

## Modelling the Income Distributions in the Czech Republic since 1992

Diana Bílková and Ivana Malá  
University of Economics, Prague

**Abstract:** The goal of this article is to study incomes in the Czech Republic and their development since 1992. The net annual per capita income of Czech households is analysed for all households and their respective subpopulations. Data from the microcensus 1992, 1996, 2002, and EU-SILC 2005–2008 surveys carried out by the Czech Statistical Office are used. The subpopulations are defined by a household's location (Bohemia or Moravia), and education and age of the head of the household in order to compare the distributions of the income in Bohemia and Moravia and to quantify the impact of education and age on incomes. The three-parameter lognormal distribution is chosen as a probability distribution to model the per capita income distribution for the whole population and for subpopulations. To estimate the unknown parameters, the maximum likelihood method and that of L-moments are employed. The medians of equalised incomes are given for the EU members and the average growth in the 2004–2007 period is compared. For the Czech Republic, a comparison of the medians of per capita and equalised income is made.

**Zusammenfassung:** Das Ziel dieses Artikels ist die Untersuchung des Einkommens in der Tschechischen Republik und dessen Entwicklung seit 1992. Das jährliche Nettoeinkommen tschechischer Haushalte wird für sämtliche Haushalte und einige Subpopulationen analysiert. Verwendet wurden Daten aus dem Mikrocensus 1992, 1996, 2002 und EU-SILC 2005–2008, die das Tschechische Statistikamt durchführt. Die Subpopulationen sind durch Standort des Haushalts (Böhmen oder Mähren), Ausbildung und Alter des Familienvorstands definiert, mit dem Ziel, die Einkommensverteilung in Böhmen und Mähren zu vergleichen und den Einfluss von Ausbildung und Alters auf das Einkommen zu quantifizieren. Die Verteilung des Pro-Kopf-Einkommens der Gesamtpopulation und von Subpopulationen wird durch die dreiparametrische Lognormalverteilung modelliert. Um die unbekannt Parameter zu schätzen, wurde die Maximum-Likelihood und die L-Momenten Methode angewandt. Das Medianäquivalenzeinkommen für EU Mitglieder und das Durchschnittswachstum 2004–2007 wird verglichen. Für die Tschechische Republik wird ein Vergleich der Mediane des Haushaltseinkommens und des Äquivalenzeinkommen gemacht.

**Keywords:** Lognormal Distribution, Maximum Likelihood Estimation, L-Moment Estimation.

### 1 Introduction

Wages and incomes of the population are in the centre of interest of economists in all developed countries of the world as there is a constant endeavor for objective answers to

questions relating to the standard of living. Income models (and characteristics derived from them) can be used, apart from quality of life research, for making an inter-regional or international comparison of standards of living.

Not only knowledge of the overall income distribution over the given period and country, but also income distributions amongst particular subgroups of people are worthwhile. These distributions can be used in an analysis of income differentiation or income inequalities in society. Knowledge of income distributions can also be used for other purposes. It refers to process improvements relevant to other statistical variables correlating with income. We can give examples of household expenses and facilities, time utilization, purchasing purposes as well as various positions within the scope of sociological research. Having in mind all the above mentioned applications of income distribution analysis, we work with mass quantitative or qualitative observations.

In investigating income level and differentiation, indicators of income characteristics (such as average, median, standard deviation etc.) – categorized according to the sector of economy, gender, education, age and other criteria – are useful, but more detailed knowledge of the frequency distribution of income can be beneficial in the analysis.

Within the scope of this research, data provided by the Czech Statistical Office are used. These sets of data come from microcensus 1992, 1996 and 2002 and EU-SILC 2005–2008 (these surveys include data about incomes in 2004–2007). The first sample from 1992 is part of the survey that covered the Czech and Slovak Federal Republic. This federal state split into two parts: the Czech Republic and the Slovak Republic on January 1st, 1993.

The analysis employs annual net nominal per capita income of Czech households in Czech crowns (CZK). The sample units are represented by the households whose heads are categorized according to education or age, respectively. Subpopulations in this paper are defined by a household's location (in the historical lands of Bohemia and Moravia), education (four levels) and age (10-year-long intervals) of the head of household, respectively.

In order to compare the development of incomes in households in the Czech Republic and in Europe, the medians of equalized (not per capita) incomes from European Union members (data from Eurostat) for 2004–2007 are shown.

In probability modelling, when the analysed income distribution is described by a probability model, a suitable probability distribution must be fitted and its unknown parameters have to be estimated. The choice of a proper distribution is the first step in a successful analysis. If the above choice is inappropriate, misleading results can be obtained. In this text, the three-parameter lognormal distribution is used. The appropriateness of this choice has been widely discussed in the literature from both theoretical and empirical point of view (see Johnson, Kotz, and Balakrishan, 1994; Kleiber and Kotz, 2003). As for the data for the Czech Republic, the fit of this distribution can be found in Bartošová (2006), Bartošová and Bína (2009) or Bílková (2008). A more detailed description with the use of a finite mixture of lognormal distributions is used for the Czech Republic in Pavelka (2009), for the United Kingdom in Flachaire and Nunez (2007). A variety of other probability distributions (sometimes called income distributions) is often encountered in the literature dealing with the topic. The generalized lambda distribution was used for incomes in the Czech Republic in Pacáková and Sipková (2007). The Dagum

distribution is frequently used with good results. Loglogistic, generalized beta and Pareto distributions (for high incomes) are also successfully used in numerous resources dealing with the issue of wages or incomes (Dagum, 1990, 1997; McDonald, 1984; Parker, 1997; Kleiber and Kotz, 2003).

For an estimation of unknown parameters, the maximum likelihood method is usually employed, maximizing the logarithm of a likelihood function with respect to unknown parameters (a numerical procedure has to be applied as no closed formulas are known for estimators). In this paper, the results of this widespread estimation method are compared with those obtained from the use of so-called L-moments. L-moments and their use for the estimation of unknown parameters are briefly introduced. The differences between estimates (and quantities derived from these fits) detected by both methods are shown in Figures 5 to 7 and commented on in the text.

## 2 Methodology

### 2.1 Three-Parameter Lognormal Distribution

To model the distribution of annual net income per capita we use the three-parameter lognormal distribution with unknown parameters  $\mu$ ,  $\sigma$  and  $\theta$ . These parameters have a simple meaning, as the distribution and their values can be easily interpreted. If a random variable  $X$  has a three-parameter lognormal distribution,  $\text{LN}(\mu, \sigma^2, \theta)$ , then a variable  $\log(X - \theta)$  has a normal distribution with expectation  $\mu$  and variance  $\sigma^2$ . The parameter  $\theta$  is the theoretical minimal value of  $X$ . The distribution is described by the probability density function

$$f(x; \mu, \sigma^2, \theta) = \frac{1}{\sqrt{2\pi}\sigma(x - \theta)} \exp \left\{ -\frac{(\log(x - \theta) - \mu)^2}{2\sigma^2} \right\}, \quad x > \theta, \quad (1)$$

where  $\mu, \theta \in \mathbb{R}$  and  $\sigma^2 > 0$ .

The corresponding cumulative distribution function (cdf) can be written as

$$F(x; \mu, \sigma^2, \theta) = \begin{cases} 0 & x \leq \theta, \\ \Phi \left( \frac{\log(x - \theta) - \mu}{\sigma} \right) & x > \theta, \end{cases}$$

where  $\Phi(\cdot)$  denotes the cdf of a standard normal distribution. In the text we use the first common and second central moment of the distribution given by

$$E(X) = \theta + e^{\mu + \sigma^2}, \quad \text{var}(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1). \quad (2)$$

The quantiles are

$$x_p = \theta + \exp(\mu + \sigma u_p), \quad 0 < p < 1.$$

Here  $x_p$  denotes a 100 $p$  % quantile of the lognormal distribution, and  $u_p$  is the 100 $p$  % quantile of a standard normal distribution.

## 2.2 Methods of Estimation

Parameters of the lognormal distribution are estimated with the use of the maximum likelihood method and that of L-moments. Suppose that  $X$  is a random variable and the sample of  $n$  observations from the distribution of  $X$  is  $X_1, \dots, X_n$ . The ordered sample is denoted by  $X_{1:n}, \dots, X_{n:n}$  with  $X_{1:n} \leq \dots \leq X_{n:n}$ .

Maximum likelihood estimates  $\hat{\mu}$ ,  $\hat{\sigma}$ , and  $\hat{\theta}$  of the three unknown parameters in (1) were calculated by a numerical optimization procedure (see Cohen and Whitten, 1980) where the logarithm of the likelihood function based on (1) is maximized with respect to the parameters.

The method of L-moments is less known and part of the estimation is briefly described. The estimates of the unknown parameters  $\mu$ ,  $\sigma$ , and  $\theta$  obtained by this method are denoted by  $\tilde{\mu}$ ,  $\tilde{\sigma}$ , and  $\tilde{\theta}$ . Theoretical values of the L-moments are linear combinations (L stands for “linear”) of the expectations of the order statistics. By ordering we mean ordering of the sample elements from the smallest to the largest value. The  $k$ -th population L-moment  $\lambda_k$ , for  $k = 1, \dots, n$ , is defined as

$$\lambda_k = \frac{1}{k} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} E(X_{k-j:n}), \quad (3)$$

where  $E(X_{j:n})$  denotes the expectation of the  $j$ -th order statistic from an independent sample of size  $n$ . It can be seen from (3) that the first L-moment  $\lambda_1$  is equal to the mean of  $X$ . It can be shown (Hosking, 1990) that if  $\lambda_1$  is finite, all L-moments exist and are finite.

The sample  $k$ -th L-moment is the sample version of (3), i.e.

$$l_k = \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{1}{k} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} X_{i_k-j:n}, \quad (4)$$

where  $i_1, \dots, i_k$  and  $X_{i_1}, \dots, X_{i_k}$  are ordered subsamples of size  $k$  from the ordered sample of the size  $n$  ( $\binom{n}{k}$  subsamples). Instead of formula (4), the sample L-moments can be evaluated with the use of a more convenient formula for computation (see Hosking, 1990, or Hosking and Wales, 1997) as

$$l_k = \sum_{j=0}^{k-1} (-1)^{k-1-j} \binom{k-1}{j} \binom{k-1+j}{j} b_j, \quad (5)$$

where

$$b_j = \frac{1}{n} \sum_{i=j+1}^n \frac{(i-1)(i-2)\dots(i-j)}{(n-1)(n-2)\dots(n-j)} X_{i:n}, \quad j = 0, \dots, k-1. \quad (6)$$

In Hosking (1990) relations between the parameters  $\mu$ ,  $\sigma$ ,  $\theta$  and the first three theoretical L-moments  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  are given and values of the L-moments are evaluated with the use of the distribution parameters. Moreover, the inverse relationship between the first three sample L-moments  $l_1$ ,  $l_2$ ,  $l_3$  and the maximum likelihood estimates  $\tilde{\mu}$ ,  $\tilde{\sigma}$ ,  $\tilde{\theta}$  are also derived

in Hosking (1990). In the following, in order to estimate the parameters of lognormal distributions using L-moments, the first three sample L-moments are evaluated according to formula (5). Then estimates of the unknown parameters in the lognormal cdf are found as (Hosking, 1990)

$$\begin{aligned}\tilde{\sigma} &= 0.999281z - 0.006118z^3 + 0.000127z^5, \\ \tilde{\mu} &= \log\left(l_2/(2\Phi(\tilde{\sigma}/\sqrt{2}) - 1)\right) - \tilde{\sigma}^2/2, \\ \tilde{\theta} &= l_1 - \exp(\tilde{\mu} + \tilde{\sigma}^2/2),\end{aligned}$$

where

$$z = \sqrt{\frac{8}{3}} \Phi^{-1}\left(\frac{1+t_3}{2}\right) = \sqrt{\frac{8}{3}} u_{\left(\frac{1+t_3/l_2}{2}\right)}.$$

Here,  $u_p$  denotes the 100p % quantile of the standard normal distribution and  $t_3 = l_3/l_2$  is the sample third L-moment ratio. More about L-moments can be found, for example, in Smithers and Schulze (2001) or Kyselý and Pícek (2007). All computations (and numerical method applications) were made in Excel and R 2.13.1. To describe the fitted distribution, estimates of the parameters are plugged into the above results.

### 2.3 Characteristics of Differentiation

There are various characteristics to describe the variability of incomes (or differentiation of incomes) – variance, standard deviation, coefficient of variation or Gini index (Kleiber and Kotz, 2003). Here, only the standard deviation and a coefficient of variation is used. As L-moments are of interest, we give a few comments on the relationship between two- and three-parameter lognormal distribution and characteristics of differentiation.

If the two-parameter lognormal distribution is used, it is clear from (2) (if we take  $\theta = 0$ ) that the coefficient of variation depends only on the parameter  $\sigma^2$  and equals  $\sqrt{\exp(\sigma^2) - 1}$ . In various articles (Kleiber and Kotz, 2003), formulas for the Gini coefficient can be found in the form of  $\text{erf}(\sigma/2)$ , where erf is the error function, or equivalently as  $2\Phi(\sigma/\sqrt{2}) - 1$ . Unfortunately, in the case of the three-parameter model this doesn't hold and both characteristics depend on all three parameters. For the coefficient of variation we have from (2)

$$V = \frac{\exp(\mu + \sigma^2/2)\sqrt{\exp(\sigma^2) - 1}}{\theta + \exp(\mu + \sigma^2/2)}.$$

Formula (3) for  $k = 2$  can be rewritten as

$$\lambda_2 = \frac{1}{2}\text{E}(X_{2:2} - X_{2:1}) = \frac{1}{2}\text{E}(|X_1 - X_2|)$$

and we conclude that the Gini mean difference is  $2\lambda_2$  (see also Hosking, 1990) and the Gini coefficient can be evaluated as  $\lambda_2/\lambda_1$ . Moreover, with the use of Hosking's paper we obtain for the Gini coefficient  $G$  of the three-parameter lognormal distribution the result

$$G = \frac{\exp(\mu + \sigma^2/2)\text{erf}(\sigma/2)}{\theta + \exp(\mu + \sigma^2/2)}.$$

Here, Gini coefficients are not included but from previous considerations the usefulness of L-moments in evaluating these characteristics is clear.

## 3 Data Analysis and Results

### 3.1 Per Capita Income in the Czech Republic

In the paper, we use particular data about Czech households from microcensus (files for 1992, 1996, and 2002) and from the Living Conditions Survey, a national module of the European Union Statistics on Income and Living Conditions (EU-SILC), from 2005, 2006, 2007, and 2008. Net annual household income per capita (nominal income) is the analysed variable. These surveys were carried out by the Czech Statistical Office.

For the sake of completeness, we use data from the microcensus, focusing, however, on EU-SILC figures. The aim of the annual survey is to gather representative data on the income distribution for the whole population and for various household types, housing conditions (housing quality and affordability), household durables, labour, financial and health conditions of adults living in private households (CZSO, Czech Statistical Office). The total net income of a household is the sum of all monetary and non-monetary earnings of all members of the household after tax and deductions. EU-SILC data are collected by means of questionnaires with different units of reference – dwelling unit, household and personal questionnaires (for all members over 16 years of age). Social benefits targeted at particular households, annuity income and the value of goods produced directly by the household through either private or professional activity are covered by a household questionnaire. Income from employment (permanent and temporary jobs), sickness, old-age, unemployment and social benefits claimed by individuals as well as other incomes (such as capital income, sales of property, insurance claims, etc.) are recorded in personal questionnaire forms.

Data for various types of income of Czech households were collected together with other characteristics such as the status of the head of household, the number of employed household members and dependent children, household type, the age and education of the head of household, the size of municipality and region, etc. Marriages with or without children are always headed by men, regardless of their economic activities. In lone-parent families or non-family households, the primary criterion for determination of the head of household status was economic activity and the secondary criterion was the income of household members (CZSO). For our analysis, net income per capita was evaluated as the ratio of the total annual net income of a household to the number of household members. This variable is the ratio of two random variables, the value having been evaluated for each household in the sample and data being treated as a random sample (not stratified the way it was). Furthermore, the households were divided into groups according to their location (Bohemia, Moravia), the household head's education (primary, secondary, complete secondary, and tertiary education) and age (grouped into ten-year intervals: under 30, 30–40, 40–50, 50–60, over 60). In the following text, “income” is used instead of “annual net income per capita in CZK”.

Sample characteristics of all the analysed samples are given in Table 1. There are very different sample sizes (from 4,352 households in 2006 to 28,148 households in 1996), precise numbers of households in subgroups being displayed in Table 2. We can see from the table that almost all samples are large, the smallest value being 118 household heads with complete secondary education in 2004.

Table 1: Sample characteristics, inflation rate and CZK/EUR exchange rate.

Year	1992	1996	2002	2004	2005	2006	2007
Sample size	16,233	28,148	7,973	4,351	7,483	9,675	11,294
Mean	35,247	68,286	105,030	111,024	114,945	123,806	132,877
Median	31,000	57,700	89,204	97,050	100,640	108,744	117,497
Standard deviation	19,364	51,102	83,598	77,676	74,503	74,578	73,982
Coeff. of variation	0.549	0.748	0.795	0.699	0.648	0.602	0.556
Coeff. of skewness	7.8	17.6	17.1	14.9	10.3	7.7	6.9
1 % quantile	13,521	21,246	31,201	29,909	33,429	35,383	40,795
5 % quantile	18,498	32,100	45,777	47,657	49,914	55,914	63,888
10 % quantile	21,292	37,935	54,366	58,120	60,832	68,147	76,571
25 % quantile	25,900	47,550	73,464	79,600	82,998	90,000	97,160
75 % quantile	39,298	76,550	115,966	124,068	128,000	138,000	148,937
90 % quantile	51,977	105,100	163,281	171,833	174,904	189,505	202,327
95 % quantile	63,864	130,138	206,601	215,307	217,366	231,913	246,600
99 % quantile	102,627	223,642	333,915	343,287	368,429	372,504	391,094
Inflation rate	—	—	1.358	1.029	1.019	1.025	1.028
Exchange rate	—	—	30.812	31.904	29.784	28.343	26.300

Source: own computations, CZSO and CNB

In the analysed period the level of incomes increased markedly as both characteristics of location (the arithmetic average and the median in CZK) increased in time (mean income from 35,000 CZK in 1992 to almost 133,000 CZK in 2008, the median from 31,000 CZK to more than 117,000 CZK). Table 1 also offers a review of income differentiation. The characteristics of absolute variability (standard deviation in CZK) show a rising trend at first (the period between 1992 and 2002) and later fluctuate around a constant value. The characteristic of relative variability (the coefficient of variation in %) also increases between 1992 and 2002, but in the following years it shows a decreasing trend. A certain diversity in the light of income distribution skewness is further palpable from Table 1. The positive skewness of distribution of net per capita household incomes over all the analysed periods is shown. From this it follows that lower per capita incomes predominate in comparison with higher incomes in the given sets of data. It is a typical situation not only for income distributions, but also for wage distributions. In the last two rows of Table 1, the inflation rate and average rate of CZK/EUR are shown. The median value of per capita income in CZK increased from 2004 to 2007 by 21.1 %, the mean value by 19.7 % and inflation was 7.4 %. If we use the exchange rate from Table 1, the increase was 46.8 % for the median and 45 % for the mean. It reflects the strengthening of the Czech crown that is obvious from Table 1.

The development of chosen sample percentiles is shown in Figure 1. The above mentioned positive skewness is obvious from the gap between the graph lines. Lower quartile (under the median) curves are denser; those above the median are rarer. The line that separates 1 % of the highest incomes (99 % quantile) is markedly distant from the others.

Moreover, the interval in which 50 % of middle per capita incomes occur can be

Table 2: Sample sizes of the analysed samples and subsamples.

	Set	Year						
		1992	1996	2002	2004	2005	2006	2007
Land	Czech Republic	16,233	28,148	7,973	4,351	7,483	9,675	11,294
	Bohemia	9,923	22,684	5,520	2,775	4,692	6,086	7,074
	Moravia	6,310	5,464	2,453	1,576	2,791	3,589	4,220
Education	Primary	9,302	15,891	3,480	553	940	1,183	1,385
	Secondary	4,646	3,172	2,493	3,186	5,460	7,168	8,371
	Complete secondary	1,951	6,356	1,129	118	282	266	319
	Tertiary	334	2,729	871	494	801	1,058	1,219
Age	≤ 29	1,680	2,809	817	413	627	649	827
	30–39	3,035	4,718	1,398	716	1,247	1,620	1,655
	40–49	3,829	6,348	1,446	738	1,249	1,609	1,863
	50–59	2,621	5,216	1,642	919	1,581	2,051	2,391
	≥ 60	5,068	9,057	2,670	1,565	2,779	3,746	4,558

Source: own computations

specified from Table 1. For example, in 1992 50 % of middle per capita incomes occurred in the interval from 25,900 CZK to 39,298 CZK (the interval length of 13,400 CZK), while in 2007 50 % of per capita incomes occurred in the interval from 97,160 CZK to 148,937 CZK (the length of more than 51 thousand CZK). This length of the interval of fifty percent of medium incomes (interquartile range) is gradually gaping (see also Figure 1 and the three lines in the middle of the figure – for lower, median and upper quartile).

The first decile can be used as an upper bound for low incomes; the ninth decile, on the contrary, is used in the text as a lower bound of high incomes. In 1992 ten per cent of households had their income lower than or equal to 21,292 CZK. This bound increased gradually to the amount of 76,571 CZK in 2007. In the light of determination of 10 percent of the highest per capita incomes, in 1992 ten % of the highest incomes were equal to or higher than 51,977 CZK. This bound also increased gradually in time to the amount of 202,327 CZK in 2007. Marginal percentiles (1 % and 99 %) are used for the characterization of minimum and maximum incomes. In the interval given by these values, there are 98 % of middle incomes and only 1 % of the lowest and 1 % of the highest incomes are excluded. The development of the above mentioned quantities is noticeable in Figure 1.

The median is frequently used as a characteristic of the location of the sample instead of the mean in case of a skewed distribution. As a supplement to Figure 1, sample medians are shown for subpopulations defined by a household's location (Figure 2) and the education and age of the head of household (Figures 3 and 4 respectively). The line for the whole sample is shown in each figure in order to indicate a relative position of subgroups towards the whole sample. The values of the medians are given in Table 8. No obvious differences in the level of per capita incomes between the populations of Bohemia and Moravia are noticeable in Figure 2, although the level of incomes is a little higher in Bohemia compared to Moravia during the analysed period. The level of income has a

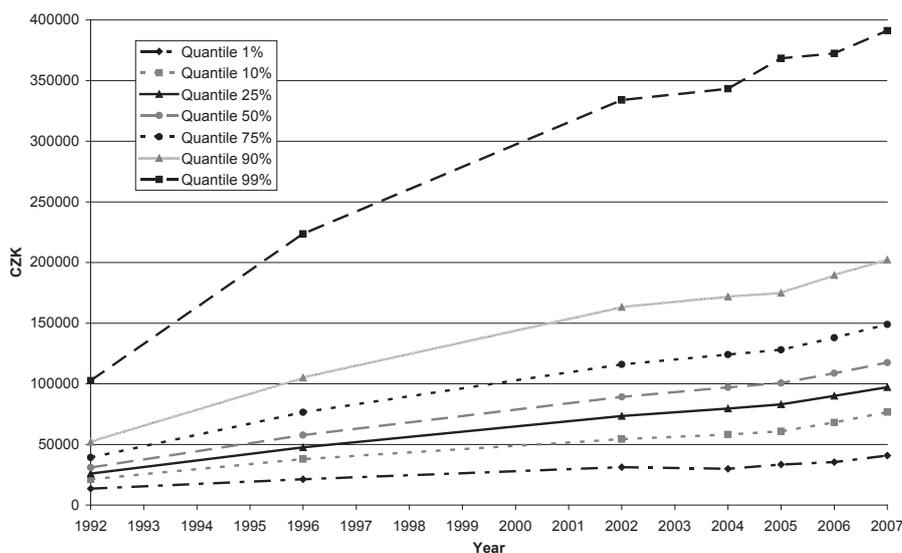


Figure 1: Sample percentiles (1 %, 10 %, 25 %, 50 %, 75 %, 90 %, 99 %) in CZK.

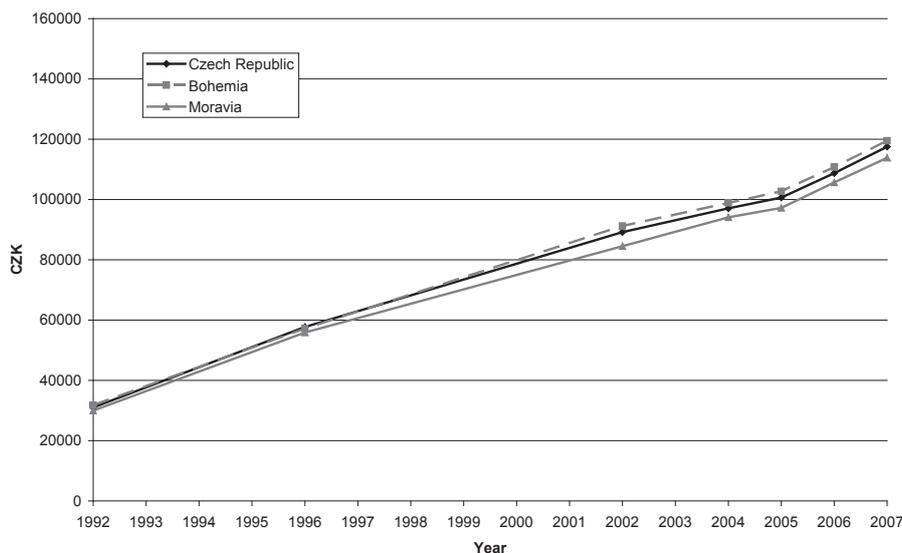


Figure 2: Sample medians of annual net per capita income (historical lands).

markedly increasing trend over the whole period in both historical lands (with the exception of 2004–2005) and the dynamics of this upward trend deserve a thorough analysis. If the analysed 16-year time interval is divided into three periods, 1992–2002, 2002–2004 and 2004–2007, the development can be characterized by means of the annual difference in the median income. In the first period the median income increased by 11.1 % a year on average in Bohemia (from 31,779 CZK in 1992 to 91,200 CZK in 2002) and by 10.9 % in Moravia (from 29,994 CZK in 1992 to 84,514 CZK in 2002). During the next period the rate of growth falls slightly, but it is necessary to have in mind that only two observations (the initial level in 2002 and the ending value in 2004) are registered. The median per

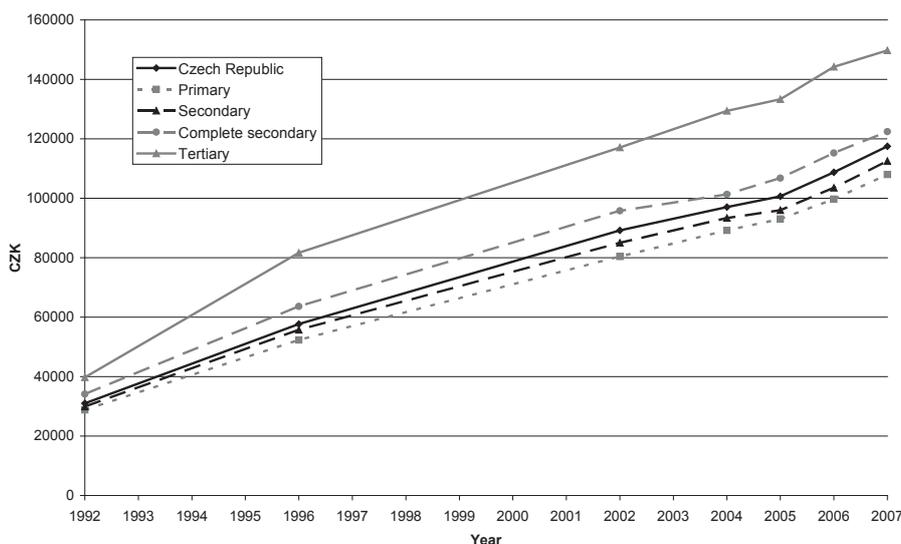


Figure 3: Sample medians of net per capita income (education of the head of household).

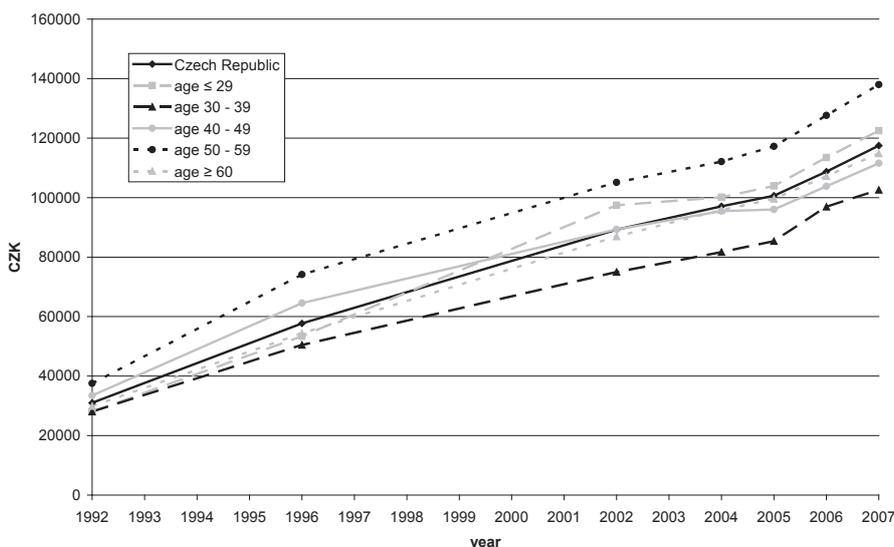


Figure 4: Sample medians of annual net income per capita (age of the head of household).

capita income in Bohemia grew by 4.1 % a year on average (from 91,200 CZK in 2002 to 98,861 CZK in 2004) while in Moravia it rose by 5.5 % on average (from 84,514 CZK in 2002 to 94,092 CZK in 2004). It means that in this period the rate of growth was higher in Moravia than in Bohemia. During the last period the growth rate of net annual per capita incomes rises again, the median of incomes increasing by 6.5 % a year on average in Bohemia (from 98,861 CZK in 2004 to 119,508 CZK in 2007) while in Moravia by 6.6 % (from 94,092 CZK in 2004 to 113,826 CZK in 2007).

In Figure 2, the positive impact of education is obvious. We would like to note that only education of the head of household is taken into account, other household members’

Table 3: Estimates of parameters for the whole samples,  $\chi^2$  statistics with 27 df.

Method	Param.	Year						
		1992	1996	2002	2004	2005	2006	2007
Maximum likelihood	$\mu$	10.384	10.995	11.438	11.503	11.542	11.623	11.703
	$\sigma$	0.390	0.424	0.459	0.665	0.446	0.435	0.421
	$\theta$	-0.3	52.2	73.5	7.6	-8.8	-42.3	-171.2
	$\chi^2$	1,227	2,197	1,060	524	995	1,067	1,199
L-moments	$\mu$	9.696	10.343	10.819	11.028	11.040	11.112	11.163
	$\sigma$	0.700	0.545	0.773	0.675	0.677	0.440	0.654
	$\theta$	14,491	25,362	37,685	33,738	36,606	40,327	45,634
	$\chi^2$	811	1,742	1,535	866	999	959	1,220

Source: own computations

education (that of a spouse in particular) is not reflected in the analysis. The median per capita income in 2007 for the head of household with tertiary education (149,751 CZK) is 122 percent of the median income for complete secondary education, the respective medians being 122,398 CZK, 112,508 CZK and 107,986 CZK for secondary complete, secondary incomplete and primary education.

It is evident from Figure 4 that the net per capita income of a household depends on the age of its head. The level of income gradually rises during the whole survey period. The level of per capita income of the households whose head is between 30 and 40 years of age is lower than that of the households whose head is 30 years old or younger. The reason evidently lies in the fact that young people up to 30 get ahead in career and start a family later; at that time their income decreased markedly. The level of income rises with age up to 50–60 years which is the last age group formed by the majority of employed people. Most people in the group over 60 retire and the income goes down again. Households with the head in the 50–60 age group have the highest per capita income, whereas members of the 30–40 year of age group have the lowest median of per capita income. Some of these observations are surprising, as we would expect a better position of households with younger male heads. This is the reason for analysing equalised incomes instead of per capita incomes. Equalised income takes into account potential expenditure sharing in households, especially in those with children. If we use per capita income, households with children (usually with 20–40 or 50-year-old heads) are disadvantaged when we put weight 1 to each member of a household. If equalised incomes are used, weights depend on the structure of a household and weight 1 is assigned only to the head of household.

The three-parameter lognormal distribution was fitted into the sets of data described above. Estimates of three parameters of the lognormal distribution for the whole sample for each year obtained by both methods of estimation mentioned above are given in Table 3. Differences between estimates, especially to estimate the shift parameter  $\theta$ , are apparent in the table. It is clear that even in the case of large samples, the estimates of parameters depend strongly on the method of estimation. According to Figure 5, however, characteristics of the level evaluated on the basis of these estimates are similar.

From our analysis, we conclude that estimation errors (not given in the text) are rather high when the shift parameter  $\theta$  is estimated in comparison with other two parameters.

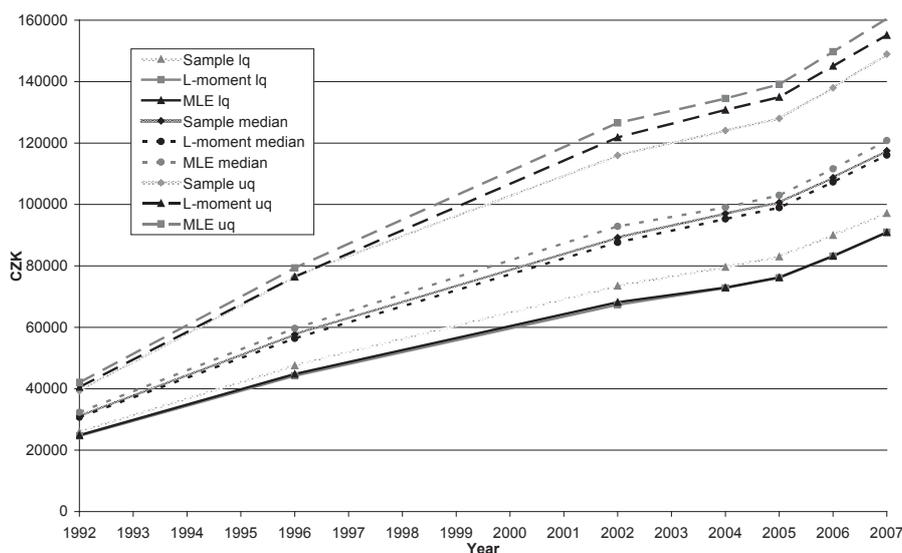


Figure 5: Sample and estimated quartiles (lower quartile, median, upper quartile).

Very different values of the parameter  $\theta$  are shown in Table 3. The values of theta for maximum likelihood estimates are even negative and estimates via L-moments are greater than minimal incomes, coinciding roughly with the third percentile. Because of a very close contact of the lower tail of the lognormal density curve with the horizontal axis, the negative parameter theta does not seem to be a problem for a good fit of the model. Due to its estimates being negative, however, the value of theta has no interpretation. The estimates of the parameter  $\mu$  are smaller in the case of maximum likelihood estimation; both methods point to an increase of these parameters in time. The variance parameter  $\sigma^2$  is remarkably greater for L-moments estimation.

The fitted models may be compared with the use of various techniques; in this text, the  $\chi^2$  statistic of a goodness-of-fit test was employed. Because of the time development of income level, different evaluation intervals (31 in each year) were used for different values of income. The  $\chi^2$  statistic depends strongly on the choice of intervals for small values of income because of very different estimates of the shift parameter  $\theta$ . No method of estimation seems to be superior for all analysed years. It coincides with the literature, as for large samples maximum likelihood estimates have an optimal theoretical (asymptotic) property. Estimates with the use of L-moments are sometimes more efficient than maximum likelihood estimates, see Hosking (1990). For all years  $p$ -values for the goodness-of-fit test equal zero, so they are not given in Table 3.

In Figures 6 and 7, the estimated lognormal densities are shown. Estimated modes are not given in the text, but from these two figures the development of this (estimated) characteristic of the income level is obvious. The modes shift from the left (1992) to the right (2007) and values of densities in the mode decrease with time. The curves are similar in their shapes, but as the scale of both figures is the same, we observe different values of estimated densities.

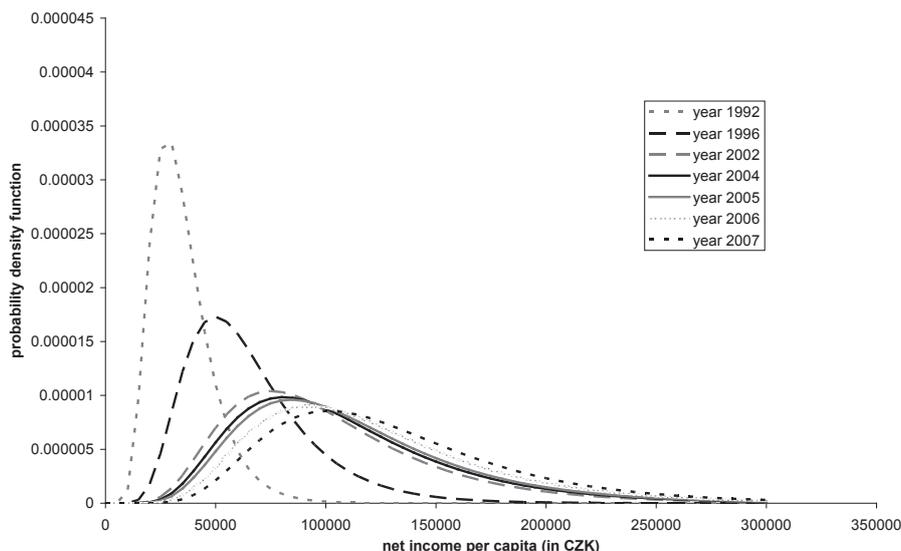


Figure 6: Estimated lognormal densities for annual net per capita income (CZK) for 1992, 1996, 2002, 2004–2007 (maximum likelihood method).

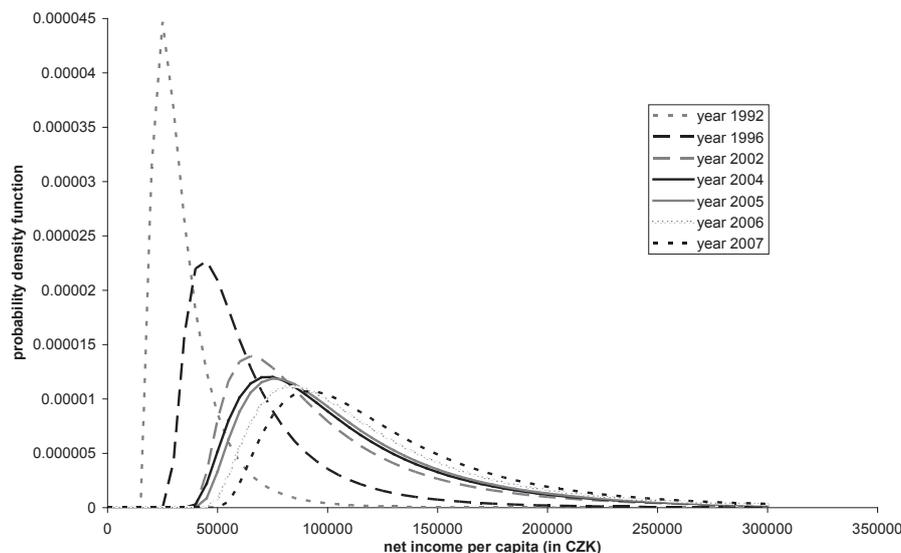


Figure 7: Estimated lognormal densities for annual net per capita income (CZK) (L-moment method).

### 3.2 Equivalised Income in the European Union

This chapter offers a comparison of nominal income levels in the European Union countries (in EUR) between 2004 and 2007 (and a more recent value for 2009 in Figure 7).

Table 4: Medians of equalised net annual income in 2004–2007 (in EUR).

Country	Year			
	2004	2005	2006	2007
EU (all 27 countries)	12,975	13,188	13,900	14,598
EU (15 old countries)	15,465	15,575	16,545	17,282
EU (12 new countries)	—	—	3,255	3,868
Austria	18,001	17,854	18,156	19,011
Belgium	16,581	17,213	17,566	17,985
Bulgaria	—	1,383	1,481	2,171
Cyprus	13,157	14,536	16,014	16,765
Czech Republic	4,233	4,802	5,423	6,068
Denmark	22,124	22,663	23,341	24,161
Estonia	2,981	3,639	4,448	5,547
Finland	17,496	18,345	18,703	19,815
France	15,946	16,209	16,441	18,984
Germany	6,393	15,663	17,777	18,309
Greece	9,417	9,850	10,200	10,800
Hungary	3,447	3,850	3,936	4,400
Ireland	18,798	19,757	22,065	22,995
Italy	14,352	14,524	15,011	15,639
Latvia	2,204	2,534	3,350	4,832
Lithuania	2,058	2,534	3,276	4,169
Luxembourg	28,396	29,480	29,892	30,917
Malta	8,047	8,747	9,100	9,558
Netherlands	17,001	17,263	18,244	19,522
Poland	2,533	3,111	3,502	4,155
Portugal	7,195	7,311	7,573	8,143
Spain	10,600	11,480	12,038	12,950
Romania	—	—	1,657	1,953
Slovenia	8,797	9,317	9,907	10,893
Slovakia	2,830	3,313	3,972	4,792
Sweden	17,499	17,991	18,845	20,573
United Kingdom	18,540	19,512	21,014	18,923

Source: Eurostat

The income that is analysed here is not a net annual household income per capita as in Subsection 3.1, but the net annual household income per economic unit, or the equalised income. In the EU methodology, the head of household's weight value is 1, other adult household members and children having weights of 0.5 and 0.3 respectively (an alternative weighing system is that of OECD). It means that a household with two parents and two children has 4 members (for per capita income evaluation  $4 \times 1 = 4$  in the denominator) but only 2.1 ( $= 1 + 0.5 + 2 \times 0.3$ ) when equalised income is evaluated. Even greater is the gap in the case of a household with a single parent and 3 children, where the number of members is 4 again, but the number of economic units is 1.9 ( $= 1 + 3 \times 0.3$ ). As the total net income is the same for both the approaches (per capita and per economic

Table 5: Average growth of the net annual equivalised income median in 2004–2007.

Countries	EU 27	EU 15	EU 12	Czech Republic
Average growth rate	1.040	1.038	1.188 (2006–2007)	1.128 (2004–2007) 1.119 (2006–2007)

Source: own computations

Table 6: Average growth of the net annual income median in 2004–2007.

EU 15		EU 12	
Country	Average growth rate	Country	Average growth rate
Austria	1.018	Bulgaria	1.253*
Belgium	1.027	Cyprus	1.084
Denmark	1.030	Czech Republic	1.128
Finland	1.042	Estonia	1.230
France	1.060	Hungary	1.085
Germany	1.038	Latvia	1.299
Greece	1.047	Lithuania	1.265
Ireland	1.069	Malta	1.059
Italy	1.029	Poland	1.179
Luxembourg	1.029	Romania	1.179**
Netherlands	1.047	Slovenia	1.074
Portugal	1.042	Slovakia	1.192
Spain	1.069		
Sweden	1.055		
United Kingdom	1.007		

Source: own computations

\* only years 2005–2007, \*\* only years 2006–2007

unit), the resulting income per an economic unit is greater for equalised than for per capita income. All values in this part are taken from the Eurostat webpage (EUROSTAT).

Table 4 contains the medians of net annual household income per economic unit in 2004–2007. Uniform survey methodology was employed in the whole European Union, allowing a direct comparison of nominal income in different countries. Standard division of the EU countries into three groups (EU27: all present member countries, EU15: the “old” members before 31th December 1995, and EU12: the “new countries”) is used first of all. Then all the EU countries are dealt with separately. All values of the medians are in EUR. Particular members of these groups are shown in Table 6.

Tables 5 and 6 represent the average growth rate of the nominal median of incomes of the European Union states in 2004–2007. From Table 4 and Figure 7 we can observe a markedly lower level of nominal income in virtually all former communist countries. Despite this, it is just in the case of these countries that the average growth rate of nominal income is the highest, as we can see from Tables 5 and 6.

In Table 5, the average annual income medians are compared for the above mentioned



Figure 8: Median of net annual income in 2008 (in EUR).

groups of EU countries and the Czech Republic. Higher values of the rate in the Czech Republic or in the group of “new” countries compared to the “old” EU members are obvious.

In Table 6, the old EU countries are in the left column and the new ones on the right. Again, higher values are apparent for new members with the highest average rate for Bulgaria (+25.3 % in two years) or Estonia (with +23 %). Due to the fact that the last year in Tables 5 and 6 is 2007, no effects of the economic crisis are reflected in the statistics.

In Figure 7, the medians of equivalised annual net income are shown for 2008, black triangles indicating the “old” and grey squares the “new” countries. The greater the distance from the centre, the higher the median of income. We can see that the new countries are closer and the old ones further from the centre (with the exception of Cyprus, whose median is comparable to that of old countries), the highest and smallest medians for Luxembourg and Romania respectively being noticeable.

The present paper puts emphasis on the Czech Republic. In Table 7, the medians of both per capita and equivalised incomes for the analysed years (the former taken from our computations, the latter from Eurostat data) are compared. As expected, income per economic unit tends to be greater (it is equal only for one-member households), which is in accordance with higher medians in the second row of the table. The increase in medians of income per economic unit and per capita is 43.3 % and 46.9 %, respectively.

Table 7: Medians of net annual household income (in EUR).

	Year			
	2004	2005	2006	2007
Net annual per capita income	3,042	3,379	3,837	4,468
Net annual equivalised income	4,233	4,802	5,423	6,068

Source: Eurostat, CNB, own computations

Table 8: Medians of annual net per capita income in the Czech Republic (in CZK).

	Set	Year				Average growth rate
		2004	2005	2006	2007	
Country	Czech Republic	97,050	100,640	108,744	117,497	1.066
	Bohemia	98,861	102,719	110,754	119,508	1.065
	Moravia	94,092	97,242	105,669	113,826	1.066
Education	Primary	89,200	93,000	99,720	107,986	1.066
	Secondary	93,374	96,000	103,500	112,508	1.064
	Complete secondary	101,316	106,800	115,248	122,398	1.065
	Tertiary	129,389	133,296	144,205	149,751	1.050
Age	≤ 29	100,120	103,881	113,470	122,503	1.070
	30–39	81,699	85,355	96,930	102,573	1.079
	40–49	95,447	95,992	103,800	111,522	1.053
	50–59	112,100	117,209	127,625	137,984	1.072
	≥ 60	95,800	99,600	107,185	114,800	1.062

Source: own computations

## 4 Conclusions

In this article, the net annual per capita income is investigated for the period from 1992 to 2007 during which there were the two different states in question – the Czech and Slovak Federal Republic (1992) and the Czech Republic (1993–2007). Sample distributions and their lognormal models are discussed from various points of view.

The three-parameter lognormal distribution was used as a model for incomes, two methods of estimation of the unknown parameters – the maximum likelihood and L-moment methods – were employed. This approach enables us to compare their accuracy. According to the estimated densities, both methods give similar results, point estimates of parameters, however, being surprisingly different. The obtained level characteristics are also very similar, regardless of the method used.

The worst estimate (measured by standard error) is obtained for the shift parameter  $\theta$ . It is rather difficult to interpret its values as they are negative for the maximum likelihood method and greater than the minimum value in the sample for the L-moment method.

The appropriateness of the choice of the lognormal as a model was rejected by a  $\chi^2$  test for all models. But the large-sized sample test power being so high, the test reveals the smallest differences between sample distribution and the model that are neither substantial nor important in practical use. We hope that the distribution is acceptable for the analysis

as it is frequently used with satisfactory results in the literature dealing with incomes and wages, such a choice being underpinned by relevant theoretical reasons as well (e.g. Gibrat's law of proportionate effects, Kleiber and Kotz, 2003).

The development of distributions and their various characteristics are examined in order to analyse the change of income distributions in the transformation from centrally planned into market economy. It is obvious that income disparities have emerged, a group with markedly high incomes having developed since 1992. The level of income distributions increases distinctively, the skewness of income distributions gradually declining. The amount of net household per capita income depends explicitly on the age and education of the head of household. It can be noted that net per capita income rises with the highest achieved education, but we cannot claim categorically that it rises with the age of the head of household. Households whose head is between 30 and 40 years (maybe with small children) and over 60 years of age (usually those of pensioners) have the lowest net per capita income. Upper and lower bounds determining the net per capita income gap move away from each other in time. During the whole survey period, the obvious growth of annual net per capita income level is registered (in the light of both total and separated sets, according to residence, age and education), a remarkable inflation rate being taken into account. Nominal income was focused on exclusively, its fast growth having been caused by inflation and the real income increase. The nominal income growth was faster than inflation, the real income increase being obvious. This observation is quantified in the text.

Income per capita in CZK – despite the widespread and recommended use of equivalised income – is analysed in the present research paper. The use of income per economic unit can surpass some strange or surprising results in our analysis and it can ensure better results. At the end, we present medians of equivalised income in EUR for the EU members. European states from the group of 15 original members have a noticeably higher nominal level of income per economic unit, but we can observe a lower growth rate of these incomes. On the other hand, twelve new member states have a substantially lower median of income but higher growth rates.

For the Czech Republic, we compare income per capita with income per unit in Euro in order to illustrate the relation between these two quantities. Different growth rates of median income for the Czech Republic in Euro and CZK may have been caused by the development of the exchange rate of CZK to EUR.

## Acknowledgements

This paper was supported by the means of institutional support of a long-term conceptual advancement of science and research at the Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic.

## References

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

- Bartošová, J., and Bína, V. (2009). Modelling of income distribution of Czech households in years 1996–2005. *Acta Oeconomica Pragensia*, 17, 3-18.
- Bílková, D. (2008). Application of lognormal curves in modeling of wage distributions. In *7<sup>th</sup> International Conference APLIMAT 2008* (p. 341-351). Bratislava: Slovak University of Technology.
- Cohen, A. C., and Whitten, J. B. (1980). Estimation in the three-parameter lognormal distribution. *Journal of American Statistical Association*, 75, 399-404.
- Dagum, C. (1990). Generation and properties of income distribution functions. In *Income and wealth distribution, inequality and poverty: Proceedings of the Second International Conference on Income Distribution by Size: Generation, Distribution, Measurement and Applications, University of Pavia, Italy, September 23-30, 1989* (p. 1-17).
- Dagum, C. (1997). A systemic approach to the generation of income distribution models. *Journal of Income Distribution*, 6, 105-126.
- Flachaire, E., and Nunez, O. (2007). Estimation of the income distribution and detection of subpopulations: an explanatory model. *Computational Statistics and Data Analysis*, 51, 3368-3380.
- Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52, 105-124.
- Hosking, J. R. M., and Wales, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. New York: Cambridge University Press.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 1* (2nd ed.). New York: Wiley.
- Kleiber, C., and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. New York: Wiley-Interscience.
- Kyselý, J., and Pícek, J. (2007). Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic. *Climate Research*, 33, 243-255.
- McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 52, 647-665.
- Pacáková, V., and Šipková, L. (2007). Generalized lambda distributions of household incomes. *E+M Ekonomie a Management*, 10, 98-107.
- Parker, S. C. (1997). The distribution of self-employment income in the United Kingdom, 1976-1991. *The Economic Journal*, 107, 455-466.
- Pavelka, R. (2009). Application of density mixture in the probability model construction of wage distributions. *Applications of Mathematics and Statistics in Economy: AMSE 2009, Uherské Hradiště*, 341-350.
- Smithers, J. C., and Schulze, R. E. (2001). A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments. *Journal of Hydrology*, 241, 42-52.

Authors' address:

Diana Bílková and Ivana Malá

Department of Statistics and Probability

Faculty of Informatics and Statistics

University of Economics in Prague

Sq. W. Churchill 1938/4

130 67 Prague, Czech Republic

E-Mails: [bilkova@vse.cz](mailto:bilkova@vse.cz), [malai@vse.cz](mailto:malai@vse.cz)

Web Page: <http://www.vse.cz/>