EQUILIBRIUM
Quarterly Journal of Economics and Economic Policy
Volume 13 Issue 4 December 2018
p-ISSN 1689-765X, e-ISSN 2353-3293
www.economic-policy.pl

ORIGINAL PAPER

Citation: Łukasiewicz, P., Karpio, K., & Orłowski, A. (2018). Two-component structure of household income distributions in Poland. *Equilibrium. Quarterly Journal of Economics and Economic Policy, 13*(4), 603–622. doi: 10.24136/eq.2018.029

Contact to corresponding author: piotr_lukasiewicz@sggw.pl, Department of Informatics, Warsaw University of Life Sciences (SGGW), Nowoursynowska 159, 02-776 Warsaw, Poland
Received: 23 May 2017; Revised: 4 June 2018; Accepted: 27 June 2018

Piotr Łukasiewicz
Warsaw University of Life Sciences (SGGW), Poland

Krzysztof Karpio
Warsaw University of Life Sciences (SGGW), Poland

Arkadiusz Orłowski
Warsaw University of Life Sciences (SGGW), Poland

Two-component structure of household income distributions in Poland

JEL Classification: *C51; C52; D31*

Keywords: *income distribution; Pareto model; power law*

Abstract
Research background: Studies of the structures of the income distributions have been performed for about 15 years. They indicate that there is no model which describes the distributions in their whole range. This effect is explained by the existence of different mechanisms yielding to low-medium and high incomes. While more than 97% of the distributions can be described by exponential or log-normal models, high incomes (about 3% or less) are in agreement with the power law.
Purpose of the article: The aim of this paper is an analysis of the structure of the household income distributions in Poland. We verify the hypothesis about two-part structure of those distributions by using log-normal and Pareto models.
Methods: The studies are based on the households' budgets microdata for years 2004–2012. The two-component models are used to describe the income distributions. The major parts of the distributions are described by the two parametric log-normal model. The highest incomes are described by the Pareto model. We also investigate the agreement with data of the more complex models, like Dagum, and Singh-Madalla.

Findings & Value added: One has showed that two or three parametric models explain from about 95% to more than 99% of income distributions. The poorest agreement with data is for the log-normal model, while the best agreement has been obtained for the Dagum model. However, two-part model: log-normal for low-middle incomes and Pareto model for the highest incomes describes almost the whole range of income distributions very well.

Introduction

The studies of structures of the income distributions have been performed for about 15 years. They indicate that there is no one model which describes the whole ranges the distributions. This issue is explained by the existence of different mechanisms yielding to low-medium and high incomes. This effect has been observed for the distributions of incomes in the USA, United Kingdom, Germany and Japan. In the majority of studies incomes are best described by lognormal model with power law tail. One also notices the stability of this structure in time. First works regarding fat tails of income distributions were published in econophysical literature. The authors of (Levy & Solomon, 1997, pp. 90–94) analyze the data from the 1996 Forbes 400 list of the richest people in the US. The obtained results confirm that wealth is distributed according to the power law with exponent equal 1.36. Okuyama *et.al.* (1999, pp. 125–131) performed the studies which showed that income distributions of Japanese firms are the subject to Zipf's low (power low with exponent equal to 1). Suoma (2001, pp. 463–470) studied Japanese income distributions for years 1887–1998. He showed that two-part model, lognormal with power law tail is the universal structure describing distributions of personal incomes in Japan. The author investigated the negative correlations between the value of the power law exponent and the prices of various assets, especially a land price index and the Tokyo Stock Price Index (TOPIX). Nirei and Souma (2004, pp. 61–68) continued research and proposed dynamic stochastic model explaining power low tails. Dragulescu and Yakovenko (2001, pp. 213–221) studied the income distributions in United Kingdom (1994–1999) and in individual US states (1998). They showed that income distributions had two-part structure. They were: exponential in low-middle part and power law for the highest incomes. Nirei and Souma (2007, pp. 440–459) studied income distributions in Japan and the US based on the individual income tax returns data from 1960 to 1999. They confirmed the hypothesis about the two-part structure of income distributions. The authors described the left-central part of the distributions by the exponential model and the top 1% of incomes — by the power law model. Clementi and Gallegati (2005, pp. 3–14) investigated income distributions of households in the US (1980–

2001), United Kingdom (1990–2001) and Germany (1990–2002). A low-middle income group was approximated by the lognormal function and a high income group by the power law function in all studied countries. The fits were stable, but the parameters' values varied very much. The authors of another work (Clementi & Gallegati, 2005, pp. 427–438) studied personal income distributions for Italy in the years 1977–2002. They also confirmed two-part stable structure of income distributions: log-normal with power law tail. They showed that fluctuations of shapes of the income distributions were related to the business cycle phases experienced by the Italian economy.

The latest works may point to the more complex structure of the income distributions. Jagielski and Kutner (2010, pp. 615–618) studied total incomes of Polish households in 2003 and 2006. The studies were based on the Household's Budgets Survey data and independently on data regarding wealth taken from rank of the 100 richest Poles. The authors showed that the Polish income distributions may have a three-part structure: lognormal distribution in the case of poor households, Pareto law with exponent equal to 3 for middle-income households and Pareto law with exponent equal to about 1 in the case of wealth of the richest Poles. In a paper (Jagielski & Kutner, 2013, pp. 2130–2138), the authors merge Eurostat Survey on Income and Living Conditions data with income data evaluated based on the wealth of billionaires in the EU published by the Forbes ('The World's Billionaires' rank). They obtain empirical distribution with a three-part structure which, in turn, they approximate by the proposed model.

In this paper, we conduct temporal studies of the income distributions in Poland. Our studies are based on the households' budgets microdata from 2004 to 2012. Our main aim is to verify the hypothesis about two-part structure of the income distributions. In the first step, we assume the structure is of the form: log-normal with power law tail and we study its characteristics vs. time. It is a well-known fact that the log-normal model does not describe the whole range of income distributions. Therefore, we also investigate the agreement of more complex (Dagum and Singh-Maddala) models with empirical data.

Data

Data from the Household Budget Survey (HBS) project from 2004 to 2012 have been used in this work. The HBS studies are performed by Central Statistical Office (CSO) in Poland each year. The HBS data, before being made accessible, are processed by CSO, which usually takes about 1.5–2 years. The 2012 data were the newest data accessible to authors at the time of studies. One selected subset of data containing microdata about the available monthly incomes in the households. Household's available income is a sum of household's gross incomes from various sources reduced by all income taxes, as well as by social security and health insurance taxes. The available income comprises: wages and salaries, incomes from farms, self-employment, properties, rents, various social benefits (including retirement pensions and pensions), and other incomes (e.g. alimonies). Available income is allocated to expenditures and savings increase.

The data contain a number of zero or negative incomes. They represent between 5.7 and 7.4 per mille depending on year. They occur for some of the households which gain incomes from business related activities (farms, self-employment, rents, and others). Non-positive incomes indicate lack of income or loss according to accounting balance. They arise when the costs of business activity are greater than incomes, so they are not simply incomes of households. On the other hand, the statistical methods used in this analysis (models of incomes) can describe only positive incomes. The two above issues caused the non-positive incomes have been removed from data.

The income of each household has been recalculated into the annual income in thousands PLN. In order to analyze households with different number of persons together, we recalculated the total income of each household into the income per person. Thus, the households described by their incomes per person are the objects being studied in this work.

The basic statistics of income variables for years from 2004 to 2012 are listed in Table 1. The data is characterized by high values of standard deviation and skewness, which is related to the presence of high incomes and significant right-side asymmetry. In order to evaluate the models, the data have been grouped into income classes of the same width. The numbers of classes are between 250 and 300, and the widths are between about 0.9 and 1.5 thousand PLN depending on the year. The number of classes and their widths are different in various years, because the spread of income data varies significantly from year to year (see Table 1). In order to compensate for the different widths of classes, the percentage of objects in each class was divided by class' width providing empirical density in the class. One also constructed empirical cumulative distributions based on the detailed data to evaluate power models and to present the results. The empirical cumulative distribution is defined:

$$F_{emp}(x_i) = \frac{k_i}{N}$$

where data x_i, $i = 1, ..., N$ are sorted ascending, k_i is rank of income x_i.

Research methodology

We study the agreement between the selected models of incomes and the empirical distributions. We take into account three commonly used models: lognormal, Dagum, and Singh-Maddala. These distributions come from the more general family of distributions called generalized beta of the second kind (McDonald, 1984, pp. 647–663; McDonald & Xu, 1995, pp. 133–152). Exponential model (Dragulescu & Yakovenko, 2001, pp. 213–221) describes well the US personal income data, but is not suitable for describing income distributions of Polish households. Probability density function (pdf) of lognormal distribution is:

$$f_{LN}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \qquad (1)$$

where $x > 0$, while the σ parameter fulfills the condition $\sigma > 0$. The μ and σ parameters are interpreted as mean value and standard deviation of incomes logarithms respectively. Cumulative density function (cdf) of lognormal distribution is not an elementary function but can be expressed by the cdf Φ of the standard normal distribution:

$$F_{LN}(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right). \qquad (2)$$

Lognormal distribution has been often used to describe distributions of wages and incomes (Aitchison, & Brown, 1957, p. 176). Pdf of Dagum distribution (Dagum, 2008, pp. 3–25) is described by the equation:

$$f_D(x) = \frac{\alpha\beta\delta}{x^{\beta+1}(1+\alpha x^{-\beta})^{\delta+1}}, \qquad (3)$$

where $x > 0$, while the parameters fulfill the conditions: $\alpha > 1$ and $\beta, \delta > 0$. The pdf of Singh-Manddala distribution (Singh & Manddala, 1976, pp. 963–970) can be expressed by the formula:

$$f_{SM}(x) = \frac{\alpha\beta\delta}{x^{-\beta+1}(1+\alpha x^{\beta})^{\delta+1}}, \tag{4}$$

where $x > 0$ and $\alpha, \beta, \delta > 0$, $\beta\delta > 1$. Cumulative density functions of the above models are described by the equations:

$$F_D(x) = (1+\alpha x^{-\beta})^{-\delta}, \tag{5}$$

$$F_{SM}(x) = 1-(1+\alpha x^{\beta})^{-\delta}. \tag{6}$$

Studies performed in various countries show that models (3) and (4) exhibit high conformance with empirical distributions of incomes (Bandourian *et al.*, 2002, p. 47; Dagum & Lemmi, 1988, pp. 123–157; Kleiber, 1996, pp. 265–268). They were adopted successfully for describing other kinds of data (Brzeziński, 2014, pp. 362–368). They are universal, as they may describe zero- as well as one-modal distributions (see: Łukasiewicz *et al*., 2012, pp. B82–B85). Curves (3) and (4) have ‘fat tails’, which is their advantage because empirical distributions are significantly extended in the range of incomes exceeding average. The Dagum model is very often used in studies of incomes (see i.e. Łukasiewicz & Orłowski, 2003, pp. 122–130; Łukasiewicz & Orłowski, 2004, pp. 146–151; Quintano & D’Agostino, 2006, pp. 525–546).

The Pareto model Type I (Pareto, 1896–97, p. 430) has been used to describe the highest incomes (right tails of the distributions). This model, also known as a power law, contains one parameter and its pdf and cdf functions are of the forms:

$$f_P(x) = \alpha x_m^{\alpha} x^{-\alpha-1}, \tag{7}$$

$$F_P(x) = 1 - x_m^{\alpha} x^{-\alpha}, \tag{8}$$

where $x \geq x_m$ and $\alpha > 0$. A limit value of income is indicated by x_m. The pdf and cdf functions are equal to 0 for $x < x_m$.

All the models have been evaluated by means of the nonlinear least square method utilizing Levenberg-Marquardt algorithm. Therefore, the coefficients of the models are estimated by minimizing the function:

$$SSE(\theta) = \sum_{i=1}^{n} (y_i - f(x_i;\theta))^2 , \tag{9}$$

where θ is a vector of the model's parameters.

The models (1) — (3) were evaluated based on the grouped data. In the case of the model (7), the obtained results were unstable because of the small number of counts for the highest incomes. The Pareto model's parameter α has been evaluated based on the cumulative data using the function (8). The limit values of x_m were evaluated for each model and year after estimating the functions (1) — (3). The x_m was determined as the income above which the model's residuals start rising.

Results

The lognormal, Dagum, and Singh-Maddala models of incomes were fitted to the empirical distributions. The results are listed in Table 2. There are standard errors of the parameters' estimators in brackets. The columns contain: values of nonlinear coefficient of determination $R^2 = 1 - SSE$, limit values x_m and values of the theoretical cdf: $F(x_m)$. The latter is a percentage of the income distribution (percentage of households) explained by the model.

All the evaluated models describe empirical data very well. They are characterized by the high coefficients of determination and very small errors of their parameters. The values of R^2 are similar to one another for all models. The smallest values of R^2 are observed for the lognormal model, which describes the smallest part of the incomes distributions: from 94.9% to 98.3%, depending on year. On the other hand, log-normal model doesn't explain from 1.7% (2004) to 5.1% (2010) of income distribution, Singh-Maddala: from 0.8% (2008) to 4.1% (2012), and Dagum: from 0.1% (2006) to 2.4% (2012). Models' functions are plotted in Fig. 1 for years: 2007, 2009 and 2012. There are complementary cumulative density functions (ccdf) in the figure. They are also known as 'tail distributions', and given by the equations:

$$\overline{F}_{emp}(x_i) = 1 - F_{emp}(x_i) \text{ and } \overline{F}(x) = 1 - F(x) . \tag{10}$$

In order to emphasize the differences between the empirical and theoretical distributions, the plots are on the log-log scale. The values of annual incomes (in thousands PLN) are on the horizontal axis, while the percentages of households are on the vertical axis.

In the next step, tails of the empirical distributions have been approximated by the Pareto model. The tails have been defined in this work as incomes satisfying the inequality $x \geq x_m$. They are the top parts of the incomes distributions, which are not explained by the models considered in this work. Further on, we will take into account and discus two cases: when the limit values x_m have been determined for: (i) lognormal model and (ii) Dagum model. For those two models one obtained the minimum and the maximum values of x_m respectively (see Table 2). In the case of Singh-Maddala model values of x_m are slightly bigger than in (i), while the values of estimation parameters for Pareto model are similar to those in (i). Because of that, these values are omitted in this paper. The results of the estimations of Pareto model are presented in Tables 3 and 4, while the plots of the Pareto functions are in Fig. 2 and 3.

In the case (i), the log-normal model does not explain 1.7% ÷ 2.1% of income distributions for years from 2004 to 2006. In the next years, the tails of the distributions are bigger: 3.8% ÷ 5.1%. The quality of the Pareto model's fits is very high (slightly lower in 2008 and 2010). We also observe very small errors of the α parameter. The Pareto exponent has the value of 3.04 in the first year (2004), and has values below 3.00, in the range 2.65 ÷ 2.93, for the successive years (2005–2012). The power law exponent is very stable in time, its changes are small and the values are around 2.80.

The results are more dispersed in the case (ii) than in the case (i). The Dagum model does not explain only about 0.1% of income distributions in 2005, 2006, and 2008 and from 0.2% to 0.5% in years 2004, 2009–2011. One can state that the Dagum model describes almost whole range of the income distributions in the majority of years. Such an excellent agreement of this model with income data is emphasized in the empirical studies (Bandourian *et al*., 2002, p. 47; Kleiber, 1996, pp. 265–268). The Dagum model is characterized by the high flexibility and agree with various income distributions. This is because the Dagum model possesses the property of the weak Pareto law. That means (3) converges to the Pareto model (7) for incomes sufficiently high (Dagum, 2008, pp. 3–25). The right part of the Dagum model visible in Fig. 1 is approximately compatible with power law (7).

The ranges of income distributions not explained by the model are narrow. They are wider in the years 2007 and 2012, and are equal to 1.3% and

2.4%, respectively. These far parts of tails have been approximated by the power models (Table 4). These models are characterized by lower qualities of fits and greater errors than in the case (i). This is due to the smaller number of empirical points and bigger dispersions at the right tail-ends. In 2004 Pareto exponent is equal to 3.21, and in 2005–2012 it assumes values 1.76 ÷ 2.75. The obtained results are more dispersed than in the case (i).

Discussion

At the beginning, we want to discuss the issue related to the reliability of the highest incomes. It is well known that the highest incomes are understated by the interviewed earners. The problem exists in every household budget surveys, regardless of the country. As a result, the right tail of the income distribution is doubtful, and an evaluation of Pareto model can be less reliable. The extension of the right tail of the income distribution using the wealth of the 100 richest Poles, proposed by Jagielski and Kutner (2013, pp. 2130–2138), does not solve this problem. This extension of data leads to adding another society group which probably is not covered by survey studies. The empirical distribution gains the addition tail and modeling of such distribution requires sophisticated models and methods. This issue extends beyond the scope of the analysis presented in this paper, which is based on the representative's samples of Central Statistical Office.

The incomes provided by the richest earners are understated, but it is not known what the scale of this phenomenon is. That's why it is impossible to determine the true incomes for such individual households. Let us return to the reliability of our evaluations of the Pareto model. Let us assume that for the richest households the provided x and true x_t incomes are related to each other in the following way: $x = w \cdot x_t$, where $0 < w < 1$ denotes the scale of the understatement of incomes for $x > x_m$ (see formula (7)). Replacing $x = w \cdot x_t$ in equation (7) we obtain:

$$f_P(x) = w^{-\alpha-1} f_P(x_t)\,.$$

Income density function depending on the new variable x_t (true income) has the form:

$$f_P(x_t) = (1/w^{-\alpha-1})\, f_P(x)\,,$$

where the factor in brackets normalizes the model. Thus the evaluated α parameter for understated incomes remains the same as for true income (at the assumption that both incomes are proportional to each other). In other words, the slopes of lines showed in Fig. 2 and 3 will be the same.

The Dagum model describes almost the whole range of income distributions for majority of years according to presented above analysis. Its agreement with data exceeds those obtained for models log-normal and Singh-Maddala. Unexplained parts of the distributions are narrow. Some weakness of the model is that its parameters do not have straightforward economical interpretation. The lognormal and Pareto models merged together explain the income distributions very well too. Such a composite model possesses the following parameters: μ, σ, and α, having simple and direct interpretation. The μ and σ parameters are interpreted as the mean value and standard deviation of incomes logarithms, respectively. Another parameter α (Pareto exponent) reflects the slope of the right tail of the income distribution in log-log scale. The values of σ and α reflect economic inequalities (concentration of incomes) in both parts of the distribution. The performed analysis required an evaluation of the cut-off log-normal model (for $0 < x < x_m$). The obtained values of estimates are very similar to those presented in Table 2. Referring to the left part of the distribution (for $x < x_m$), we observe that the value of μ parameter is systematically growing up for years 2004–2012. This behavior reflects the increase of income logarithms. The σ parameter is equal to about 0.60 in 2004–2006. Starting from 2007 σ parameter is very stable being equal to about 0.56. Although the mean log-income increases the dispersion of log-income becomes slightly lower in 2007 and remains the same onward. In the right part of the distribution ($x \geq x_m$) the α exponent has values of about 2.8 for all the analyzed years, except the 2004, when it is slightly bigger. The values of α are surprisingly stable starting from 2005, similarly to log-normal dispersion. The changes of σ and α estimators in the 2004–2012 years are presented in Fig. 4.

Economic inequalities are usually measured by Gini index G, which depends only on σ and α (see Dagum, 2006, pp. 235–268). If the income distribution is described by the log-normal function then $G_{LN} = \xi^2 \Phi(\sigma/\sqrt{2})$ (ξ is the parameter related to the cut-off of the distribution, which is close to 1). For the Pareto distribution of incomes the Gini index is expressed by $G_P = 1 / (2\alpha - 1)$. Its values calculated for both parts of the income distributions are stable in time. The values of Gini index in 2007–2012 are about 0.295 for low-middle incomes, and they fluctuate between 0.206 and 0.233 for the highest incomes. These numbers point to significantly higher income inequalities for the low-middle incomes. On one the hand, the highest incomes are hardly concentrated. Let's note that the overall empirical values of Gini index for the whole income distributions are above 0.33.

The incomes in European countries were studied by Clementi and Gallegati (2005, pp. 3–14). They studied the incomes in Germany (1990–2002) based on the German Socio-Economic Panel data and the UK (1991–2001), using British Household Panel Survey. Additionally, they also investigated incomes in US (1980–2001) based on The Panel Study of Income Dynamics data. The analysis was based on the household post-government equivalent incomes. They showed that about 97%–99% of the studied populations can be explained by log-normal model and the remaining 3%–1% of the populations by Pareto model. The obtained values of the Pareto exponent are more dispersed than in our results: $1.63 \div 2.14$ for Germany, $3.47 \div 5.76$ for UK, and $1.10 \div 3.34$ for US. In (Clementi & Gallegati, 2005, pp. 427–438) the authors studied personal income distributions in Italy from 1977 to 2002. The analyzed data came from the Survey on Household Income and Wealth and were made publicly available by the Bank of Italy. The authors obtained average values of parameters' estimators for the log-normal distribution $\mu = 3.50$, $\sigma = 0.32$, and average Pareto exponent $\alpha = 2.74$ for years 1993–2002.

For comparison, average values of power low exponent α for personal income distributions in Japan, UK, and USA are following: Japan (1887–1998) $\alpha = 2.0$ (Suoma, 2001, pp. 463–470), UK (1994–1999): $2.0 \div 2.3$, US (1998): 1.7 ± 0.1 (Yakovenko & Dragulescu, 2001, pp. 213–221), and US (1983–2001): $1.4 \div 1.8$ (Silva & Yakovenko, 2005, pp. 304–310).

The studies of Clementi and Gallegati based on a survey on household data are the most similar ones to our work. The Pareto exponent for Poland (2004–2012) obtained within our studies has the value of about 2.8 and is similar to that for Italy (2002) which is equal to 2.7. At the same time, a dispersion of the log-incomes in the low-middle part of the distribution in is significantly larger for Poland. The value of σ is equal to 0.6 for Poland (2004), while it is 0.32 for Italy (2002). The Pareto exponent for Poland is higher than that for Germany (on average below 2.0) and significantly lower than α for UK (on average above 4.0). To summarize, income inequalities for the highest incomes at the beginning of the XXI centaury were at the same level in Poland and in Italy ($G \approx 0.22$). At the same time, they were significantly lower than in Germany ($G \approx 0.33$) and higher than in the UK ($G \approx 0.14$).

Conclusions

The income distributions of Polish households in 2004–2012 were studied in this paper. It has evaluated an agreement with data of three models of incomes: log-normal, Dagum, and Singh-Maddala. The lowest goodness-of-fits were observed for lognormal model, while the best fits were for the Dagum one. None of the analyzed models were able to describe the distributions in their whole ranges with sufficiently high precision. The biggest discrepancies were observed in right tail-ends of the distributions.

The Dagum model describes almost the whole range of the income distributions in the majority of years, unexplained tails are short. That's why the tails cannot be described with sufficiently high precision by the Pareto model. The Power law exponents are not stable in time and have relatively big errors.

The lognormal models with power law tails are unexpectedly stable throughout all years. This two-part model will be analyzed in future studies. The method of estimation of the four-parameter model of the form:

$$f(x) \propto \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right), & \text{for } x \le \delta \\ \alpha \delta^{\alpha} x^{-\alpha-1}, & \text{for } x > \delta \end{cases}$$

will be investigated. The boundary between two formulas δ will be evaluated together with the remaining parameters based on the empirical data. It seems the two above formulas could be merged into the one model. The values of derivatives of both functions at δ should be equal to each other. This would yield to the smooth transition from one formula to the other. There are no actual results of income studies for many European countries. To fill this gap, the wide and detailed analysis of shapes of the income distributions in European countries will be performed.

References

Aitchison, J., & Brown, J. A. C. (1957). *The lognormal distribution with special reference to its use in economics*. New York: Cambridge University Press.

Bandourian, R., McDonald, J., & Turley, R. S. (2002). A comparison of parametric models of income distribution across countries and over time. *Luxembourg Income Study Working Paper, 305*. doi: 10.2139/ssrn.324900.

Brzeziński, M. (2014). Empirical modeling of the impact factor distribution. *Journal of Informetrics, 8*. doi: 10.1016/j.joi.2014.01.009.

Clementi, F., & Gallegati, M. (2005). Pareto's law of income distribution: evidence for Germany, the United Kingdom, and the United States. In A. Chatterjee, S. Yarlagadda & B. K. Chakrabarti (Eds.). *Econophysics of wealth distributions*. Springer-Verlag. doi: 10.1007/88-470-0389-X_1.

Clementi, F., & Gallegati, M. (2005). Power law tails in the Italian personal income distribution. *Physica A*, *350*. doi: 10.1016/j.physa.2004.11.038.

Dagum, C. (2008). A new model of personal income distribution: specification and estimation. In D. Chotikapanich (Ed.). *Modeling income distributions and Lorenz curves*. Springer. doi: 10.1007/978-0-387-72796-7_1.

Dagum, C., & Lemmi, A. (1989). A contribution to the analysis of income distribution and income inequality and a case study: Italy. In D. J. Slottjee (Ed.). *Research on economic inequality, 1*. Greenwich CT: JAI Press.

Dagum, C. (2006). Wealth distribution models: analysis and applications. *Statistica*, *61*(3).

Dragulescu, A. A., & Yakovenko, V. M. (2001). Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A*, *299*. doi: 10.1016/S0378-4371(01)00298-9.

Jagielski, M., & Kutner R. (2010). Study of households' income in Poland by using the statistical physics approach. *Acta Physica Polonica A, 117*(4). doi: 10.12693/APhysPolA.117.615.

Jagielski, M., & Kutner, R. (2013). Modelling of income distribution in the European Union with the Fokker–Planck equation. *Physica A*, *392*(9). doi: 10.1016/j.physa.2013.01.028.

Kleiber, C. (1996). Dagum vs. Singh-Maddala income distributions. *Economics Letter*, *53*(3). doi: 10.1016/S0165-1765(96)00937-8.

Levy, M., & Solomon, S. (1997). New evidence for the power-law distribution of wealth. *Physica A, 242*. doi: 10.1016/S0378-4371(97)00217-3.

Łukasiewicz, P., & Orłowski A. J. (2004). Probabilistic models of income distributions. *Physica A*, *344*(1-2). doi: 10.1016/j.physa.2004.06.106.

Łukasiewicz, P., & Orłowski, A. J. (2003). Probabilistic models of income distributions of Polish households. *Quantitative Methods in Economic Research III*.

Łukasiewicz, P., Karpio, K., & Orłowski, A. J. (2012). The models of personal incomes in USA. *Acta Physica Polonica A, 121*(2B). doi: 10.12693/APhysPol A.121.B-82.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, *52*(3). doi: 10.2307/1913469.

McDonald, J. B., & Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics, 66*(1-2). doi: 10.1016/0304-4076(94) 01612-4.

Nirei, M., & Souma, W. (2004). Income distribution and stochastic multiplicative process with reset events. In M. Gallegati, A. P. Kirman & M. Marsili (Eds.). *The complex dynamics of economic interaction*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-17045-4_9.

Nirei, M., & Souma, W. (2007). A two factor model of income distribution dynamics. *Review of Income and Wealth*, *53*(3). doi: 10.1111/j.1475-4991.2007.00242.x.

Okuyama, K., Takayasu, M., & Takayasu, H. (1999). Zipf's law in income distribution of companies. *Physica A*, *269*. doi: 10.1016/S0378-4371(99)00086-2.

Pareto, V. (1896-97). *Cours d'economie politique*. Lausanne: F. Rouge.

Quintano, C., & D'Agostino, A. (2006). Studying inequality in income distribution of single-person households in four developed countries. *Review of Income and Wealth*, *52*(4). doi: 10.1111/j.1475-4991.2006.00206.x.

Silva, A. C., & Yakovenko, V. M. (2005). Temporal evolution of the "thermal" and "superthermal" income classes in the USA during 1983–2001. *Europhysics Letters, 69*(2). doi: 10.1209/epl/i2004-10330-3.

Singh, S. K., & Manddala, G. S. (1976). A function for size distribution of incomes. *Econometrica*, *44*(5). doi: 10.2307/1911538.

Suoma, W. (2001). Universal structure of the personal income distribution. *Fractals*, *09*(04). doi: 10.1142/S0218348X01000816.

Annex

Table 1. Summary statistics of households' incomes per person in Poland during the period 2004–2012

Statistic	**Year**								
	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
N	32,054	34,569	37,282	37,131	37,107	37,031	37,189	37,099	37,129
Min	0.01	0.02	0.03	0.01	0.01	0.01	0.01	0.01	0.01
Max	228.8	505.2	420.0	854.3	1655.2	635.2	2309.5	809.5	867.2
Mean	10.12	10.33	11.20	12.44	14.14	15.22	16.33	16.83	17.59
Std. deviation	7.82	8.55	8.92	11.35	15.01	12.16	19.27	14.01	14.99
Skewness	5.18	10.86	7.92	19.47	44.80	10.05	56.47	12.22	13.04
Kurtosis	69.23	406.05	195.88	1071.64	4224.42	320.19	5921.2	442.75	479.43
Percentile 0.10	3.68	3.80	4.20	4.84	5.55	5.97	6.32	6.55	6.84
Percentile 0.25	5.56	5.68	6.25	7.05	8.10	8.77	9.26	9.68	10.14
Percentile 0.50	8.44	8.59	9.38	10.38	11.80	12.80	13.53	14.07	14.80
Percentile 0.75	12.35	12.60	13.63	14.80	16.85	18.21	19.33	20.11	21.00
Percentile 0.90	17.73	18.00	19.24	21.10	24.00	26.01	27.84	28.65	29.91
Percentile 0.95	22.44	22.80	24.51	26.99	30.09	33.00	36.00	36.73	37.98
Percentile 0.99	38.01	39.40	41.72	48.00	52.14	56.92	60.70	62.50	66.74

Source: own calculations based on the HBS data.

Table 2. Estimations of the lognormal (*LN*), Dagum (*D*), and Singh-Maddala (*SM*) models. There are standard errors of the parameters' estimators in the brackets. The symbols μ, σ are the parameters of the lognormal model

Year	Model	$\alpha(\mu)$	$\beta(\sigma)$	δ	R^2	x_m	$F(x_m)$
2004	*LN*	2.149 (0.004)	0.612 (0.003)		0.995	31.59	0.983
	D	838.0 (76.1)	2.995 (0.030)	0.787 (0.016)	0.998	59.04	0.997
	SM	0.003 ($< 10^{-4}$)	2.502 (0.016)	1.421 (0.041)	0.998	36.16	0.989
2005	*LN*	2.166 (0.003)	0.608 (0.002)		0.995	29.83	0.979
	D	1285.0 (112.8)	3.124 (0.029)	0.735 (0.014)	0.996	72.26	0.999
	SM	0.0028 ($< 10^{-4}$)	2.531 (0.015)	1.378 (0.036)	0.997	34.88	0.986
2006	*LN*	2.250 (0.002)	0.593 (0.002)		0.996	33.10	0.981
	D	1220.0 (113.4)	3.045 (0.031)	0.810 (0.016)	0.997	102.40	> 0.999
	SM	0.0020 ($< 10^{-4}$)	2.609 (0.015)	1.327 (0.034)	0.998	42.55	0,991
2007	*LN*	2.349 (0.002)	0.562 (0.002)		0.996	27.47	0.956
	D	1884.0 (109.3)	3.132 (0.019)	0.851 (0.010)	0.998	42.11	0.987
	SM	0.0011 ($< 10^{-4}$)	2.784 (0.010)	1.241 (0.019)	0.998	32.35	0.972
2008	*LN*	2.479 (0.002)	0.556 (0.001)		0.996	32.36	0.962
	D	2060 (73.0)	3.050 (0.011)	0.927 (0.007)	0.999	103.10	0.999
	SM	0.0007 ($< 10^{-4}$)	2.860 (0.007)	1.155 (0.010)	0.999	54.93	0.992
2009	*LN*	2.562 (0.002)	0.559 (0.002)		0.995	33.35	0.954
	D	2649.0 (168.3)	3.041 (0.019)	0.916 (0.012)	0.998	95.81	0.998
	SM	0.0006 ($< 10^{-4}$)	2.776 (0.010)	1.298 (0.020)	0.999	37.59	0.968
2010	*LN*	2.616 (0.001)	0.559 (0.001)		0.996	34.64	0.949
	D	2611.0 (96.0)	2.995 (0.011)	0.955 (0.007)	0.999	73.30	0.995
	SM	0.0005 ($< 10^{-4}$)	2.831 (0.006)	1.186 (0.011)	0.999	57.74	0.989
2011	*LN*	2.660 (0.002)	0.555 (0.002)		0.994	36.75	0.952
	D	5603.0 (261.5)	3.176 (0.013)	0.850 (0.008)	0.999	80.34	0.996
	SM	0.0004 ($< 10^{-4}$)	2.836 (0.009)	1.217 (0.016)	0.999	42.98	0.969

Table 2. Continued

Year	Model	$\alpha(\mu)$	$\beta(\sigma)$	δ	R^2	x_m	$F(x_m)$
2012	*LN*	2.707 (0.003)	0.557 (0.002)		0.992	40.27	0.959
	D	7823.0 (567.3)	3.222 (0.022)	0.814 (0.010)	0.998	47.70	0.976
	SM	0.0004 ($< 10^{-4}$)	2.753 (0.012)	1.386 (0.026)	0.998	40.27	0.959

Source: own calculation based on the HBS microdata.

Table 3. Estimations of the Pareto model. Values of limits x_m have been set for lognormal model. Values s_α indicate standard errors of parameters' estimators α

Year	x_m	α	s_α	R^2
2004	31.59	3.036	0.008	0.996
2005	29.83	2.887	0.006	0.996
2006	33.10	2.865	0.004	0.998
2007	27.47	2.669	0.003	0.998
2008	32.36	2.649	0.007	0.989
2009	33.35	2.928	0.003	0.998
2010	34.64	2.732	0.008	0.983
2011	36.75	2.911	0.005	0.994
2012	40.27	2.761	0.005	0.996

Source: own calculations based on the HBS data.

Table 4. Estimations of the Pareto model. Values of limits x_m have been set for Dagum model. Values s_α indicate standard errors of parameters' estimators α

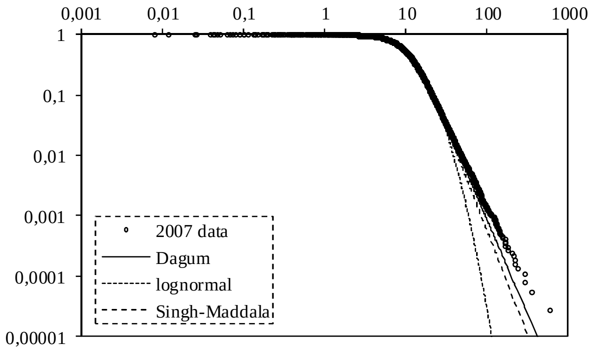
Year	x_m	α	s_α	R^2
2004	59.04	3.205	0.043	0.983
2005	72.26	2.542	0.046	0.984
2006	102.40	2.252	0.060	0.986
2007	42.11	2.567	0.085	0.994
2008	103.10	1.756	0.050	0.960
2009	95.81	2.438	0.040	0.979
2010	73.30	1.902	0.026	0.963

Table 4. Continued

Year	x_m	α	s_α	R^2
2011	80.34	2.747	0.036	0.969
2012	47.70	2.714	0.007	0.994

Source: own calculations based on the HBS data.

Figure 1. Complementary cumulative density functions of the lognormal, Dagum and Singh-Maddala models for years: 2007, 2009 and 2012 in log-log scale. The horizontal axis: annual income in thousands PLN, the vertical axis: percentage of the households



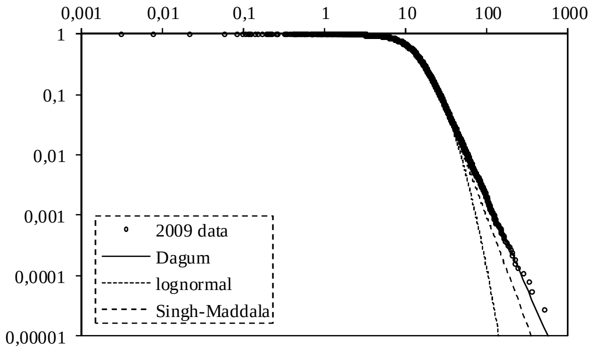
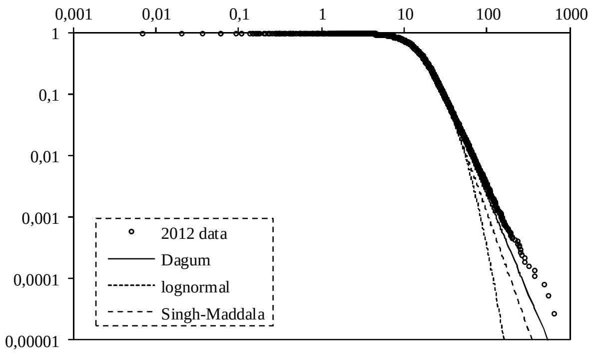
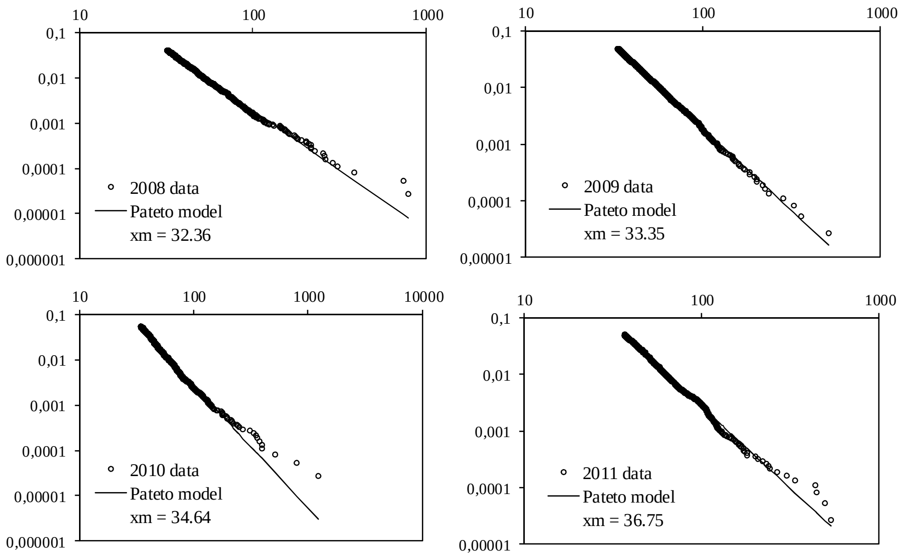


Figure 1. Continued



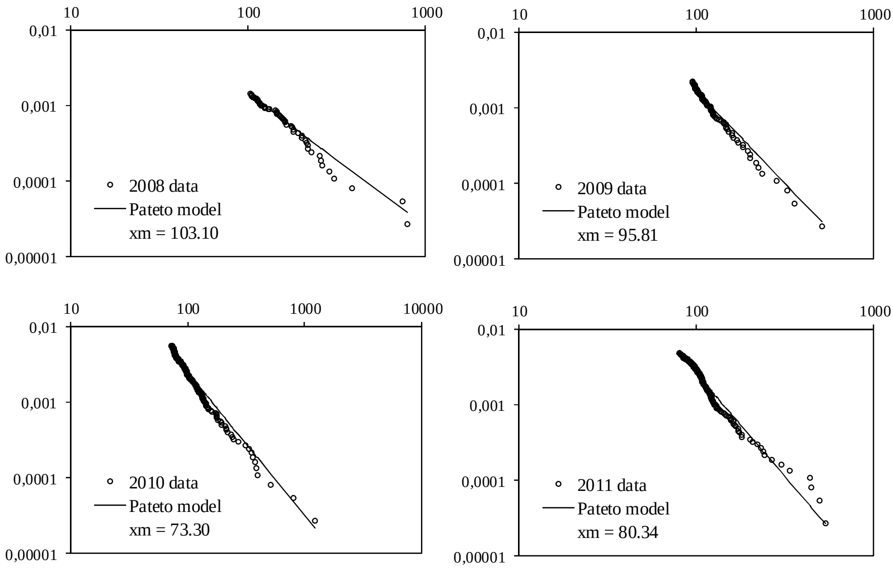
Source: own preparation based on the HBS data.

Figure 2. Tails of income distributions and Pareto model fits for: 2008, 2009, 2010 and 2011 in log-log scale. Values of limits x_m have been set for lognormal model. The horizontal axis: annual income in thousands PLN, the vertical axis: percentage of the households



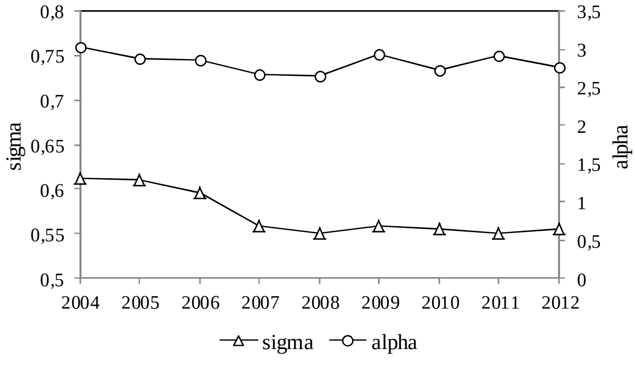
Source: own preparation based on the HBS data.

Figure 3. Tails of income distributions and fits of Pareto model for 2008, 2009, 2010 and 2011 in log-log scale. Values of limits x_m have been set for Dagum model. The horizontal axis: annual income in thousands PLN, the vertical axis: percentage of the households



Source: own preparation based on the HBS data.

Figure 4. Changes of values α and σ estimators of parameters for 2004–2012 years



Source: own preparation based on the HBS data.