [22] advanced an exponential type distribution of individual income similar to the Boltzmann–Gibbs distribution of energy in statistical physics. Chatterjee et al. [23] discussed an ideal gas model of a closed economic system where total money and agents' number are fixed.

* Corresponding author. Fax: +55 21 2562 7368.
E-mail addresses: fernanda@chami.com.br (F. Chami Figueira), newtonjunior@ibge.gov.br (N.J. Moura Jr.), mbr@if.ufrj.br (M.B. Ribeiro).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2010.10.014

Clementi et al. [24–26] proposed the *k*-generalized distribution as a descriptive model for the size distribution of income, based on considerations of statistical physics. Willis and Mimkes [27] used log-normal and Boltzmann distributions to argue in favor of a separate treatment of waged and unwaged income. Moura Jr. and Ribeiro [28] showed that the Gompertz curve combined with the Pareto power law provide a good descriptive model for the whole income distribution and where the exponential appears as an approximation for the middle portion of the individual income data. In this model the Gompertz curve represents the overwhelming majority of the economically less favorable part of the population, whereas the Pareto law describes the richest part.

Regarding the related phenomenon of wealth distribution, related because income and wealth are not the same quantity and, therefore, should not be confused (see Ref. [28] and Section 4), Solomon [29] argued that a power law wealth distribution implies Levy-flights returns, whereas Bouchaud and Mézard [30] reached a Pareto power law wealth distribution in a model containing exchange between individuals and random speculative trading. Solomon and Richmond [31] used a generalized Lotka–Volterra model to show that the wealth distribution among individual investors fulfills a power law, Repetowicz et al. [32] studied a model of interacting agents that allows agents to both save and exchange wealth, Coelho et al. [33] revealed the existence of two distinct power law regimes in wealth distribution, one for the super-rich and another with smaller Pareto exponents for the top earners in income data sets, and Scafetta et al. [34] used a two-part function stochastic model to discuss trade and investment dynamics of a society stratified into two distinct classes (more on this in Section 4). Further references on income and wealth distribution can be found in Ref. [22], as well as in Refs. [28,35].

The aim of this paper is to discuss further the model advanced by Moura Jr. and Ribeiro [28]. We show here that this combined model, named *Gompertz–Pareto distribution* (GPD), provides a simple way of modeling income distribution since it is formed by simple functions and is fully characterized by three positive parameters which can be determined by linear data fitting. We discuss simple consistency tests in order to ascertain whether or not the results produced by the model can recover basic features of the original distribution, namely the Lorenz curves, the Gini coefficients and the percentage share of the Gompertzian population relative to the total income of the country. We conclude that the GPD is consistent and provides a coherent and conveniently very simple way of modeling income data.

The GPD is a power law tailed distribution and, as such, it is likely to have a larger set of applications than just income distribution. This is so because a very wide range of observed phenomena in physical, biological and social sciences are known to be described by power law tailed distributions. For instance, in physical sciences this is the case of galaxy distribution [36,37], relativistic cosmology [38–43] and turbulence [44]. In human activities these distributions are found in citation of scientific papers [45], intensity of wars [46] and their military and civilian casualties [47,48], population of cities [49] and stock prices [50]. In biological sciences, power law tailed distributions were found in botany [51], genomics [52] and branching networks of biological systems [53]. Refs. [54,55] provide several other examples of physical, biological and social systems exhibiting power law tailed distributions. The Gompertz curve is known to be a good descriptor of population dynamics, mortality rate and growth processes in biology (see [28], and the references therein). Therefore, a system whose distribution is characterized by the combination of the Gompertz curve and a power law tail suggests that growth may possibly be one of the main dynamical components of its underlying complex system dynamics.

The plan of the paper is as follows. In Section 2 we review the basic equations for modeling income distribution data. Section 3 presents the equations for the GPD of individual income and extends the model to describe the most basic descriptive tools used to measure income inequality, namely the Lorenz curve and the Gini coefficient. We also discuss how the GPD has an exponential type behavior in its middle part. Section 4 applies the model to the income data of Brazil from 1981 to 2007 and also presents new results not available in Ref. [28]. Consistency checks are provided by re-obtaining the Lorenz curves, Gini coefficients and the percentage share of the Gompertzian part of the distribution. These are derived from the parameters of the model and compared with the original, not model based, equivalent results. It is shown that the results coming from the GPD present self-consistency. Section 5 ends the paper with the conclusions.

2. Basic equations

This section reviews very briefly the most essential quantities and functions necessary for the analytical description of the individual income distribution. We followed the comprehensive treatment provided by Ref. [2], although a slightly different notation and normalization was adopted to match similar choices made in Ref. [28].

Let us define $\mathcal{F}(x)$ to be the *cumulative income distribution* giving the probability that an individual receives an income less than or equal to x. Then the *complementary cumulative income distribution* $F(x)$ gives the probability that an individual receives an income equal to or greater than x. It follows from these definitions that $\mathcal{F}(x)$ and $F(x)$ are related as follows,

$$\mathcal{F}(x) + F(x) = 100, \tag{1}$$

where the maximum probability was taken as 100%. Here x is a normalized income, obtained by dividing the nominal, or real, income values by some suitable nominal income average [28]. If both functions $\mathcal{F}(x)$ and $F(x)$ are continuous and have continuous derivatives for all values of x, we have that,

$$\mathrm{d}\mathcal{F}(x)/\mathrm{d}x = f(x), \qquad \mathrm{d}F(x)/\mathrm{d}x = -f(x), \tag{2}$$

and

$$\int_0^\infty f(x)\mathrm{d}x = 100, \tag{3}$$