STATI

MODELLING OF INCOME DISTRIBUTION OF CZECH HOUSEHOLDS IN THE YEARS 1996–2005

Jitka Bartošová, Vladislav Bína*

Introduction

The choice of this paper's topic is based on current questions related to post-Velvet Revolution development of the Czech Republic's economy. Many important economic indexes have been changing throughout the shift from a planned to a market economy. The transformation process and the subsequent globalisation process in the EU influences the living standards of individuals and households. Current disposable income is often preferred to current consumption as a proxy for the standard of living of each member of the population (Chaudhuri and Ravallion, 1994), since it is an ex-ante measure of available resource free from preferences and constraints that could affect consumption decisions. After appropriate adjustments, income is variously interpreted as a measure of the productivity of labour and welfare of the society. Income of individuals and households is an extensively studied subject in economics.

The planned form of economy prior to the Velvet Revolution was characterised by its high homogeneity of population income across social classes. After 1989, however, significant changes in income distribution of population have been occurring due to the transformation from a planned to a market economy. The impact of regional and demographic factors rises; incomes of households of pensioners and the unemployed deserve a particular attention. The significant changes that the transformation process brings about lead, among other things, to impairment of the current statistical model and discontinuation of time lines of the distribution characteristics. This situation requires a certain precaution with respect to previously valid research methods and also

conclusions that were based on them. Household income distribution in some other transitional countries (e.g., Slovakia, Poland, and Hungary) experiences a similar situation.

Analyses and modelling of the state and development of incomes in transitional economies are the main objective of several PhD theses. Modelling of income distribution in Slovakia in 2002 by virtue of generalised lambda distribution (GLD) is the theme of the thesis by Sipková (2005), while the theses by Bartošová (2006a) and Bílková, (1996) are devoted to the complex analysis of properties of income distribution in the Czech Republic and its modelling by virtue of three-parameter log-normal distribution.

1. Advanced Methods of Analysis and Modelling of Living Standards

There are a lot of economical and statistical publications published in specialised and scientific journals and proceedings (domestic and international) dedicated to the topic of social situation of population and households. Several of them also contain results of professional statistical analyses based on available data on social situations and at the same time provide valuable information about applicability and methods of such analyses. Contemporary research aims in particular at analysis of the income development dynamics, its stability, and detection of important factors affecting the income levels. Some works concentrated on the description of income development dynamics are Paap and van Dijk (1998), Di Prete and McManus (2000); its stability in EU countries is considered in Longford and Pittau (2006). The articles by Kneip and Utikal (2001), and Pittau (2004) focus on identification of the most influential factors and regional diversities in the EU.

Present income distributions are rather complicated and thus many authors use more general classes of models for their modelling (see e.g. Longford and Pittau, 2006). Theoretical foundations of the modern mathematical and statistical classification methods (modelling with the use of mixtures, hidden Markov models, generalised linear and additive models) can be found in McCullagh and Nelder (1994), and McLachlan and Peel (2000). Many authors give priority to modelling with the use of kernel density estimates (Kneip and Utikal, 2001), which allows them to choose a suitable degree of curve smoothing and thus construct detailed models of density function. The progressive method of modelling based upon the use of properties of the generalised lambda distribution quantile function (RS GLD) (see Ramberg and Schmeiser, 1974, pp. 78-82; Sipková, 2005, pp. 90-164; Pacáková and Sodomová, 2003, pp. 30-44) or the generalised Pareto distribution (see Luceno, 2006) is also widely applied nowadays.

1.1 Income Models

Income models may be used directly for assessment of the living standard or for comparison of living standards among individual areas or nations. For the purpose of statistical analyses of the living standard, the focus must be only on measurable elements of the living standard. In order to quantify this element of the living standard of a population which is directly dependent on income, it is necessary to elicit the correct level and

structure of the population income, i.e., to find suitable statistical models of income distribution for individual social classes as well as for the population as a whole, irrespective of social classes. Knowledge of the statistical model, which in many cases represents a simple approximation of complicated survey sample distribution, and knowledge of the development trends of its parameters may be used to predict the behaviour of a particular variable in the upcoming period.

For the purpose of construction of the statistical model, it is important to first find a theoretic distribution function that would describe well the empirical frequency distribution, as well as to choose suitable methods of parameter estimates of the models. The methods used so far have been based on the assumption of homogeneity of a population's income inside and among social classes. The wage differentiation process, which is currently occurring in transitional economies, results in increasing inhomogeneity of income distribution. These changes require verification of methods used for the description of the state of income distribution and construction of its parametric model, or alternatively, choice of new methods. It must be noted that methods for estimation of distribution characteristics and parameters of the chosen model must satisfy not only the requirement of consistency and efficiency, but also that of robustness.

For the purpose of the choice of a parameter model of household income distribution, logical criteria and experience with behaviour of the given element will be given priority over the level of congruity of the model with empirical frequency distribution. This is not to mean, of course, that the chosen model should not at the same time reflect empirical income distribution to the highest possible extent. Important discrepancies always require the conduct of analyses for reasons of such discrepancies, or alternatively a change of the model (see Bartošová, 2006b).

To express a theoretical income distribution, it is possible to use some known distributions such as log-normal distribution, Pareto or Weibull distribution, and some others, or to approximate the empirical distribution by means of one of the theoretical curves within Pearson's or Johnson's system (see Pearson, 1895, pp. 343-414; Johnson, 1949, pp. 149-176). For the purpose of analytical illustration of empirical income distribution, a suitable convergent line may also be used. In economics, the log-normal model is the most widely used model of distribution of wages and income. The biggest "competitor" of the log-normal distribution in modelling of population income distribution is the Pareto distribution. The log-normal curve corresponds to empirical distribution in the central part; however, it diverges in extremes remarkably. In contrast, the Pareto curve represents a suitable model of income distribution in extremes (see Johnson, Kotz and Balakrishnan, 1994).

Empirical income distribution of all households, irrespective of social classes, is complicated and forms a mixture of two (or more) one-peaked curves (see Bartošová and Bína, 2007). This circumstance that is evidenced by increasing inconsistency between the empirical income distributions with the model may also be observed in the income distribution of some social classes.

Economic quantities, such as income, wages, turn-out, profits, expenses etc., are bounded below by nonnegative values. In the past the three-parameter log-normal distribution with parameters μ, σ^2 and γ, where γ is the theoretical minimum, represented a good approximation of income distribution. Therefore, the probability distribution function of the chosen model is determined by the following relation

$$f(x; \mu, \sigma^2, \gamma) = \begin{cases} (2\pi\sigma^2)^{-\frac{1}{2}}(x-y)^{-1}e^{-\frac{(ln(x-y)-\mu)^2}{2\sigma^2}} & \text{if } x \in (\gamma, \infty) \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $f(x; \mu, \sigma^2, \gamma)$ – probability distribution function,

        μ, σ^2, γ – parameters of the model.

1.2 Construction of Income Models

The basic aim for the construction of the theoretic model is its maximum correspondence to the empirical distribution (Bartošová, 2006a; Bartošová, 2006b; Bílková, 1996). Consequently, sufficient flexibility and elasticity belongs among conditions for the choice of the model. A three-parameter log-normal model is often used for modelling empirical income distribution. Since the sample files of households incomes in the years 1996–2005 are sufficiently large for the construction of log-normal models with parameters μ, σ^2 and γ, the maximum likelihood method was applied. It is based on the searching for an argument of the likelihood function suprema

$$\arg \sup_{(\vec{\theta})} L\left(\vec{\theta} \,\middle|\, x_1, ..., x_n\right), \tag{2}$$

where $\{x_1, ..., x_n\}$ – values of net annual financial incomes of households in particular groups,

        n – sample size.

The system of likelihood equations for the estimation of a parameter vector $\vec{\theta} = (\mu, \sigma^2, \gamma)$ is derived by maximisation of the respective likelihood function and can only be solved numerically. Significant simplification can be achieved by substituting the maximum likelihood estimates $\hat{\mu}(\gamma)$ and $\hat{\sigma}^2(\gamma)$ into the respective log-likelihood function. Thus the reduced formula for the log-likelihood function is obtained:

$$\widetilde{\ell}(\gamma) = -n\left[\hat{\mu}(\gamma) + \frac{1}{2}\ln\hat{\sigma}^2(\gamma)\right], \tag{3}$$

where $\hat{\mu}(\gamma), \hat{\sigma}^2(\gamma)$ – maximum likelihood estimates of parameters for chosen value of parameter γ,

        γ – chosen value of theoretical minimum in the model,

        n – sample size.

The maximum likelihood estimate of parameter ă of the three-parameter log-normal distribution is calculated numerically using the following two methods:

- search for the maximum of the modified log-likelihood function $\widetilde{\ell}(\gamma)$

- search for the minimum of the likelihood ratio (see Andĕl, 1985; Bartošová, 2006a)

$$LR(\mu,\sigma^2,\gamma \mid n) = 2[\ell(\vec{p} \mid n) - \ell(\vec{\pi}(\mu,\sigma^2,\gamma)|n],\tag{4}$$

where $\vec{p}$ – vector of income empirical probability,

$\vec{\pi}(\mu,\sigma^2,\gamma)$ – vector of income theoretical probability,

$\ell(\vec{p}|n), \ell(\vec{\pi}(\mu,\sigma^2,\gamma)|n)$ – corresponding log-likelihood functions.

Considering the character of the particular feature, we obtained the estimate by searching for the maximum value of the function $\widetilde{\ell}(\gamma)$ (minimum of the function $LR(\gamma|n)$) in the interval $\langle -x_{max}, x_{min})$, where x_{min} and x_{max} are the minimum and maximum values of incomes. The task was solved by iteration method on a grid, which is refined in each iteration step (see Bartošová and Bína, 2008). The corresponding iteration procedure was implemented by a script in the R language.

Since we treat only finite samples, the maximum likelihood estimates are not guaranteed to show a sufficient quality. More detailed information about the accuracy of the maximum likelihood estimates of the parameter vector $\widehat{\vec{\theta}} = (\widehat{\mu},\widehat{\sigma}^2,\gamma)$ in a log-normal model of income distribution is presented in the dissertation thesis by Bartošová (2006a).

The level of conformity of the empirical household income distribution to the log-normal model was quantified by the likelihood ratio. The *LR* statistic was chosen because the method of maximum likelihood and its modifications were used to estimate parameters of the model (see Bartošová, 2006b). The results are also influenced by the number of classes where data are gathered during the calculation. The problem of the optimal number of classes *m* is the subject of many papers. In this case we chose $m = 15 \cdot \sqrt[5]{(n/100)^2}$, which is suitable for a sufficiently large sample, i.e., for *n > 80* (see Williams, 2001).

1.3 Non-parametric Estimate of Density of Income Distribution

In order to consider the quality of the parametrical model, non-parametrical density estimates can also be used. Since the income distribution is continuous, we used a kernel estimate to fit its density:

$$\widehat{f}_E(x) = \frac{1}{n \cdot h} \cdot \sum_{i=1}^{n} K\left(\frac{X_i - x}{h} \right),\tag{5}$$

where K – kernel of estimate,
H – curve smoothing coefficient,
N – sample size.

For the estimate construction the cosine kernel was used, which is defined as follows:

$$
K\left(\frac{X_i - x}{h}\right) =
\begin{cases}
\frac{1}{2}\cos\left(\frac{X_i - x}{h}\right) & \text{if } \left|\frac{X_i - x}{h}\right| \leq \frac{\pi}{2} \\
\\
0 & \text{otherwise,}
\end{cases}
\tag{6}
$$

Since the income distribution is bounded on the left side, the cosine kernel was chosen (the commonly used normal kernel is not bounded). The decision was motivated by the effort to restrict undesirable nonzero density estimates for negative values of incomes (see Bartošová, 2006a).

2. Modelling of Income Distribution of Czech Households in 1996-2005

Data files from the years 1996–2005 were used for the modelling. Newer data were unavailable at the time of study. Samples of household incomes used for the modelling were collected in two different studies: the Mikrocensus and the SILC. The Mikrocensus sample survey is a periodical sample survey conducted once every 3–5 years since 1957. After the Czech Republic's accession to the EU, the former Mikrocensus was replaced by the SILC survey.

In 1996 (2002), a sample Mikrocensus survey was conducted in approximately 1% (0.3%) of the Czech households, which represented about 28,000 (8,000) households at that time. In 2005, a sample SILC survey was made in 0.15% of the Czech households, which represented about 4,000 households. The existence of a complete non-aggregated sample set enabled us to gain quality estimates of parameters for our distribution models. For the purposes of this research, the following data were chosen:

■ the social class of the head of the household,

■ the net income of the household (CZK per year).

In connection with the economical transformation being in progress, new sources of income have arisen and social structure of sample sets has been changing accordingly. Prior to the Velvet Revolution, households were divided into classes of workers, co-operative farmers, employees, and pensioners. The social structure of the sample sets of household incomes from the years 1996–2005 is shown in Table 1 and Figures 1 and 2. It reflects the changes in the sources of income as well as in the percentage representation of the classes.

Table 1
Structure of data sets in the years 1996 and 2005

| Social class	1996		2002		2005	
	Size	%	Size	%	Size	%
Employees	15,966	56.72	4,094	51.35	2,148	49.37
Self-employed	1,879	6.68	757	9.49	391	8.99
autoPensioners with EA	1,156	4.11	278	3.49	178	4.09
Pensioners without EA	8,651	30.73	2,533	31.77	1,425	32.75
Unemployed	260	0.92	172	2.16	131	3.01
Others	236	0.84	139	1.74	78	1.79
Total	28,148	100.00	7,973	100.00	4,351	100.00

Source: Mikrocensus 1996 and 2002, SILC 2005, and own R computations.

Figure 1
Structure of data sets in the years 1996–2002
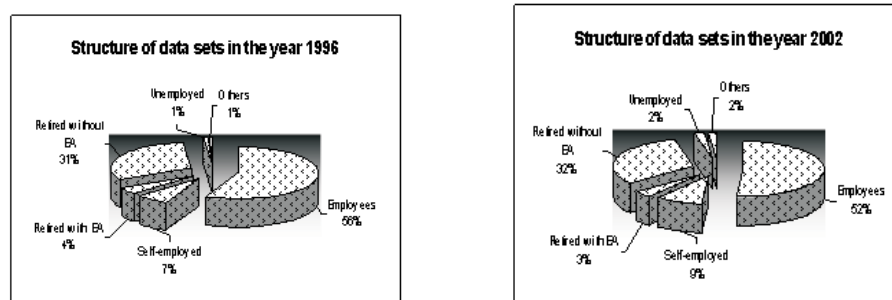
Source: Mikrocensus 1996 and 2002, and own R computations.

Figure 2
Structure of data sets in the year 2005.
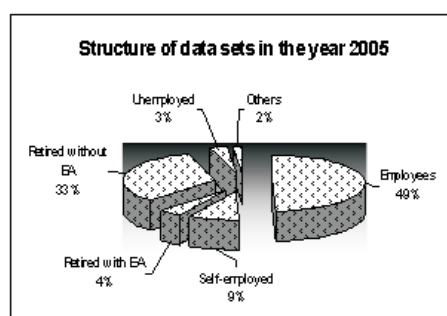
Source: SILC 2005 and own R computations.

It is apparent from Table 1 (Figures 1 and 2) that the two largest social classes, i.e., employees and pensioners without economically active members, constituted 87.45% (83.12%, 82.12%, respectively) of all the households in the Czech Republic in 1996 (2002, 2005, respectively). New social classes that emerged due to the post-Velvet Revolution transformation, i.e., households of self-employed, unemployed, and other households, constituted only about 8.44% (13.39%, 13.79%, respectively) altogether. That is the reason why the new classes cannot significantly influence the total income distribution of households. Consequently, the character of the total income distribution of households will mainly be determined by the manner of the income distribution in the three major classes.

2.1 The results obtained

For the sake of clarity, the behaviour the of log-likelihood function belonging to the estimates is also shown (see Figures 3 and 4). Although the shapes of the curves slightly differ, the uniqueness of the function maximum can be seen in all of them and it facilitates the identification of exactly one solution to the maximisation task on the given interval. The examples show the estimation of the log-normal model parameters for **all households' income distribution** in the years 1996 and 2005. Figures 3 and 4 show that the estimates faithfully converge to a single result. Whether the designed procedure effectively and faithfully converges to the unique result can be seen from the illustration.

Table 2 shows the values of the likelihood ratio *LR* for the constructed log-normal models with two and three parameters. Greater agreement of the empirical distribution with the model was achieved in all social groups for the three-parameter log-normal models. The use of the three-parameter log-normal models with the above described iterative estimation of the parameter γ led to the improvement of the model validity. The method of likelihood ratio minimisation was used in the iterative procedure in order to estimate the value of the parameter γ.

The estimates of the three parameters of the log-normal models of incomes in the Czech households in the years 1996–2005 are listed in Tables 3–5. For the consideration of the model validity, the estimates of μ, σ^2 and γ are supplemented by the values of the *LR* statistic and the 95% quantiles of $\chi^2_{0,95}(m-k-l)$, where $k=3$ is the number of estimated parameters.

Figure 3
Behaviour of the reduced form of the log-likelihood function – all households
(Incomes per household in 1996–2002)
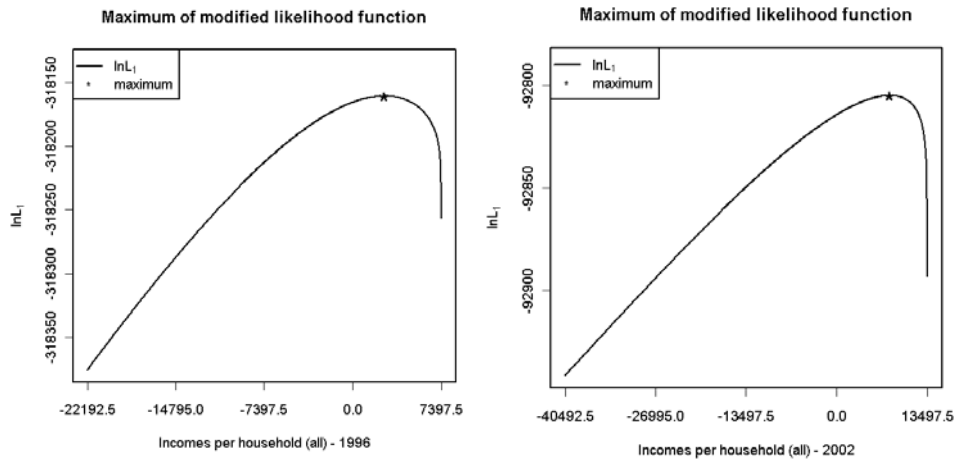
Source: Mikrocensus 1996 and 2002, and own R computations.

Figure 4
Behaviour of the reduced form of the log-likelihood function – all households
(Incomes per household in 2005)

Source: SILC 2005 and own R computations.

11
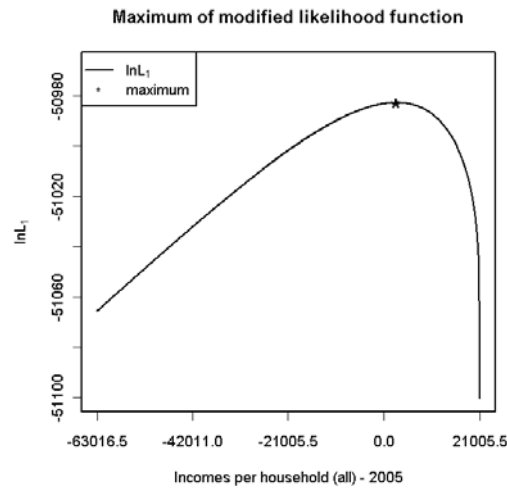
Table 2
Comparison of conformity of empirical distribution with two- and three-parameter log-normal models (Incomes per household in 1996–2005)

Social class	LR (1996)		LR (2002)		LR (2005)	
	$\gamma = 0$	γ_LR	$\gamma = 0$	γ_LR	$\gamma = 0$	γ_LR
Employees	477.35482	349.95501	76.101208	73.521025	59.723929	51.369312
Self-employed	99.910424	99.876105	52.367792	52.358131	27.306059	27.260379
Pensioners with EA	32.614093	21.366174	25.863125	16.459187	21.809623	21.642089
Pensioners without EA	3,874.4552	3,698.5442	1,174.2735	1,113.5243	565.20513	535.01713
Unemployed	29.293800	23.023990	15.046472	15.045804	7.3784923	7.2391322
Others	26.871509	25.315383	11.035980	10.590493	13.784844	12.339053
Total	3,239.8009	3,224.4095	790.27457	756.7107	398.93645	397.66954

Source: Mikrocensus 1996 and 2002, SILC 2005, and own R computations.

Table 3
Estimates of the parameters for the models (Incomes per household in 1996)

Social class	Parameter			Statistic LR	Quantile $\chi^2_{0.95}(m-4)$	Number of classes m
	μ	σ^2	γ			
Employees	12.23851	0.15431	-30,133	349.95501	135.48018	114
Self-employed	12.22920	0.34700	1,062	99.876105	60.480887	48
Pensioners with EA	11.78215	0.19035	36,141	21.366174	50.998460	40
Pensioners without EA	10.99314	0.22218	13,331	3,698.5442	107.52174	89
Unemployed	11.50930	0.23588	-21,716	23.023990	28.869230	22
Others	11.07883	0.51136	4,239	25.315383	27.587112	21
Total	11.75616	0.40419	4,343	3,224.4095	167.51430	143

Source: Mikrocensus 1996 and own R computations.

Table 4
Estimates of the parameters for the models (Incomes per household in 2002)

Social class	Parameter			Statistic LR	Quantile $\chi^2_{0.95}(m-4)$	Number of classes m
	μ	σ^2	γ			
Employees	12.46541	0.20951	-11,159	73.521025	81.381015	66
Self-employed	12.56097	0.32957	1,316	52.358131	43.772972	34
Pensioners with EA	11.99530	0.27834	87,257	16.459187	30.143527	23
Pensioners without EA	11.37340	0.24821	28,477	1,113.5243	68.669294	55

Unemployed	11.51230	0.39123	297	15.045804	24.995790	19
Others	11.45312	0.40731	8,999	10.590493	22.362032	17
Total	12.07828	0.41810	12,933	756.7107	104.13874	86

Source: Mikrocensus 2002 and own R computations.

Table 5
Estimates of the parameters for the models (Incomes per household in 2005)

Social class	Parameter			Statistic LR	Quantile $\chi^2_{0,95}(m\text{-}4)$	Number of classes m
	μ	σ^2	γ			
Employees	12.61504	0.17561	-34,033	51.369312	64.001112	51
Self-employed	12.70748	0.34039	-4,562	27.260379	33.924438	26
Pensioners with EA	12.47647	0.10223	29,636	21.642089	24.995790	19
Pensioners without EA	11.40137	0.29192	36,890	535.01713	54.572228	43
Unemployed	11.47304	0.45997	6,215	7.2391322	22.362032	17
Others	11.85294	0.26867	-27,916	12.339053	18.307038	14
Total	12.20097	0,38028	4,818	397.66954	83.675261	68

Source: SILC 2005 and own R computations.

We can infer from Tables 3–5 that the values of LR and $\chi^2_{0,95}(m-4)$ are comparable in most cases. A strong discrepancy between the empirical distribution and the model only appears **for pensioners without economically active members** and **all households** (see Bartošová, 2006c). In both the above cases, the LR statistics significantly exceed the value of the corresponding χ^2 quantile. In 1996, $LR_R \approx 30 \cdot \chi^2_{0,95}(m-4)$ **for pensioners without economically active members** and $LR_A \approx 20 \cdot \chi^2_{0,95}(m-4)$ for **all households**; in 2002, $LR_R \approx 20 \cdot \chi^2_{0,95}(m-4)$ and $LR_A \approx 20 \cdot \chi^2_{0,95}(m-4)$; and finally, in 2005, $LR_R \approx 10 \cdot \chi^2_{0,95}(m-4)$ and $LR_A \approx 7 \cdot \chi^2_{0,95}(m-4)$ (see Figure 5). Those income sets have a bimodal distribution (see Figures 6–8) and could not be modelled using simple parametrical models. Modelling of such mixtures is the topic of our previous paper (Bartošová and Bína, 2007).

Figures 6 and 7 below show the three-parameter log-normal models, the empirical densities and the kernel estimates of the theoretical density of income distribution of **pensioners without economically active members** and **all households** in the years 1996–2005. The proposed iterative procedure was used for the estimation of the log-normal model parameters. That allows the estimation of the theoretical minimum using the likelihood ratio minimisation. The cosine kernel was used for the construction of the kernel estimates of density. The above figures show that the kernel estimates of the theoretical density of the households of **pensioners without economically active members** are obviously bimodal, while this effect is less significant for **all households**.

Figure 5

Comparison of conformity of empirical distribution with three-parameter log-normal income models (Incomes per household in 1996–2005)
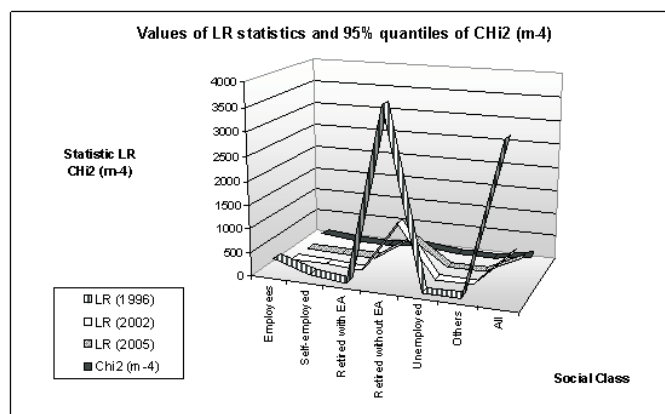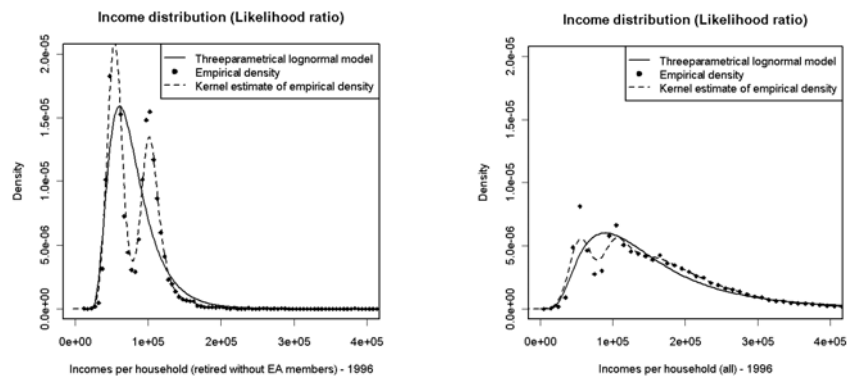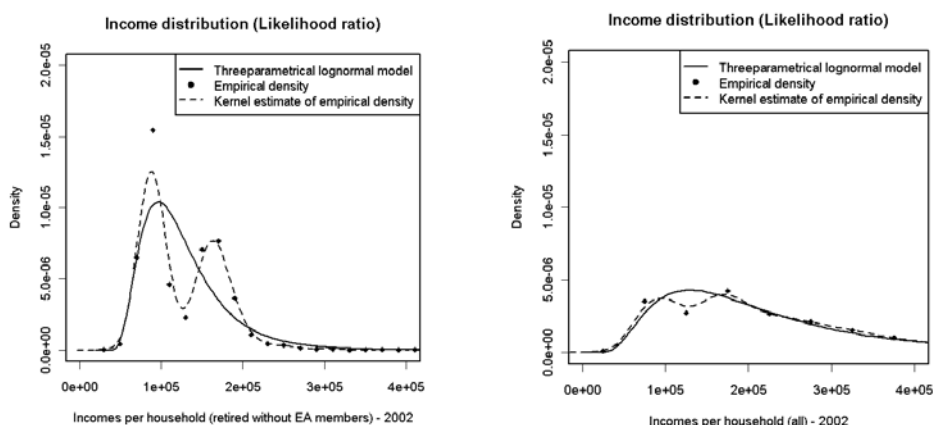

Figure 6

Three-parameter model, empirical density and kernel estimate (Pensioners without EA members and all households in 1996)
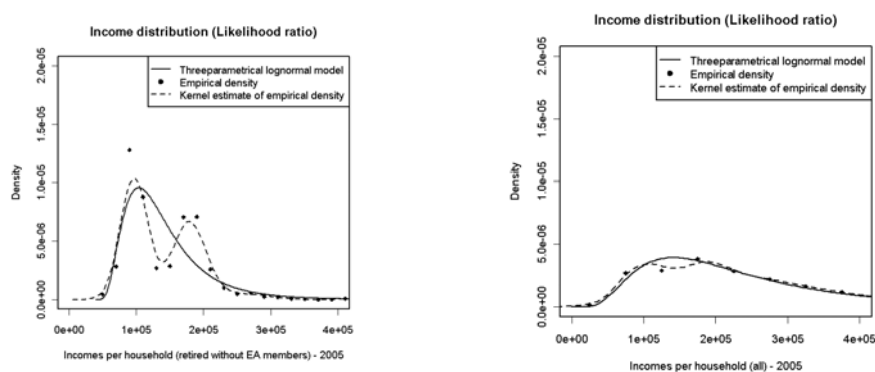

Source: Mikrocensus 1996 and own R computations.

Figure 7
Three-parameter model, empirical density and kernel estimate (Pensioners without EA members and all households in 2002)

Source: Mikrocensus 2002 and own R computations.

Figure 8
Three-parametrical model, empirical density and kernel estimate (Pensioners without EA members and all households in 2005)

Source: SILC 2005 and own R computations.

One advantage to the approximation of the empirical distribution by a suitable parametrical model is the possibility of the prompt calculation of the basic characteristic estimates. Table 6 lists the values of mean and median annual incomes of the Czech households in the years 1996–2005, and Table 7 contains the relevant values of the standard deviation and the interquartile range (in Czech crowns). Those values were estimated using the above mentioned three-parameter log-normal models.

15

Table 6
Estimates of the chosen characteristics of location (Incomes per household in the years 1996–2005)

Social class	1996		2002		2005	
	Mean	Median	Mean	Median	Mean	Median
Employees	193,801	175,800	278,487	247,998	295,296	264,480
Self-employed	248,738	201,100	340,373	285,476	393,784	321,380
Pensioners with EA	180,256	168,250	273,682	252,305	305,632	295,364
Pensioners without EA	79,955	67,900	126,830	108,000	140,485	121,000
Unemployed	91,480	77,900	123,599	102,457	125,981	107,784
Others	85,086	59,450	125,551	100,616	135,308	116,708
Total	160,066	137,900	230,007	190,923	245,901	205,410

Source: Mikrocensus 1996 and 2002, SILC 2005, and own R computations.

Table 7
Estimates of the chosen characteristics of variability (Incomes per household in the years 1996–2005

Social class	1996		2002		2005	
	St. Dev.	IQR	St. Dev.	IQR	St. Dev.	IQR
Employees	103,980	101,800	173,069	153,879	156,388	165,426
Self-employed	205,304	139,000	245,684	186,049	320,449	226,934
Pensioners with EA	68,800	78,900	108,650	109,237	92,955	101,647
Pensioners without EA	37,707	48,550	56,106	76,668	71,655	84,116
Unemployed	70,670	64,150	101,922	84,513	82,209	104,290
Others	97,358	47,250	89,983	71,583	113,670	80,989
Total	113,804	113,200	170,654	162,310	177,059	177,270

Source: Mikrocensus 1996 and 2002, SILC 2005, and own R computations.

Conclusions

In most cases the presented results of modelling using the three-parameter model show an agreement with the empirical distribution. Only in the cases of **pensioners without economically active members** and **all households** the distribution is bimodal and a significant discrepancy shows. In all surveyed cases, the three-parameter models were better than the two-parameter one. The use of the three-parameter models always led to an improvement in the validity of the constructed models. The likelihood ratio minimisation was used for estimating g in the iterative procedure. Similar results could be achieved using the log-likelihood function maximisation.

The obtained results show that the estimates of the theoretical minimum in the three-parameter log-normal models are often negative. However, since the parameters μ and γ are negatively correlated, the effects of the negative estimates of the parameter γ are compensated for by the higher values of the parameter μ estimates. Thus, all the constructed log-normal models can well approximate the given empirical distribution in spite of the fact that their theoretical minima differ.

The analysis of the structure of households yields three facts. Firstly, it can be seen that three social classes – the most numerous before the Velvet Revolution (workers, employees, and pensioners without economically active (EA) members) – have kept their dominance, and the income distribution in the newly emerged social classes has a minimal influence on the total income distribution of households. Secondly, it can be seen that for the majority of the social classes the log-normal distribution can be considered a suitable model of household income distribution. Thirdly, the analysis shows the ageing of the population and the increasing influence of the income distribution of **pensioners without economically active members**. In this class, there is a significant discrepancy between the empirical distribution and the model. The same is true of **all households**, which is mainly influenced by this class. These income sets have a bimodal distribution and could not be modelled using simple parametrical models. In both cases, the bimodal nature of the income distribution could be overcome by splitting the files into two subgroups: households with one member and households with multiple members.

References

ANDĚL, J. 1985. *Statistické metody*. Praha : Matfyzpres, 1985.

BARTOŠOVÁ, J. 2006a. Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990. [Doktorská disertační práce]. Praha : VŠE, 2006.

BARTOŠOVÁ, J. 2006b. Logarithmic-normal model of household income distribution in the Czech Republic after 1990. *Forum Statisticum Slovacum*. 2006, no. 3. Slovak Statistical and Demographic Society, s. 3–10.

BARTOŠOVÁ, J. 2006c. Logarithmic-Normal Model of Income Distribution in the Czech Republic. *Austrian Journal of Statistics*. 2006, vol. 35, no. 2&3, s. 215–222.

BARTOŠOVÁ, J.; BÍNA, V. 2007. Mixture models of household income distribution in the Czech Republic. In KOVÁČOVÁ, M. (ed.). 6th International Conference APLIMAT 2007, Part I. Slovak University of Technology, Bratislava, s. 307–316.

BARTOŠOVÁ, J.; BÍNA, V. 2008. Iterative Procedure for the Construction of Three parametrical Lognormal Income Distribution Model. In 7th International Conference APLIMAT 2008 [CD-ROM]. Slovak University of Technology, Bratislava, s. 1037–1042.

BÍLKOVÁ, D. 1996. Příjmová rozdělení: modelování v letech 1956 a 1992 a předpovědi pro roky 1995 a 1997. [Doktorská disertační práce]. Praha : VŠE, 1996.

BÍLKOVÁ, D. 2008. Application of Lognormal Curves in Modeling of Wage Distributions. In 7th International Conference APLIMAT 2008 [CD-ROM]. Slovak University of Technology, Bratislava, s. 341–352.

DI PRETE, T. A.; MCMANUS, A. 2000. Family change, employment transition, and welfare state: household income dynamics in the United States and Germany. *American Sociological Revue*. 2000, vol. 65, s. 343–370.

CHAUDHURI, S.; RAVALLION, M. 1994. How well do static indicators identify the chronically poor. *Journal of Public Economic*. 1994, vol. 53, s. 367–394.

KNEIP, A.; UTIKAL, K. J. 2001. Inference for density families using functional principal component analysis. *Journal of the American Statistical Association.* 2001, vol. 96, no. 454, Theory and Methods, s. 519–533.

LONGFORD, N. T.; PITTAU, M. G. 2006. Stability of household income in European countries in the 1990s. *Computational Statistics & Data Analysis*. 2006, vol. 51, s. 1364–1383.

LUCENO, A. 2006. Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis.* 2006, vol. 51, s. 904–917.

MACULLAGH, P.; NELDER, J. A. 1994. *Generalized Linear Models*. London : Chapman and Hall, 1994.

MCLACHLAN, G.; PEEL, D. 2000. *Finite mixture models*. New York : John Wiley & Sons, 2000.

PAAP, R.; VAN DIJK, H. K. 1998. Distribution and mobility of wealth of nation. *European Economic Revue.* 1998, vol. 42, s. 1269–1293.

PACÁKOVÁ, V.; SODOMOVÁ, E. 2003. Modelling with quantile distribution functions. *Ekonomika a informatika.* 2003, č. 1, s. 30–44.

PITTAU, M. G. 2004. Regional income distributions in the European Union. *Oxford Bulletin Economic Statistics.* 2004, vol. 67, s. 135–161.

RAMBERG, J.; SCHMEISER, B. 1974. An approximate method for generating asymmetric random variables. *Communications of the ACM.* 1974, vol. 17, no. 2, pp. 78–82.

WILLIAMS, D. 2001. *Weighing the Odds, A Course in Probability and Statistics*. Cambridge : Cambridge University Press, 2001.

MODELLING OF INCOME DISTRIBUTION OF CZECH HOUSEHOLDS IN THE YEARS 1996–2005

Abstract: After the 1989 Velvet Revolution, the transformation to a market economic system, and mainly the formation of new income sources and the process of a significant differentiation in wages, has caused crucial changes to the income distribution. There have been systematic changes in the model parameters as well as changes manifesting the increasing numbers of discrepancies between empirical and theoretical income distributions: these have caused a contamination of the model. This paper concentrates on the verification of the validity of the statistical model of income distribution used in the Czech Republic at present. In order to consider as precisely as possible the suitability of log-normal distribution as the proper model for income distribution used in the Czech Republic, the optimisation of parameters is necessary. The accuracy of the model construction is crucial and is given, to some extent, by the sensitivity and adaptability of procedures used for estimation of the model parameters. The paper shows methods usable for the achievement of this goal and presents results obtained by their application on data files from the years 1996–2005.