Proceedings of the 4th Polish Symposium on Econo- and Sociophysics, Rzeszów, Poland, May 7–9, 2009

# Study of Households' Income in Poland by Using the Statistical Physics Approach

M. Jagielski* and R. Kutner

Division of Physics Education, Institute of Experimental Physics

Faculty of Physics, University of Warsaw, Smyczkowa 5/7, PL-02-678 Warsaw, Poland

At the end of 19th century Vilfredo Pareto, as the first tried by using power-laws to describe wealth and income distributions in society. We applied early works of Pareto as well as Gibrat (i.e. laws of Pareto and rules of proportionate growth, respectively). Furthermore, we used recent and advanced models: the Generalised Lotka–Volterra model and collision models. By using empirical data for annual income of Polish households, e.g. for years 2003 and 2006, the comparison with these theoretical models was successfully made. The surprisingly good agreements with Pareto distribution were obtained, where Pareto exponents near the cubic law were found for middle class. For the low class very good agreement with prediction of the cumulative log–normal distribution was gained. Hence, it was possible to establish the border between low and middle society levels. The same was possible for the border between high and middle classes as the ranking for the former follows (to some extent) the Zipf law.

PACS numbers: 89.20.−a, 89.65.Gh

## 1. Introduction

Since almost two decades, physics oriented approaches have been developed and applied to explain economic phenomena and processes [1–5]. In this work we compared empirical data for annual income of Polish households with predictions of several theoretical models.

These models are mainly based on theories of random diffusion processes within microcanonical and canonical ensembles. For example, they reveal the income borders between the low and middle society classes as well as middle and high ones. We hope that they will be helpful in understanding how wealth or income is generated and accumulated.

In our analysis we used the data from Polish Central Statistical Office. They are referring to disposable income, that is "Available income less other expenditures. Disposable income is designated for expenditures on consumer goods and services and for an increase in savings"[†].

## 2. Models of income distributions and results

In order to analyse Polish income data we constructed empirical cumulative distribution function

$$\Pi(m) = \frac{1}{N}\sum_{i=1}^{\mathrm{N}} I\left(m_i > m\right), \tag{1}$$

where $m$ is an annual disposable income, $N$ is the number of observations and $I(m_i > m)$ is an indicator function with value 1 if $m_i > m$ or with value 0 if $m_i \le m$. Therefore, for fixed value $m$, we find the number of households (indexed by $i$, $i = 1, 2, \ldots, N$) whose income is greater than $m$ and normalise it by number of observations. We construct the histogram, where the single counting was done for every 1000 PLN step by starting from 0.

At first, we fitted to empirical cumulative distribution the weak Pareto law given by formula [6] (plots in log–log scale, see Fig. 1):

$$\Pi(m) \approx (m/m_0)^{-\alpha}. \tag{2}$$

Here $m_0$ is a scaling factor and $\alpha$ is a Pareto exponent whose values are extracted from shift and slope of empirical data, respectively.

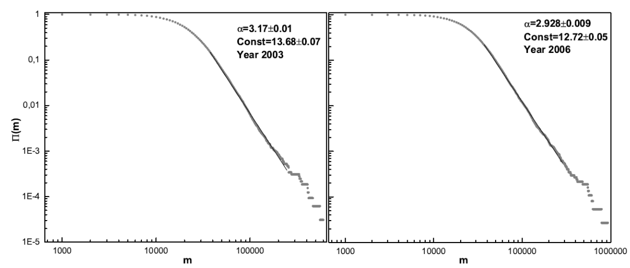


Fig. 1. Fit of the weak Pareto law (solid line) to the Polish households' income empirical data set (dots) for 2003 ($\alpha = 3.17 \pm 0.01$) and 2006 ($\alpha = 2.928 \pm 0.009$).

* corresponding author; e-mail: zagielski@interia.pl

† Polish Central Statistical Office — http://www.stat.gov.pl/

Let us note that results obtained for 2000 are very similar [7]. For above given plots Const is a constant value obtained from fitting procedure (in log–log scale)‡.

As it is seen from Fig. 1, the weak Pareto law describes very well the “bulk" of analysed distributions. In order to get such fits we deleted approximately. 20–40 extremal observations connected with the richest households. Deleted observations represent approximately 0.06% of the population and we analyse them by different way.

Obtaining the value of the Pareto exponent plays a crucial role in analysing the instability in the Pareto macro-economy — it enables to determine the “phase" of the system (the system can be found in one of the two phases — the surplus phase or the deficit phase). For large systems the instability may lead to wealth condensation [8].

We can also gain the Pareto exponent by the alternative approach analysing the rank of households, which is a graph of household income depending on its place in the rank. The corresponding graphs on a logarithmic scale can be found in Fig. 2.

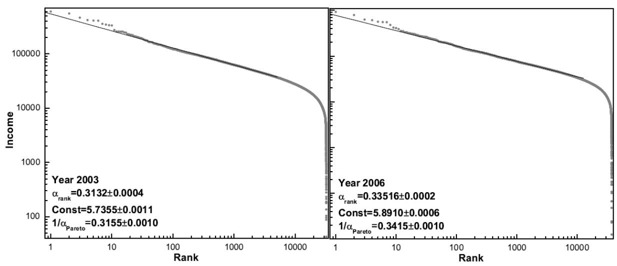


Fig. 2. Rank of the Polish households' income: solid lines are obtained by fit to empirical data (dots) — year 2003 ($1/\alpha_{\mathrm{rank}} = 3.193 \pm 0.004$) and 2006 ($1/\alpha_{\mathrm{rank}} = 2.984 \pm 0.002$).

The richest households are described in a logarithmic scale by a straight line with slope parameter $\alpha_{\mathrm{rank}}$. It turns out that the inverse of the Pareto exponent obtained in the case of empirical cumulative distribution function of annual income of households is nearly the same as $\alpha_{\mathrm{rank}}$ value but burdened by greater dispersion. This points out for the compliance of both applied methods.

We also fitted the cumulative log–normal distribution resulting from rules of proportionate growth [9]:

$$\Pi(m) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\ln(m - m_0) - \mu}{\sqrt{2}\sigma}\right)\right]. \tag{3}$$

Results, again on a logarithmic scale are presented in Fig. 3.§

The cumulative log–normal distribution function quite well describes poor-income households. The fact that the cumulative log–normal distribution shows a good agreement with empirical data result directly from Rules of Proportionate Growth. In this model was assumed that changes in income are small, which is indeed justified in the case of poor households [9].

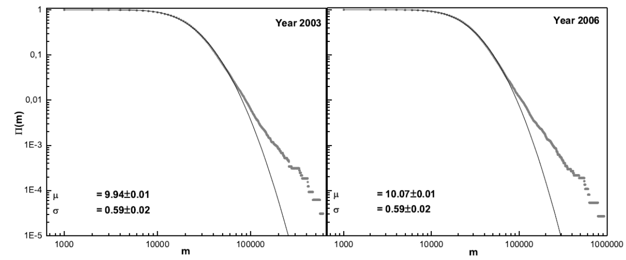


Fig. 3. Fit of the log–normal cumulative distribution function (solid line) to the Polish households' income empirical data set (dots) for 2003 and 2006.

In turn, the weak law of Pareto is appropriate to depict income of middle households. Analysis of Figs. 1 and 3 raises the question on the value of income at the point of intersection of the cumulative log–normal distribution and the weak Pareto law. This point gives a conventional and sufficiently precise border between poor and middle-income households. It may be determined graphically by reading the coordinates of the point of intersection of the both graphs drawn for the estimated parameters. Thus, the annual income limit which is the distinction between the poor and middle-income households is approximately: 35 000 PLN in 2003 and 36 000 PLN in 2006 [7], which seems to be reasonable.

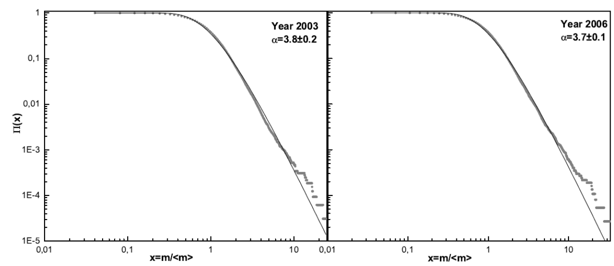


Fig. 4. Fit of generalised Lotka–Volterra cumulative distribution function (solid line) to the Polish households' income empirical data set (dots) for 2003 ($\alpha = 3.8 \pm 0.2$) and 2006 ($\alpha = 3.7 \pm 0.1$).

We also obtained a good agreement between our empirical data and the cumulative distribution function given by the generalised Lotka–Volterra model (Fig. 4) [2, 3]:

‡ From Eq. (2) we easily find $m_0 = \mathrm{e}^{\frac{\mathrm{const}}{\alpha}}$. Hence, $m_0 = 74 \pm 2$ PLN for 2003 and $m_0 = 77 \pm 2$ PLN for 2006.

§ The parameter $m_0$ is put here 0, since (to some extent) it does not affect the estimation of other parameters [7]; in order to make a fit we used here and further in this work the Levenberg–Marquardt algorithm.

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

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

Although generalised Lotka–Volterra model does not describe so well the cumulative distribution of annual households income as the cumulative distribution of log–normal and Pareto one, however, observed differences are relatively small. An important advantage of this model is the ability to characterise the empirical distributions 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.

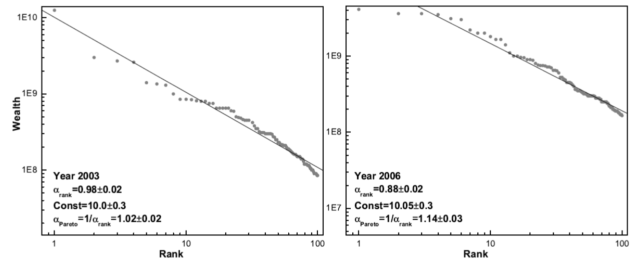


Fig. 5. Rank of the wealth of 100 richest Poles (solid line is fit, and dots are empirical data) — year 2003 ($\alpha_{\mathrm{Pareto}}(\equiv \alpha) = 1.02 \pm 0.02$) and 2006 ($\alpha_{\mathrm{Pareto}}(\equiv \alpha) = 1.14 \pm 0.03$).

So far we did not analyse the richest households income, since these observations were not numerous enough to be subjected to any statistical description. This problem can be solved indirectly by analysing the wealth of the 100 richest Poles in the form of rank¶ (Fig. 5, log–log scale).

In accordance with earlier conclusions, if the rank of the richest on a logarithmic scale can be described by the linear relationship, then the Pareto exponent will be: $\alpha_{\mathrm{Pareto}} = 1/\alpha_{\mathrm{rank}}$. The Pareto exponent values are close to unity, which is consistent with the theoretical result obtained in the collision model with distributed savings [1, 4]. Thus, it is expected that enrichment of middle class is realised (from a formal point of view) by decreasing of the Pareto exponent. The mechanism of acquiring income by the richest is based on the existence of market competition [10]. In this case, units forming households are mostly owners of competing firms, whose revenues are reported by the Zipf law [11].

¶ Data on the wealth of the 100 richest Poles in 2003 and 2006 were obtained from the website of the *Wprost Magazine* — `http://100najbogatszych.wprost.pl/`

## 3. Conclusions and inspirations

In this paper we analysed empirical cumulative distribution functions for annual income of households in Poland, for example, for years 2003 and 2006. It turned out that they can be very well described by cumulative log–normal distribution for poor households, as well as the weak Pareto law for the middle-income households. The point of intersection of the two distributions may establish the border between lower and middle society levels. Because the Pareto exponent describing the middle--income households can be obtained either by analysis of cumulative distributions or rank, we determined its value for the wealthiest individuals of Polish society. Thus, the cumulative annual income of households can be described over the entire range by:

- the cumulative log–normal distribution, in the case of poor households,
- the weak Pareto law with exponent equal to about 3, in the case of middle-income households,
- the weak Pareto law with exponent equal to about 1 (the Zipf law [10]), in the case of the richest households.

The obtained results opened the way for further research. First of all, it would be interesting to carry out a similar analysis in the case of provinces and comparing it with the results obtained for the whole country. Also studies of the revenue of societies in other European countries would provide valuable information, which could serve the construction of more sophisticated theoretical models taking into account the territorial diverse of wealth of societies.

## Acknowledgments

We wish to thank Didier Sornette and Armin Bunde for helpful comments and suggestions. Also, we would like to thank Polish Central Statistical Office for providing the Polish households' income empirical data set for years 2000**, 2003 and 2006.

## References

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

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

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

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

** In this work we did not use the empirical data set for 2000 since they are very similar to ones for 2003.

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

[6] 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 131 (2006).

[7] M. Jagielski, M.Sc. Thesis, Warsaw University, Warszawa 2009 (in Polish).

[8] M. Burda, J. Jurkiewicz, M.A. Nowak, *Acta Phys. Pol. B* **34**, 87 (2003).

[9] M. Armatte, *Mathématiques et sciences humaines* **129**, 5 (1995).

[10] K. Yamamoto, S. Miyazima, H. Yamamoto, T. Ohtsuki, A. Fujihara, in: *Practical Fruits of Econophysics: Proceedings of the Third Nikkei Econophysics Symposium*, Eds. H. Takayasu, Springer, Tokyo 349 (2006).

[11] K. Okuyama, M. Takayasu, H. Takayasu, *Physica A* **269**, 125 (1999).