Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Physica A 344 (2004) 146–151
PHYSICA A
www.elsevier.com/locate/physa

Probabilistic models of income distributions

Piotr Łukasiewicz[a], Arkadiusz Orłowski[b,c,*]

[a]Katedra Ekonometrii i Informatyki, SGGW, ul. Nowoursynowska 166, 02-787 Warszawa, Poland
[b]Instytut Fizyki, Polska Akademia Nauk, Aleja Lotników 32, 02-668 Warszawa, Poland
[c]Katedra Ekonometrii i Informatyki, SGGW, ul. Nowoursynowska 166, 02-787 Warszawa, Poland

Received 15 December 2003
Available online 23 July 2004

Abstract

The results showing relationships between distributions of individual incomes and incomes of two-earners households are presented. It is shown that individual incomes are very well described by Dagum's distributions. Income distributions in two quite different countries (USA and Poland) are studied for comparison. Obtained results show very striking and interesting differences.

PACS: 89.65.−s; 89.65.Cd

Keywords: Income distributions

1. Introduction

The preliminary results of income distributions of households in two different countries (USA and Poland) are presented. We focus on individual incomes (incomes of one-earner households) and on incomes of two-earners households. Obviously the income of a two-earners household is just the sum of individual incomes. The aim of this work is to examine if there is any relationship between a distribution of individual

*Corresponding author. Instytut Fizyki PAN & Katedra Ekonometrii i Informatyki SGGW, Aleja Lotników 32/46, Warsaw 02-668, Poland. Tel.: +48-22-8436601; fax: +48-22-8430926.
E-mail address: orlow@ifpan.edu.pl (A. Orłowski).

doi:10.1016/j.physa.2004.06.106

incomes and a distribution of incomes in two-earners households. We present result that concern two countries: USA and Poland. In the case of the USA such a relationship has recently been investigated in Ref. [1]. For a description of individual incomes the authors of that paper used the exponential model. In this work we use another model of income distributions which seems to provide a better fit to experimental data, especially for Polish households. Full results of our investigations relating individual distributions of incomes in the USA can be found at an internet site [2]. They include all the years 1975–2002. Due to the lack of space, in this paper we restrict ourselves to the year 2000. Results concerning distributions of incomes of Polish households for the year 2000 are an extension of our previous paper [3].

2. Models of individual income distributions

To indicate the best possible model for distributions of individual incomes in the USA, we start with four candidates: exponential model, Weibull's model, Dagum's model and Singh–Maddala's model. In the case of individual incomes in Poland we have studied the adjustment of Dagum's, Singh–Maddala's, gamma and log-normal models [3]. The level of adjustment of the last two models turned out to be very low and we do not consider them in this paper. The probability density functions of examined models can be described by the following formulas:

- Dagum's model:

$$f_D(x) = \frac{abcx^{b+1}}{(1+ax^{-b})^{c+1}}, \quad a>1,\ b>0,\ c>0 .$$

- Singh–Maddala's model:

$$f_{SM}(x) = \frac{abcx^{b-1}}{(1+ax^{b})^{c+1}}, \quad a>0,\ b>0,\ c>0,\ bc>1 .$$

- Exponential model:

$$f_E(x) = \frac{1}{a}\exp\left(-\frac{x}{a}\right), \quad a>0 .$$

- Weibull's model:

$$f_W(x) = \frac{b}{a}\, x^{b-1} \exp\left(-\frac{x^b}{a}\right), \quad a>0,\ b>0 .$$

Data of American households are taken from an internet site (file *mar00supp*) [4]. In the case of Polish households the data come from the Main Statistical Office (GUS) in Warsaw, Poland.

The parameters of models were estimated using maximum likelihood techniques. In the case of incomes of American households (because of large number) the data were grouped into interval series with 1800 classes of the same spread of 0.6 k$. In the case of incomes of Polish households direct data of individual households were used. The vector $\hat{\theta}$ of models parameters was estimated by maximizing the logarithm of the likelihood function. To find a degree of adjustments of a theoretical distribution to the empirical one we have calculated the sum of squared errors and sum of absolute errors, which are described by the following formulas:

$$SSE = \sum_{i=1}^{k} \left[\frac{n_i}{n} - p_i(\hat{\theta})\right]^2, \qquad SAE = \sum_{i=1}^{k} \left|\frac{n_i}{n} - p_i(\hat{\theta})\right| .$$

In addition, we use the following measures:

$$W_r = \left(1 - \frac{1}{2} SAE\right) \times 100\%, \qquad W_p = \sum_{i=1}^{k} \min\left(\frac{n_i}{n}, p_i(\hat{\theta})\right) \times 100\% .$$

In the case of a correct choice of the theoretical distribution a large agreement among empirical c_i and theoretical q_i quantiles should be observed. Therefore, one more measure of the agreement was applied, namely the square of the correlation coefficient ρ between quantiles c_i and q_i. The results of estimation of model parameters are given in Tables 1 and 2.

Results show, that among examined models, the best fitting to empirical data is provided by Dagum's model. This model well describes distributions of incomes both in American and Polish households. Adjustment of Dagum's model to empirical data is presented in Figs. 1 and 2. We accept, therefore, the Dagum's model for theoretical distributions of individual incomes in both the USA and Poland.

3. Incomes of households with two earners

Income z in two-earners household is, of course, a sum of two individual incomes x and y. Let us accept, that Z is a random variable with values z, and X and Y are random variables with values x and y, respectively. Therefore $Z = X + Y$. To examine the relationship between individual incomes (incomes of one-earner

Table 1
Individual incomes in the USA in the year 2000

Model	$\hat{a}$	$\hat{b}$	$\hat{c}$	SSE	SAE	W_r	W_p	ρ^2
Exponential	28.591	—	—	0.0008	0.1268	93.66	93.63	0.9843
Weibull	26.535	0.9806	—	0.0010	0.1392	93.04	92.97	0.9856
Dagum	29474.0	2.7434	0.3138	0.0003	0.0709	96.45	96.17	0.9987
Singh–Maddala	0.0053	1.1228	5.1245	0.0005	0.0899	95.50	95.67	0.9956

Results of estimation of parameters and adjustment of models to empirical data.

Table 2
Individual incomes in Poland in the year 2000

Model	$\hat{a}$	$\hat{b}$	$\hat{c}$	SSE	SAE	W_r	W_p	ρ^2
Dagum	12.7312	2.8498	1.4202	0.0008	0.0992	95.04	95.04	0.9987
Singh–Maddala	0.0369	1.6582	0.6955	0.0010	0.1039	94.81	94.80	0.9983

Results of estimation of parameters and adjustment of models to empirical data.

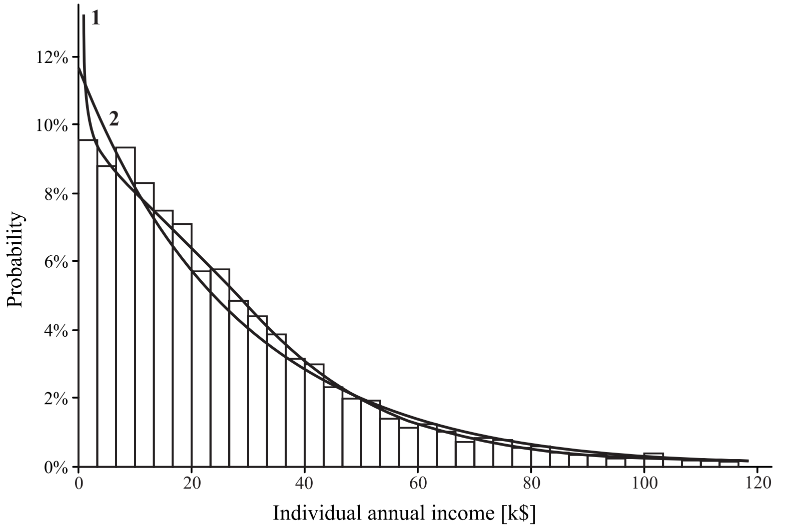

Fig. 1. Probability distribution of individual income (USA in 2000). Dagum's model (1) and exponential model (2).

households), and incomes of two-earners household we construct a convolution of individual distributions. If X and Y are independent random variables with probability density distributions f_X and f_Y then the distribution of Z is given by the formula

$$f(z) = (f_X * f_Y)(z) = \int_R f_X(t) f_Y(z-t)\,\mathrm{d}t\,.$$

In the case of American households the result of fitting the convolution to empirical data is presented in Fig. 3. A very good agreement is seen. In the case of Polish households a much less satisfactory result is presented in Fig. 4.

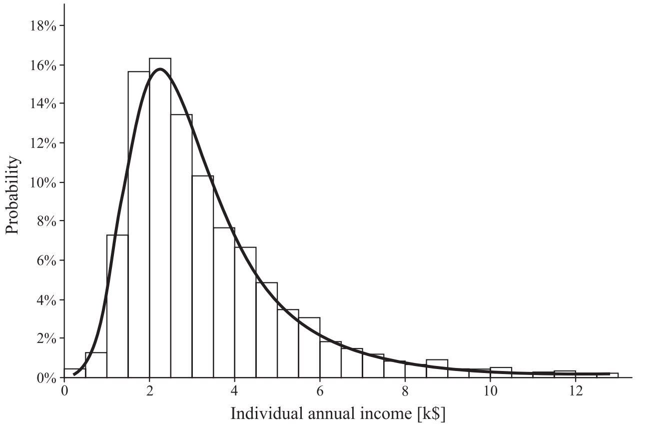

Fig. 2. Probability distribution of individual income (Poland in 2000) and fit of Dagum's model.

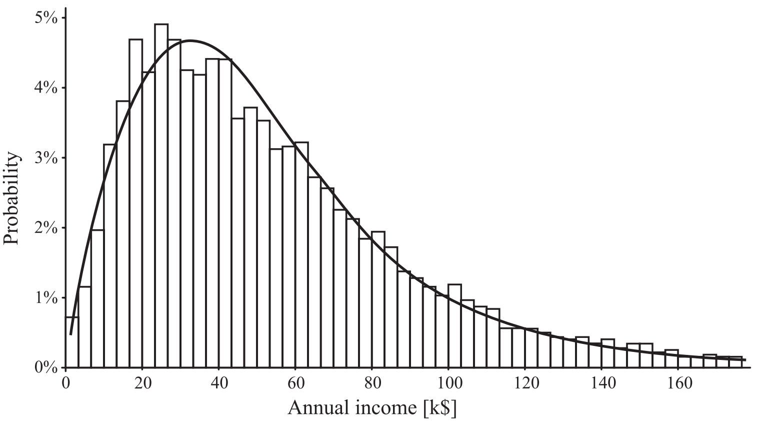

Fig. 3. Probability distribution of income for households with two earners (USA in 2000). Convolution of individual Dagum's distributions.

4. Conclusions

We have shown that Dagum's distribution describes very well the distributions of individual incomes in both the USA and Poland. We confirm, however, the results of Ref. [1] that the American incomes are well-approximated by an exponential distribution function. The distribution of incomes of two-earners household in the USA can be constructed as a convolution of the individual income distributions. This suggests that incomes of members of households in the USA are statistically

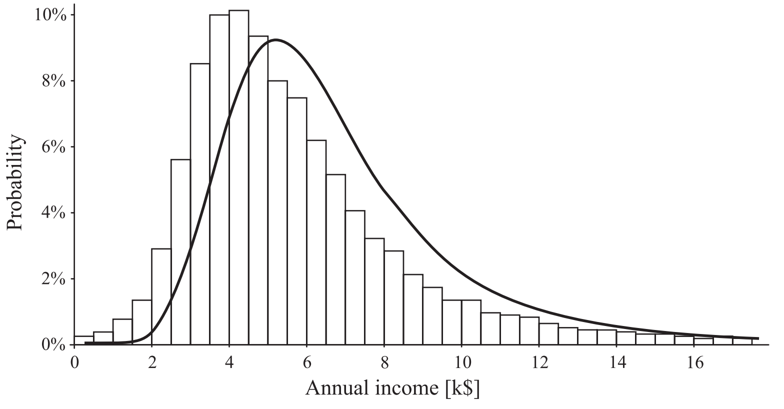

Fig. 4. Probability distribution of income for households with two earners (Poland in 2000). Convolution of individual Dagum's distributions. 1 USD = 4.1432 PLN (December 29, 2000).

independent. It is not, however, the case of the Polish households—the result are completely different. The incomes of members of households in Poland seems to be strongly statistically dependent. A further analysis of this dependence should deliver more information about distributions of incomes in Polish families. To check the generality of the obtained results, in a forthcoming paper we plan to apply our analysis to other European and non-European countries.

References

[1] A. Dragulescu, V.M. Yakovlenko, Evidence for the exponential distribution of income in the USA, Eur. Phys. J. B 20 (2000) 585–589.
[2] http://www.alpha.sggw.waw.pl/~lukasiewiczp/.
[3] P. Łukasiewicz, A. Orłowski, Probabilistic models of income distributions of Polish households, in: A. Orłowski (Ed.), Quantitative Methods in Economic Research, vol. III, Wydawnictwo SGGW, Warszawa, 2003, pp. 122–130 (in Polish).
[4] The USA Census Data; http://ferret.bls.census.gov/.