Study and Modeling of Households' Income in Different Countries and Regions in Europe

Maciej Jagielski[1], Ryszard Kutner[1], Didier Sornette[2]

[1]Department of Physics, University of Warsaw, Warsaw, Poland

[2]Department of Management, Technology and Economics, ETH, Zürich, Switzerland

Since almost two decades, physics oriented approaches have been developed and applied to explain economic phenomena and processes. We plan to perform extended study which covers all European countries and characteristic regions. We aim to consider both static and dynamic properties of developing and developed countries and regions. The economic and social situation in the last decade, not only in Europe, stimulate research within the subject defined by the title of our project. The study was already initiated at the end of 19th century by Vilfredo Pareto who tried, by using power laws, to describe wealth and income distributions in very different societies. These quantities are the basic economic and social ones which determine the household activities. Moreover, they make possible to define the society classification.

So far, we have compared empirical data for annual income of households in Poland 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 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".

• Complementary cumulative distribution function

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

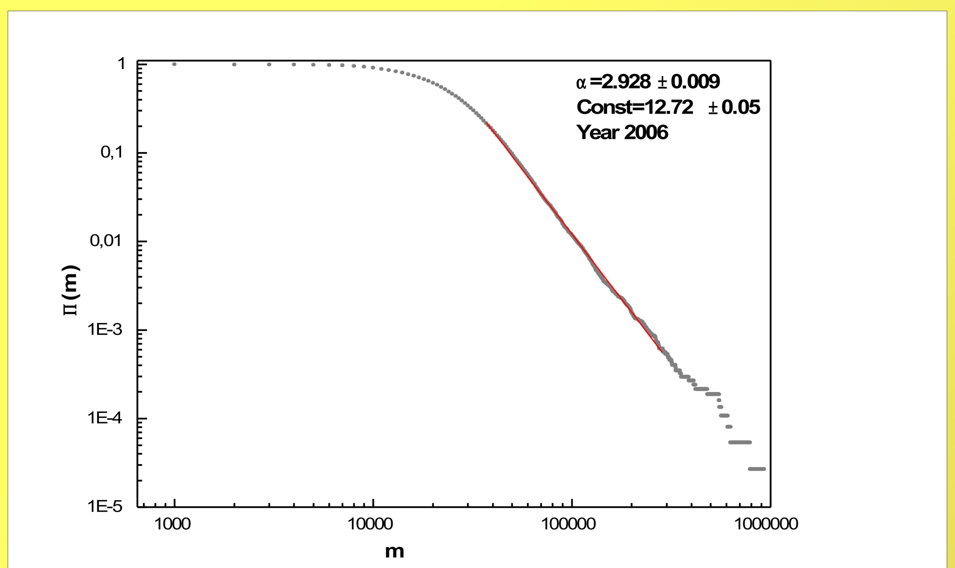
$$\Pi(m) = \frac{1}{N}\sum_{i=1}^{N} I(m_i > m)$$

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 $m_i \le m$. So, 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.

• The weak law of Pareto

At first, we fitted to empirical cumulative distribution the weak Pareto law given by formula (plot in log-log scale)

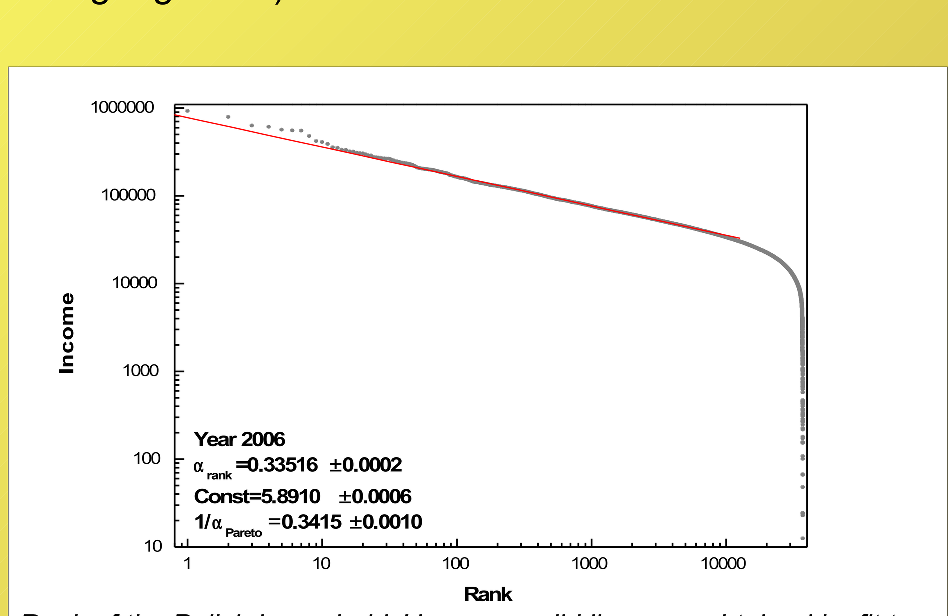
$$\Pi(m) \approx (m/m_0)^{-\alpha}$$



Fit of the weak Pareto's law (solid line) to the Polish households' income empirical data set (dots) for 2006 (α = 2.928 ±0.009).

As it is seen the weak Pareto law very good describes the "bulk" of analysed distributions.

We can also gain 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 (plot in log-log scale).



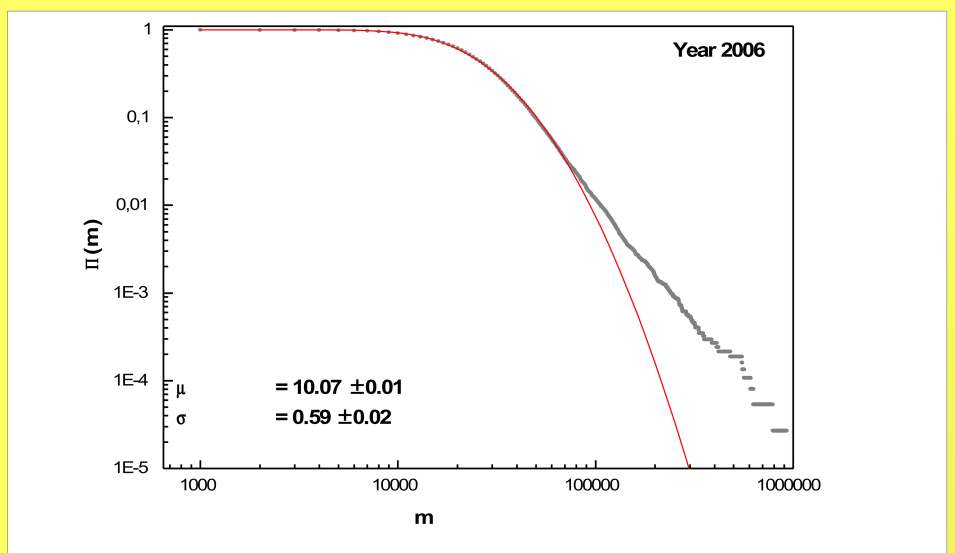
Rank of the Polish households' income: solid lines are obtained by fit to empirical data (dots) – year 2006 ($1/\alpha_{rank}$=2.984±0.002).

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

• Rules of Proportionate Growth

We also fitted the cumulative log-normal distribution resulting from Rules of Proportionate Growth (log-log scale):

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



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

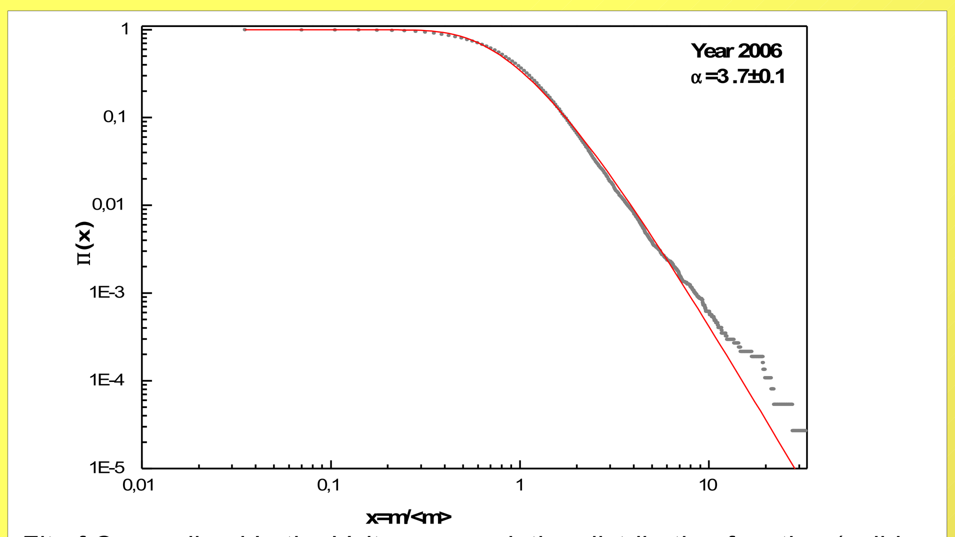
The cumulative log-normal distribution function quite well describes poor-income households. The fact that the cumulative log-normal distribution shows a good agreement with empirical data result directly from Rules of Proportionate Growth. In this model was assumed that changes in income are small, which is indeed justified in the case of poor households. 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. Thus, the annual income limit which is the distinction between the poor and middle-income households is approximately: 36 000 PLN in 2006, which seems to be reasonable.

• Generalised Lotka-Volterra model

We also obtained a good agreement between our empirical data and the cumulative distribution function given by the Generalised Lotka-Volterra:

$$\Pi(m) = 1 - \frac{\Gamma\left(\alpha, \frac{\alpha - 1}{x}\right)}{\Gamma(\alpha)}$$

here α is the shape parameter, which describes fitted function and $x = m/\langle m \rangle$ is the relative income of households (where $\langle m \rangle = \sum m_i$). Plot in log-log scale.

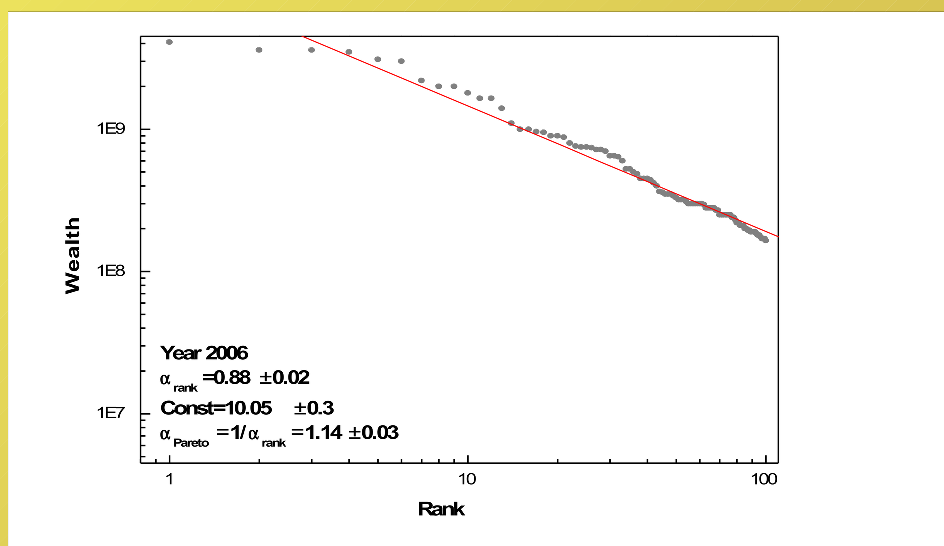


Fit of Generalised Lotka-Volterra cumulative distribution function (solid line) to the Polish households' income empirical data set (dots) for 2006 (α=3.7±0.1).

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.

• Rank of the 100 richest Poles

We analysed the richest households income. Sine these observations weren't numerous enough to be subjected to any statistical description we solved the problem indirectly by analysing the wealth of the 100 richest Poles in the form of rank (log-log scale).



Rank of the wealth of 100 richest Poles (solid line is fit, and dots are empirical data) - year 2006 (α_{Pareto}(=α) =1.14±0.03).

The Pareto exponent values are close to unity, what is consistent with the theoretical result obtained in the collision model with distributed savings. Thus, it is expected that enrichment of middle class is realised (from a formal point of view) by decreasing of Pareto exponent.

•Things to be done …

In the project we will mainly focus on:

- Comparison of empirical complementary cumulative distribution for households in different countries and regions in Europe with Pareto's law,
- Comparison of empirical complementary cumulative distribution for households in different countries and regions in Europe with theoretical models: Rules of Proportionate Growth, Generalized Lotka-Volterra model, collision models (this models refer to statistical physics and collision of particles; in this process they exchange energy. In economy we can consider households "particles" which interact with each other and exchange money "energy"),
- Verification of applied models by various statistical tests and methods like: uniformly most powerful unbiased test for null hypothesis of the Pareto distribution against the lognormal,
- Comparison of results with predictions of models based on computer simulations and then depending on the results we will try to use computer simulations to answer the question about the mechanism of enrichment for societies,
- Developing more sophisticated theoretical models taking into account the territorial diverse of wealth of societies,
- Analysis of rank of the richest households,
- Establishing the income border between low and middle society levels and between high and middle classes,
- Finding the distribution function that best describes empirical data sets.

•Bibliography

[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).
[5] M. Jagielski, R. Kutner R, *Acta Phys. Pol. A*, 2010, in print.
[6] P. Łukasiewicz, A. Orłowski, *Physica A* **344**, 146 (2004).
[7] 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.
[8] M. Jagielski, *Badanie zamożności gospodarstw domowych w Polsce metodami egzotycznej i tradycyjnej fizyki statystycznej*, MSc. Thesis, Warsaw University, Warszawa 2009.
[9] M. Burda, J. Jurkiewicz, M. A. Nowak, *Acta Phys. Pol. B* **34**, 87 (2003).
[10] M. Armatte, *Mathematiques et sciences humaines* **129**, 5 (1995).
[11] 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, 2006, p. 349.
[12] K. Okuyama, M. Takayasu, H. Takayasu, *Physica A* **269**, 125 (1999).