Physica A 392 (2013) 2130–2138

Contents lists available at SciVerse ScienceDirect

Physica A

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

Modelling of income distribution in the European Union with the Fokker–Planck equation

Maciej Jagielski *, Ryszard Kutner

Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Hoża 69, PL-00681 Warszawa, Poland

ARTICLE INFO

Article history:
Received 16 August 2012
Received in revised form 15 November 2012
Available online 18 January 2013

Keywords:
Income distribution
Langevin equation
Yakovenko model

ABSTRACT

Herein, we applied statistical physics to study incomes of three (low-, medium- and high-income) society classes instead of the two (low- and medium-income) classes studied so far. In the frame of the threshold nonlinear Langevin dynamics and its threshold Fokker–Planck counterpart, we derived a unified formula for description of income of all society classes, by way of example, of those of the European Union in years 2006 and 2008. Hence, the formula is more general than the well known formula of Yakovenko et al.. That is, our formula well describes not only two regions but simultaneously the third region in the plot of the complementary cumulative distribution function vs. an annual household income. Furthermore, the known stylised facts concerning this income are well described by our formula. Namely, the formula provides the Boltzmann–Gibbs income distribution function for the low-income society class and the weak Pareto law for the medium-income society class, as expected. Importantly, it predicts (to satisfactory approximation) the Zipf law for the high-income society class. Moreover, the region of medium-income society class is now distinctly reduced because the bottom of high-income society class is distinctly lowered. This reduction made, in fact, the medium-income society class an intermediate-income society class.

1. Introduction

For over two decades, physics oriented approaches have widely been developed to explain different economic processes [1–7] (and Refs. therein). Those approaches aim at formulating well fitted unbiased indicators of social and economic phenomena. One of their key issues is the income of society analysis using methods of statistical physics. The main goal of this economic issue is to unravel and describe mechanisms of societies' enrichment or impoverishment.

The first successful attempt in this socio-economic field was made by the legendary economist and sociologist Vilfredo Pareto [7–9]. He demonstrated that the distribution functions of individual incomes in different countries within stable economy are universal being manifested by a power law. This law is called the weak Pareto law. He emphasised that this law could not resemble the distribution functions obtained if the gain and accumulation of income were random. As a possible origin of this law, Pareto indicated a self-similarity structure of societies.

Pareto's economic discoveries initiated attempts of analytical descriptions of incomes of the societies and inspired an avalanche of related research works [2,3,5–7,10–24]. Among them, particularly significant are those of the economist Robert Gibrat [7,10,11]. He found that the complementary cumulative distribution function of the Pareto distribution is insufficient to describe empirical data within the whole range of the income. Trying to find a functional form that could account for these data, he proposed a rule called the Rule of Proportionate Growth [10,11] (see below for details).

* Corresponding author. Tel.: +48 22 5533229; fax: +48 22 5532410.
E-mail address: zagielski@interia.pl (M. Jagielski).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2013.01.028

Furthermore, the income of societies was analysed by David Champernowne, who constructed a stochastic model simulating the Pareto power law [12] and also by Benoit Mandelbrot who described several useful properties of random variables subject to the Pareto distribution [8,9].

In the recent decade, a large number of studies were performed aiming at construction of models, which (to some extent) would well replicate the observed complementary cumulative distribution functions of individual incomes. Among them, the most significant seems to be the Clementi–Matteo–Gallegati–Kaniadakis approach [17], the Generalised Lotka–Volterra Model [5–7], the Boltzmann–Gibbs law [13–16], and the Yakovenko et al. model [2,3]. Very recently, a mathematical model similar to that of Yakovenko et al. has been developed [21]. It involves complex economic justification for microscopic stochastic dynamics of wealth. However, none of the above attempts to find an analytical description of the income structure solves the principal challenges, which concern:

(i) the description of the annual household incomes of all society classes (i.e. the low-, medium-, and high-income society classes) by a single unified formula and
(ii) the problem regarding corresponding complete microscopic (microeconomic) mechanism responsible for the income structure and dynamics.
In our considerations presented herein, we used Boltzmann–Gibbs law and Yakovenko et al. model to derive a uniform analytical formula describing income of all three society classes.

2. Extended Yakovenko et al. model

In accord with an effort outlined above, we compared the empirical data of the annual household incomes in the European Union (EU), including Norway and Iceland, with predictions of our theoretical approach proposed herein. This approach is directly inspired by the Yakovenko et al. model. By using the same assumptions, however, we generalised this model to solve the principal challenges (i) and (ii) indicated above.

We used data records from the Eurostat Survey on Income and Living Conditions (EU-SILC) [25], by way of example for years 2006 and 2008 [26,27] (containing around 150 and 200 thousand empirical data points, respectively). However, these records contain (as all other records) only few data points concerning the high-income society class, i.e. the third region in the plot of the complementary cumulative probability distribution function vs. annual household income. To consider the high-income society class systematically, we additionally analysed the effective income of billionaires in the EU by using the Forbes 'The World's Billionaires' rank [28]. The term 'billionaire' used herein is equivalent (as in the US terminology) to the term 'multimillionaire' used in the European terminology. Since we consider wealth and income of billionaires in euros, we recalculated US dollars to euros by using the mean exchange rate at the day of construction of the Forbes 'The World's Billionaires' rank.

We were able to consider incomes of three society classes thanks to the following procedure.

(i) Firstly, we selected EU billionaires' wealth from the Forbes 'The World's Billionaires' rank, for instance, for four successive years 2005–2008.
(ii) Secondly, we calculated the corresponding differences between billionaires' wealth for the successive years. We assumed that their incomes are, in fact, proportional to these differences. For instance, we calculated the billionaire incomes for year 2006 by taking the difference between their wealth in years 2006 and 2005. We made it analogously for year 2008. However, we took into account only billionaires who gained effective incomes (neglecting those, who suffered from income losses).
(iii) Subsequently, having so calculated incomes for the high-income society class, we joined them (separately for years 2006 and 2008) with the corresponding EU-SILC datasets. By using so completed datasets, we then constructed the initial empirical complementary cumulative distribution function for years 2006 and 2008. For that, we used the well known Weibull recipe (see below for details). However, this direct approach shows a wide gap of incomes inside the high-income society class resulting in a horizontal line of the complementary cumulative distribution function. This gap separates the first segment belonging to the high-income society class, consisting of all data points taken from the EU-SILC dataset (only 8 for 2006 and 6 for 2008), from the second segment, consisting of remaining data points (76 for 2006 and 96 for 2008), which also belong to the high-income society class but are taken from the Forbes dataset.
(iv) In the final step, we eliminated this gap by adopting the assumption that the empirical complementary cumulative distribution functions (concerning the whole society) have no horizontal segments. That is, we assumed that statistics of incomes is a continuous function of income. Hence, we were forced to multiply the billionaire incomes from Forbes dataset by the properly chosen common proportionality factor. This factor was equal to 1.0×10^{-2} for both years, as we assumed the requirement of full overlap of the first (above mentioned) segment by the second segment. This assumption leads to a unique solution (up to some negligible statistical error) for this proportionality factor. We found that this factor was only a slowly-varying function of time (or years).

Hence, we received the data record containing already a sufficient number of data points for all society classes, including the high-income society class. Although the Forbes empirical data only roughly estimate the wealth of billionaires, they quite well establish the billionaires' rank, thus sufficiently justifying our approach. This is because our purpose is to classify billionaires to concrete universality class rather than finding their total incomes. Our procedure of linking data from two different bases does not violate this universality class.