Chinese Business Review, ISSN 1537-1506
November 2011, Vol. 10, No. 11, 974-984

Application of Pareto Distribution in Wage Models*

Diana Bílková
University of Economics in Prague, Prague, Czech Republic

This paper deals with the use of Pareto distribution in models of wage distribution. Pareto distribution cannot generally be used as a model of the whole wage distribution, but only as a model for the distribution of higher or of the highest wages. It is usually about wages higher than the median. The parameter *b* is called the Pareto coefficient and it is often used as a characteristic of differentiation of fifty percent of the highest wages. Pareto distribution is so much the more applicable model of a specific wage distribution, the more specific differentiation of fifty percent of the highest wages will resemble to differentiation that is expected by Pareto distribution. Pareto distribution assumes a differentiation of wages, in which the following ratios are the same: ratio of the upper quartile to the median; ratio of the eighth decile to the sixth decile; ratio of the ninth decile to the eighth decile. This finding may serve as one of the empirical criterions for assessing, whether Pareto distribution is a suitable or less suitable model of a particular wage distribution. If we find only small differences between the ratios of these quantiles in a specific wage distribution, Pareto distribution is a good model of a specific wage distribution. Approximation of a specific wage distribution by Pareto distribution will be less suitable or even unsuitable when more expressive differences of mentioned ratios. If we choose Pareto distribution as a model of a specific wage distribution, we must reckon with the fact that the model is always only an approximation. It will describe only approximately the actual wage distribution and the relationships in the model will only partially reflect the relationships in a specific wage distribution.

Keywords: Pareto distribution, Pareto coefficient, estimation methods for parameters, least squares method, wage distributions

Pareto Distribution

The question of income and wage distributions and their models is quite extensively treated in the statistical literature (Bartošová, 2006; Bartošová & Bína, 2009; Bílková, 2007; Dutta, Sefton, & Weale, 2001; Majumder & Chakravarty, 1990; McDonald & Snooks, 1985; McDonald, 1984; McDonald & Butler, 1987).

* **Acknowledgment:** The paper was supported by grant project IGS 24/2010 called "Analysis of the Development of Income Distribution in the Czech Republic Since 1990 to the Financial Crisis and Comparison of This Development With the Development of the Income Distribution in Times of Financial Crisis, According to Sociological Groups, Gender, Age, Education, Profession Field and Region" from the University of Economics in Prague.

Diana Bílková, Ing./Dr., Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics in Prague.

Correspondence concerning this article should be addressed to Diana Bílková, Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics in Prague, Sq. W. Churchill 1938/4, Prague 3, Czech Republic, post code: 130 67. E-mail: bilkova@vse.cz.

Pareto distribution is usually used as a model of the distribution of the largest wages, not for the whole wage distribution. In this article, we will consider using the Pareto distribution to model wages higher than median.

The 100·P% quantile of the wage distribution will be denoted by x_P, $0 < P < 1$. This value represents the upper bound of 100·P% lowest wages and also the lower bound of 100(1 – P) % highest wages. A particular quantile (denoted as x_{P0}) which will be the lower bound of some small number of the highest wages is usually set to be the maximum wage. If the following formula (1) holds for any quantile x_P, the wage distribution is Pareto distribution.

$$\frac{x_{P0}}{x_P} = \left(\frac{1-P}{1-P_0}\right)^b \tag{1}$$

The parameter b of the Pareto distribution (1) is called the Pareto coefficient. It can be used as a characteristic of differentiation of 50% highest wages.

We will now consider a pair of quantiles x_{P1} and x_{P2}, $P_1 < P_2$. It follows from equation (1) that:

$$\frac{x_{P0}}{x_{P1}} = \left(\frac{1-P_1}{1-P_0}\right)^b \tag{2}$$

and

$$\frac{x_{P0}}{x_{P2}} = \left(\frac{1-P_2}{1-P_0}\right)^b \tag{3}$$

From what we can derive for the rate of x_{P2} to x_{P1} that:

$$\frac{x_{P2}}{x_{P1}} = \left(\frac{1-P_1}{1-P_2}\right)^b \tag{4}$$

The rate $\frac{x_{P2}}{x_{P1}}$ is an increasing function of the Pareto coefficient b. If the rate of quantiles increases, the relative differentiation of wages increases too. If only absolute differences between quantiles increase, only the absolute differentiation of wages increases.

It follows from the equation (1) that once the values x_{P0} and b are chosen, we can determine the quantile x_P for any chosen P or the other way around for any value x_P we can find the corresponding value of P. In the first case, it is advantageous to write the equation (1) as:

$$x_P = \frac{x_{P0}}{\left(\frac{1-P}{1-P_0}\right)^b} \tag{5}$$

or after logarithmic transformation as:

$$\log x_P = \log x_{P0} - b[\log(1 - P) - \log(1 - P_0)] \tag{6}$$

in the second case:

$$1 - P = (1 - P_0)\sqrt[b]{\frac{x_{P0}}{x_P}} \tag{7}$$

or after logarithmic transformation as:

$$\log(1-P) = \log(1-P_0) + \frac{1}{b}(\log x_{P0} - \log x_P) \tag{8}$$

The equations (2)-(4) will after logarithmic transformation have the following forms:

$$b = \frac{\log \frac{x_{P0}}{x_{P1}}}{\log \frac{1-P_1}{1-P_0}} \tag{9}$$

$$b = \frac{\log \frac{x_{P2}}{x_{P1}}}{\log \frac{1-P_1}{1-P_2}} \tag{10}$$

It follows from the equation (9) that instead of the Pareto coefficient b we can use any other quantile x_{P1} of the Pareto distribution and it follows from the equation (10) that the Pareto coefficient b can be calculated using any known quantiles x_{P1} and x_{P2}. Then we can also determine the value x_{P0} using the formulas:

$$x_{P0} = x_{P1}\left(\frac{1-P_1}{1-P_0}\right)^b \tag{11}$$

$$x_{P0} = x_{P2}\left(\frac{1-P_2}{1-P_0}\right)^b \tag{12}$$

The model characterized with the relationship (1) will be practically applicable if the following is known:

- The value of the quantile that characterizes the assumed wage maximum and the value of the Pareto coefficient b;
- The value of the quantile that characterizes the assumed wage maximum and the value of any other quantile;
- The values of any two quantiles of the Pareto distribution.

Any two quantiles can be written as x_P and x_{P+k}, where $0 < k < 1 - P$. Using the equation (4), we can derive for the rate of these two quantiles:

$$\frac{x_{P+k}}{x_P} = \left(\frac{1-P}{1-P-k}\right)^b \tag{13}$$

The rate (13) will be equal for such pairs of quantiles for which the following formula holds:

$$\frac{1-P}{1-P-k} = c \tag{14}$$

where c is a constant, i.e., the rate will be the same for all pairs of quantiles for which:

$$k = \frac{c-1}{c}(1-P) \tag{15}$$

We will use the constant $c = 2$ in equation (15) and we will choose gradually $P = 0.5$; 0.6; 0.8. Then using the equation (13) we can show the equality of rates of some frequently used quantiles:

$$\frac{x_{0.75}}{x_{0.5}} = \frac{x_{0.8}}{x_{0.6}} = \frac{x_{0.9}}{x_{0.8}} \tag{16}$$

From the relationship (16) we can conclude that Pareto distribution assumes such a wage differentiation for which the rate of the upper quartile to median is the same as:

- The rate of the 8th to the 6th decile;
- And as the rate of the 9th to the 8th decile.

If in a particular case, the observed differences of the rates of the above mentioned quantiles are negligible, Pareto distribution will be an appropriate model of the considered wage distribution. In the case, the differences are quite material, the approximation of the considered wage distribution with Pareto distribution will be more or less inappropriate. More about the theory of Pareto distribution is described in statistical literature (Forbes, Evans, Hastings, & Peacock, 2011; Johnson, Kotz, & Balakrishnan, 1994; Kleiber & Kotz, 2003; Krishnamoorthy, 2006).

Parameter Estimates

If the Pareto distribution is chosen as a model for a particular distribution we have to keep in mind that this model is only an approximation. The wage distribution will be only approximated and the relations derived from the model will also hold for the "true distribution" only approximately. Which relations will hold more precisely and for which the precision will be lower will be mostly dependent on the method of parameter estimates.

There are many possibilities to choose from. In the following text the quantiles of Pareto distribution will be denoted as x_P and the quantiles of the observed wage distribution will be denoted as y_P.

First we need to decide which quantile to choose as x_{P0}. It this article we will assume that $x_{P0} = x_{0.99}$. From the equation (1) we can see that the considered Pareto distribution will be defined by the equation:

$$\frac{x_{0.99}}{x_P} = \left(\frac{1-P}{0.01}\right)^b \tag{17}$$

Then we need to determine the value $x_{0.99}$ and the value of the Pareto coefficient b. Because it is necessary to estimate the values of two parameters we need to choose two equations to estimate from.

A natural choice is the equation $x_{P0} = y_{P0}$; that is in our case $x_{0.99} = y_{0.99}$. As the other equation we set a quantile x_{P1} equal to the corresponding observed quantile, i.e., $x_{P1} = y_{P1}$. In this case, the parameters of the model will be:

$$x_{P0} = y_{P0} \tag{18}$$

and using equation (9):

$$b = \frac{\log \frac{y_{P0}}{y_{P1}}}{\log \frac{1-P_1}{1-P_0}} \tag{19}$$

We can get different modifications using different choice of the maximum wage and the second quantile. If we use equation $x_{0.99} = y_{0.99}$ and we use the median in the second equation, i.e., $x_{0.5} = y_{0.5}$ we get a model with parameters:

$$x_{0.99} = y_{0.99} \tag{20}$$

$$b = \frac{\log \frac{y_{0.99}}{y_{0.5}}}{\log \frac{0.5}{0.01}} \tag{21}$$

Another possibility is setting any two quantiles of the model equal to the quantiles of the observed distribution:

$$x_{P_1} = y_{P_1} \tag{22}$$

$$x_{P_2} = y_{P_2} \tag{23}$$

Using the formula (10), we get the following parameter estimates:

$$b = \frac{\log \frac{y_{P2}}{y_{P1}}}{\log \frac{1-P_1}{1-P_2}} \tag{24}$$

and from equations (11) and (12) we get:

$$x_{P0} = y_{P1}\left(\frac{1-\mathrm{P}_1}{1-\mathrm{P}_0}\right)^b = y_{P2}\left(\frac{1-P_2}{1-P_0}\right)^b \tag{25}$$

With this alternative we can also get numerous modifications depending on the choice of quantiles y_{P1} and y_{P2} that are used.

The third possibility is based on the request that $x_{P0} = y_{P0}$ and that the rate of some other two quantiles of the Pareto distribution x_{P2}/x_{P1} is equal to the rate y_{P2}/y_{P1} of correspoding quantiles of the wage distribution observed. In this case we will estimate the parameters using equation (10):

$$x_{P0} = y_{P0} \tag{26}$$

$$b = \frac{\log \frac{y_{\mathrm{P2}}}{y_{P1}}}{\log \frac{1-\mathrm{P}_1}{1-P_2}} \tag{27}$$

In this case, notwithstanding that $x_{P2}/x_{P1} = y_{P2}/y_{P1}$ holds, the equality of quantiles itself, $x_{P1} \neq y_{P1}$ and $x_{P2} \neq y_{P2}$, does not hold. In this case, we can also arrive to numerous modifications depending on what maximum wage is chosen and what quantiles y_{P1} and y_{P2} are chosen.

For all of the above methods the equality of two characteristics of the model and the observed distribution was required. There are also different approaches to the parameter estimates.

The least squares method is frequently used for the Pareto distribution parameter estimates as well. We will consider the following quantiles of the observed wage distribution $y_{P1}, y_{P2}, \ldots, y_{Pk}$ and corresponding quantiles of the

Pareto distribution $x_{P1}, x_{P2}, \ldots, x_{Pk}$. The model distribution will be most precise when the sum of squared differences:

$$\sum_{i=1}^{k}(y_{P_i} - x_{P_i})^2 \tag{28}$$

is minimized. In this case closed formula solution does not exist. Therefore sum of squared differences of logarithms of quantiles is often considered:

$$\sum_{i=1}^{k}(\log y_{P_i} - \log x_{P_i})^2 \tag{29}$$

Minimizing the objective function (29), it is possible to derive the following estimates:

$$b = \frac{k \sum_{i=1}^{k} \log y_{P_i} \log \frac{1-P_0}{1-P_i} - \sum_{i=1}^{k} \log y_{P_i} \sum_{i=1}^{k} \log \frac{1-P_0}{1-P_i}}{k \sum_{i=1}^{k} \log^2 \frac{1-P_0}{1-P_i} - \left(\sum_{i=1}^{k} \log \frac{1-P_0}{1-P_i} \right)^2} \tag{30}$$

$$\log x_{P_0} = \frac{\sum_{i=1}^{k} \log y_{P_i}}{k} - b \frac{\sum_{i=1}^{k} \log \frac{1-P_0}{1-P_i}}{k} \tag{31}$$

In the case we use this estimating method, it is needed to keep in mind that the equality of model quantiles and observed quantiles is not guaranteed for any P. Again we can arrive to different results depending on what quantiles $y_{P1}, y_{P2}, \ldots, y_{Pk}$ are used for the calculations. Furthermore the parameter estimates also depend on the choice of the maximum wage.

Characteristics of the Appropriateness of the Pareto Distribution

For the application of Pareto distribution as a model of the wage distribution, it is crucial that the model fits the observed distribution as close as possible. It is important that the observed relative frequencies in particular wage intervals are as close to the theoretical probabilities assigned to these intervals by the model as possible.

It is needed to note that the same parameter estimation method does not always lead to the best results. It is of particular importance in "what direction" is the observed wage distribution different from Pareto distribution. Pareto distribution assumes such wage differentiation that the relations (16) hold. With real data we can encounter many different situations:

$$\frac{y_{0.75}}{y_{0.5}} < \frac{y_{0.8}}{y_{0.6}} < \frac{y_{0.9}}{y_{0.8}} \tag{32}$$

$$\frac{y_{0.75}}{y_{0.5}} > \frac{y_{0.8}}{y_{0.6}} > \frac{y_{0.9}}{y_{0.8}} \tag{33}$$

$$\frac{y_{0.75}}{y_{0.5}} < \frac{y_{0.9}}{y_{0.8}} < \frac{y_{0.8}}{y_{0.6}} \tag{34}$$

$$\frac{y_{0.75}}{y_{0.5}} > \frac{y_{0.9}}{y_{0.8}} > \frac{y_{0.8}}{y_{0.6}} \tag{35}$$

$$\frac{y_{0.8}}{y_{0.6}} < \frac{y_{0.75}}{y_{0.5}} < \frac{y_{0.9}}{y_{0.8}} \tag{36}$$

$$\frac{y_{0.8}}{y_{0.6}} > \frac{y_{0.75}}{y_{0.5}} > \frac{y_{0.9}}{y_{0.8}} \tag{37}$$

It follows from equations (32)-(37) that the observed distributions will more or less systematically differ from the Pareto distribution. In the case of equation (32) the differentiation of the observed wage distribution is higher; in the case of equation (33) the differentiation will be lower than in the case of Pareto distribution. Some bias occurs in cases equations (34)-(37) as well (but cannot be so specified). Systematical bias should be a signal for potential adjustment of the model which could be based for example on adding one or more parameters into the model. These adjustments usually lead to more complicated models. Therefore, the above mentioned bias is often neglected and simple models are preferred even though they lead to some bias.

Wage Distribution of Males and Females in the Czech Republic in 2001-2008

The data used in this article is the gross monthly wage of male and female in CZK in the Czech Republic in the years 2001-2008. Data were sorted in the table of interval distribution with opened lower and upper bound in the lowest and highest interval respectively. The source is the web page of the Czech statistical office. The following quantiles were calculated (see Table 1).

Table 1

Median $y_{0.50}$ (in CZK), 6th Decile $y_{0.60}$ (in CZK), Upper Quartile $y_{0.75}$ (in CZK), 8th Decile $y_{0.80}$ (in CZK), 9th Decile $y_{0.90}$ (in CZK) a 99th Percentile $y_{0.99}$ (in CZK) of Gross Monthly Wages in the Czech Republic in the Years 2001-2008 (Total and Split up to Male and Female Separated)

	Year	$y_{0.50}$	$y_{0.60}$	$y_{0.75}$	$y_{0.80}$	$y_{0.90}$	$y_{0.99}$
Total	2001	12,502	14,042	16,987	18,254	23,319	44,921
	2002	15,545	17,125	20,215	22,193	27,754	47,172
	2003	16,735	18,458	22,224	23,797	29,590	47,719
	2004	17,709	19,557	23,077	24,849	31,082	56,369
	2005	18,597	20,566	24,470	26,328	33,292	56,852
	2006	19,514	21,564	25,675	27,693	35,230	57,326
	2007	20,987	23,227	27,590	29,900	37,892	66,395
	2008	22,310	24,696	29,553	31,769	40,541	68,828
Males	2001	14,152	15,781	19,037	20,697	26,264	46,781
	2002	16,985	18,667	22,604	24,199	31,101	48,047
	2003	18,240	20,116	24,145	26,041	34,564	48,417
	2004	19,344	21,321	25,306	27,286	34,819	57,514
	2005	20,281	22,446	26,822	28,989	37,211	57,808
	2006	21,199	23,460	28,090	30,525	39,381	58,104
	2007	22,933	25,366	30,284	32,663	42,815	70,522
	2008	24,498	27,115	32,343	35,105	46,375	72,338
Females	2001	10,770	12,187	14,655	15,700	18,904	37,526
	2002	13,746	15,181	17,727	18,903	23,291	43,339
	2003	14,831	16,453	19,281	20,628	24,637	44,883
	2004	15,642	17,303	20,293	21,560	25,776	50,776
	2005	16,454	18,211	21,426	22,804	27,503	52,508
	2006	17,311	19,202	22,530	23,966	29,082	54,054
	2007	18,390	20,392	24,024	25,924	31,338	58,649
	2008	19,399	21,600	25,558	27,215	33,405	63,628

From Table 2, we can see that, with the exception of male in the year 2003, 2007 and 2008, all other wage distributions have lower differentiation than Pareto distribution. The systematical error occurred also in the case of male in the year 2003, 2007 and 2008. It follows from the empirical criterion (16) and from Table 2 that in all cases the differences between the rates of the considered quantiles are negligible and therefore Pareto distribution can be used as the model of the distribution.

The 99th percentile will be considered as a characteristic of the maximum wage. The parameters of the Pareto distribution are estimated using the above described methods.

First we consider the conditions $x_{P0} = y_{P0}$ a $x_{P1} = y_{P1}$ and we chose median as the second quantile, i.e., $x_{0.99} = y_{0.99}$ and $x_{0.5} = y_{0.5}$. We estimate the parameter b using the formula (21). The summary of the parameter estimates is in Table 3.

Table 2

The Rates of Quantiles y_{75}/y_{50}, y_{80}/y_{60} and y_{90}/y_{80} of the Wage Distributions in the Years 2001-2008 and Its Relations

	Year	$\frac{y_{0.75}}{y_{0.50}}$	$\frac{y_{0.80}}{y_{0.60}}$	$\frac{y_{0.90}}{y_{0.80}}$	Relations between quantile rates
Total	2001	1.358815	1.299910	1.277456	(21.2)
	2002	1.300422	1.295897	1.250612	(21.2)
	2003	1.327994	1.289216	1.243457	(21.2)
	2004	1.303112	1.270608	1.250812	(21.2)
	2005	1.315815	1.280162	1.264514	(21.2)
	2006	1.315734	1.284213	1.272161	(21.2)
	2007	1.314623	1.287295	1.267291	(21.2)
	2008	1.324653	1.286403	1.276118	(21.2)
Males	2001	1.345148	1.311556	1.268936	(21.2)
	2002	1.330847	1.296386	1.285203	(21.2)
	2003	1.323680	1.294561	1.327273	(21.5)
	2004	1.308222	1.279734	1.276084	(21.2)
	2005	1.322543	1.291532	1.283632	(21.2)
	2006	1.325086	1.301135	1.290146	(21.2)
	2007	1.320542	1.287669	1.310810	(21.4)
	2008	1.320230	1.294671	1.321037	(21.5)
Females	2001	1.360723	1.288227	1.204113	(21.2)
	2002	1.289624	1.245137	1.232163	(21.2)
	2003	1.300019	1.253747	1.194319	(21.2)
	2004	1.297375	1.246052	1.195526	(21.2)
	2005	1.302189	1.252237	1.206076	(21.2)
	2006	1.301488	1.248134	1.213470	(21.2)
	2007	1.306362	1.271283	1.208841	(21.2)
	2008	1.317491	1.259954	1.227448	(21.2)

Next we apply the conditions $x_{P1} = y_{P1}$ and $x_{P2} = y_{P2}$ and we choose 6th and 9th decile for y_{P1} and y_{P2}. We use the formulas (24) and (25) to estimate the parameters. The summary of the parameter estimates is in Table 3.

Parameters of the Pareto distribution can also be estimated using the equations $x_{P0} = y_{P0}$ and $x_{P2}/x_{P1} = y_{P2}/y_{P1}$. We choose the 9th and 6th decile in the rate y_{P2}/y_{P1}. In this case we use the relations (26) and (27) to estimate the parameters. The summary of the parameter estimates is also in Table 3.

Table 3

Estimated Parameters of Pareto Distribution for Different Choices of the Estimation Equations

		Equations used					
		$x_{0.99} = y_{0.99}$, $x_{0.5} = y_{0.5}$		$x_{0.6} = y_{0.6}$, $x_{0.9} = y_{0.9}$		$x_{0.99} = y_{0.99}$, $\frac{x_{0.9}}{x_{0.6}} = \frac{y_{0.9}}{y_{0.6}}$	
		Parameter estimates		Parameter estimates		Parameter estimates	
	Year	x_{P0}	b	x_{P0}	b	x_{P0}	b
Total	2001	44,921	0.326952	54,143	0.365843	44,921	0.365843
	2002	47,172	0.283758	61,890	0.348293	47,172	0.348293
	2003	47,719	0.267846	64,800	0.340425	47,719	0.340425
	2004	56,369	0.295969	67,096	0.334192	56,369	0.334192
	2005	56,852	0.299456	74,095	0.347455	56,852	0.347455
	2006	57,326	0.275468	79,614	0.354083	57,326	0.354083
	2007	66,395	0.294405	85,426	0.353045	66,395	0.353045
	2008	68,828	0.287978	92,352	0.357552	68,828	0.357552
Males	2001	46,781	0.305624	61,207	0.367449	46,781	0.367449
	2002	48,047	0.265814	72,613	0.368246	48,047	0.368246
	2003	48,417	0.249540	84,934	0.390464	48,417	0.390464
	2004	57,514	0.278536	78,632	0.353784	57,514	0.353784
	2005	57,808	0.267749	86,165	0.364658	57,808	0.364658
	2006	58,104	0.257739	93,098	0.373653	58,104	0.373653
	2007	70,522	0.287153	102,142	0.377610	70,522	0.377610
	2008	72,338	0.276777	113,087	0.387128	72,338	0.387128
Females	2001	37,526	0.319087	39,196	0.316679	37,526	0.316679
	2002	43,339	0.293539	47,418	0.308749	43,339	0.308749
	2003	44,883	0.283055	48,172	0.291217	44,883	0.291217
	2004	50,776	0.300989	49,971	0.287505	50,776	0.287505
	2005	52,508	0.296625	54,551	0.297414	52,508	0.297414
	2006	54,054	0.291062	57,954	0.299456	54,054	0.299456
	2007	58,649	0.296461	63,977	0.309955	58,649	0.309955
	2008	63,628	0.303636	68,917	0.314516	63,628	0.314516

In the end we also estimate the parameters of the Pareto distribution using the least squares method. We use the relations (30) and (31). In this method, we choose 5th, 6th, 7th, 8th and 9th deciles of the observed wage distribution, i.e., $k = 5$. Parameters estimated using the least squares method are summarized in Table 4.

Table 4

Parameters Estimated Using the Least Squares Method

Year	Parameter estimates					
	Total		Males		Females	
	x_{P0}	b	x_{P0}	b	x_{P0}	b
2001	56.562	0.379911	63.774	0.379912	42.520	0.341047
2002	64.026	0.358469	73.770	0.372825	49.188	0.320682
2003	67.219	0.351034	85.080	0.391617	51.125	0.309187
2004	69.311	0.344615	80.310	0.360986	52.763	0.303849
2005	76.310	0.356935	88.251	0.372535	57.413	0.312826
2006	81.721	0.362626	95.225	0.381012	60.917	0.315022
2007	88.022	0.362359	103.405	0.383183	67.572	0.325878
2008	94.849	0.366387	114.131	0.391293	72.463	0.330659

The values of the sum of absolute differences of observed and theoretical absolute frequencies of all intervals calculated for all cases considered wage distributions are in Table 5. In the case of the theoretical frequencies at first we determined theoretical probabilities using the formula (8). From these, we determined theoretical absolute frequencies.

Table 5

Sums of the Absolute Differences of the Observed and Theoretical Frequencies

	Year	Equations used			
		$x_{0.99} = y_{0.99}$ $x_{0.5} = y_{0.5}$	$x_{0.6} = y_{0.6}$ $x_{0.9} = y_{0.9}$	$x_{0.99} = y_{0.99}$ $\frac{x_{0.9}}{x_{0.6}} = \frac{y_{0.9}}{y_{0.6}}$	Least squares method
Total	2001	37,459	23,255	85,795	23,859
	2002	51,358	27,327	171,404	31,658
	2003	73,388	36,520	204,535	39,722
	2004	103,625	64,422	249,348	66,249
	2005	167,946	69,930	353,661	68,679
	2006	157,094	68,849	426,442	69,104
	2007	268,740	260,786	322,437	262,224
	2008	282,396	253,373	372,117	257,050
Males	2001	20,603	10,089	56,291	9,959
	2002	33,576	19,711	111,796	20,298
	2003	47,909	23,576	96,863	23,747
	2004	60,241	32,457	178,858	33,076
	2005	81,505	35,349	220,276	36,321
	2006	96,789	37,737	250,764	37,653
	2007	140,965	138,678	202,143	139,428
	2008	135,960	133,953	173,262	135,089
Females	2001	24,256	23,926	23,687	21,270
	2002	23,697	16,716	42,148	18,595
	2003	37,215	30,902	40,237	30,011
	2004	45,429	41,416	45,460	40,957
	2005	51,793	41,615	52,493	41,449
	2006	58,014	41,137	74,302	41,812
	2007	138,241	128,854	150,258	127,313
	2008	140,955	132,125	155,071	131,224

Conclusions

The appropriateness of particular modifications of the Pareto distribution can be evaluated comparing the theoretic and empirical frequencies. It is possible to compare both the absolute and relative differences between the theoretic and observed empirical distributions. In this article we used the absolute differences. The values sums these differences are in Table 5. The values seem to be relatively high. The question of appropriateness of a given theoretic wage distribution in the case of large samples was described in statistical literature (Bílková, 2007). Some more general conclusions can be made from the values of the absolute differences of observed and theoretic distributions.

With the exception of the wage distribution of women in 2001, the worst results are achieved using the equation $x_{0.99} = y_{0.99}$ and setting the ratio of other two quantiles of the Pareto distribution $x_{0.9}/x_{0.6}$ equal to the ratio $y_{0.9}/y_{0.6}$ of the corresponding empirical quantiles. This fact is less obvious for female distribution and most obvious for total distribution. This is also due to the larger sample size of the total sample (in comparison with the sample size of the sub-groups of male and female). Again with the exception of the wage distribution of women in 2001 the second worst model is the estimate based on the equations $x_{0.99} = y_{0.99}$ and $x_{0.5} = y_{0.5}$. This fact is again less obvious for female distribution and most obvious for total distribution. In the case of the wage distribution of women in 2001, the worst estimate is based on the equations $x_{0.99} = y_{0.99}$ and $x_{0.5} = y_{0.5.}$ In the case of the total group is the third worst (second best) method the least squares method (with the exception of 2005). The best results are achieved with the method based on the equations $x_{0.6} = y_{0.6}$ and $x_{0.9} = y_{0.9}$. In the case of the total wage distribution in 2005 is the third worst method based on the equations $x_{0.6} = y_{0.6}$ and $x_{0.9} = y_{0.9}$ and the best method is the least squares method. In the case of the wage distribution of male (with the exception of the years 2001 and 2006), the third worst (second best) results are again achieved using the least squares method. The best results are achieved with the method based on the equations $x_{0.6} = y_{0.6}$ and $x_{0.9} = y_{0.9}$. In the years 2001 and 2006 (set of men) is the third worst method the method based on the equations $x_{0.6} = y_{0.6}$ and $x_{0.9} = y_{0.9}$ and the best is the least squares method. In the case of the female group (with the exception of the years 2001, 2002 and 2006) is the third worst (second best) method based on the equations $x_{0.6} = y_{0.6}$ and $x_{0.9} = y_{0.9}$ and the most precise results are achieved with the least squares method. In the years 2001, 2003, 2004 and 2005 was for the group of women the most precise the least squares method. The very best method for the group of male in 2001 was the least squares method. In this case other methods had much higher values of the above mentioned sum of absolute differences.

From the above described comparison, it is obvious that the simplest parameter estimating methods can be in the case of the Pareto distribution competing with more advanced methods.

References

Bartošová, J. (2006). Logarithmic-normal model of income distribution in the Czech Republic. *Austrian Journal of Statistics, 35*(23), 215-222.

Bartošová, J., & Bína, V. (2009). Modelling of income distribution of Czech households in years 1996-2005. *Acta Oeconomica Pragensia, 17*(4), 3-18.

Bílková, D. (2007). Modeling of income distributions using lognormal distribution. In *10th International Scientific Conference AMSE—Applications of matematics and statistics in Ekonomy*. Poprad-Tatry, Slovakia, August 29-September 1, 2007, CD.

Dutta, J., Sefton, J. A., & Weale, M. R. (2001). Income distribution and income dynamics in the United Kingdom. *Journal of Applied Econometrics, 16*(5), 599-617.

Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011). *Statistical diatributions* (4th ed.) (p. 212). New Jersey: John Wiley & Sons.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions* (2nd ed.) (p. 756). New York: John Wiley & Sons.

Kleiber, Ch., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences* (p. 332). New Jersey: John Wiley & Sons.

Krishnamoorthy, K. (2006). *Handbook of statistical distributions with applications* (p. 346). Boca Rato: Chapman & Hall/CRC Press.

Majumder, A., & Chakravarty, S. R. (1990). Distribution of personal income: Development of a new model and its application to U. S. income data. *Journal of Applied Econometrics, 5*(2), 189-196.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica, 52*(3), 647-665.

McDonald, J. B., & Butler, R. J. (1987). Some generalized mixture distributions with an application to unemployment duration. *The Review of Economics and Statistics, 69*(2), 232-240.

McDonald, J. B., & Snooks, G. D. (1985). The determinants of manorial income in domesday England: Evidence from Essex. *The Journal of Economic History, 45*(3), 541-556.