ANALYSIS OF THE DISTRIBUTION OF INCOME IN RECENT YEARS IN THE CZECH REPUBLIC BY REGION

Jana Langhamrová – Diana Bílková

Abstract

This paper deals with the comparing of the development of the sample characteristics by the income distribution. Data for this research come from a survey of the Czech Statistical Office Microcensus (2002) and SILC (2005-2009). The studied variable is the annual net household income per capita (in CZK). We researched 84 income distributions. We used for each of the income distribution a model distribution. For purpose of construction of these theoretical distributions has been used three-parametric lognormal curve. Moment method of point estimation of parameters was used in estimating the parameters of the lognormal curve. The paper also deals with the development of probability density curves of income distribution in time. Furthermore, trend analysis was used to study the development of parameters of lognormal curve, on which basis, income distribution predictions were made for next year by region. Using the predicted values of the parameters of considered lognormal distribution forecasts of income distributions were constructed for 2010 (interval frequency distribution).

Key words: Income distribution, lognormal distribution, forecasts of income distribution

JEL Code: C13, C16

Introduction

The wealth and living standards of people living in the country or region reflect among other things, the amount of their income. Analysis of income distributions is therefore one way of assessing the population's living standards. Comparison of the income distribution can be performed on inter-regional or international level.

Information obtained from the analysis of income distribution can be used in setting state tax burden, or determining the amount of social benefits.

1 Model selection

When we construct the model of income distribution, we need to make a kompromise between the requirement of a sufficient number of parameters, which is good in terms of flexibility and adaptability to the actual shape of the distribution, but the model can not contain too many parameters, because the model is less stable in time and space, and it is difficult to interpret.

Lognormal distribution is one of the most frequently used distribution in modelling the of income distributions.

Model parameters are estimated on the basis of a random sample, in our case the method of moments. When we use the method of moments, we have not guaranteed maximal efficiency of estimate. However, due to the large sample size in the case of the income distribution, we do not solve this problem. Moments of higher order including our characteristic of skew are sensitive to inaccuracies on both ends of the distribution.

Probability model provides us with detailed information about the population and is therefore qualitatively very valuable result.

1.1. Three-parametric lognormal distribution

Random variable X has three-parametric lognormal distribution LN(μ,σ^2,θ) with parameters μ, σ^2 a θ, where $-\infty < \mu < \infty$, $\sigma^2 > 0$ a $-\infty < \theta < \infty$, if its probability density function $f(x;\, \mu,\sigma^2,\theta)$ has the form

$$f(x;\, \mu,\sigma^2,\theta) = \frac{1}{\sigma(x-\theta)\sqrt{2\pi}}\, e^{-\frac{[\ln(x-\theta)-\mu]^2}{2\sigma^2}}, \quad x > \theta, \tag{1}$$

$$=0, \text{ otherwise.}$$

Random variable

$$Y = \ln(X - \theta) \tag{2}$$

has normal distribution N(μ,σ^2) and random variable

$$U = \frac{\ln(X-\theta)-\mu}{\sigma} \tag{3}$$

has standard normal distribution N(0;1). Parametr μ is the expected value of random variable (2) and parametr σ^2 is the variance of the random variance. Parametr θ represents the theoretical minimum of the random variable X. The income distribution is possible that the value of the parameter θ is negative, i.e. three-parametric lognormal curve is often the beginning of its course gets below zero. However, due to the fact that the curve has initially very close contact with the x-axis, it does not interfere good agreement the model with the actual distribution.

The basic moment characteristic of the level of the random variable X, having three-parametric lognormal distribution, is a expected value of this random variable

$$E(X) = \theta + e^{\mu + \frac{\sigma^2}{2}}. \tag{4}$$

The quantile characteristic of the level is 100 P% quantile of the random variable for which, the value of the distribution function of random variable X at point 100 P% quantile is equal to P

$$F(x_P) = P, \tag{5}$$

where $0 < P < 1$. 100 P% quantile of the random variable X having three-parametric lognormal distribution is given by

$$x_P = \theta + e^{\mu + \sigma u_P}, \tag{6}$$

where u_P is 100 P% quantile of the standard normal distribution N(0;1). Substituting into relation (6) P= 0.5, we get 50% quantile of the random variable X having three-parametric lognormal distribution this is a median of the random variable

$$\tilde{x} = \theta + e^{\mu}. \tag{7}$$

Another characteristic of the level of the random variable X having three-parametric lognormal distribution is a mode of the random variable

$$\hat{x} = \theta + e^{\mu - \sigma^2}. \tag{8}$$

The basic moment characteristic of variability of the random variable X having three-parametric lognormal distribution is a variance of the random variable

$$D(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1). \tag{9}$$

Another moment characteristic of variability of the random variable X having three-parametric lognormal distribution is the standard deviation of the random variable

$$\sqrt{D(X)} = e^{\mu + \frac{\sigma^2}{2}} \sqrt{e^{\sigma^2} - 1}. \tag{10}$$

Characteristic of the relative variability of the random variable *X*, which has three-parametric log-normal distribution is the coefficient of variation of this random variable. It is a dimensionless characteristic of variability

$$V(X) = \frac{e^{\mu + \frac{\sigma^2}{2}} \sqrt{e^{\sigma^2} - 1}}{\theta + e^{\mu + \frac{\sigma^2}{2}}}. \tag{11}$$

Among the moment characteristics of the shape of the random variable *X* having three-parametric lognormal distribution is a coefficient of skewness

$$\beta_1(X) = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1} \tag{12}$$

and a coefficient of kurtosis of this random variable

$$\beta_2(X) = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3. \tag{13}$$

2 Estimation of the parameters of lognormal distribution using the method of moments

To estimate the parameters of three-parametric lognormal distribution we use the method of moments.

In the method of moments, we give equality to the sample moment and theoretical moment of the distribution. We can combine moments about the common and central moments. This method of estimating parameters is to use very simple but also very inaccurate. Significantly inaccurate is the estimate of the theoretical variance of the random variable *X*, its selective counterpart. The use of the method of moments for estimation of parameters is not a bad thing in the case of the income distribution, because we work with large-scale samples.

In the case of the method of moments parameter estimation we give equality to the sample arithmetic mean $\bar{x}$ to the expected value of the random variable *X* and the sample second central moment m_2 we give equality to the variance of random variable *X*. The third equation is obtained so that we give equality to the sample third central moment m_3 with the theoretical third central moment of the random variable *X*. We get a set of the methods of moments

$$\bar{x} = \tilde{\theta} + e^{\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}}, \tag{14}$$

$$m_2 = e^{2\tilde{\mu} + \tilde{\sigma}^2}(e^{\tilde{\sigma}^2} - 1), \tag{15}$$

$$m_3 = e^{3\tilde{\mu} + \frac{3}{2}\tilde{\sigma}^2}(e^{\tilde{\sigma}^2} - 1)^2(e^{\tilde{\sigma}^2} + 2). \tag{16}$$

We obtain from equations (15) and (16)

$$b_1^2 = m_3^2 \cdot m_2^{-3} = (e^{\tilde{\sigma}^2} - 1)(e^{\tilde{\sigma}^2} + 2)^2, \tag{17}$$

and here we obtain from the system of the moments equations (14) to (16) the moment estimates of the parameters of the three-parametric lognormal distribution

$$\tilde{\sigma}^2 = \ln\left[\sqrt[3]{1 + \frac{1}{2}b_1^2 + \sqrt{\left(1 + \frac{1}{2}b_1^2\right)^2 - 1}} + \sqrt[3]{1 + \frac{1}{2}b_1^2 - \sqrt{\left(1 + \frac{1}{2}b_1^2\right)^2 - 1}} - 1\right], \tag{18}$$

$$\tilde{\mu} = \frac{1}{2}\ln\frac{m_2}{e^{\tilde{\sigma}^2}(e^{\tilde{\sigma}^2} - 1)}, \tag{19}$$

$$\tilde{\theta} = \bar{x} - e^{\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}}. \tag{20}$$

3 Data

Data were obtained from a survey of the Czech Statistical Office Microcenzus (2002) and SILC - European survey on income and living conditions (2005-2009). Different length of the interval between 2002 and 2005 and between other years is caused by a change of methodology of statistical surveys. Observed variable is a net annual household income per capita (in CZK).

4 Results

4.1 Development in 2002-2009

Using the lognormal distribution with three parameters and the method of moments were modelled the following income distributions for each region of the Czech Republic in 2002, 2005-2009.

In Figures 1-6 are shown the probability density functions of net household income per capita in all regions for each year separately. This graphs show that during the period increased in all regions the average net household income per capita. The highest average net household income was over the whole period in the Capital Prague Region and the lowest average net household incomes were recorded in the Zlin Region in 2002 and 2007,

Figure 1: Probability density functions of net annual household income per capita according to region in 2002

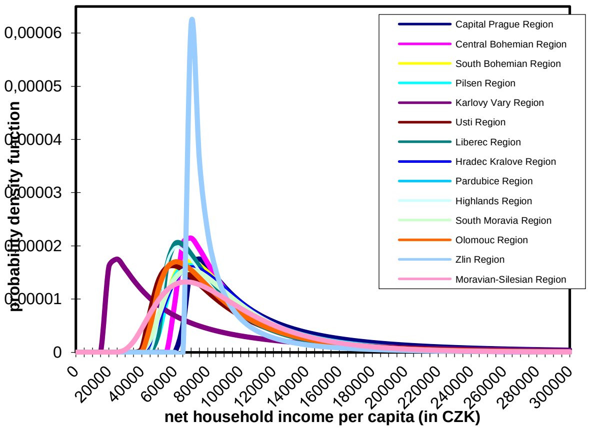


Figure 2: Probability density functions of net annual household income per capita according to region in 2005

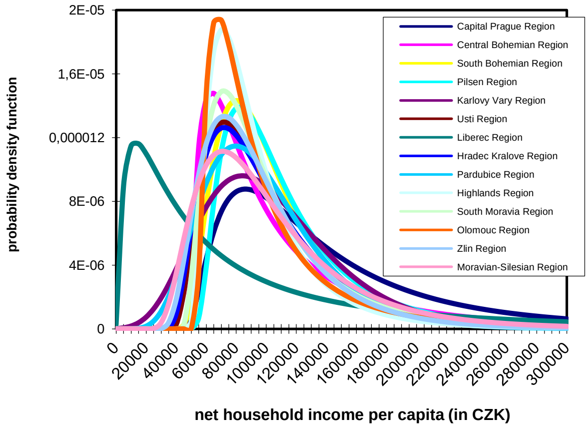


Figure 3: Probability density functions of net annual household income per capita according to region in 2006

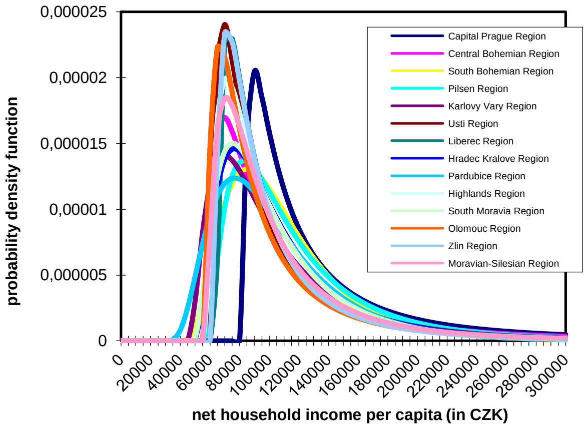


Figure 4: Probability density functions of net annual household income per capita according to region in 2007

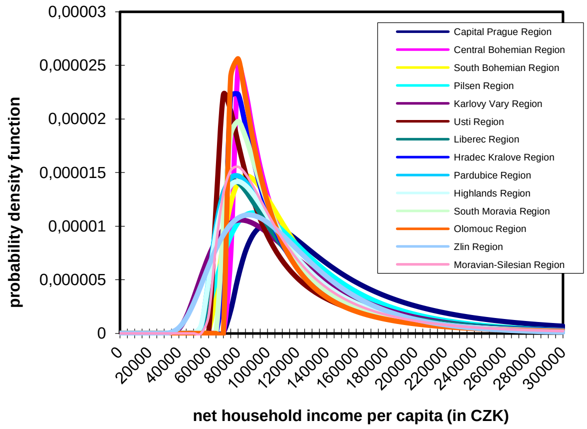


Figure 5: Probability density functions of net annual household income per capita according to region in 2008

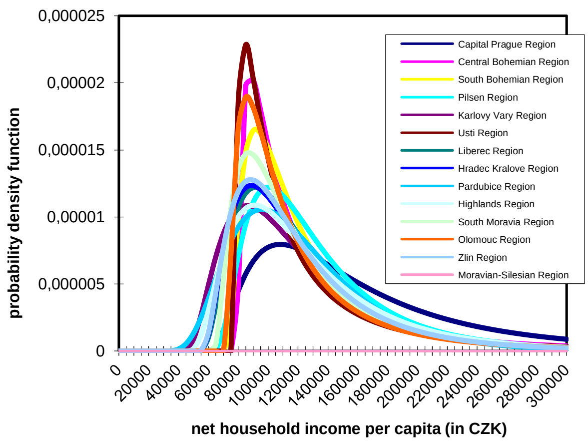
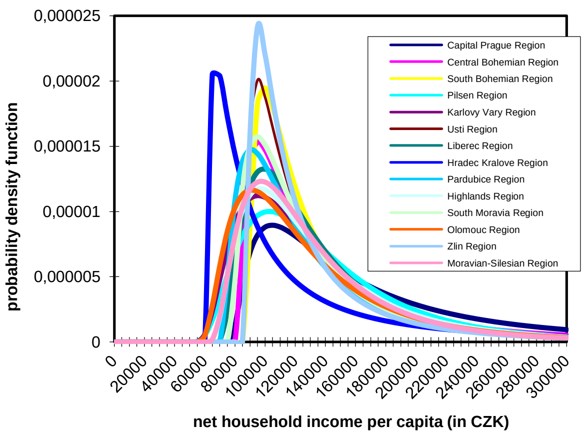


Figure 6: Probability density functions of net annual household income per capita according to region in 2009



in the Highlands Region in 2005, the Olomouc Region in 2006 and 2009 and in the Pardubice Region 2008.

Then we selected two regions – Capital Prague Region and Karlovy Vary Region, to which will be further described development in the period 2002-2009.

Table 1 lists some selected characteristics and the estimated values of the parameters of the three-parametric lognormal distribution for the Capital Prague Region.

From these values it is clear that the arithmetic mean is increasing over the entire period from the original value of 137,015 in 2002 up to 193,211 in 2009.

The highest variability in net annual household income was recorded in the Capital Prague Region in 2002, while lowest in 2008.

Table 1: Sample characteristics of net annual household income per capita and corresponding estimates of parameters of three-parametric lognormal curves in Capital Prague Region in 2002-2009

Region		Sample characteristics				Parameter estimates		
Capital Prague Region	Arithmetic mean	Standard deviation	Variance	Coefficient of Variation	Skewness	μ	σ^2	θ
2002	137,015	122,255	14,946,193,319	89.23 %	9.5	10.546	1.322	63,326.97
2005	149,426	96,804	9,370,990,951	64.78 %	3.5	11.281	0.598	42,469.45
2006	153,111	132,476	17,549,828,689	86.52 %	12.1	10.423	1.497	81,994.11
2007	162,198	95,659	9,150,707,163	58.98 %	4.1	11.106	0.704	67,531.66
2008	173,725	99,965	9,993,048,244	57.54 %	3.2	11.398	0.547	56,551.52
2009	193,211	146,187	21,370,724,011	75.66 %	5.7	11.206	0.936	75,772.45

Source: data Microcensus, data SILC, own calculations

Table 2 contains the estimated parameters of the three-parametric lognormal distribution for the Karlovy Vary Region and the arithmetic mean is also increasing over the period. The table shows that the value of the parameter θ is negative in 2005, which means that the probability density function is in its beginning in negative values but in negative values is this curve very close adherence to the horizontal axis.

The arithmetic mean reached in the Karlovy Vary Region in 2002 110,785. In 2005, dropped arithmetic mean to 104,151 and from this year is increasing to134,247 in 2009.

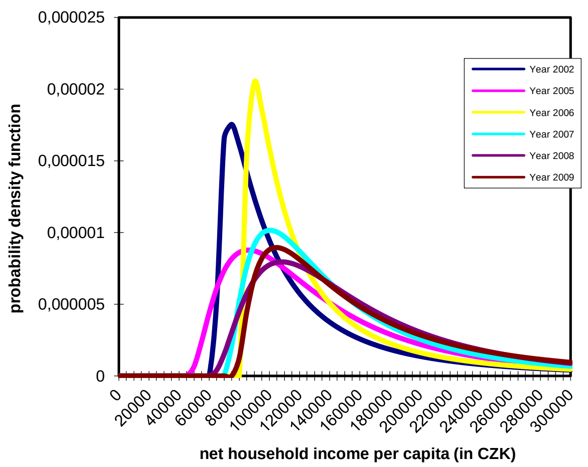
The highest variability of income was amounted the Karlovy Vary Region in 2002, while the lowest variability has in 2009.

Table 2: Sample characteristics of net annual household income per capita and corresponding estimates of parameters of three-parametric lognormal curves in Karlovy Vary Region in 2002-2009

Region		Sample characteristics				Parameter estimates		
Karlovy Vary Region	Arithmetic mean	Standard deviation	Variance	Coefficient of Variation	Skewness	μ	σ^2	θ
2002	110,785	201,733	40,696,268,747	182.09 %	16.0	10.606	1.709	15,930.45
2005	104,151	47,084	2,216,867,611	45.21 %	1.0	11.84	0.099	-41,591.5
2006	110,158	61,733	3,810,957,261	56.04 %	3.6	10.801	0.617	43,334.56
2007	118,596	58,343	3,403,938,580	49.20 %	2.2	11.242	0.346	27,903.42
2008	125,245	60,804	3,697,110,089	48.55 %	2.4	11.172	0.399	38,424.69
2009	134,247	60,201	3,624,147,468	44.84 %	2.5	11.124	0.418	50,713.84

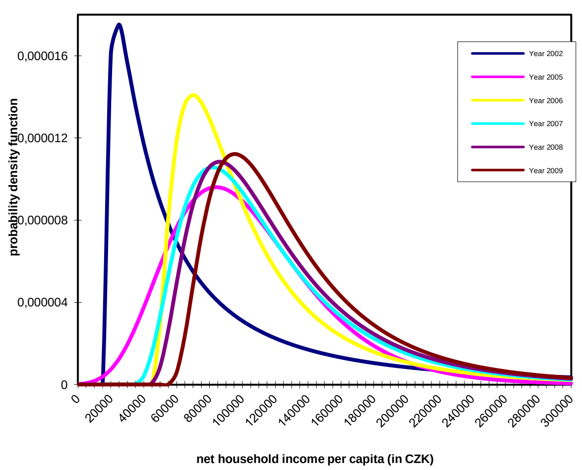
Source: data Microcensus, data SILC, own calculations

Figure 7: Probability density functions of net annual household income per capita in Capital Prague Region in 2002 - 2009



Source: data Microcensus, data SILC, own calculations

Figure 8: Probability density functions of net annual household income per capita in Karlovy Vary Region in 2002 - 2009



Source: data Microcensus, data SILC, own calculations

In Figure 7 are shown the probability density functions for the Capital Prague Region in the period 2002-2009 and there is noticeable that between 2002 and 2006, the probability density functions were more kurtosis than in other years. We can observe that the probability density function shifts to the right every year, it suggesting the fact that in this region is increasing number of people with higher net incomes.

Figure 8 shows the probability density functions for the Hradec Králové Region in the period 2002-2009. Again, it is evident that in 2002 the probability density function was more kurtosis than in subsequent years. The curve of the probability density function is in this case slowly moves to the right.

4.2 Prediction for 2010

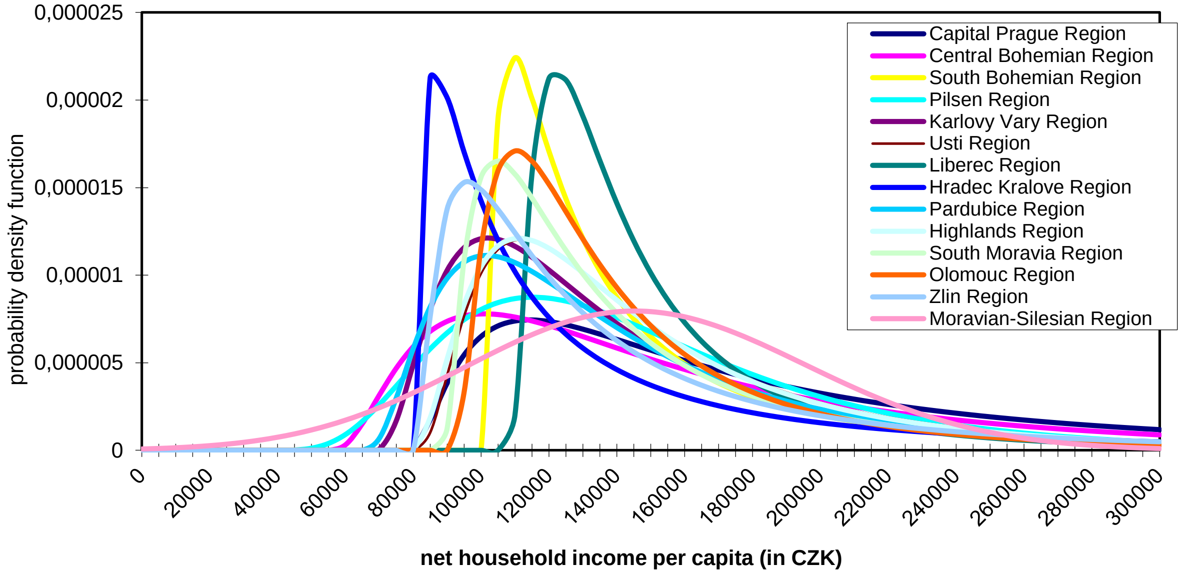
Table 3: Forecasts of sample characteristics of net annual household income per capita for 2010 and corresponding estimates of parameters of lognormal curves

		Sample characteristics				Parameter estimates		
Region	Arithmetic mean	Standard deviation	Variance	Coefficient of Variation	Skewness	μ	σ^2	θ
Capital Prague Region	215,810	165,206	27,293,022,436	76.55	5.3	11.392	0.888	77,697.25
Central Bohemian Region	174,562	112,589	12,676,282,921	64.5	3.6	11.391	0.624	53,629.65
South Bohemian Region	156,349	88,589	7,848,010,921	56.66	8.8	10.297	1.262	100,666.80
Pilsen Region	149,107	59,932	3,591,844,624	40.19	1.6	11.599	0.218	27,618.83
Karlovy Vary Region	143,164	62,759	3,938,692,081	43.84	3.0	10.992	0.512	66,416.33
Usti Region	156,450	67,988	4,622,368,144	43.46	3.3	10.975	0.569	78,834.43
Liberec Region	153,420	49,169	2,417,590,561	32.05	4.5	10.339	0.773	107,895.00
Hradec Kralove Region	158,616	167,148	27,938,453,904	105.38	16.1	10.412	1.714	80,245.44
Pardubice Region	146,469	68,374	4,675,003,876	46.68	3.0	11.078	0.512	62,854.78
Highlands Region	153,440	64,410	4,148,648,100	41.98	3.1	10.985	0.531	76,533.84
South Moravia Region	150,309	74,444	5,541,909,136	49.53	5.3	10.595	0.888	88,073.58
Olomouc Region	147,091	57,239	3,276,303,121	38.91	4.1	10.584	0.71	90,782.74
Zlin Region	152,037	95,853	9,187,797,609	63.05	6.4	10.668	1.025	80,338.12
Moravian-Silesian Region	146,990	52,865	2,794,708,225	35.97	0.1	14.277	0.001	-1,439,55

Source: data Microcensus, data SILC, own calculations

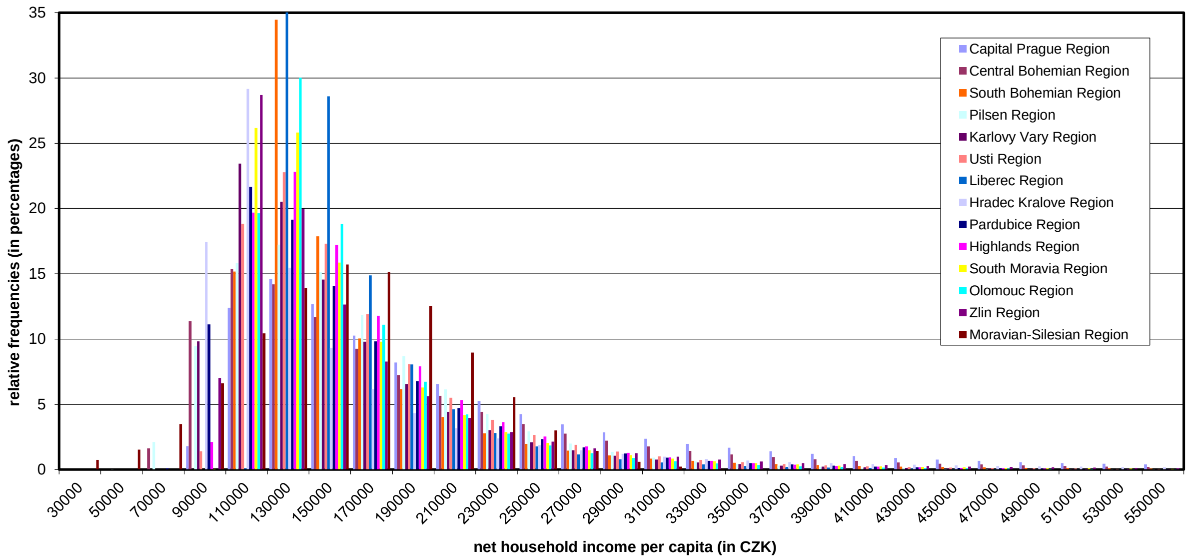
Further, the trend analysis was calculated of the development of the parameters of the three-parametric lognormal curves and on the basis of the parameters were constructed lognormal probability density functions and histogram for 2010.

Figure 11: Forecasts of probability density functions of net annual household income per capita according to region in 2010



Source: data Microcensus, data SILC, own calculations

Figure 12: Forecasts of histogram of relative frequencies (in percentages) of net annual household income per capita according to region in 2010



Source: data Microcensus, data SILC, own calculations

As shown in Table 3, the highest average net household income reaches the Central Bohemian Region, the lowest average net household income, according to predictions has Usti Region. The coefficient of variation should be the highest in the Pardubice Region, the lowest in the Hradec Králové Region.

In Figure 11 we see that the predicted probability density functions of the Hradec Kralove Region, the Liberec Region and South Bohemian Region are much more kurtosis than others predicted distributions. In these regions should be more people with lower incomes.

The Figure 12 shows a histogram of predicted relative frequencies of net household income per capita by region. In Figure 12, for example, can be read that 35 % of people in the Liberec Region would reach an net household income from 130,000 – 150,000 CZK.

Conclusion

The lognormal distribution is one of the most frequently used in modeling income distributions. The calculated probability model provides important detailed information about the population and it is qualitatively very valuable result.

On the basis of the analysis we see that throughout the period increases in all regions the average net household income per capita. The highest average net income was over the whole period in the Capital Prague Region. Based on the prediction of future development in 2010 reached the highest average net income the Central Bohemian Region, while the lowest average net income has the Usti Region.

The results for individual regions confirm that significantly changes the charakter of the income distribution, there are increasing differences in a wage differentiation, and it is increasing the number of people with high incomes.

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.

References

Aitchison, J. and Brown, J.A.C. (1957). The Lognormal Distribution with Special Reference to Its Uses in Economics. Cambridge University Press, Cambridge.

Bartošová, J. (2006). Logarithmic-Normal Model of Income Distribution in the Czech Republic. Austrian Journal of Statistics, 35 (23), 215 − 222.

Bílková, D. (2008). Application of Lognormal Curves in Modeling of Wage Distributions. Journal of Applied Mathematics, 1 (2), 341 − 352.

Cohen, A.C. and Whitten, J.B. (1980). Estimation in the Three-parameter Lognormal Distribution. Journal of American Statistical Association, 75, 399 − 404.

Contact

Jana Langhamrová

University of Economics Prague, Department of Statistics and Probability

nám. W. Churchilla 4, Praha, Czech Republic

xlanj18@vse.cz

Diana Bílková

University of Economics Prague, Department of Statistics and Probability

nám. W. Churchilla 4, Praha, Czech Republic

bilkova@vse.cz