

First Evidence of Interdependences between Incomes of Family Members

P. ŁUKASIEWICZ*, K. KARPIO AND A.J. ORŁOWSKI

Faculty of Applied Informatics and Mathematics, WULS-SGGW, Nowoursynowska 159, 02-776 Warsaw, Poland

In this paper we study the relations between personal incomes and incomes of families with two adults in USA. We describe family income distributions using the simple two-parametric model. Assuming incomes of spouses are statistically independent of each other we obtain theoretical exponential income distributions for males and females. We show that these distributions are not coincident with distributions constructed based on the personal data. Obtained results indicate on statistical dependence between incomes of males and females in the families. We track changes and trends in data for years from 2001 to 2016.

DOI: [10.12693/APhysPolA.133.1441](https://doi.org/10.12693/APhysPolA.133.1441)

PACS/topics: 89.65.Gh, 88.05.Lg

1. Introduction

A distribution of personal incomes is the base structure in the studies of incomes. Its properties directly determine shape and properties of the family and household income distributions. The latter are secondary distributions because incomes of families or households are the sum of personal incomes of its members. It would be interesting to show the relations between personal and family income distributions. We study this problem analyzing USA income data from 2001 to 2016.

Personal income distribution in USA was studied widely. The primary result was presented in [1], where one proved that 99% of personal income distribution is in accordance with exponential law $P(x) = (1/p) \exp(-x/p)$, where $x > 0$ is income and $p = \langle x \rangle$. The top 1% of the distribution is in accordance with the power law. Such a structure of the personal income distribution has been observed for many years [2]. Similar results were obtained for the distribution of wealth in the United Kingdom [3]. The exponential model has a high compatibility with the empirical distributions in USA. That conclusion was brought in [4, 5], where authors compared results yielded by this model and 2–3 parametric models: Weibull, Dagum and Singh–Maddala, used in economic studies. Other alternative theoretical distributions were presented in [6–8]. These models interpolate between the exponential law for low incomes and the power law for high incomes.

Let the random variables X_1, X_2 with densities f_1, f_2 indicate on personal incomes of persons in the families with two adults. If X_1 and X_2 are independent then random variable $X_1 + X_2$ denoting total income of the family will have the density

$$f_1 * f_2(x) = \int_0^x f_1(x-t)f_2(t) dt.$$

If X_1, X_2 have the same exponential distribution P ,

then income distribution of families with two adults has the form

$$P_2(x) = (x/p^2) \exp(-x/p). \quad (1)$$

This result was presented in [1]. Additionally, in [9] authors also showed that incomes of earners in the families with two adults are uncorrelated to each other.

Instead of starting the analysis from personal income one may start from personal income distributions of males and females. They have exponential shape but differ significantly from each other (e.g. average income of males and females in 2016 are 55k\$ and 34k\$ (thousands of dollars), respectively). Convolution of exponential functions gives the model

$$P_2^*(x) = \frac{a_M a_F}{a_M - a_F} [\exp(-a_F x) - \exp(-a_M x)], \quad (2)$$

where a_F, a_M are parameters of exponential personal income distributions for males and females. They are equal to mean empirical incomes of males and females. Models (1) and (2) describe well about 94% of the empirical distribution which corresponds to about 77% of the total income.

In this paper we fit model (2) to the empirical income distributions of families with two adults (mostly marriages). We obtain estimators of parameters denoted by \hat{a}_1, \hat{a}_2 and compare them with the a_F, a_M parameters. We show that incomes of spouses X_1, X_2 are interdependent. We track trends of changes of the model parameters from 2001 to 2016.

2. Data and model estimation

Data being analyzed in this paper contain information about personal and family incomes in USA from 2001 to 2016. Files with data have been collected within the project Current Population Survey (CPS) [10], see also [11]. The CPS is monthly survey of about 50,000 households conducted by the Bureau of the Census on behave of the Bureau of Labor Statistics. The CPS is the primary source of labor force statistics in USA. We studied two variables: “total persons income” and “total family income”. Both incomes are after-tax and they are expressed in k\$. The data undergone preliminary selection:

*corresponding author

zero values (lack of data) have been eliminated; monthly incomes have been recalculated into annual ones. The final number of analyzed items, depending on year, was about 130,000 for persons and about 80,000 for families.

Model (2) was evaluated by using a nonlinear least squares method

$$\sum_{j=1}^n [\bar{F}(x_j|\theta) - \bar{F}_0(x_j)]^2,$$

where \bar{F} and \bar{F}_0 are complementary cumulative distribution functions (ccdf), theoretical and empirical respectively. Evaluation of cumulative distribution functions allowed us to obtain stable parameters estimates. We computed goodness of fit using the root mean squared error (RMSE) for ccdf and the sum squared error (SSE) for pdf

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n [\bar{F}(x_j|\theta) - \bar{F}_0(x_j)]^2},$$

$$\text{SSE} = \sum_{j=1}^n [f(x_j|\theta) - f_0(x_j)]^2.$$

The coefficient of determination R^2 has been evaluated separately for ccdf and pdf. Measurement of the fit's goodness performed based on the probability distributions yields the more precise results.

Empirical income distributions have two-part structure: a main part, covering about 98%–99% of objects (persons, families) and a 2%–1% tail. Model (2) explains the main part of the studied distributions. That is why we fitted model with normalizing parameter to the truncated distributions ($0 < x < x_{max}$). The x_{max} was evaluated using the iterative method. The starting value of x_{max} corresponded to 95% of the distribution. Next, the x_{max} was increasing till the minimum value of RMSE was reached.

3. Results

The results of estimations containing values and errors of the parameters are presented in Table I. The goodness of the fits in each year is close to one another. Average values of measures for years from 2001 to 2016 are summarized in Table II. Coefficients of determination R^2 were calculated for ccdf and pdf. Theoretical and empirical pdfs and ccdfs are presented in Fig. 1, for 2006 and 2015 years. Errors of data are indicated on the histogram bars. Evaluated functions explain from 98.0% to 98.7% of family income distributions and 90%–92% of total family incomes. Models are characterized by the high consistency with the empirical distributions and the small errors of their estimations.

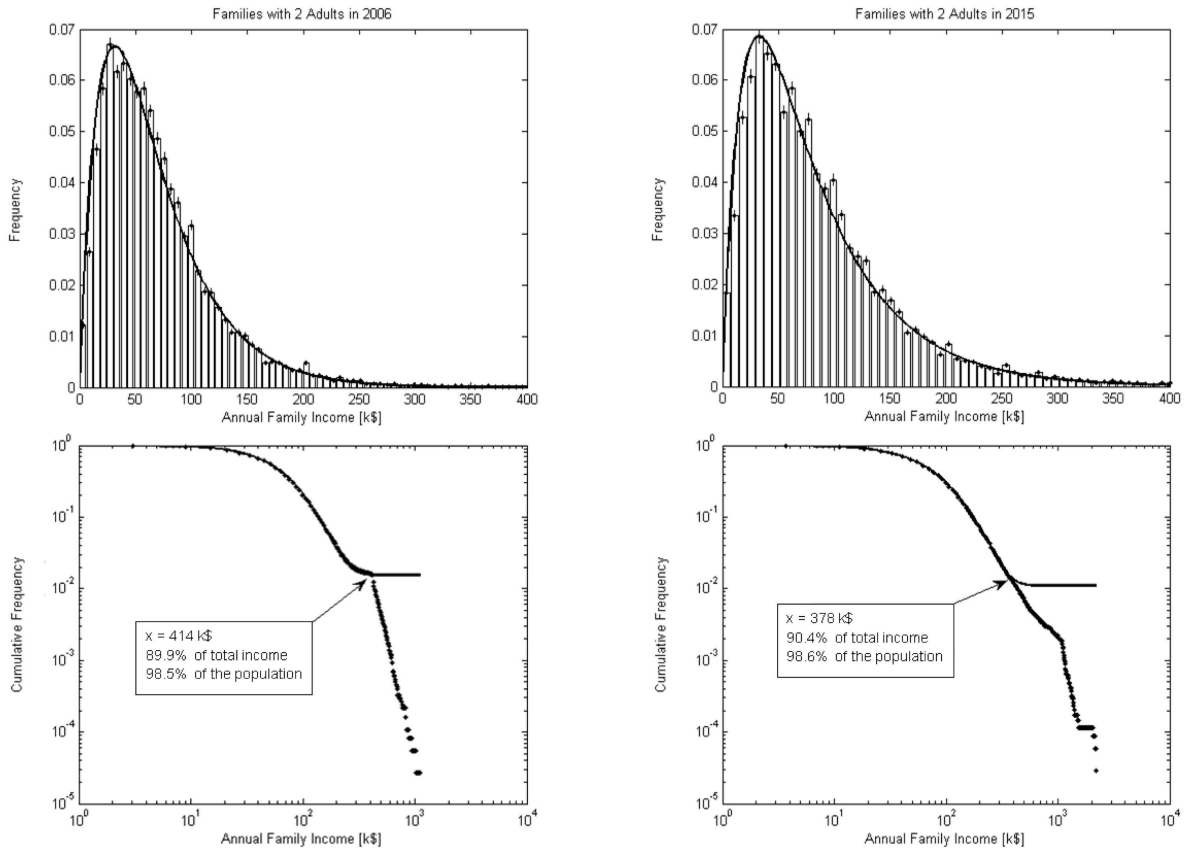


Fig. 1. Income distributions of families with 2 adults in 2006 (left parts) and 2015 (right parts) with fitted model (2). Top plot: probability distributions, bottom plot: cumulative distributions.

TABLE I

Estimators of the family model parameters and their errors.

Year	\hat{a}_1		\hat{a}_2	
2001	35.397	(0.540)	20.028	(0.461)
2002	35.030	(1.355)	25.010	(1.215)
2003	35.249	(1.088)	24.791	(0.985)
2004	39.593	(0.593)	22.584	(0.508)
2005	41.064	(0.763)	22.354	(0.645)
2006	42.891	(0.718)	23.803	(0.617)
2007	45.055	(0.540)	23.457	(0.445)
2008	44.573	(0.860)	27.112	(0.738)
2009	45.595	(0.942)	27.026	(0.801)
2010	48.586	(0.579)	22.784	(0.462)
2011	52.394	(0.535)	21.672	(0.409)
2012	55.294	(0.681)	19.473	(0.498)
2013	56.673	(0.553)	20.648	(0.405)
2014	57.917	(0.581)	19.646	(0.416)
2015	63.267	(0.651)	19.037	(0.444)
2016	64.321	(0.857)	19.488	(0.583)

TABLE II

Average values of goodness of the fit measures.

ccdf		pdf	
RMSE	R^2	SSE	R^2
0.00360	0.99983	0.00054	0.98599

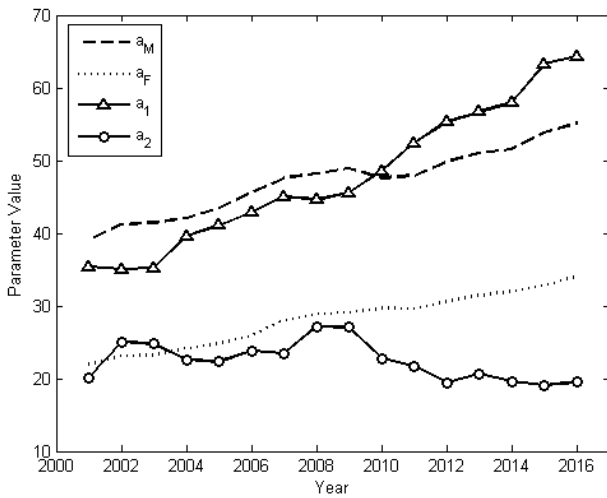


Fig. 2. Values of theoretical \hat{a}_1, \hat{a}_2 and empirical a_M, a_F parameters vs. year.

Evaluated parameters \hat{a}_1, \hat{a}_2 define theoretical exponential distributions of X_1 and X_2 for which model (2) explains income distribution for families with two adults. This result is correct when variables X_1, X_2 are independent of each other. Model (2) is coincident with whole main part of the family income distributions (excluding tail). Changes of \hat{a}_1, \hat{a}_2 are presented in Fig. 2 for years

from 2001 to 2016. The values of a_M, a_F are presented, too. Error bars are the size of the data points.

Values of evaluated theoretical parameters \hat{a}_1, \hat{a}_2 differ significantly from each other, similarly to the empirical parameters a_M, a_F . In 2001, 2004–2010 values of \hat{a}_1, \hat{a}_2 were close to the values of a_M, a_F . Starting from 2010 we observe opposite directions of their changes and increasing discrepancies with empirical parameters. Differences between a_M, a_F and \hat{a}_1, \hat{a}_2 testify about the statistical dependence of personal incomes of males and females in the families.

Model (2) with parameters a_M, a_F describes family income distribution up to the certain income x_0 only. During the years 2001–2010, when a_M, a_F are close to the theoretical \hat{a}_1, \hat{a}_2 the range of compliance is wider than during the years 2011–2016. This issue was shown in Fig. 3 for years 2006 and 2015. We can conclude that for family income below x_0 , incomes of spouses are

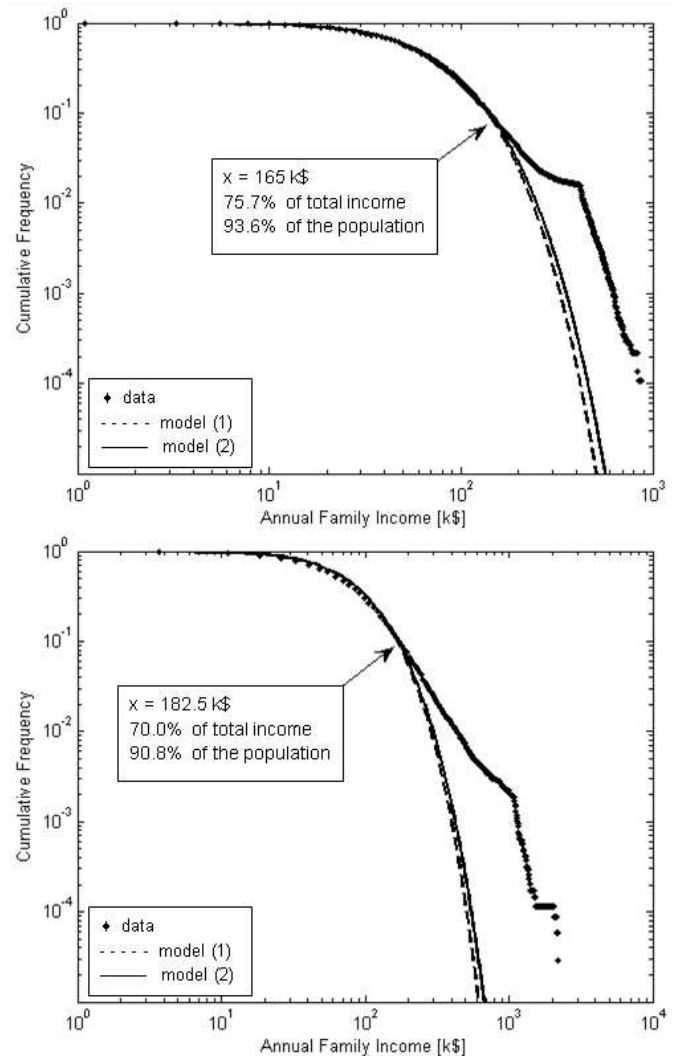


Fig. 3. Income distributions of families with 2 adults in 2006 and 2015. The dashed and solid lines indicate model (1) with parameter p and model (2) with parameters a_M, a_F , respectively.

independent of each other, so they are also uncorrelated. This is in agreement with the observation made in [9]. One can suppose, when family income is above x_0 then the incomes of the spouses are correlated with each other. This behavior causes observed interdependences between variables X_1 and X_2 . During years 2001–2010 x_0 assumes values about 150 k\$ ÷ 190 k\$ ($\approx 94\% \div 95\%$ of the income distribution) and x_0 is equal to about 180 k\$ ÷ 210 k\$ ($\approx 91\% \div 93\%$) in 2011–2016.

4. Conclusion

Earlier studies, based on USA data, showed that income distribution for families with two adults can be constructed by convoluting personal incomes. An agreement of the convolution with data indicates on independence of personal incomes of family members. The independence in turn indicates on the lack of correlations between incomes. In this paper we use opposite approach. Instead of convoluting personal income distributions we answer the questions how the personal income distributions should look like if the incomes are uncorrelated. Two-parametric model of income distribution for families with two adults was evaluated in this work. Assuming incomes of family members are independent of each other, model parameters were calculated. We showed that they differ from parameters of personal incomes for males and females and the differences increase after 2010. The differences indicate on interdependences between incomes of family members. We showed that observed interdependences are present only for sufficiently high incomes. The threshold values of family incomes below which incomes of family members are uncorrelated was indicated. It is interesting

that correlations are present only for high total incomes of families. At this stage of the analysis we cannot conclude about the personal incomes of family members. Future studies will concern quantitative description of the observed correlations. The analysis of income variances seems to be helpful here. The detailed knowledge about correlations will allow us to effectively use simulation methods.

References

- [1] A.A. Drăgulescu, V.M. Yakovenko, *Europ. Phys. J. B* **20**, 585 (2001).
- [2] A.C. Silva, V.M. Yakovenko, *Europhys. Lett.* **69**, 304 (2005).
- [3] A.A. Dragulescu, V.M. Yakovenko, *Physica A* **299**, 213 (2001).
- [4] P. Łukasiewicz, A.J. Orłowski, *Physica A* **344**, 146 (2004).
- [5] P. Łukasiewicz, K. Karpio, A.J. Orłowski, *Acta Phys. Pol. A* **121**, B-82 (2012).
- [6] A. Banerjee, V.M. Yakovenko, *New J. Phys.* **12**, 075032 (2010).
- [7] M. Jagielski, R. Kutner, M. Pęczkowski, *Acta Phys. Pol. A* **121**, B-47 (2012).
- [8] M. Jagielski, R. Kutner, *Physica A* **392**, 2130 (2013).
- [9] A.A. Dragulescu, V.M. Yakovenko, in: *Modeling of Complex Systems: Seventh Granada Lectures*, Eds. P.L. Garrido, J. Marro, AIP Conf. Proc. No. 661 AIP, New York 2003, p. 180.
- [10] [The USA Census Data](#).
- [11] P. Łukasiewicz, K. Karpio, A.J. Orłowski, *Quantitat. Meth. Econ.* **XIII**, 96 (2012).