VOLUME 4, 2011

ISSN 2045-3345

ANALYSIS OF INCOMES IN THE CZECH REPUBLIC USING THE LOGNORMAL DISTRIBUTION

Jakub NEDVĚD, Ivana MALÁ
University of Economics, Prague, Czech Republic

ABSTRACT

In the paper the three parametric lognormal distribution is used as a model of hourly wage distribution in the Czech Republic. Data from the Average Earnings Information System from 2000-2010 are analysed. Random samples from the second quarter of 2009 survey are used and separate models are constructed for business and non-business spheres of economy. Three methods of estimation were used to estimate unknown parameters: maximum likelihood method, moment method and quantile method. Quality of fits was compared and the impact of the method of estimation on quality of fit was quantified. Characteristics of location and variability were estimated and their values were compared with sample values. The results of Average Earnings Information System have been published since 2000. In the text lognormal model was fit with the use of selected quantiles from 2000 and 2010.

JEL CLASSIFICATION & KEYWORDS

■ C13 ■ C16 ■ INCOME ANALYSIS ■ LOGNORMAL DISTRIBUTION ■ CZECH REPUBLIC

INTRODUCTION

Research into incomes and wages is very important especially in economics and politics, but income, wages and related problems are very interesting topic even for general public. Knowledge of income and its distribution is very important in effort to quantify income inequality in society or quality of life. In this case it adds objective point of view to largely subjective topic. Distributions of incomes of different type (and their characteristics of level and variability) can be used in order to make international or interregional comparisons or to compare incomes in subgroups of population defined according to various socio-economical or demographic characteristics as gender, education, location, age or job. The predictions of future development can be based on such analyses. Moreover the structure of incomes is frequently of interest and an effect of different taxation on the structure of income can be analysed.

It is important that incomes are not uniformly distributed in the society. We can quantify this non-homogeneity by detailed analysis of income distribution. Probability models enable researchers to derive detailed properties and relations. In the simplest probability models a suitable probability distribution is used as a model for income or wage distribution. More complicated models are frequently used in huge literature dealing with the topic (see Kleiber, Kotz (2003) for many references).

Different data about incomes and wages are available in the Czech Republic. Data concerning Czech households are included in annual survey Results of the Living Conditions Survey (a national module of the European Union Statistics on Income and Living Conditions (EU-SILC)). The survey is organised by the Czech Statistical Office (CZSO).

In this text net wages per hour are analysed with the use of data from regular quarterly survey the Average Earnings Information System that has been carried out by a private agency (TREXIMA, spol. s r.o.) on behalf of the Ministry of Labour and Social Affairs of the Czech Republic since 1992.

The three-parametric lognormal distribution is used as a model for hourly wages in the Czech Republic in 2009 (random samples from the survey are used) and 2000 and 2010 (published quantiles are used). The appropriateness of this distribution from various points of view is discussed for example in Keiber, Kotz (2003). Distributions of incomes and wages are positively skewed and lognormal distribution meets this property. If we suppose lognormal distribution of a random variable, it means that logarithm of its values is normally distributed. Three-parametric distribution in addition to the mean and variance of the logarithms allows a choice (guess or estimation) of minimal value of original values. As the income distribution is usually highly skewed, the mean loses its ability to objectively describe the distribution (mean is strongly affected by rare but very high values) and then the median seems to be better characteristic of the level of income. In the text mode is used as a useful characteristics of location.

Lognormal distribution is one of many distributions that are frequently used in analyzing incomes. "Two-parametric lognormal distribution fits well over a large part of middle income range, but gives a poor fit at the tails. However, in the middle income range it exaggerates skewness. Pareto distribution provides an excellent fit to the upper tail of the income distribution, but the fit over the whole range of income is poor. Gamma distribution provides a better fit than lognormal at the tails. In the middle range, both lognormal and gamma exaggerate skewness, but the tendency is more marked in case of lognormal. Dagum distribution performs better than lognormal and gamma distributions" (Chakravarty, Majumder (1990)). The distribution is supposed to be acceptable but it is outperformed by other frequently used income distributions as Dagum, generalized lognormal, gamma or lambda distributions. Wages in the Czech Republic are analysed in Bílková (2008, 2011), Pavelka (2009) or Jurajda (2011). In the articles of Pavelka a Jurajda more detailed description of Average Earnings Information System can be found.

Lognormal distribution

In this text lognormal distribution is used as distribution for the modeling of per hour wages. For the analyses of incomes and another statistics attributes which are correlated with incomes three-parametric lognormal distribution is mainly used. Two-parametric and distribution with four and more parameters are also used in practice but not as frequently. This distribution can be also used as a distribution of components in mixture models. In this case the distribution of the mixture of lognormal components is not lognormal (Pavelka (2009)). Generalized lognormal distribution (Keiber, Kotz (2003)) is supposed to be very useful in the modeling of incomes. Parameters of three-parametric lognormal distribution have straightforward interpretation. A random variable X has three-parametric lognormal distribution with parameters θ, μ, σ^2 if a random variable Z = ln (X - θ) has normal distribution with expected value μ and variance σ^2. For the analysed hourly wage it means that parameter μ is expected value and parameter σ^2 variance of logarithms of these values. Probability density function of the distribution is given by formula

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

In (1) parameter θ means a theoretical minimum of variable X. If parameter θ equals zero, then the random variable X has two-parametric lognormal distribution. Distribution function of variable X with density function (1) is given by formula

$$F(x) = \Phi\left(\frac{\ln(x-\theta)-\mu}{\sigma}\right), \quad (2)$$

where ϕ is a distribution function of standard normal distribution. From (2) a formula for evaluation of quantiles can be easily derived.

Estimation of parameters

The most frequently used methods of estimation of unknown parameters are moment method, quantile method and maximum likelihood method. We need three equations to find out the estimated values of unknown parameters of three-parametric lognormal distribution and each method creates these equations differently (Johnson at all. (1994)).

Suppose that X_i, i = 1,…,n is a random sample from lognormal distribution. The maximum likelihood estimates of unknown parameters have theoretical optimal asymptotic properties. The logarithm of likelihood function based on density function (1) is maximized with respect to estimates $\hat{\mu}, \hat{\sigma}^2$ of parameters μ,σ^2 by estimates (Johnson at all. (1994))

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i - \theta), \quad (3)$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\ln(X_j-\theta)-\hat{\mu}\right)^2. \quad (4)$$

The most difficult problem is to estimate the value of parameter θ. For a large sample the minimum in the sample can be taken as an estimate. Cohen's method combines method of maximum likelihood with the quantile method and parameter θ is estimated as 100/(n+1)% sample quantile.

The method of moments is relatively simple but it could be quite inaccurate. This method equates sample and theoretical moments. The point estimates $\theta^+,\mu^+,\sigma^{2^+}$ of three unknown parameters are given by formulas

$$\theta^+ = \overline{x} - e^{\mu^+ + \frac{\sigma^{2^+}}{2}}, \quad (5)$$

$$\mu^+ = \frac{1}{2}\ln\frac{m_2}{e^{\sigma^{2^+}}\left(e^{\sigma^{2^+}}-1\right)}, \quad (6)$$

$$\sigma^{2^+} = \ln\left[\sqrt[3]{1+b_2+\sqrt{(1+b_2)^2-1}}+\sqrt[3]{1+b_2-\sqrt{(1+b_2)^2-1}}-1\right], \quad (7)$$

where $b_2 = \frac{1}{2}\left(e^{\sigma^{2^+}}-1\right)\left(e^{\sigma^{2^+}}+2\right)^2$ and $m_2 = \overline{x^2}-\overline{x}^2$.

The method of quantiles is also relatively simple. This method uses three quantiles (to estimate three parameters) – the most frequently median and two symmetric 100α% and 100(1-α)% quantiles for a chosen probability α. Equations (8),(9) and (10) give the point estimates $\theta^*,\mu^*,\sigma^{2*}$

$$\sigma^{2^*} = \left[\ln\frac{x_\alpha^S - x_{0.5}^S}{x_{0.5}^S - x_{1-\alpha}^S} / u_\alpha\right]^2, \quad (8)$$

$$\mu^* = \ln\frac{x_\alpha^S - x_{1-\alpha}^S}{e^{\sigma^* u_\alpha} - e^{-\sigma^* u_\alpha}}, \quad (9)$$

$$\theta^* = x_{0.5}^S - e^{\mu^*}, \quad (10)$$

where x_P^S is 100P% sample quantile.

Average Earnings Information System

The Average Earnings Information System (AEIS) is a quarterly employer survey carried out by a private agency (TREXIMA, spol. s r.o.) on behalf of the Ministry of Labour and Social Affairs since 1992 (see TREXIMA, AEIS1). The AEIS is based on the stratified random sampling which has been fully in accordance with the European Structure of Earnings Survey guidelines since 2006 and it provides precise and reliable information on wages in Czech firms. The data cover companies of all size categories and industries, except the public budgetary sector. Only economic subjects with more than 10 employees are sampled in the AEIS. The data contain hourly wages, age, education and detailed information about occupation of the employee. The wage records are taken from personnel databases of firms. An average hourly wage of an employee is quarterly evaluated as total cash compensation including bonuses and other special payments divided by total hours worked for that quarter. Data are classified by KZAM-R[1] levels (CZSO).

Companies in the Czech Republic are divided into two subgroups – business and non-business sectors and the technique of sampling differs in these sectors of economy. Non-business sector is surveyed every half-year and it is a comprehensive (non-sample) survey. It covers about 14,550 economic subjects with approximately 660 thousand employees. Business sector is investigated selectively (stratified sample) quarterly. The sample covers about 3,500 economic subjects with more than 1.3 million employees.

Business and non-business sectors data sets have different characteristic. The statistical characteristics of data from non-business sector are affected only by non-sampling errors because it is the comprehensive survey. The statistical characteristics of data set from business sector are affected by both sampling and non-sampling errors.

In the text we use two random samples of the size 10,000 from data of the AEIS survey from 2009. First data set contains 10,000 entries of hourly wage (in CZK) from the business sector and the second data set contains 10,000 values of hourly wage (in CZK) from the non-business sector.

On the website of the survey (AEIS2) have been published results of AEIS surveys since 2000. Descriptive characteristics on the website mentioned above include median, mean, the first and the third quartiles, the first and the ninth deciles of hourly wage. We use these values from 2010.

This paper uses data from AEIS to compare results (and quality) of different lognormal models. All hourly wages in the models are nominal in the Czech koruna (CZK). The exchange rate of CZK to Euro was 27.142 in 2009 (January-

[1] Since 2011 is used classification CZ-ISCO.

June), average exchange rates for 2000 and 2010 were 35.609 and 25.290 (CNB). Moreover inflation rate from 2000 to 2010 was 1.28 (CZSO).

Methods and Results

To the samples from 2009 five lognormal models for both business and non-business sectors were fitted:

- Model 2_ml uses two-parametric lognormal distribution with maximum likelihood estimates of parameters,
- model 3_m uses moment method to estimate parameters of three-parametric lognormal distribution,
- model 3_ql uses quantile method with α=0.1 (first and last deciles are used in formulas (7), (8), (9)),
- model 3_qII uses quantile method with α=0.25 (lower quantile and upper quantile are used in (7), (8), (9)),
- model 3_ml uses maximum likelihood method and parameter is evaluated by numeric minimization of the statistics S evaluated according to (12).

In order to compare quality of fits we use intervals of the length 5 CZK and we compare theoretical and empirical frequencies of intervals. The empirical frequency is given by number of entries in the corresponding interval in the data set from the business or non-business sector. The theoretical frequency of interval i is given by formula

$$n\pi_i = n\left[\Phi\left(\frac{\ln(x_i + \nu/2 - \theta) - \mu}{\sigma}\right) - \Phi\left(\frac{\ln(x_i - \nu/2 - \theta) - \mu}{\sigma}\right)\right], \quad (11)$$

where x_i is the middle value of the i-th interval, v is the interval length (v = 5 CZK) and n is the sample size (n = 10,000) and estimated values of parameters are used. Theoretical frequencies (11) must be complete with frequencies of the first and the last intervals[2] . Quality of various models is compared by a statistic S which is evaluated as a sum of absolute deviations of theoretical and empirical frequencies[3]

$$S = \sum_{i=1}^{k} |n_i - n\pi_i|, \quad (12)$$

where k is number of intervals.

To compare distributions of hourly wage in 2000 and 2010, the results of AEIS (median and first and ninth deciles) were used to fit lognormal model with the quantile method of estimation.

All computations were performed in MS Excel. [4]

Table 1: Non-business sector, second quarter of 2009. Estimates of parameters and characteristics of level and variability (CZK), values of S (12).

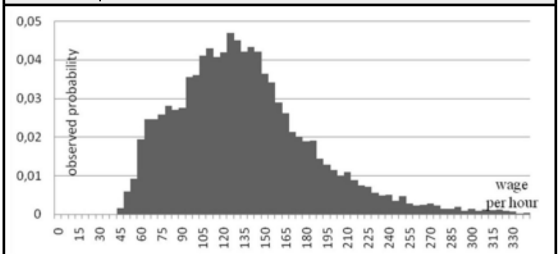
Model		2_ml	3_m	3_ql	3_qII	3_ml	Sample
Method		maximum likelihood	Moment	quantile	quantile	maximum likelihood	/
Comment		/	9980 values[4]	$\alpha = 0.1$	$\alpha = 0.25$	θ by MS Excel	/
Parameters	θ	/	4.2690	-76.3609	-181.0605	-25.4797	/
	μ	4.8500	4.8111	5.3286	5.7393	5.0422	
	σ^2	0.1484	0.1520	0.0582	0.0218	0.1020	
$E(X)$		137.58	136.83	135.87	133.18	137.43	137.74
$\sqrt{D(X)}$		55.03	53.70	51.95	46.61	53.39	57.82
Mode	$\hat{x}$	110.12	109.81	118.13	123.09	114.31	125-130
Median	$x_{0.5}$	127.74	127.13	129.78	129.78	129.33	129.78
S		1,344	1,359	1,320	1,277	1,255	/

Source: own computations

Results for non-business sector

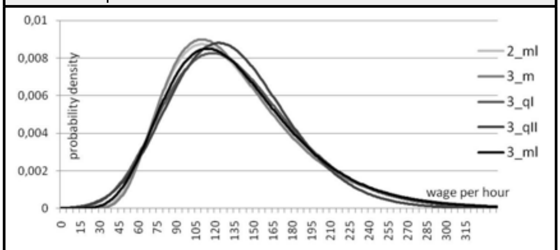
In this part five models mentioned above were fitted into 10,000 observations from non-business sphere. Table 1 presents estimates of parameters, estimated characteristics of location and variability of fitted lognormal distributions and values of the statistics S defined in (12) together with descriptive characteristics of the data set. There are very different values of parameters, especially of the parameter of variance. Moreover in three models the value of parameter theta is negative, in the model with maximum likelihood estimates even -181. This model (3_ml) has the lowest value of S statistic, but all the models give quite similar values of S. All models underestimate not only standard error (as it is common if lognormal distribution is used) but also the location characteristics. Although the parameters are very different, estimates of characteristics of location and variability given in the table are closed.

Figure 1: Histogram of wage per hour (CZK) in non-business sector in the 2nd quarter of 2009



Source: own calculations

Figure 2: Estimated probability densities for non-business sector for the 2nd quarter of 2009



Source: own calculations

The worst model is model 3_m which uses moment method of estimation, however its quality was improved by excluding the highest 20 values. Model 3_qII provides relatively good fit. Model 3_qII uses quartiles to estimate the unknown parameters, these values are published by AEIS[5] and the model can be constructed for every year with the use of published sample quantiles (without original data).

The histogram of wage per hour is shown in Figure 1 and fitted probability densities are given in Figure 2. We can see in Figure 1 a local extreme in the left part of the histogram. This paper models the income distribution by one curve of

[2] Then the sum of the theoretical frequency equals 1.

[3] We don't use the chi-square statistic because we operate with large sample. In this case we get almost always the rejection of χ^2 at the conventional level of significance.

[4] The highest 20 values weren't used to estimate the parameters of model 3_m. The calculation of models showed that approximately 1-2 percent of the highest values make the model worse. This paper doesn't find out the most quality model but compares the quality of different lognormal models so we don't need to leave out these values. But by estimating the parameters with moment method was the S statistic so bad that these 20 values weren't used to get the value of S statistic of model 3_m closer to the S statistics of other models.

[5] The quartiles and the 1st and the 9th deciles for the years 2000 – 2010 are published on http://www.ispv.cz/cz/Vysledky-setreni/Archiv.aspx (AEIS2)

the lognormal distribution, more detailed analysis could use two or more curves in a mixture (Pavelka 2009). In this approach a data set is divided into two or more subgroups, each subgroup is modeled by a (lognormal) probability density and the final model is constructed as a weighted average of component densities. Lower values of wage per hour (the left part of the histogram) in non-business sector should be modeled separately and the final model would be probably better. Nevertheless all models seem to be suitable for the modeling of the wage distribution of non-business sector.

We can see from the figure that negative values of parameter theta don´t make estimated densities inapplicable, as values of the densities are closed to 0 for negative and very small positive values of x. In the case of negative estimates of θ its value are not interpretable.

Results for business sector

In this part the same five models were fitted into 10,000 observations from business sphere. Table 2 presents estimates of parameters, the estimated values of the theoretical characteristics and values of statistic S together with sample characteristics. The best fit gives the model 3_qII which uses quantile method of estimation of the parameters and chooses sample quartiles for the estimation. The second model which uses the quantile method of estimation of the parameters (3_qI) is also better than the model 3_ml. The estimated theoretical characteristics of mean, variance and modus are again underestimating sample characteristics. In this case all values of estimated parameters are closed and no negative value of parameter theta occurs. From it follows that characteristics derived from fitted models are similar.

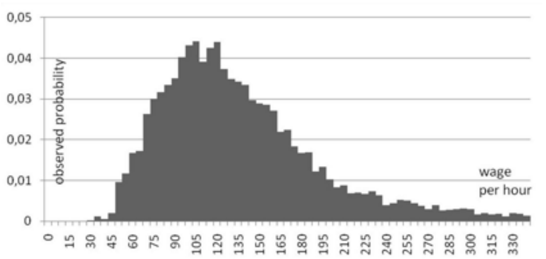
Table 2: Business sector, second quarter of 2009. Estimates of parameters and characteristics of level and variability (CZK), values of S (12)

Model		2_ml	3_m	3_qI	3_qII	3_ml	Sample
Method		maximum likelihood	moment	quantile	quantile	maximum likelihood	/
Comment		/	/	$\alpha = 0.1$	$\alpha = 0.25$	θ by MS Excel	/
Parameters	θ	/	27.1020	27.7359	23.9340	21.6933	/
	μ	4.8726	4.5942	4.5935	4.6313	4.6653	
	σ^2	0.2073	0.3708	0.3314	0.2595	0.3061	
$E(X)$		144.93	146.16	144.40	140.80	145.46	146.16
$\sqrt{D(X)}$		69.57	79.77	73.13	63.62	74.07	79.77
Mode	$\hat{x}$	106.19	95.37	98.70	103.12	101.71	105-110
Median	$x_{0.5}$	130.66	126.01	126.58	126.58	127.89	126.58
S		1,475	1,605	1,280	1,086	1,303	/

Source: own computations

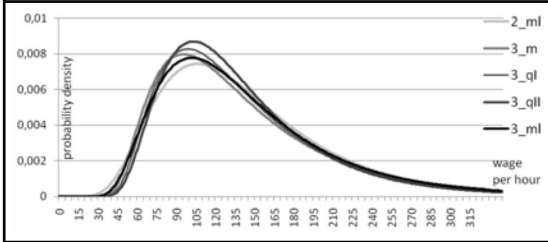
The observed frequencies of wage per hour are shown in Figure 3 and the fitted probability densities are shown in Figure 4. The estimated densities are closed for all models (except for two-parametric lognormal model) and all of them seem to be suitable for the modeling of the wage distribution of business sector especially in central part of values.

Figure 3: Histogram of wage per hour (CZK) in business sector in the 2nd quarter of 2009



Source: own computations

Figure 4: Estimated probability densities for business sector for the 2nd quarter of 2009

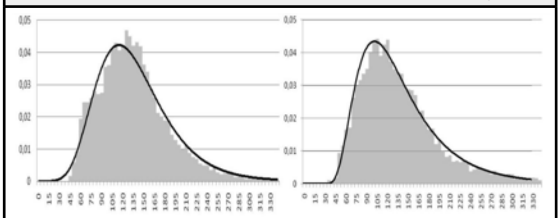


Source: own computations

In the Figure 5 best fitted densities are given for both sectors together with histograms of samples in the same scale in order to demonstrate quality of fits and differences in the distributions between business and non-business spheres.

From the models in Figure 5 percentages of people with hourly wage under average wage were evaluated as 60% in business sector and 56% in non-business sector. To compare with samples, there were 63% of values under the mean in the sample of business sector and 57% in the sample of non-business sector.

Figure 5: Comparison of the best models and sample wage distributions in business (left) and non-business sector (right)



Source: own computations

Analysis of hourly wage distribution of business sector in 2000 and 2010

ISAI has published quarterly selected quantiles of wage per hour in business and non-business sector since 2000 on its web site (TREXIMA, AEIS2). In this part we use quantile method of estimation of unknown parameters to fit a lognormal model to data for the fourth quarters of 2000 and 2010.

In the Table 3 the values of published quantiles (median, first and last deciles) are given together with the mean value. All wages in the text are nominal, in the Table 3 are also given real values evaluated with the use of inflation rate 1.28 (CZSO).

Table 3: Quantiles of income distribution (in CZK) of business sector

Period	quantiles			mean
	1st decile	median	9th decile	
4th quarter 2000	45,1	76,94	134,08	88,29
4th quarter 2010 real	54,63	96,98	187,58	118,76
4th quarter 2010 nominal	70,11	124,46	240,72	152,41

Source: TREXIMA, ISPV2, CZSO

Table 4 presents estimates of parameters and basic characteristics for the models of wage distributions for 2000 and 2010. Comparing Table 4 with Table 3 we can see that both models underestimate the mean. Sample and estimated medians are equal, because sample median is used for the estimation of parameters and on the web site

AEIS are not given sample modes to be compared with estimated values. Expected value of nominal hourly wage has increased by 70 per cent; 28 per cent was caused by increasing of price level. Characteristics of variability quantify wage inequality. Variance has nearly quadrupled during ten years and it can be interpreted as relatively large increase of non-uniformity of wages. It is also clear from Figure 6 that the model for 2000 seems to be more concentrated than the model for 2010 which shows increasing the number of people with higher wage per hour. Coefficient of variation increased from 45% in 2000 to 55% in 2010.

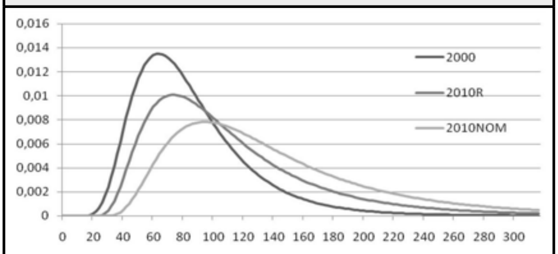
Table 4: Estimates of parameters and estimated characteristics for lognormal models in business sector for 2000 and 2010

		2000	2010R	2010NOM
Parameters	σ^2	0,2082	0,352	0,352
	μ	4,2754	4,3761	4,6256
	θ	5,0294	17,4523	22,3968
Characteristics	Mode	63,42	73,38	94,17
	$x_{0,5}$	76,94	96,98	124,46
	E (X)	84,83	112,29	144,1
	$\sqrt{D(X)}$	38,39	61,61	79,06
	Skewness	1,55	2,22	2,22
	Kurtosis	7,58	12,9	12,9

Source: TREXIMA, CZSO

From fitted models the percentages of people with hourly wage under expected value (average hourly wage) were evaluated. The value has increased from 59% in 2000 to 62% in 2010. This fact was expected due to the increasing of inequality.

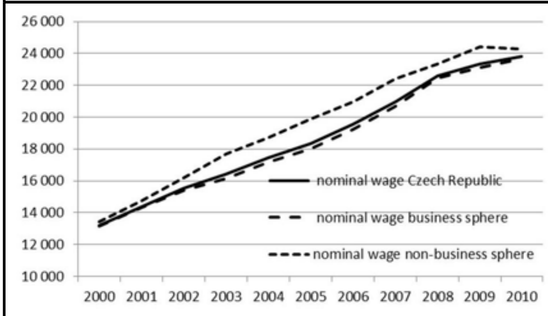
Figure 6: Estimated densities of wage distribution of business sector for 2000 and 2010



Source: own calculations

In Figure 7 average gross nominal wage (month, full time equivalent) in the Czech Republic and its development since 2000 is shown. Business and non-business spheres are distinguished and overall values for the Czech Republic are shown. The values in the Figure are based on different sources than data from previous parts of this text. Moreover data dealing with public budgetary sector are included. For the whole period of ten years nominal wages in the non-business sector are higher than in business sector, from 2009 to 2010 the gap seems to decrease and the difference is 576 CZK in 2010 in comparison with 1307 CZK in 2009 and the highest difference 1858 CZK in 2005. The converging of average nominal wages in business and non-business sector (or decreasing of the gap between them) may be caused by Czech government austerity politics and its direct influence on wages in non-business sector. There was an annual decline of average gross nominal monthly wage in non-business sector for the first time in last ten years from 2009 to 2010.

Figure 7: Average gross nominal monthly wage (CZK) in the Czech Republic, 2000-2010



Source: CZSO

Conclusion

In the text three-parametric lognormal distribution is used as a model of wage distribution in the Czech Republic. Hourly wages are analysed separately for business and non-business sector with satisfactory results especially in central part of values of wages. This fact corresponds with theoretical properties of lognormal distribution. Quality of fitted models depends on the method of estimation of unknown parameters even for large analysed samples (10,000 observations) and values of estimated parameters are very different (especially for non-business sphere). For non-business sector the best fit provided method of maximum likelihood, for business sector the best model was obtained with the use of quantile method. For both fits quantile method gave better fit with the use of quartiles than deciles which does not correspond with theoretical assumption (Johnson et all, (1994)). The percentage of people with hourly wage under average is higher in business sector (60-63%), in non-business sector this value is lower by approximately 5%. In the wage distribution of non-business sector appears local extreme in left part of histogram which does not seems to be in business sector, therefore non-business sector would be better modeled by mixture of lognormal distribution.

Figure 7 was created with the use of data from the Czech Statistical Office, these values are based on different survey than Average Earnings Information System and it causes differences in results. There was higher average gross nominal monthly wage in non-business sector then in business sector in Czech Republic, but the decline of wage in non-business sector in 2010 signifies reduction of difference between business and non-business sector.

Using the three-parametric lognormal distribution, three models of income distribution are made for business sector for 2000, 2010 nominal and 2010 related to prices of 2000. The models use the quantile method of estimating the parameters and quantiles are taken from the Average Earnings Information System. The results show faster increase of characteristics of the level than inflation rate (inflation +28 per cent), 70% increase in expected value, 60% in the value of median, almost 50% in the value of modus and the growth of inequality of wage distribution in business sector between 2000 and 2010.

Models, constructed in the text, show an increase of percentage of people with below-average hourly wage from 59% in 2000 to 62% in 2010 in business sector. Parameter sigma in the model of nominal wage in 2010 equals parameter sigma in model of real wage in 2010 and the same works for skewness and kurtosis. This fact results from the formulas for these moments because these characteristics depend only on parameter sigma (Johnson et all. (1994)). The comparison of models for nominal and real wages in 2010 with the model of wages in 2000 shows that the real increase of wages during the observed period was not so great. Real increase of average hourly wage was half compared to nominal increase, real increase of median and mode was even less than one half of nominal increase. Relatively high is the real increase of standard deviation (and variance as well). Mean is undervalued by 4 CZK in 2000 and by 8 CZK in 2010 (nominal). This indicates the possibility of increasing the undervaluing of some characteristics with time when we use a three-parametric lognormal model.

Acknowledgement

The paper was supported by a project IGS 24/2010 from the University of Economics in Prague.

References

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

Bílková, Diana. Use of the L-Moments Method in Modeling the Wage Distribution. Bratislava February 1 – 4, 2011, Proceedings. In: 10th International Conference APLIMAT 2011, Bratislava: Slovak University of Technology, Bratislava, 1471–1481.

Chakravarty, Satya R. and Aditi Majumder. Distribution of Personal Income: Development of a new Model and its Application to U.S. Income Data. Journal of Applied Econometrics5/2 Apr. – Jun., 1990: 189-196.

CNB. Czech National bank. www.cnb.cz.

CZSO. Czech Statistical Office. www.czso.cz.

AEIS1. Informační system o průměrném výdělku. http://www.ispv.cz/.

AEIS2. Informační system o průměrném výdělku. http://www.ispv.cz/cz/Vysledky-setreni/Archiv.aspx.

Johnson, Norman L., Narayanaswamy Balakrishnan and Samuel Kotz. Continuous Univariate Distributions. Vol. 1. New York: John Wiley & Sons, 1994.

Jurajda, Štěpán. Regional Divergence and Returns to Schooling. http://home.cerge-ei.cz/jurajda/Jurajda.pdf

Kleiber, Christian and Samuel Kotz. Statistical Size Distributions in Economics and Actuarial Sciences. New York: Wiley-Interscience, 2003.

Nedvěd, Jakub. Využití logaritmicko-normálního rozdělení při analýze příjmů. Bachelor thesis. University of Economics, Prague, 2011.

Pavelka, Roman. Application of density mixture in the probability model construction of wage distributions, Applications of Mathematics and Statistics in Economy: AMSE 2009, Uherské hradiště, 2009, 341-350, 2009.

Titterington, D.M., Smith, A.F., Makov, U.E. Statistical analysis of finite mixture distributions, Wiley, 1985.

TREXIMA. TREXIMA spol. s r.o. www.trexima.cz.