ISSN 2045-3345

RESEARCH JOURNAL OF ECONOMICS, BUSINESS AND ICT

ESTIMATING PARAMETERS OF LOGNORMAL DISTRIBUTION USING THE METHOD OF L-MOMENTS

Diana BÍLKOVÁ
University of Economics, Prague, Czech Republic

ABSTRACT

Commonly used statistical procedure to describe the observed statistical sets is to use their conventional moments or cumulants. An alternative approach is based on the use of other characteristics, which we call L-moments. L-moments are analogous to conventional moments, but they are based on linear combinations of order statistics, i.e., L-statistics. Using L-moments is theoretically preferable to the conventional moments and consists in the fact that L-moments characterize a wider range of distribution. When estimating from sample L-moments, L-moments are more robust to the presence of outliers in the data. Experience also shows that, compared to conventional moments, L-moments are less prone to bias of estimation. Parameter estimates obtained using L-moments are mainly in the case of small samples often even more accurate than estimates of parameters made by maximum likelihood method. This paper deals with the use of L-moments in the case of large data sets of income distribution (individual data) and wage distribution (data are ordered to the form of interval frequency distribution of extreme open intervals). The data for this research concern the Czech Republic and has been obtained from the Czech Statistical Office. Three-parametric lognormal curves were used as the model in all cases.

JEL CLASSIFICATION & KEYWORDS

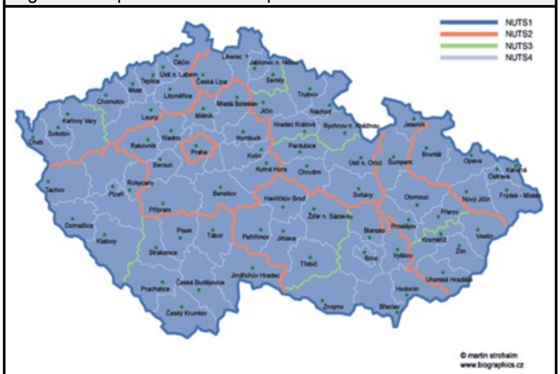
■ C13 ■ C16 ■ LOGNORMAL DISTRIBUTION ■ PARAMETERS ■ ESTIMATION ■ L-MOMENTS

INTRODUCTION

The applicability of the estimates of income and wage distribution is that it provides the possibility of linking the considerations relating to income and wage differentiation with socio-political considerations, in which it is not mostly enough to estimate development of the average income and wage, but it is necessary to estimate the proportions of workers with low, middle and high incomes and wages or it is necessary to estimate the proportions of workers in all income or wage groups. Knowledge of models of income and wage distribution is also used for example in assessing the population\ s living standards or at inter-area and international comparisons of living standards. In the field of statistics, we see many more using the knowledge of the income and wage distribution.

From the statistical literature it is well-known use of L-moments in connection with the data from the field of hydrology and meteorology (for example rainfall). In such cases, there are generally relatively small data sets. This paper deals with the use of L-moments in the case of large data sets. There are the data of two types, namely, individual data on net annual household income per capita in CZK (years 1992, 1996 and 2002 – statistical survey Microcensus and years 2005, 2006, 2007 and 2008 – statistical survey SILC), and second, data sorted into a form of interval frequency distribution, these data refer to gross monthly wage in CZK (from official web page of the Czech Statistical Office). In both cases we compare the accuracy of the method of L-moments with an accuracy of other methods of parameter estimation. Income data come from the statistical surveys SILC and Microcensus of the Czech Statistical Office, while the wage data come from official website of the Czech Statistical Office. Three-parametric lognormal distribution was used as the basic parametric distribution. Accuracy of the method of L-moments was compared with the accuracy of other methods of parameter estimation, such as moment method, quantile method, maximum likelihood method. Figure 1 is a map of the Czech Republic.

Figure 1: Map of the Czech Republic



Source: www.biographics.cz

Methods

Three-Parametric Lognormal Distribution

The essence of lognormal distribution is treated in detail form example in Aitchison and Brown (1957). Use of lognormal distribution in connection with wage or income distributions is described in Bartošová (2006) or Bílková (2008).

Random variable X has three-parametric lognormal distribution $LN(\mu,\sigma^2,\theta)$ 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

$$f(x; \mu,\sigma^2,\theta) = \frac{1}{\sigma\,(x-\theta)\sqrt{2\pi}}\, e^{-\frac{[\ln(x-\theta)-\mu]^2}{2\sigma^2}}, \quad x > \theta,$$
$$= 0, \text{ jinak.} \qquad (1)$$

Random variable

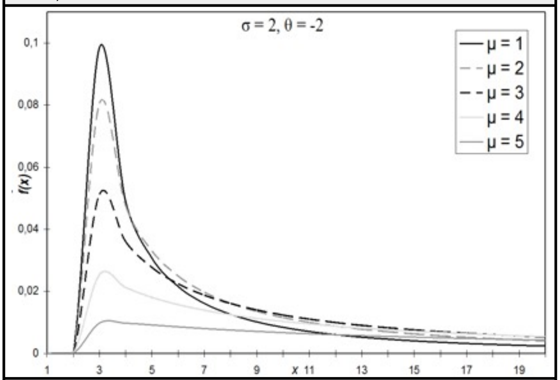
$$Y = \ln(X - \theta) \qquad (2)$$

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

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

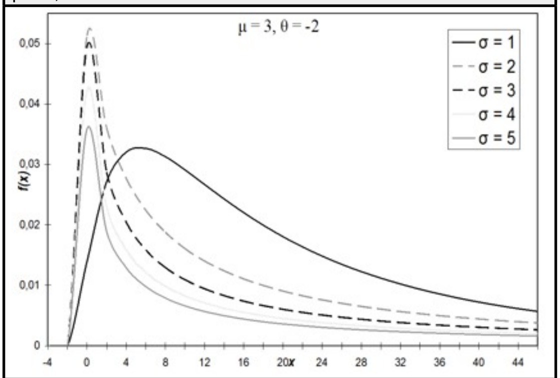
has a standardized normal distribution N(0, 1). The parameter μ is the expected value of random variable (2) and the parameter σ^2 is the variance of this random variable. Parameter θ is the theoretical minimum of random variable X. Figures 2 and 3 represent the probability density functions of three-parametric lognormal curves depending on the values of their parameters.

Figure 2: Probability density function for the values of parameters σ = 2, θ = −2



Source: Own research

Figure 3: Probability density function for the values of parameters μ = 3, θ = −2



Source: Own research

The expected value (4) is the basic moment location characteristic of a random variable X having three-parametric lognormal distribution

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

100 P% quantile is the basic quantile location characteristic of a random variable X

$$x_P = \theta + e^{\mu + \sigma u_P}, \quad (5)$$

where $0 < P < 1$ and u_P is 100 P% quantile of standardized normal distribution. Substituting into the relation (5) P = 0.5, we get 50% quantile of three-parametric lognormal distribution, which is called median

$$\tilde{x} = \theta + e^{\mu}. \quad (6)$$

The Median (6) divides the range of values of random variable X on the two equally likely parts. The mode (7) of random variable X is another often used location characteristic of three-parametric lognormal distribution

$$\hat{x} = \theta + e^{\mu - \sigma^2}. \quad (7)$$

The variance (8) of random variable X is a basic variability characteristic of three-parametric lognormal distribution

$$D(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1). \quad (8)$$

Standard deviation (9) is the square root of the variance and it represents another moment variability characteristic of the considered theoretical distribution

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

The coefficient of variation (10) is characteristic of the relative variability of this distribution and we get it by dividing the standard deviation to the expected value of the distribution

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

It is a dimensionless characteristic of variability.

The coefficient of skewness (11) and the coefficient of kurtosis (12) belong to basic moment shape characteristic of the distribution

$$\beta_1(X) = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}, \quad (11)$$

$$\beta_2(X) = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3. \quad (12)$$

Methods of Point Parameter Estimation

Question of parameter estimation of three-parametric lognormal distribution is already well developed in statistical literature, see for example Cohen and Whitten (1980). We can use various methods to estimate the parameters of three-parametric lognormal distribution. We give as an example: moment method, quantile method, maximum likelihood method, method of L-moments, Kemsley's method, Cohen's method or graphical method.

Moment Method

The essence of moment method of parameter estimation lies in the fact that we put the sample moments and the corresponding theoretical moments into equation. We can combine the general and the central moments. This method of estimating parameters is indeed very easy to use, but is very inaccurate. In particular, the estimate of theoretical variance by its sample counterpart is very inaccurate. In the case of wage and income distribution, however we work with large sample sizes, and therefore the use of moment method of parameter estimation may not be a hindrance in terms of efficiency of estimators.

In the case of moment method of parameter estimation of three-parametric lognormal distribution we put the sample arithmetic mean $\overline{x}$ equal to the expected value of random variable X and we put the sample second central moment equal to the variance of random variable X. Furthermore, we put equal the sample third central moment m_3 with a theoretical third central moment of random variable X and we get a third equation. We obtain a system of moment equations

$$\overline{x} = \tilde{\theta} + e^{\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}}, \quad (13)$$

$$m_2 = e^{2\tilde{\mu} + \tilde{\sigma}^2}(e^{\tilde{\sigma}^2} - 1), \quad (14)$$

$$m_3 = e^{3\tilde{\mu} + \frac{3}{2}\tilde{\sigma}^2}(e^{\tilde{\sigma}^2} - 1)^2(e^{\tilde{\sigma}^2} + 2). \quad (15)$$

We obtain from equations (14) and (15)

$$b_1^2 = m_3^2 \cdot m_2^{-3} = (e^{\tilde{\sigma}^2} - 1)(e^{\tilde{\sigma}^2} + 2)^2, \quad (16)$$

and therefore we also gain the moment parameter estimation of three-parametric lognormal distribution from the system of equations (13) to (15)

$$\tilde{\sigma}^2 = \ln\left[\sqrt[3]{1 + \frac{1}{2}b_1^2 + \sqrt{\left(1 + \frac{1}{2}b_1^2\right)^2 - 1}} + \sqrt[3]{1 + \frac{1}{2}b_1^2 - \sqrt{\left(1 + \frac{1}{2}b_1^2\right)^2 - 1}} - 1\right], \quad (17)$$

$$\tilde{\mu} = \frac{1}{2} ln \frac{m_2}{e^{\tilde{\sigma}^2}(e^{\tilde{\sigma}^2}-1)}, \quad (18)$$

$$\tilde{\theta} = \overline{x} - e^{\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}}. \quad (19)$$

Quantile method and Kemsley's method

Quantile method of parameter estimation of three-parametric lognormal distribution is based on the use of three sample quantiles, namely there are 100 · P1% quantile, 100 · P2% quantile and 100 · P3% quantile, where P2 = 0,5 and P3 = 1 – P1, and thus

$$u_{P_2} = 0 \text{ a } u_{P3} = -u_{P1}.$$

We create a system of quantile equations by substituting to (5)

$$x^V_{P1} = \theta^* + e^{\mu^* + \sigma^* u_{P1}}, \quad (20)$$

$$x^V_{0,5} = \theta^* + e^{\mu^*}, \quad (21)$$

$$x^V_{(1-P1)} = \theta^* + e^{\mu^* - \sigma^* u_{P1}}, \quad (22)$$

where x^V_{P1}, $x^V_{0,5}$ a $x^V_{(1-P1)}$ are the corresponding sample quantiles. We obtain quantile parameter estimations of three-parametric lognormal distribution from the system of quantile equations (20) to (22).

$$\sigma^{2*} = \left[\frac{\ln \dfrac{x^V_{P1} - x^V_{0,5}}{x^V_{0,5} - x^V_{(1-P1)}}}{u_{P1}} \right]^2, \quad (23)$$

$$\mu^* = \ln \frac{x^V_{P1} - x^V_{(1-P1)}}{e^{\sigma^* u_{P1}} - e^{-\sigma^* u_{P1}}}, \quad (24)$$

$$\theta^* = x^V_{0,5} - e^{\mu^*}. \quad (25)$$

The sample median can be replaced by the sample arithmetic mean. Then we solve a similar system of equations as in the case of quantile method. This method is called Kemsley's method.

Maximum likelihood method and Cohen's method

If the value of the parameter θ in known, the likelihood function is maximized when the likelihood parameter estimations of three-parametric lognormal distribution have the form

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \ln(x_i - \theta)}{n}, \quad (26)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} [\ln(x_i - \theta) - \hat{\mu}]^2}{n}. \quad (27)$$

If the value of parameter θ is not known, this problem is considerably more complicated. If the parameter θ is estimated based on its sample minimum

$$\hat{\theta} = x^V_{\min}, \quad (28)$$

the likelihood function is unlimited. Maximum likelihood method is therefore sometimes combined with the Cohen's method. In this procedure, we put the smallest sample value to equality with 100 · (n + 1)$^{-1}$ % quantile

$$x^V_{\min} = \hat{\theta} + e^{\hat{\mu} + \hat{\sigma} u(n+1)^{-1}}. \quad (29)$$

Equation (29) is then combined with a system of equations (26) and (27).

Method of L-Moments

Question of L-moment is described in detail for example in Hosking and Wales (1997). We will assume that X is a real random variable with the distribution function F(x) and quantile function x(F) and $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$ are the order statistics of the random sample of the size n selected from the distribution X. Then the r-th L-moment of the random variable X is defined as

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad r = 1, 2, 3, \ldots. \quad (30)$$

The letter “L” in the name “L-moments” is to stress the fact that r-th L-moment λr is a linear function of the expected order statistics. Natural estimate of the L-moment λr based on the observed sample is furthermore a linear combination of the ordered values, i.e. the so called L-statistics. The expected value of the order statistic is of the form

$$EX_{j:r} = \frac{r!}{(j-1)! \cdot (r-j)!} \cdot \int x[F(x)]^{j-1} \cdot [1-F(x)]^{r-j} \, dF(x). \quad (31)$$

If we plough the equation (31) in the equation (30), we get after some operations

$$\lambda_r = \int_0^1 x(F) ^* P_{r-1}(F) \, dF, \quad r = 1, 2, 3, \ldots, \quad (32)$$

where

$$P^*_r(F) = \sum_{k=0}^{r} p^*_{r,k} F^k \quad (33)$$

and

$$p^*_{r,k} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}, \quad (34)$$

w h e r e $P^*_r(F)$ represents r-th shifted Legender's polynom which is related to the usual Legender’s polynoms. Shifted Legender's polynoms are orthogonal on the interval (0,1) with a constant weight function. The first four L-moments are of the form

$$\lambda_1 = EX = \int_0^1 x(F) \, dF, \quad (35)$$

$$\lambda_2 = \frac{1}{2} E(X_{2:2} - X_{1:2}) = \int_0^1 x(F) \cdot (2F - 1) \, dF, \quad (36)$$

$$\lambda_3 = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) = \int_0^1 x(F) \cdot (6F^2 - 6F + 1) \, dF, \quad (37)$$

$$\lambda_4 = \frac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) = \int_0^1 x(F) \cdot (20F^3 - 30F^2 + 12F - 1) \, dF. \quad (38)$$

Details about the L-moments can be found in Guttman (1993) or Hosking (1990). The coefficients of the L-moments are defined as

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, 5, \ldots. \quad (39)$$

L-moments λ_1, λ_2, λ_3, …, λ_r and coefficients L-moments τ_1, τ_2, τ_3, …, τ_r can be used as the characteristics of the distribution. L-moments are in a way similar to the conventional central moments and coefficients of L-moments are similar to the moment ratios. Especially L-moments λ_1 and λ_2 and coefficients of the L-moments τ_3 and τ_4 are considered to be characteristics of the location, variability and skewness.

Using the equations (35) to (37) and the equation (39), we get the first three L-moments of the three parametric lognormal distribution LN(μ, σ^2, θ), which is described e.g. in Hosking (1990). The following relations are valid for these L-moments

$$\lambda_1 = \theta + \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad (40)$$

$$\lambda_2 = \exp\left(\mu + \frac{\sigma^2}{2}\right) \cdot \operatorname{erf}\left(\frac{\sigma}{2}\right), \tag{41}$$

$$\tau_3 = \frac{6\pi^{-1/2}}{\operatorname{erf}\left(\frac{\sigma}{2}\right)} \cdot \int_0^{\sigma/2} \operatorname{erf}\left(\frac{x}{\sqrt{3}}\right) \cdot \exp(-x^2)\, dx, \tag{42}$$

where erf(z) is the so called error function defined as

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \cdot \int_0^z e^{-t^2}\, dt. \tag{43}$$

Now we will assume that $x_1, x_2, \ldots, x_n$ is a random sample and $x_{1:n} \le x_{2:n} \le \ldots \le x_{n:n}$ is the ordered sample. The r-th sample L-moment is defined as

$$l_r = \binom{n}{r}^{-1} \cdot \sum_{1 \le i_1 \le i_2 \le \ldots \le i_r \le n} \sum \ldots \sum r^{-1} \cdot \sum_{k=0}^{r-1} (-1)^k \cdot \binom{r-1}{k} \cdot x_{i_{r-k}:n}, \quad r = 1, 2, \ldots, n. \tag{44}$$

We can write specifically for the first four sample L-moments

$$l_1 = n^{-1} \sum_i x_i, \tag{45}$$

$$l_2 = \frac{1}{2}\binom{n}{2}^{-1} \cdot \sum_{i>j} \sum (x_{i:n} - x_{j:n}), \tag{46}$$

$$l_3 = \frac{1}{3}\binom{n}{3}^{-1} \cdot \sum_{i>j>k} \sum \sum (x_{i:n} - 2x_{j:n} + x_{k:n}), \tag{47}$$

$$l_4 = \frac{1}{4}\binom{n}{4}^{-1} \cdot \sum_{i>j>k>l} \sum \sum \sum (x_{i:n} - 3x_{j:n} + 3x_{k:n} - x_{l:n}). \tag{48}$$

Sample L-moments can be used similarly as the conventional sample L-moments because they characterize basic properties of the sample distribution and estimates the corresponding properties of the distribution from which were the data sampled. They might be also used to estimate the parameters of this distribution. In these cases, L-moments are often used instead of the conventional moments because as linear functions of the data are less sensitive on the sample variability or the error size in the case of the presence of the extreme values in the data than the conventional moments. Therefore it is assumed that the L-moments provide more precise and robust estimates of the characteristics or parameters of the population probability distribution.

Let us denote the distribution function of the standard normal distribution as Φ, then Φ^{-1} represents the quantile function of the standard normal distribution. The following relation holds for the distribution function of the three parametric lognormal distribution LN(μ, σ^2, θ)

$$F = \Phi\left[\frac{\ln(x-\theta) - \mu}{\sigma}\right]. \tag{49}$$

The coefficients of the L-moments (39) are then commonly estimated using the following estimates

$$t_r = \frac{l_r}{l_2}, \quad r = 3, 4, 5, \ldots. \tag{50}$$

The estimates of the three parametric lognormal distribution can then be calculated as

$$z = \sqrt{\frac{8}{3}} \cdot \Phi^{-1}\left(\frac{1+t_3}{2}\right), \tag{51}$$

$$\hat{\sigma} \approx 0{,}999\ 281\, z - 0{,}006\ 118\, z^3 + 0{,}000\ 127\, z^5, \tag{52}$$

$$\hat{\mu} = \ln\left[\frac{l_2}{erf\left(\frac{\hat{\sigma}}{2}\right)}\right] - \frac{\hat{\sigma}^2}{2}, \tag{53}$$

$$\hat{\theta} = l_1 - exp\left(\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right). \tag{54}$$

More on L-moments is for example in Kyselý and Picek (2007), Smithers and Schulze (2001) or Ulrych, Velis, Woodbury and Sacchi (2000).

Appropriateness of the Model

In the next stage, we have to check the suitability of the model or we have to choose one model from several alternatives. Therefore we have to construct a criterion. This criterion can be for example sum of absolute differences of observed and theoretical frequencies of all intervals

$$S = \sum_{i=1}^{k} |n_i - n\pi_i| \tag{55}$$

or the popular χ^2 criterion

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n\pi_i)^2}{n\pi_i}, \tag{56}$$

where ni is the observed frequency in i-th interval, πi is theoretical probability that a particular observation falls in to i-th interval, n is the total size of the dataset, n · πi is the theoretical frequency in i-th interval, i = 1, 2, ..., k, and k is the number of intervals.

The question of suitability of a particular distribution is not a common statistical problem of testing the null hypothesis "H_0: The sample comes from the assumed distribution" versus the alternative hypothesis "H_1: non H_0", because in the case of using the goodness of fit test for income and wage distributions, we often encounter the „large n" problem, i.e. when we work with large datasets, the test tends to reject almost every null hypothesis. There are two reasons for this. First the power of this test is for large samples (for a given confidence level) too high and it takes in to account even the smallest differences of the real wage distribution and the model. The other reason is the principle of the test construction it self. Since small differences are out of our scope, the approximate fit of the curve is sufficient – we „just borrow the model (the curve)". In these cases, the use of the χ^2 criterion is rather limited. Therefore the interpretation of the results can be more or less arbitrary and we have to take the advantage of the experience and logical analysis. Other details of the theory of lognormal distribution can be found for example in Bílková (2008).

Outputs

Table 1 contains calculated values of sample L-moments, the estimated parameters of the three-parametric lognormal distribution obtained using the L-moment method and the sum of absolute deviations of the observed and theoretical frequencies that the model assumes S. Table 1 refers to the distribution of the net annual household income per capita. Table 2 presents the same for the distribution of gross monthly wage. For comparison, Table 3 contains the estimated parameters of the three-parametric lognormal distribution, which were acquired by moment method of parameter estimation and the sum of all deviations of the observed and theoretical frequencies for all intervals S, both for the distribution of the net annual household income per capita and for the distribution of the gross monthly wage. The moment method of parameter estimation is described in detail for example in Bílková (2008).

We can see from Table 3 that the value of the parameter θ (beginning of the distribution) can be negative. This means that the initially course of this curve gets into negative territory. This does not interfere with a good agreement of the model with the actual distribution due to the fact that the curve is initially very close contact with the horizontal axis. Parameter θ cannot give any interpretation for its negative values. It should be noted here that the purpose of this study

Table 1: Sample L-moments and estimated parameters of the lognormal distribution using the method of L-moments − distribution of net annual household income per capita

Year	Sample L-moments l_1	l_2	l_3	Estimated parameters μ	σ^2	θ	S
1992	35,246.51	7,874.26	2,622.14	9,696	0,49	14,491.687	1,904.506
1996	66,121.92	16,237.54	5,685.45	10,343	0,545	25,362.753	1,616.537
2002	105,029.89	27,978.40	10,229.62	10,819	0,598	37,685.637	625,662
2005	111,023.71	28,340.18	9,113.57	11,028	0,455	33,738.911	570,824
2006	114,945.08	28,800.68	9,286.18	11,04	0,458	36,606.903	1,336.021
2007	123,806.49	30,126.11	9,530.57	11,112	0,44	40,327.610	2,333.984
2008	132,877.19	31,078.96	9,702.45	11,163	0,428	45,634.578	2,639.240

Source: Own research

Table 2: Sample L-moments and estimated parameters of the lognormal distribution using the method of L-moments − distribution of gross monthly wage

Year	Sample L-moments l_1	l_2	l_3	Estimated parameters μ	σ^2	θ	S
2002	17,437.49	4,251.48	1,267.44	9,238	0,388	4,952.259	134,844
2003	18,663.18	4,524.95	1,251.90	9,402	0,332	4,364.869	135,841
2004	19,697.57	5,001.34	1,586.09	9,313	0,442	5,872.138	252,002
2005	20,738.14	5,262.93	1,636.67	9,392	0,424	5,908.390	260,423
2006	21,803.28	5,454.74	1,738.23	9,393	0,447	6,795.207	277,559
2007	23,882.83	6,577.65	2,627.93	9,222	0,724	9,349.280	429,282
2008	25,477.59	6,993.72	2,737.94	9,319	0,693	9,719.297	455,574

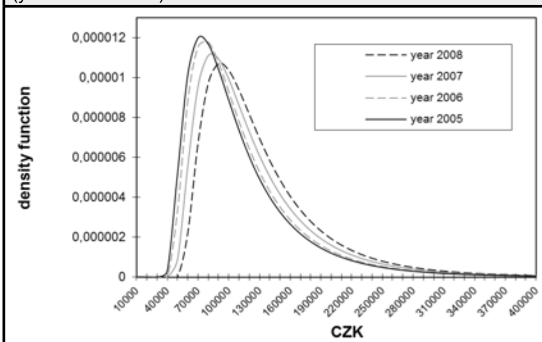
Source: Own research

Table 3: Estimated parameters of the lognormal distribution using the moment method − distribution of net annual household income per capita and distribution of gross monthly wage

Year	Income μ	σ^2	θ	S	Year	Wage μ	σ^2	θ	S
1992	8,883	1,083	22,284.335	2,985	2002	9,492	0,264	2,311.688	114,691
1996	9,154	1,334	45,269.967	4,161	2003	9,698	0,155	2,993.514	157,301
2002	9,668	1,327	66,925.879	2,418	2004	9,779	0,221	-25,695	226,646
2005	9,71	1,287	73,299.950	1,478	2005	9,906	0,193	-1,339.601	225,479
2006	9,976	1,177	71,936.249	2,281	2006	9,979	0,18	-1,805.527	248,955
2007	10,242	1,079	73,575.417	2,736	2007	9,734	0,377	3,509.924	332,148
2008	10,328	1,044	80,180.795	2,848	2008	9,851	0,345	2,920.381	341,796

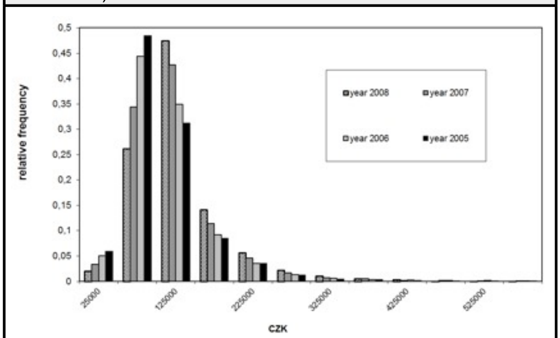
Source: Own research

Figure 6: Probability density function of the net income per capita (years 2005-2008)



Source: Own research

Figure 7: Frequency histogram of the net income per capita (years 2005-2008)



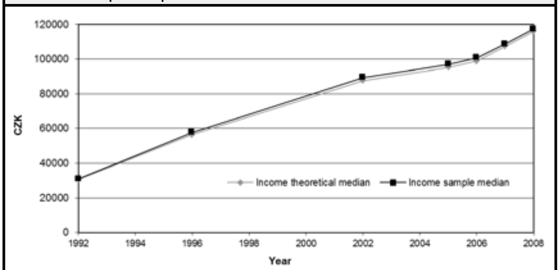
Source: Own research

is not to compare these two files with each other, but the purpose is to investigate the accuracy of parameter estimation for different types of data in terms of their arrangement within the meaning of individual data and data organized to the form of frequency distribution. Another purpose of this study is to compare the accuracy of different methods of parameter estimation with the accuracy of the L-moment method.

Figure 4 represents the development of the sample and theoretical median of the three-parametric lognormal distribution with parameters estimated using the L-moment method for the distribution of the net annual household income per capita and Figure 5 represents the same for the distribution of gross monthly wage. Figure 6 contains the development of probability density function (in the years 2005-2008) of the theoretical three-parametric lognormal distribution with the parameters estimated using the L-moment method for the distribution of the net household income per capita and Figure 7 presents the corresponding sample interval frequency distribution.

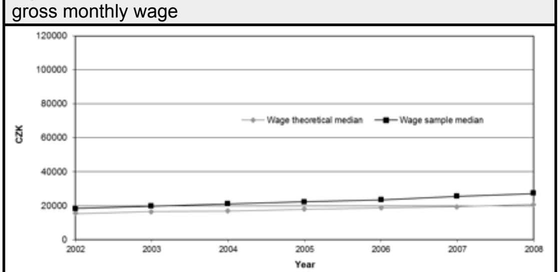
The values of well known test criterion χ^2 were also calculated, but due to the fact that in these large sample sizes, such as in the case of income and wage distribution are seen, the test power is too high that test uncovers the all very slight deviations between the sample and theoretical distribution. This test results to the rejection of the tested hypothesis about the expected theoretical distribution practically in all cases. However, we are not interested in such small deviations and approximate agreement between model and reality is sufficient. For this reason, we do not give the values of the test criterion χ^2.

Figure 4: Development of theoretical and sample median of the net income per capita



Source: Own research

Figure 5: Development of theoretical and sample median of the gross monthly wage



Source: Own research

Conclusion

We can see from Tables 1 − 3 that the values of S are considerably higher in the case of data set arranged to the form of interval frequency distribution (distribution of gross monthly wage) than in the case of individual data set (distribution of net annual household income per capita), which was expected. We can also see that the values S result essentially higher in the case of moment method of parameter estimation than in the case of L-moment method − both regarding to the set of individual data. But we cannot say the same thing in terms of data into a form of interval frequency distribution, where the value S results comparable in the case of both data sets. If we compare the accuracy of the method of L-moments with an accuracy of other methods of parameter estimation (quantile method and even the maximum likelihood method), we come to similar conclusions as to the accuracy of this method compared with the accuracy of moment method.

The L-moment method of parameter estimation gives more accurate results than other methods of parameter estimation (moment method, quantile method, maximum likelihood method) for individual data. In the case of data grouped into the form of interval frequency distribution, all four methods of parameter estimation offer comparable results. In these cases, the inaccuracies arise above all at both tails of the distribution (heavy tails). All Figures 1 − 4 are related to the L-moment method of parameter estimation and they also give an idea about the accuracy of this method.

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.

Guttman, N.B. The Use of L-moments in the Determination of Regional Precipitation Climates. Journal of Climate, 6, pp. 2309 − 2325, 1993.

Hosking, J.R.M. L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. Journal of the Royal Statistical Society (Series B), 52 (1), pp. 105 – 124, 1990.

Hosking, J.R.M., and J.R. Wales. Regional Frequency Analysis: An Approach Based on L-moments. New York: Cambridge University Press, 1957.

Kyselý, Jan, and Jan Picek. Regional Growth Curves and Improved design Value Estimates of Extreme Precipitation Events in the Czech Republic. Climate Research, 33, pp. 243 – 255, 2007.

Smithers, J.C., and R.E. Schulze. A Methodology for the Estimation of Short Duration Design Storms in South Africa Using a Regional Approach Based on L-moments. J. Hydrol., 241, pp. 42 – 52, 2001.

Ulrych, T.J., D.R. Velis, A.D. Woodbury, and M.D. Sacchi. L-moments and C-moments. Stoch. Environ. Res. Risk Asses, 14, pp. 50 − 68, 2000.