Income distribution patterns from a complete social security database

N. Derzsy [a], Z. Néda [a,b,*], M.A. Santos [c]

[a] Department of Theoretical and Computational Physics, Babeş-Bolyai University, Kogălniceanu street 1, RO-400080, Cluj-Napoca, Romania
[b] Interdisciplinary Computer Modelling Group, Hungarian University Federation of Cluj, Baba Novac street 23/2, RO-400080, Cluj-Napoca, Romania
[c] Departamento de Física e Astronomia and Centro de Física do Porto, Faculdade de Ciências, Universidade do Porto, Rua de Campo Alegre s/n, 4169-007 Porto, Portugal

ARTICLE INFO

ABSTRACT

We analyze the income distribution of employees for 9 consecutive years (2001–2009) using a complete social security database for an economically important district of Romania. The database contains detailed information on more than half million taxpayers, including their monthly salaries from all employers where they worked. Besides studying the characteristic distribution functions in the high and low/medium income limits, the database allows us a detailed dynamical study by following the time-evolution of the taxpayers income. To our knowledge, this is the first extensive study of this kind (a previous Japanese taxpayers survey was limited to two years). In the high income limit we prove once again the validity of Pareto's law, obtaining a perfect scaling on four orders of magnitude in the rank for all the studied years. The obtained Pareto exponents are quite stable with values around $\alpha \approx 2.5$, in spite of the fact that during this period the economy developed rapidly and also a financial-economic crisis hit Romania in 2007–2008. For the low and medium income category we confirmed the exponential-type income distribution. Following the income of employees in time, we have found that the top limit of the income distribution is a highly dynamical region with strong fluctuations in the rank. In this region, the observed dynamics is consistent with a multiplicative random growth hypothesis. Contrarily with previous results obtained for the Japanese employees, we find that the logarithmic growth-rate is not independent of the income.

1. Introduction

At the end of the twentieth century the research in many modern fields of science condensed around complex systems. Statistical physicists got involved in such topics of interdisciplinary research, applying classical statistical physics methods and models to understand the structure and evolution of such systems. Many physicists chose to study systems of economic nature, like trade relations, economic transactions, wealth distributions, company interdependencies, etc. These researches led to the development of econophysics, which became a modern interdisciplinary field [1] of statistical physics. In the present period of global financial and economic turmoil, this topic is of great interest for everyone who is preoccupied with understanding the complex behavior of our society and wants to speculate about the future. One interesting and much debated problem in the field of econophysics, is to understand and model the wealth and income distribution in our society. With this aim, much data has been collected and many models have been elaborated on. An excellent overview of the current standing in this field is given in the review paper of Yakovenko and Rosser [2]. Our research intends to contribute to this

* Corresponding author at: Department of Theoretical and Computational Physics, Babeş-Bolyai University, Kogălniceanu street 1, RO-400080, Cluj-Napoca, Romania. Tel.: +40 745 310531; fax: +40 264 591906.
E-mail address: zneda@phys.ubbcluj.ro (Z. Néda).

field by analyzing a huge and complete social security database from Romania. The data set contains monthly information regarding employers with head-office registered in the Cluj district, their employees and the paid salaries for years between 2001 and 2009. This codified social security dataset offers a unique possibility to study the individual income distribution and dynamics on a relatively long period. The aim of the present study is to investigate the dynamics of the income for the top taxpayers, and to find experimental evidences that would help future modeling attempts for the quite general Pareto-type distribution of their wealth.

Vilfredo Pareto, an Italian economist, observed that the wealth distribution in societies obeys a general law, that later became known as Pareto's law in honor of his work [3]. His observation states that the cumulative income or wealth distribution for the richest 3%–5% part of several countries and cities in the XV–XIX century Europe follows a power-law. The (modulus of the) exponent of this power-law is denoted by α and it is named the Pareto exponent. Nowadays, when most of the financial and economic data are available in electronic format, many recent studies confirmed his predictions with much improved statistics. Personal income and wealth distribution studies coming from India [4], Japan [5,6], UK [7], continental Europe [8], USA [9,10] and Brazil [11] support this conjecture. Indirect wealth data originating from history, such as ancient Egypt [12] or a medieval society [13], also confirm the universality of the Pareto's law. Experimental data revealed also that in the limit of low and medium wealth/income region the shape of the cumulative distribution is well-fitted by either an exponential or a log-normal function [14,15].

Measuring the wealth is a problematic task, since it includes several types of incomes and properties and, more importantly, these data are of a highly confidential nature. Thus studying wealth-distribution is more difficult and often leads to inaccurate results caused by incomplete databases. Although wealth and income are related and both of them show Pareto's principle, the distribution of the two quantities have clearly separable Pareto exponents. The distribution of wealth is usually broader than the distribution of income, or equivalently, the Pareto index for wealth distribution is usually smaller than the corresponding one for income [4]. More specifically, the measured α values for the individuals income distribution span a quite broad interval, typically in the 1.5–2.8 range, while studies focusing on the wealth distribution show a smaller Pareto index, usually in the 0.8–1.5 interval [16]. This large variation of α indicates the absence of universal scaling in this problem, a feature which modeling efforts have to take into account and reproduce.

Several theoretical models were created aiming to reproduce the observed distributions. In case of wealth distribution, the most popular models are agent-based approaches where the wealth of agents varies in a multiplicative and random manner and they can interchange money following pre-established rules. The interaction between the agents can be either local or global. Such models were successful in reproducing many features of the wealth-distribution curves [17–19]. Asset exchange models are also very popular nowadays [20–22]. Trade is the crucial ingredient of these models, and is taken into consideration by the fact that pairs of randomly chosen agents exchange part of their money while saving the remaining fraction [23–26]. For randomly distributed and quenched saving factors, a Pareto-type wealth distribution with $\alpha = 1$ exponent is found [27]. Variants of this model considering asymmetric exchanges are able to generate $\alpha < 1$ Pareto exponents as well [24]. These models are thus able to explain different Pareto index values by varying the free parameters in the wealth-exchange rule. Due to the complex structure of the underlying social networks on which the wealth-exchange is realized, researchers proposed models implementing the network approach. In such models the economic interactions between agents take place on small-world or scale-free network topologies [28,29]. A successful approach in such direction is the model based on first-degree family relations networks that successfully generates both a realistic wealth distribution and a social network topology [30]. For a more complete review of wealth/income-distribution models we recommend again the review paper of Yakovenko and Rosser [2].

The income distribution was also modeled by means of statistical physics approach. The Fokker–Planck equation can be applied for describing the time-evolution of the income distribution function [10]. In order to get a stationary solution for the income distribution, one has to postulate how the income changes in time (ΔW) as a function of the present income value W. If for the low income region, it is assumed that ΔW is independent of W (additive diffusion), while for the top income class one considers $\Delta W \propto W$ (multiplicative diffusion), one gets the right exponential distribution for the low and medium income region and the power-law distribution for the high income limit. A combination of additive and multiplicative processes has also been studied [31]; both deterministic and random [32] growth rules have been considered. Multiplicative growth is usually associated with income from bonuses, investments and capital gain; salaries can increase (decrease?) by a constant (merit bonus) or proportionally (cost of living raise, in percentage).

In order to have experimental evidence for the ΔW versus W dependence assumed in theoretical models, exhaustive data for several consecutive years are needed, where one can clearly identify and follow the income of all individuals. This type of information is not easily available—as far as we know, such studies have only been performed for two consecutive years in Japan [33,34]. The Japanese researchers concluded that the distribution of the growth rate in one year is roughly independent of income in the previous year. This finding supports the generality of Gibrat's law [35], which in its original form stated that the size of a firm and its growth rate are roughly independent of each other. In the most simple form Gibrat's law of proportionate growth leads to a log-normal distribution. In Refs. [33,34] the authors also argue, that this independency, combined with an approximate time-reversal symmetry, leads to Pareto law, but such a claim has been formally proven to be false [36]. The dataset available for us, spanning 9 consecutive years and having around half million taxpayers, offers excellent possibilities to reconsider with much better accuracy the ΔW versus W relationship, and to bring new evidences supporting or disproving the results obtained for Japan.