Proceedings of the 5th Symposium on Physics in Economics and Social Sciences, Warszawa, Poland, November 25–27, 2010

Preliminary Comparison of Households' Income in Poland with European Union and United States Ones by Using the Statistical Physics Methods

M. Jagielski[a*], R. Kutner[a] and M. Pęczkowski[b]

[a]Institute of Experimental Physics, Faculty of Physics, University of Warsaw
Smyczkowa 5/7, PL-02-678 Warsaw, Poland

[b] Faculty of Economic Sciences, University of Warsaw, Długa 44/50, PL-00-241 Warsaw, Poland

In this work we compared the empirical data of annual income of Polish and European households as well as annual income of individuals in United States (e.g. for years 2006 and 2008) with predictions of the most popular theoretical models. Particularly good agreements with Pareto distribution and prediction of the Yakovenko model were obtained. For the low society class well agreement with prediction of the cumulative exponential distribution was gained. However, it turned out that the cumulative distribution of annual income of Polish households can be described quite well by the Generalised Lotka-Volterra model.

PACS: 89.20.-a, 89.65 Gh

1. Introduction

For more than two decades, physics oriented approaches have been developed to explain economic phenomena and processes [1–6]. In this work we compared empirical data for annual income of European Union (EU, including Norway and Iceland) and Polish (PL) households with predictions of theoretical models. For a greater completeness we also analysed empirical data of annual income of individuals in United States of America (US).

These models are as follows:

- Boltzmann-Gibs formula,
- Pareto distributions,
- Yakovenko et al [1, 2] representation of the nonlinear Langevin equation,
- Generalised Lotka-Volterra model [3, 4].

We hope that such a varied approach will be helpful in understanding how wealth and income are generated and accumulated.

The principal aim of this work is to show that households' income of low and middle society classes in EU can be described analogously as it Yakovenko et al [1, 2] made for United States.

In this work we used empirical data both from the Polish Central Statistical Office and Eurostat [7, 8] concerning, for instance, annual households' income in 2006 and 2008. For additional comparison we used data from Internal Revenue Service (IRS), the government tax agency [9].

Notably, the empirical data are referring to:

- the total household gross income for EU,
- the household disposable income for Poland,
- adjusted gross income of individual person for US.

2. Models of income distributions and results

As usual, the basic tool of the analysis was empirical cumulative distribution. Income data was sorted descending from the richest to poorest household (i.e. according to the rank). Then using the ratio $\frac{k}{n+1}$, where k is the position of the household in the rank and n is the sample size, it was determined which fraction of households has income that is greater than the household's income in a given position in the rank. This distribution is much more stable than the initial probability density and ensure number of points the same as in the empirical data record.

At the first stage, we fitted to empirical cumulative distribution the weak Pareto law (for the middle-income households) given by formula [5] (plots in log-log scale in Figs. 1 and 3a):

$$\Pi(m) \approx (m/m_0)^{-\alpha}. \quad (1)$$

Here m_0 is a scaling factor and α is a Pareto exponent which value is determined by shift and slope of empirical data, respectively.

As it is seen from Figs. 1 and 3a, the weak Pareto law well describes the "bulk" of analysed distributions. Similar results were obtained for individuals in US for 2006 but with essentially smaller $\alpha = 1.36$ [2].

By taking into account only the richest households we fitted to the empirical data again the weak Pareto law (plots in log-log scale in Figs. 2 and 3b). These households are described by the weak Pareto law with an exponent α close to unity. In this case, the people forming the household are usually company owners, whose profits are described indeed by the Zipf law.

* e-mail: zagielski@interia.pl

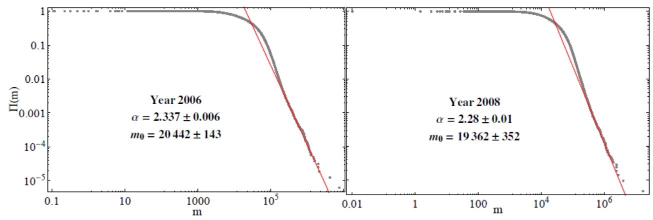

Fig. 1. Fit of the weak Pareto law (solid line) to the EU middle households' income empirical data set (dots) for 2006 ($\alpha = 2.337 \pm 0.006$) and 2008 ($\alpha = 2.28 \pm 0.01$) [7, 8].

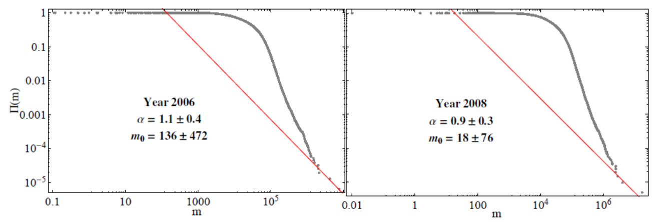

Fig. 2. Fit of the weak Pareto law (solid line) to the EU richest households' income empirical data set (dots) for 2006 ($\alpha = 1.1 \pm 0.4$) and 2008 ($\alpha = 0.9 \pm 0.3$) [7, 8].

On the basis of the kinetic theory, by solving the Fokker-Planck equation, the emergence of Pareto's law is the result of the assumption that changes in the richest households' income are proportional to income gained so far. This thesis is valid because for the richest profits come mainly from investments and capital gain. This type of stochastic process is called the multiplicative stochastic process [1, 2].

We found that the Pareto exponent α can be considered as an indicator of social inequalities. The smaller the value of α, the bigger the social stratification.

We also fitted the cumulative exponential distribution resulting from the Boltzmann-Gibs formula [11–13] (plots in log-log scale shown in Fig. 4):

$$\Pi(m) = \exp\left(-\frac{m}{T}\right). \tag{2}$$

Above distribution is characterised by a single parameter that is, an income temperature T [1], which can be interpreted as an average income per household [13].

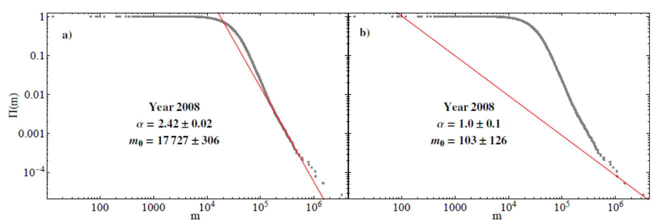

Fig. 3. Fit of the weak Pareto law (solid line) to the Polish a) middle households' ($\alpha = 2.42 \pm 0.02$) and b) richest households ($\alpha = 1.0 \pm 0.1$) income empirical data set (dots) for 2008. Results for 2006 were already published in [10].

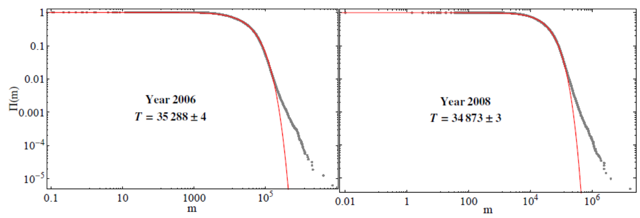

Fig. 4. The fit of exponential function (solid line) to the EU households' income empirical data set (dots) for 2006 ($T = 35288 \pm 4$ Euros) and 2008 ($T = 34873 \pm 3$ Euros) [7, 8].

The cumulative exponential distribution function quite well describes European poor-income households (cf. plots in Fig. 4). Similar results were obtained with $T = 46000\$$ for individuals in US, for instance, for 2006 [2].

On the kinetic theory approach again, by solving the Fokker-Planck equation, the emergence of the Boltzmann-Gibbs law is supported by the assumption that changes in incomes of the poor and middle-income households are independent of the income gained so far. These households receive income mainly in the form of wages and salaries, which justifies the above assumption. Stochastic process associated with that kind of mechanism of changes in income is called the additive stochastic process [1, 2].

At the second stage, we fitted to the data the cumulative distribution of the probability density

$$P(m) = c\frac{e^{-(m_0/T)\arctan(m/m_0)}}{[1+(m/m_0)^2]^{(\alpha+1)/2}}, \tag{3}$$

proposed by Yakovenko [1, 2]. The corresponding plots in a logarithmic scale were shown in Figs. 5ab and 6. The probability density (3) has not a cumulative distribution in an analytical form. Fortunately, we can by numerical way calculate and fit to empirical data the cumulative distribution. For this purpose, the parameters α and T were determined by corresponding fits of the weak Pareto law and Boltzmann-Gibbs formula, respectively The value of m_0 can be obtained by approximating the point of intersection of both cumulative distributions arising from these fits. The constant c is a normalisation constant $\int_0^\infty P(m)dm = 1$.

The model proposed by Yakovenko well describes the empirical cumulative distributions of income of households in the European Union, as well as the income of individuals in the United States. This results from the assumptions of the model, which allow the coexistence of the additive and multiplicative processes. The distribution described by equation (3) for small m behaves like an exponential distribution, while for large m as the weak Pareto law. The value of m_0 designates the transition point between these two distributions.

Furthermore, we also obtained well agreement between our empirical data for Poland and the cumulative distri-

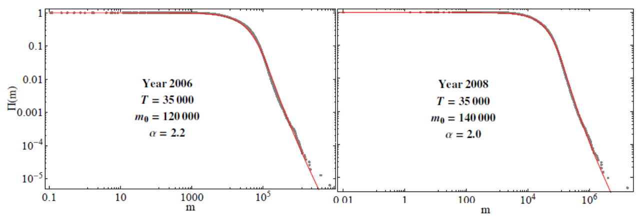

Fig. 5. Fit of the Yakovenko's model (solid line) to the EU households' income empirical data set (dots) for 2006 (T = 35000 Euros, m_0 = 120000 Euros, $\alpha = 2.2$) and 2008 (T = 35000 Euros, m_0 = 140000 Euros, $\alpha = 2.0$) [7, 8].

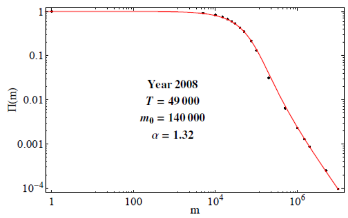

Fig. 6. Fit of the Yakovenko's model (solid line) to the United States individuals' income empirical data set (dots) for 2008 (T = 49000 Euros, m_0 = 120000 Euros, $\alpha = 1.32$) [9]. Results for 2006 can found in [2].

bution function given by the Generalised Lotka-Volterra model (cf. plots in Fig. 7) [3, 4]:

$$\Pi(x) = 1 - \frac{\Gamma(\alpha, \frac{\alpha-1}{x})}{\Gamma(\alpha)}, \qquad (4)$$

here α is the shape parameter, which describes fitted function and $x = m/ < m >$ is the relative income of households (where $< m >= \sum_{i=1}^{N} m_i$).

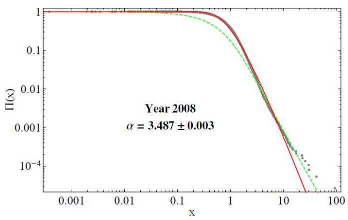

Fig. 7. Fit of Generalised Lotka-Volterra cumulative distribution function (solid line; $\alpha = 3.487 \pm 0.003$) and the Yakovenko's (dashed line; T = 41000 PLN, m_0 = 50000 PLN, $\alpha = 2.4$) model to the Polish households' income empirical data set (dots) for 2008. Results for 2006 can found in [10].

Generalised Lotka-Volterra model describes quite well the cumulative distribution of annual households' income; while Yakovenko prediction is unsatisfactory (cf. dashed line in Fig. 7). An important advantage of this model is the ability to characterise the empirical distribution using a single function. It also offers valuable theoretical approach on the microscopic level, where households income is determined by the revenue gained so far, the social security benefits (in general, redistribution of revenues in society) and the general state of economy [3, 4, 10].

3. Conclusions

In this paper we analysed empirical cumulative distribution functions of annual income of households in Poland and European Union, mainly for years 2006 and 2008. We also focused on empirical cumulative distribution function of annual income of individuals in United States for year 2008. It turned out that:

- for EU and US they can be well described by cumulative exponential distribution for poor households, as well as the weak Pareto law for the middle-income and rich households,
- for EU and US the poor and middle-income households can be described over the entire range by model proposed by Yakovenko,
- in case of Poland the Generalised Lotka-Volterra model gives a good description of the empirical cumulative distribution function almost over the entire range

The challenge is to define the distribution which covers all ranges of the empirical data set, that is, dominating by poor, middle and rich classes of society. Although, it is a question why income of the richest class is so inaccessible.

References

[1] V. M. Yakovenko, J. B. Rosser, *Rev. Mod. Phys.* **81**, 1703 (2009).

[2] A. Banerjee, V. M. Yakovenko *New J. Phys.* **12**, 075032 (2010).

[3] S. Solomon, P. Richmond, *Eur. Phys. J. B* **27**, 257 (2002).

[4] P. Richmond, S. Solomon, *Int. J. Mod. Phys. C* **12**, 333 (2001).

[5] P. Richmond, S. Hutzler, R. Coelho, P. Repetowicz, in: *Econophysics and Sociophysics: Trends and Perspectives*, Eds. B. K. Chakrabarti, A. Chakraborti, A. Chatterjee, WILEY-VCH, Weinheim, 2006, p. 131..

[6] P. Łukasiewicz, A. Orłowski, *Physica A* **344**, 146 (2004).

[7] EUSILC UDB 2006 - version 4 of March 2010.

[8] EUSILC UDB 2008 - version 2 of August 2010.

[9] Personal income data from the Statistic of Income division of IRS `http://www.irs.gov/taxstats/indtaxstats/article/0,,id=134951,00.html` .

[10] M. Jagielski, R. Kutner, *Acta Phys. Pol. A* **117**, 619 (2010).

[11] A. Chatterjee, B. K. Chakrabarti, *Eur. Phys. J. B* **60**, 135 (2007).

[12] P. Bhattacharyya, A. Chatterjee, B. K. Chakrabarti, *Physica A* **381**, 377 (2005).

[13] A. Dragulescu, V. M. Yakovenko, *Eur. Phys. J. B* **17**, 723 (2000).