ISSN 2045-3345

ANALYSIS OF THE INCOME DISTRIBUTION IN CAPITAL PRAGUE REGION IN 2002-2009 AND PREDICTION FOR 2010

Jana LANGHAMROVÁ, Diana BÍLKOVÁ
University of Economics, Prague, Czech Republic

ABSTRACT

This paper compares 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). It was used for the income distribution in Capital Region Prague 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 in the Czech Republic. Using the predicted values of the parameters of considered lognormal distribution forecasts of income distributions were constructed for 2010 (interval frequency distribution).

JEL CLASSIFICATION & KEYWORDS

■ C13 ■ C16 ■ INCOME DISTRIBUTION ■ PRAGUE ■ CZECH REPUBLIC

INTRODUCTION

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, how we can assess 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.

Model selection

When we construct the model of income distribution, it is necessary to make a compromise 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 cannot 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 distributions in modeling of income distributions. Model parameters we 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 it is for us qualitatively very valuable result.

Three-parametric lognormal distribution

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

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

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

Random variable

$$Y = \ln(X - \theta) \quad (2)$$

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

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

has standard normal distribution N (0;1). Parameter μ is the expected value of random variable (2) and parameter σ^2 is the variance of the random variance. Parameter 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

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

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

$$F(x_P) = P, \quad (5)$$

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

$$x_P = \theta + e^{\mu + \sigma u_P}, \quad (6)$$

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

ISSN 2045-3345

$$\tilde{x} = \theta + e^{\mu}. \qquad (7)$$

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

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

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

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

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

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

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

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

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

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

and a coefficient of kurtosis of this random variable

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

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

For estimation of the parameters of three-parametric lognormal distribution was in this case used 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 $\overline{x}$ mean 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

$$\overline{x} = \tilde{\theta} + e^{\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}}, \qquad (14)$$

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

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

We obtain from equations (15) and (16)

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

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

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

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

$$\tilde{\theta} = \overline{x} - e^{\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}}. \qquad (20)$$

Data

Data in this paper 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 was a net annual household income per capita (in CZK).

Development in Capital Prague Region in 2002-2009

Due to using the lognormal distribution with three parameters and the method of moments were modelled following income distributions for Capital Region Prague in 2002, 2005-2009.

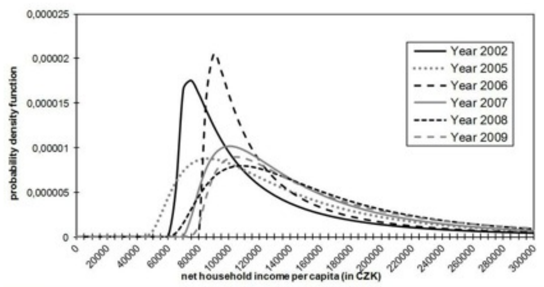
Table 1 contains some selected characteristics and the estimated values of the parameters of the three-parametric lognormal distribution for the Capital Prague Region. Arithmetic mean in this region is increasing over the entire period from the original value of 137,015 in 2002 up to 193,211 in 2009. The highest variability in net annual household income was recorded in the Capital Prague Region in 2002 and 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			
Capital Prague Region	Standard deviation	Variance	Coefficient of Variation	Skewness
2002	122,255	14	89.23 %	9,5
2005	96,804	9	64.78 %	3,5
2006	132,476	17	86.52 %	12,1
2007	95,659	9	58.98 %	4,1
2008	99,965	9	57.54 %	3,2
2009	146,187	21	75.66 %	5,7
Region	Parameter estimates			
Capital Prague Region	μ	σ 2	θ	Arithmetic mean
2002	10,546	1,322	63,326.97	137,015
2005	11,281	0,598	42,469.45	149,426
2006	10,423	1,497	81,994.11	153,111
2007	11,106	0,704	67,531.66	162,198
2008	11,398	0,547	56,551.52	173,725
2009	11,206	0,936	75,772.45	193,211
Source: data Microcensus, data SILC, own calculations				

In Figure 1 we can see the probability density functions for the Capital Prague Region in the period 2002-2009. Between 2002 and 2006, the probability density functions

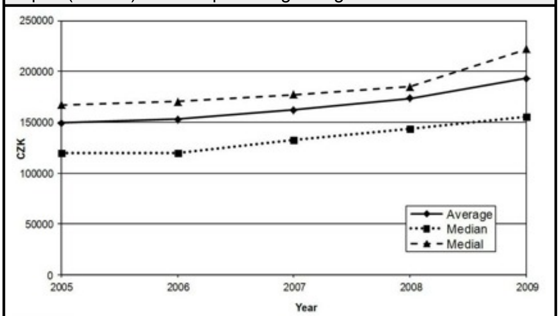
Figure 1: Probability density functions of net household income per capita in Capital Prague Region in 2002 - 2009



Source: data Microcensus, data SILC, own calculations

were more kurtosis than in other years. It can observe that the probability density function shifts to the right every year. That suggesting the fact that in this region is increasing number of people with higher net incomes. The highest modus has curve in 2006 and the second highest modus was in 2002. In these years was there greater representation of people with lower incomes.

Figure 2: Characteristics of the level of net household income per capita (in CZK) in the capital Prague region



Source: data Microcensus, data SILC, own calculations

In Figure 2 shows average values, median and medial of annual net income per capita in 2005-2009. The graph shows that all characteristics during the reporting period grew. There is clearly seen that medial reached during this period values from 167,000 in 2005 to 222,000 in 2009. The second highest value has arithmetic mean in 2005-2009. And median reached values 120,000 in 2005 to nearly 156,000 in 2009.

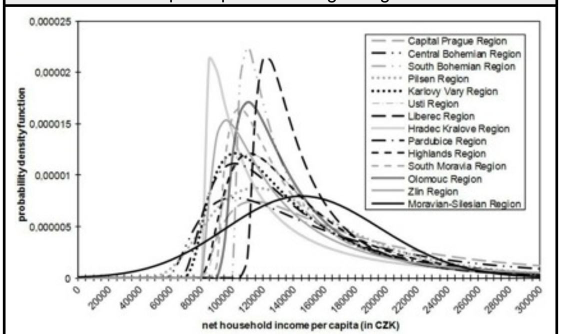
Prediction for 2010

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 3 shows forecasted basic characteristics and calculated parameters. 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 Kralove Region.

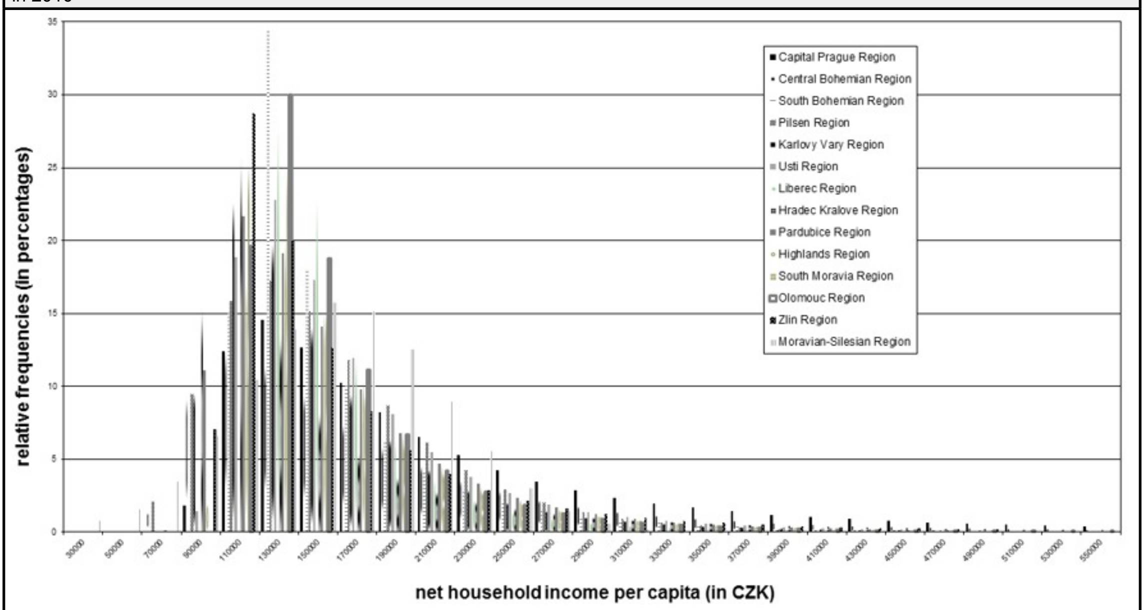
In Figure 3 we see the predicted probability density functions. In this figure Hradec Kralove Region, Liberec

Figure 3: 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 4: 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

Table 2: 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	215,81	165,206	27293022436		5,3	11,392	0,888	77,697.25
Central Bohemian Region	174,562	112,589	12676282921	76,55	3,6	11,391	0,624	53,629.65
South Bohemian Region	156,349	88,589	7848010921	56,66	8,8	10,297	1,262	100,666.80
Pilsen Region	149,107	59,932	3591844624	40,19	1,6	11,599	0,218	27,618.83
Karlovy Vary Region	143,164	62,759	3938692081	43,84	3	10,992	0,512	66,416.33
Usti Region	156,45	67,988	4622368144	43,46	3,3	10,975	0,569	78,834.43
Liberec Region	153,42	49,169	2417590561	32,05	4,5	10,339	0,773	107,895.00
Hradec Kralove Region	158,616	167,148	27938453904	105,38	16,1	10,412	1,714	80,245.44
Pardubice Region	146,469	68,374	4675003876	46,68	3	11,078	0,512	62,854.78
Highlands Region	153,44	64,41	4148648100	41,98	3,1	10,985	0,531	76,533.84
South Moravia Region	150,309	74,444	5541909136	49,53	5,3	10,595	0,888	88,073.58
Olomouc Region	147,091	57,239	3276303121	38,91	4,1	10,584	0,71	90,782.74
Zlin Region	152,037	95,853	9187797609	63,05	6,4	10,668	1,025	80,338.12
Moravian-Silesian Region	146,99	52,865	2794708225	35,97	0,1	14,277	0,001	-1

Source: data Microcensus, data SILC, own calculations

Table 3: Forecasts of percentage ratios in intervals of net household income per capita in individual regions in 2010 (in %)

Interval Lower limit	Upper limit	Central Prague Region	Central Bohemian Region	South Bohemian Region	Pilsen Region	Karlovy Vary Region	Usti Region	Liberec Region	Hradec Kralove Region	Pardubice Region	Highlands Region	South Moraia Region	Olomou Region	Zlin Region	Moravian-Slesian Region
0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0,75
30,001	50	0	0	0	0,03	0	0	0	0	0	0	0	0	0	1,54
50,001	70	0	1,63	0	2,12	0	0	0	0	0,1	0	0	0	0	3,49
70,001	90	1,81	11,38	0	9,45	9,83	1,41	0	17,44	11,12	2,13	0,06	0	7,03	6,6
90,001	110	12,4	15,38	15,18	15,84	23,43	18,84	0,11	29,17	21,66	19,7	26,17	19,63	28,7	10,44
110,001	130	14,58	14,2	34,45	17,23	20,53	22,79	35,03	15,47	19,15	22,8	25,83	30,03	19,97	13,91
130,001	150	12,66	11,7	17,86	15,13	14,55	17,3	28,59	9,34	14,07	17,2	15,86	18,8	12,65	15,72
150,001	170	10,27	9,26	10,04	11,85	9,8	11,92	14,89	6,16	9,82	11,78	9,79	11,1	8,28	15,16
170,001	190	8,2	7,24	6,17	8,69	6,56	8,07	8,05	4,32	6,79	7,91	6,28	6,72	5,64	12,55
190,001	210	6,55	5,65	4,04	6,15	4,43	5,5	4,62	3,17	4,72	5,34	4,19	4,23	3,97	8,96
210,001	230	5,26	4,42	2,78	4,26	3,03	3,8	2,8	2,4	3,31	3,65	2,89	2,75	2,88	5,55
230,001	250	4,25	3,49	1,98	2,92	2,1	2,67	1,77	1,87	2,35	2,53	2,05	1,84	2,14	3
250,001	270	3,47	2,76	1,46	1,99	1,48	1,9	1,16	1,48	1,69	1,78	1,49	1,27	1,62	1,42
270,001	290	2,86	2,21	1,1	1,35	1,06	1,38	0,79	1,2	1,23	1,28	1,11	0,89	1,25	0,59
290,001	310	2,37	1,77	0,85	0,92	0,77	1,01	0,55	0,98	0,91	0,93	0,84	0,64	0,98	0,22
310,001	330	1,98	1,44	0,66	0,63	0,56	0,75	0,39	0,82	0,68	0,68	0,64	0,47	0,78	0,07
330,001	350	1,67	1,17	0,53	0,43	0,42	0,57	0,28	0,69	0,51	0,51	0,5	0,35	0,63	0,02
350,001	370	1,41	0,96	0,43	0,3	0,31	0,43	0,21	0,58	0,39	0,38	0,39	0,26	0,51	0,01
370,001	390	1,2	0,79	0,35	0,21	0,24	0,33	0,16	0,5	0,3	0,29	0,31	0,2	0,42	0
390,001	410	1,03	0,66	0,29	0,14	0,18	0,26	0,12	0,43	0,23	0,22	0,25	0,16	0,35	0
410,001	430	0,89	0,55	0,24	0,1	0,14	0,2	0,09	0,37	0,18	0,17	0,2	0,12	0,29	0
430,001	450	0,77	0,46	0,2	0,07	0,11	0,16	0,07	0,32	0,14	0,13	0,17	0,1	0,24	0
450,001	470	0,67	0,39	0,17	0,05	0,09	0,13	0,06	0,28	0,11	0,11	0,14	0,08	0,21	0
470,001	490	0,58	0,33	0,14	0,04	0,07	0,1	0,04	0,25	0,09	0,08	0,11	0,06	0,17	0
490,001	510	0,51	0,28	0,12	0,03	0,05	0,08	0,04	0,22	0,07	0,07	0,1	0,05	0,15	0
510,001	530	0,45	0,24	0,11	0,02	0,04	0,07	0,03	0,2	0,06	0,05	0,08	0,04	0,13	0
530,001	550	0,39	0,2	0,09	0,01	0,04	0,05	0,02	0,18	0,05	0,04	0,07	0,03	0,11	0
550,001	and more	3,77	1,44	0,76	0,04	0,18	0,28	0,13	2,16	0,27	0,24	0,48	0,18	0,9	0
Total		100	100	100	100	100	100	100	100	100	100	100	100	100	100

Source: data Microcensus, data SILC, own calculations

Region and South Bohemian Region are much more kurtosis than others predicted distributions. This means that in these regions should be more people with lower incomes.

The Figure 4 shows a histogram of predicted relative frequencies of net household income per capita by region. In Figure 4, for example, can be read that 35 % of people in the Liberec Region would reach a net household income from 130,000 – 150,000 CZK. And we can also notice that Capital Prague Region has from value 270,000 to the end highest frequencies of all regions. This region is represented by a large number of people with higher incomes.

Conclusion

The lognormal distribution, which we used in this paper, is one of the most frequently used in modelling income distributions. Qualitatively very valuable result has for us the calculated probability model, which provides important detailed information about the population.

On the basis of the analysis is evident that throughout the period increases in Capital Prague Region the average net household income per capita. We can observe that the probability density function in Central Prague Region shifts to the right every year. It suggesting the fact that in this region is increasing number of people with higher net incomes.

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. From histogram with predicted values we can also see that Capital Prague Region has from value 270,000 to the end highest frequencies of all regions.

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

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

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

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