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 $LN(\mu, \sigma^2, \theta)$ with parameters $\mu$, $\sigma^2$ and $\theta$, 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

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

$= 0$, otherwise.

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

has normal distribution $N(\mu, \sigma^2)$ and random variable

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

has standard normal distribution $N(0; 1)$. Parameter $\mu$ is the expected value of random variable $(2)$ and parameter $\sigma^2$ is the variance of the random variance. Parameter $\theta$ represents the theoretical minimum of the random variable $X$. The income distribution is possible that the value of the parameter $\theta$ 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.
Another characteristic of the level of the random variable $X$ having three parametric lognormal distributions is a mode of the random variable

$$
\hat{X} = \mu + e^{\sigma^2/2}.
$$

(7)

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

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

(9)

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

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

(12)

and a coefficient of kurtosis of this random variable

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

(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 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 contains some selected characteristics and the estimated values of the parameters of the three-parametric lognormal distribution for the Capital Prague Region.

### Data

Data in this paper were obtained from a survey of the Czech Statistical Office Microcenzens (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.

<table>
<thead>
<tr>
<th>Region</th>
<th>Sample characteristics</th>
<th>Parameter estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital Prague Region</td>
<td>Standard deviation</td>
<td>Variance Coefficient of Variation</td>
</tr>
<tr>
<td>2002</td>
<td>122,255</td>
<td>14</td>
</tr>
<tr>
<td>2005</td>
<td>96,804</td>
<td>9</td>
</tr>
<tr>
<td>2006</td>
<td>132,476</td>
<td>17</td>
</tr>
<tr>
<td>2007</td>
<td>95,659</td>
<td>9</td>
</tr>
<tr>
<td>2008</td>
<td>99,965</td>
<td>9</td>
</tr>
<tr>
<td>2009</td>
<td>146,187</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Capital Prague Region</th>
<th>Parameter estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>10,546</td>
</tr>
<tr>
<td>2005</td>
<td>11,281</td>
</tr>
<tr>
<td>2006</td>
<td>10,423</td>
</tr>
<tr>
<td>2007</td>
<td>11,106</td>
</tr>
<tr>
<td>2008</td>
<td>11,398</td>
</tr>
<tr>
<td>2009</td>
<td>11,206</td>
</tr>
</tbody>
</table>

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...
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.

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 1: Probability density functions of net household income per capita in Capital Prague Region in 2002 - 2009](source)

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

![Figure 3: Forecasts of probability density functions of net annual household income per capita according to region in 2010](source)

![Figure 4: Forecasts of histogram of relative frequencies (in percentages) of net annual household income per capita according to region in 2010](source)
Table 2: Forecasts of percentage ratios in intervals of net household income per capita in individual regions in 2010 (in %)

<table>
<thead>
<tr>
<th>Lower</th>
<th>Upper</th>
<th>30,001</th>
<th>50,001</th>
<th>70,001</th>
<th>110,001</th>
<th>130,001</th>
<th>150,001</th>
<th>210,001</th>
<th>330,001</th>
<th>510,001</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>70</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Interval</th>
<th>Lower</th>
<th>Upper</th>
<th>Capital Prague Region</th>
<th>Central Bohemian Region</th>
<th>South Bohemian Region</th>
<th>Pilsen Region</th>
<th>Karlovy Vary Region</th>
<th>Usti Region</th>
<th>Liberec Region</th>
<th>Hradec Kralove Region</th>
<th>Pardubice Region</th>
<th>Liberec Region</th>
<th>Liberec Region</th>
<th>Pardubice Region</th>
<th>Masaryk-Silesian Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.81</td>
<td>0.75</td>
<td>0.75</td>
<td>1.54</td>
<td>-1</td>
</tr>
<tr>
<td>50,001</td>
<td>50</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>70,001</td>
<td>70</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

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**


