MODELLING INCOME DISTRIBUTION IN SLOVAKIA

Alena Tartaľová

Abstract

The paper presents an estimation of income distribution with application for Slovak household's income. The two functions most often used are the Pareto and the lognormal. The Pareto function fits the data fairly well towards the higher levels but the fit is poor towards the low income levels. The lognormal fits the lower income levels better but its fit towards the upper end is far from satisfactory. We described less known models of incomes - Dagum and Singh-Maddala distribution. The considered distributions are used to fit data about Slovak household's income. The distributions fits actual data remarkably well compared with the Pareto and the lognormal. We used three concepts of income definition. We compare total disposable income, total equivalised income and income per capita and show that different definitions of household's incomes leads to different estimates of income distribution and inequality indices.

Key words: Income distribution, inequality indices, total disposable income, total equivalised income, income per capita

JEL Code: C46; D31; D63

Introduction

Analysis of income distributions is useful tool for decisions in various fields of social politics and it is crucial in estimation of household's consumption.

In analysis of income distribution usually different concept is used. We will discuss the different possible definitions of incomes. First, important is what the income unit is. The income unit could be the person, the nuclear family and the household. We will analyze the data from a survey of income and living conditions of households called EU-SILC in which the household is defined as the group of people living together at the same address with common housekeeping. Analysis about incomes usually doesn't bear in mind their size. A very simple way is to obtain income per capita, but according to Coulter (1992) there exists several disadvantages of this approach. The second approach is based on the weighting the household's income by a scale rate and obtaining the equivalised income. The scales used in

EU-SILC data are the OECD scales (or Oxford scales), but the scales and their calculation is subject for discussions and there is no general agreement about which equivalence to use. In this paper we will compare models of household's income using total disposable income, equivalised income and income per capita.

Modeling of income distribution is to find a suitable probability model. From the obtained probability distribution we could estimate basic characteristics and find the quantiles for the lowest and for the highest income. In Section 1 we describe different types of functional forms of income distribution (see Kleiber and Kotz, 2003) and in the Section 2 we proposed these models for Slovak income data.

All the calculations were executed by means of freeware R available on the internet (http://cran.rproject.org/).

1 Models of Income Distribution

The study of income distribution has a long history. The probability modelling of income distribution started with the work of Italian economist Vilfredo Pareto in 1897 and his work *Cours d' economie politique*. He described a principle which states that for many events; roughly 80% of the effects come from 20% of the causes. The original observation was in connection with population wealth. Pareto noticed that 80% of Italy's land was owned by 20% of the population. He carried out several surveys on a variety of other countries and found a similar distribution. This is nowadays known as a Pareto law.

Since the work of Pareto distribution a large number of models have been introduced to describe the distribution of incomes. Distributions of incomes are usually positively skewed with a long right tail and high density at the lowest percentiles. In order to identify the suitable model of income distribution kernel estimates are used (see Tartaľová, 2010). The most frequently used in practise are Pareto and lognormal distribution. Less known are Dagum and Singh-Maddala distribution, but we will show that they have also convenient properties for fitting income distribution.

Pareto distribution

The importance of Pareto distribution in study of income distributions is due to his good fit to empirical data. However, Pareto distribution usually poses better fit for the largest and for the smallest incomes and it is not useful as a model for the whole data. We can find

various form of Pareto distribution, there are European and American version, so one should known which version is used due to interpretation of parameters. We will use definition (1), which is probability density function defined in statistical programme R. A random variable X follows a Pareto distribution, if his probability density function is

$$f_X(x) = \frac{k.\alpha^{k+1}}{x^{k+1}} \text{ , (1)}$$

Where α is location parameter and k is shape parameter.

Lognormal distribution

Lognormal distribution is convenient for modelling not only because parameters of distribution has clear economic interpretation. Parameter μ is the logarithm of the geometric mean income and σ^2 is the variance of the logarithm of income and one of the simple inequality measures, the larger σ^2, the larger the inequality measure. Two-parametric lognormal distribution fits well part of middle income range, but gives a poor fit at the tails. A random variable X follows a lognormal distribution, if his probability density function is

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

The appropriateness of this distribution from various points of view is discussed for example in Kleiber, Kotz (2003).

Dagum distribution

Camilo Dagum (in the 1970) was not satisfied with the classical statistical distributions used to summarize income data, such as Pareto or lognormal distribution. He developed distribution, named Dagum, based on log-logistic distribution (if $p=1$, then it is Burr distribution) by adding another parameter. A random variable X follows a lognormal distribution, if his probability density function is

$$f_X(x) = \frac{\alpha p x^{\alpha p-1}}{\beta^{\alpha p}\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{p+1}} \text{ , (3)}$$

Where β is the scale parameter, α and p are shape parameters.

Singh-Maddala distribution

Singh and Maddala (1976) propose a justification of the old Burr XII distribution by considering the log survival function as a richer function of x than what the Pareto does.

$$f_X(x) = \frac{\alpha p x^{\alpha-1}}{\beta^{\alpha p}\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{p+1}}, (4)$$

Dagum and Singh-Maddala distributions are closely related (see Kleiber, 1996)

$$X \sim D(\alpha, \beta, p) \Leftrightarrow \frac{1}{X} \sim SM\left(\alpha, \frac{1}{\beta}, p\right), (5)$$

This relationship permits to translate several results pertaining to the Singh-Maddala distribution into corresponding results for the Dagum distribution.

For analyzing and visualizing income inequality are several indexes used. In this article we will discuss about Gini coefficient, Atkinson and Theil's index. For visualizing income inequality is the Lorenz curve used.

Gini coefficient

The Gini coefficient is one of the most commonly used indicators of income inequality. The Gini coefficient is usually defined mathematically based on the Lorenz curve, which plots the proportion of the total income of the population (y axis) that is cumulatively earned by the bottom x% of the population (see diagram).

$$I_G = 1 - 2\int_0^1 L(x)dx \ , (6)$$

where *L(x)* is Lorenz curve. An estimator of the population Gini coefficient is

$$I_G = 1 - \frac{2}{n-1}\left(n - \frac{\sum_{i=1}^{n} i.x_i}{\sum_{i=1}^{n} x_i}\right).(7)$$

For known function of income distribution with cumulative distribution function *F*, Gini coefficient can be calculated as

$$I_G = 1 - \frac{1}{\mu}\int_0^\infty (1 - F(x))^2 \, dx = \frac{1}{\mu}\int_0^\infty F(x)(1 - F(x)) \, dx, \ (8)$$

where $\mu = E(X)$.

Atkinson index

The Atkinson Index is one of the few inequality measures that explicitly incorporates normative judgments about social welfare (Atkinson 1970). The index is derived by calculating the so-called equity-sensitive average income (ye), which is defined as that level of per capita income which if enjoyed by everybody would make total welfare exactly equal to the total welfare generated by the actual income distribution. The equity-sensitive average income is given by:

$$I_A(\varepsilon) = 1 - \left[\int_0^\infty \left(\frac{x}{\mu}\right)^{1-\varepsilon} F(x) dx \right]^{\frac{1}{1-\varepsilon}}, \ \varepsilon > 0, \ (9)$$

where $\mu = E(X)$ and ε is the parameter that controls inequality aversions.

Theil's index

Theil's index is computed as an expectation taking the estimated parameters. A measure of inequality proposed by Theil (1967) derives from the notion of entropy in information theory. The index has a potential range from zero to infinity, with lower values (greater entropy) indicating more equal distribution of income.

$$I_T = E\left[\frac{x}{\mu} \log\left(\frac{x}{\mu}\right) \right], \ (10)$$

where $\mu = E(X)$. An estimator of the population Theil's index is:

$$I_T = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{\mu} \log\left(\frac{x_i}{\mu}\right), \ (11)$$

One way to choose between the large numbers of inequality indices available is to evaluate them in terms of their properties.

2   Application to Slovak Data

Sample surveys of household's income in the Slovakia are made by the Statistical Office. After the entrance to the European Union they annually make a survey of income and living

conditions of households called EU-SILC. In this dataset are several variables using for analysis. In many published articles Sipková (2004), Sipková and Sipko (2010), Želinský, (2010) as an income unit is total disposable income or equivalised income considered. The definitions for the analyzed concept of incomes are:

Total disposable household income (variable HY020) is calculated as the sum of the components of gross personal income of all household members plus gross income components at household level (e.g. social transfers).

The equivalised disposable income (variable HX100) is the total income of a household, after tax and other deductions, that is available for spending or saving, divided by the number of household members converted into equalised adults; household members are equalised or made equivalent by weighting each according to their age, using the so-called modified OECD equivalence scale. This scale attributes a weight to all members of the household: 1.0 to the first adult; 0.5 to the second and each subsequent person aged 14 and over; 0.3 to each child aged under 14. The equivalent size is the sum of the weights of all the members of a given household.

Total disposable income per capita (variable HY020/variable HX070) which is total disposable household income divided by the number of members of households.

**Figure 1. Histogram and characteristics of total disposable income**
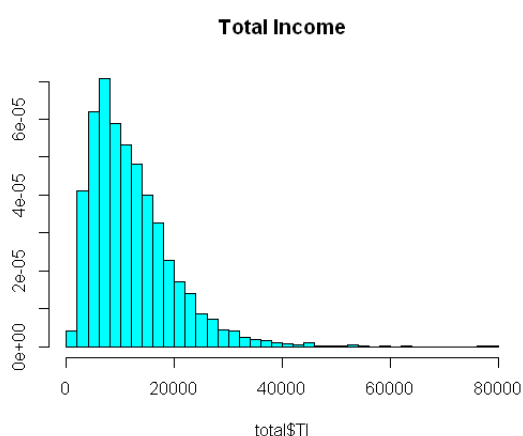
	Total income
Count	5256
Average	12127,9
Standard deviation	7869,4
Coeff. of variation	64,89%
Minimum	42,3222
Maximum	78431,7
Range	78389,4
Skewness	1,7914
Kurtosis	6,24563

Source: Own calculation based on EU-SILC

Characteristics of the samples of Slovak household's incomes in the year 2009 are presented in Figures 1.-3. The units are in Euros. There are 5256 observations. The differences between three concepts of incomes are apparent from basic characteristics. Average total household income obviously increases with household size, whereas average of per capita household

income generally decreases. Results show that total income has the highest variability (coeff. Of variation is 64,89%) and the equivalised income has the lowest variability with coefficient of variation 50.89 %.The histogram of incomes reveals right skewed distribution with extreme values on the right tail. We could suppose that distributions we considered in Section 1 are suitable for empirical data.
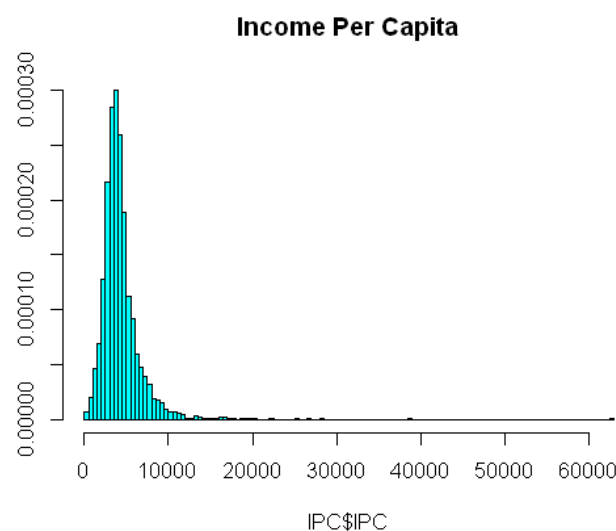
Figure 2. Histogram and characteristics of total equivalised income
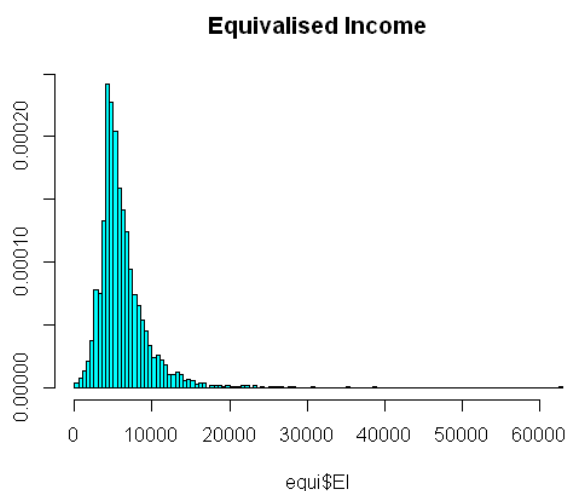
	Equivalised Income
Count	5256
Average	6090,63
Standard deviation	3099,43
Coeff. of variation	50,89%
Minimum	42,3222
Maximum	62517,9
Range	62475,6
Skewness	3,21233
Kurtosis	30,3214

Source: Own calculation based on EU-SILC

Figure 3. Histogram and characteristics of income per capita

|	*Income Per Capita*
Count	5256
Average	4282,42
Standard deviation	2357,14
Coeff. of variation	55,04%
Minimum	42,3222
Maximum	62517,9
Range	62475,6
Skewness	5,51232
Kurtosis	89,7393

Source: Own calculation based on EU-SILC

To indicate the best possible model for distribution of incomes we start with two most common used models: Pareto and lognormal distribution. We have studied also less known Dagum and Singh-Maddala distribution. The parameters of models were estimated using maximum likelihood techniques in programme R (see Table 1). We performed goodness of fit tests; the results show that among examined models we can accept Dagum and Singh-Maddala distribution. From the plots comparing estimated distribution we can see, that Dagum and Singh-Maddala distribution fit the data very good at the whole range (see Figure 4. for the lack of space there is only plot for variable total income).

Tab. 1: Results of estimation of parameters to Slovak household's incomes in 2009

Model		Total income	Equivalised Income	Income Per Capita
Pareto	*α*	0,18	0,21	0,22
	k	42,32	42,32	42,32
Lognormal	*μ*	9,20	8,61	8,25
	σ	0,66	0,48	0,49
Dagum	*α*	3,25	4,26	4,52
	β	12983,08	5829,24	4347,18
	p	0,63	0,85	0,72
Singh Maddala	*α*	2,19	3,93	3,71
	β	16735,10	5624,65	4181,13
	p	2,25	1,06	1,22

Source: Own calculation based on EU-SILC

Figure 4: Dagum (red line) and Singh-Maddala (black line) distribution fitted to the Slovak household's total incomes
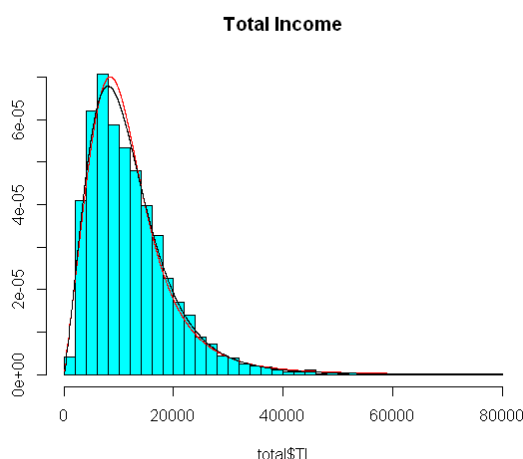
Source: Own calculation based on EU-SILC

Another picture of income distribution could be given by computing inequality indexes. We choose Gini, Atkinson and Theil index and compare results for three different definitions of incomes. The differences between the Gini indexes being quite large, the largest is for the Income per capita. According to Atkinson and Theil index the largest inequality in income is found for total income.

Tab.2: Inequality measures of Slovak household's incomes in 2009

Inequality measure	Total Income	Equivalised Income	Income Per Capita
Gini Index	0,250	0,247	0,337
Atkinson Index	0,092	0,052	0,055
Theil's Index	0,186	0,107	0,115

Source: Own calculation based on EU-SILC

Conclusion

This paper contains analysis of incomes of Slovak households in the year 2009. The analysis is based on the sample of 5256 observations from survey of income and living conditions of households called EU-SILC. We point out that concept of income definition leads to different results. From EU-SILC data three different definitions of income can be used. In this paper we have concentrated on total income, equivalised income scaled with OECD scale and income per capita scaled with the number of people in the household.

We fit income data by two commonly used models Pareto and lognormal distribution. We also introduced less known Dagum nad Singh-Maddala distribution and show that present also suitable model with good fit at the whole range. We compare fitted model for three series of data and obtain different estimates for income distribution and inequality measure. The study shows that the estimation method using per capita income and total income resulted to a higher estimate of poverty incidence in the country than for equivalised income. There is no general agreement which definition to use, but we would like to stir up discussions about it. Another topic for further research and discussion are the scales used to compute equivalised income.

Acknowledgment

References

Chotikapanich, D. (2008). *Modelling Income Distribution and Lorenz Curves*, Springer Science and Business Media LLC

Coulter, F., Cowell, F. And Jenkins, S. (1992). Differences in needs and assessment of income distributions. *Bulletin of Economic Research 44*, 77-124

Kleiber, Ch. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Science*, John Wiley&Sons, Inc., Hoboken, New Jersey

Kleiber Ch. (1996). Dagum vs. Singh-Maddala income distributions. *Economics Letters 53*, 265-268

Pacáková, V. , Sipková, Ľ. and Sodomová, E. (2004) Statistics modelling of household's incomes in the Slovak Republic, *Journal of Economics 53*,   427 -439

Sipková, Ľ. and Sipko J. (2010) Wage levels in the regions of the Slovak Republic. *SOCIALNY KAPITAL, LUDSKY KAPITAL A CHUDOBA V REGIONOCH SLOVENSKA: SCIENTIFIC CONFERENCE PROCEEDINGS*, 51-66

ŠÚ SR. 2010. EU SILC 2008, UDB verzia 26/07/2010 [databáza s mikroúdajmi]. Bratislava: Štatistický úrad SR, 2010.

Tartaľová, A. (2010). Nonparametric estimation method of probability density function. *Forum Statisticum Slovacum 5*, 250-255

Victoria-Feser, M.P. and Alaiz M.P. (1996). Modelling Income Distribution in Spain: A Robust Parametric Approach. *DARP Discussion Paper 20*, London School of Economics

Victoria-Feser, M.P. (2000). Robust methods for the analysis of income distribution, inequality and poverty. *International Statistical Review 68* (3), 277-293

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Želinský, T. (2010). Regions of Slovakia from the View of Poverty. *SOCIALNY KAPITAL, LUDSKY KAPITAL A CHUDOBA V REGIONOCH SLOVENSKA: SCIENTIFIC CONFERENCE PROCEEDINGS*, 37-50

Yee, T. (2012). *VGAM: Vector Generalized Linear and Additive Models*. R package version 0.8-7. URL http://CRAN.R-project.org/package=VGAM

Contact

Alena Tartaľová, Mgr., PhD.

Department of Applied Mathematics and Business Informatics

Faculty of Economics, TU Kosice

Nemcovej 32, 040 01 Kosice, Slovakia

alena.tartalova@tuke.sk