FOUR-PARAMETER LOGNORMAL CURVES IN WAGE DISTRIBUTION MODELS: COMPARISON WITH THREE-PARAMETER LOGNORMAL CURVES

Diana Bílková

Abstract

This paper deals with the construction of wage distribution models using four-parameter and three-parameter lognormal curves. The main objective of the research is to compare the accuracy using both types of lognormal distribution as wage distribution models. The minimum wage in a given year represents the beginning of four-parameter and three-parameter lognormal curves. The estimates of the remaining three, respectively two parameters are constructed using the quantile method. The chi-square test criterion is used to evaluate the accuracy of the models obtained. In almost all wage distributions, the four-parameter lognormal models yielded more accurate results than the three-parameter lognormal models. The results in terms of accuracy of both lognormal curves are just as accurate in only a few cases. However, the differences in the accuracy of the four-parameter and three-parameter wage models are not critical. Gross monthly nominal wage is the main research variable. There are data in the form if interval frequency distribution.

Key words: models of wage distribution, lognormal curves, quantile method of parameter estimation, chi-square test criterion

JEL Code: C13, C46, C18

Introduction

The topic of wages, incomes and other indicators of labour market appears almost continuously in the economic professional literature, and that is and will be still permanently topical. An example is several following publications. The issue of income stability is solved in (Bartošová & Longford, 2014). The issue of income inequality is researched in (Bartošová & Bína, 2018; Langhamrová & Šimpach, 2012, and Sipková & Sipko, 2012). An important field is the wage level, which is examined in (Marek, 2018; Marek, Doucek & Nedomová, 2018, and Maryška, Nedomová & Doucek, 2017). The issue of employment and unemployment is researched in (Pavelka & Löster, 2014, and Pavelka & Löster, 2013). The authors (Pechrová & Šimpach, 2013) research regional competitiveness and employment. Models of wage or income distributions are examined in (Malá, 2017, and Malá, 2016).

This paper deals with the construction of wage distribution models using four-parameter lognormal curves. The main objective of this study is to compare the accuracy of wage distribution models constructed using four-parameter lognormal curves and three parameter lognormal curves. The quantile method of parameter estimation is used to obtain the parameter estimations in both cases. The beginning of lognormal curves is the minimum wage in the year in both cases, four-parameter and three-parameter lognormal distributions. The chi-square test criterion is used to evaluate the accuracy of the models obtained. Thus, the main scientific hypothesis consists in the statement that using the quantile method of parameter estimation, four-parameter lognormal curves result in more accurate models than three-parameter lognormal models.

Tab. 1: Sample sizes of employees by unit size (thousands of employees)

	Year				
Unit size	2014	2015	2016	2017	2018
less than 10 employees	527.5	571.4	570.5	576.5	593.0
from 10 to 49 employees	738.0	693.9	702.0	697.2	674.5
from 50 to 249 employees	795.7	811.7	830.5	842.0	866.4
from 250 to 999 employees	657.6	677.2	697.6	714.8	725.5
from 1,000 to 4,999 employees	657.6	1,340.3	494.8	518.3	525.6
more than 5,000 employees	305.6	298.7	409.6	320.4	329.4

Source: www.czso.cz

Tab. 2: Development of the minimum wage (in CZK) in 2014–2018

Year	2014	2015	2016	2017	2018
Minimum wage	8,500	9,200	9,900	11,000	12,200

Source: www.mpsv.cz

The data for this research and sample sizes come from the official website of the Czech Statistical Office (CSO). Data include wages of employees of the Czech Republic. Gross monthly nominal wage in CZK is the main researched value. The relevant annual CSO data are in the form of interval frequency distribution with opened extreme intervals. The data are sorted according the company size and they cover the period from 2014 to 2018, see

Table 1. Table 2 shows the development of minimum wage in the period. A total of 30 wage distributions were researched.

The data include employees in both business and non-business sectors of the economy. The wage is paid to an employee for work done in the private corporate (business) sphere, while the salary is earned in the state budgetary (non-business) sector. Within the present study, both wages and salaries are under the umbrella term of "wage".

1 Theory and Methods

1.1 Four-parameter Lognormal Distribution

The random variable X has a four-parameter lognormal distribution with parameters μ, σ^2, θ and τ, where $-\infty < \mu < \infty$, $\sigma^2 > 0$, $-\infty < \theta < \tau < \infty$, if its probability density function is

$$f(x;\,\mu,\,\sigma^2,\,\theta,\,\tau) = \frac{(\tau-\theta)}{\sigma\cdot(x-\theta)\cdot(\tau-x)\cdot\sqrt{2\pi}}\cdot\exp\left[-\frac{\left(\ln\frac{x-\theta}{\tau-x}-\mu\right)^2}{2\,\sigma^2}\right], \qquad \theta < x < \tau, \tag{1}$$

$$= 0, \qquad \text{jinak.}$$

Lognormal distribution with parameters μ, σ^2, θ and τ will be marked LN (μ, $\sigma2$, θ, τ). The probability density function of the four-parameter lognormal distribution can take very different shapes depending on the values of the parameters of this distribution. The distribution can be also double-peak for $\sigma^2 > 2$ and $|\mu| < \sigma^2\cdot\sqrt{(1-2/\sigma^2)} - 2\tanh^{-1}\sqrt{(1-2/\sigma^2)}$. Figs. 1–3 show the probability density function shapes of four-parameter lognormal distribution depending on parameter values.

If the random variable X has four-parameter lognormal distribution with parameters μ, σ^2, θ and τ, then random variable

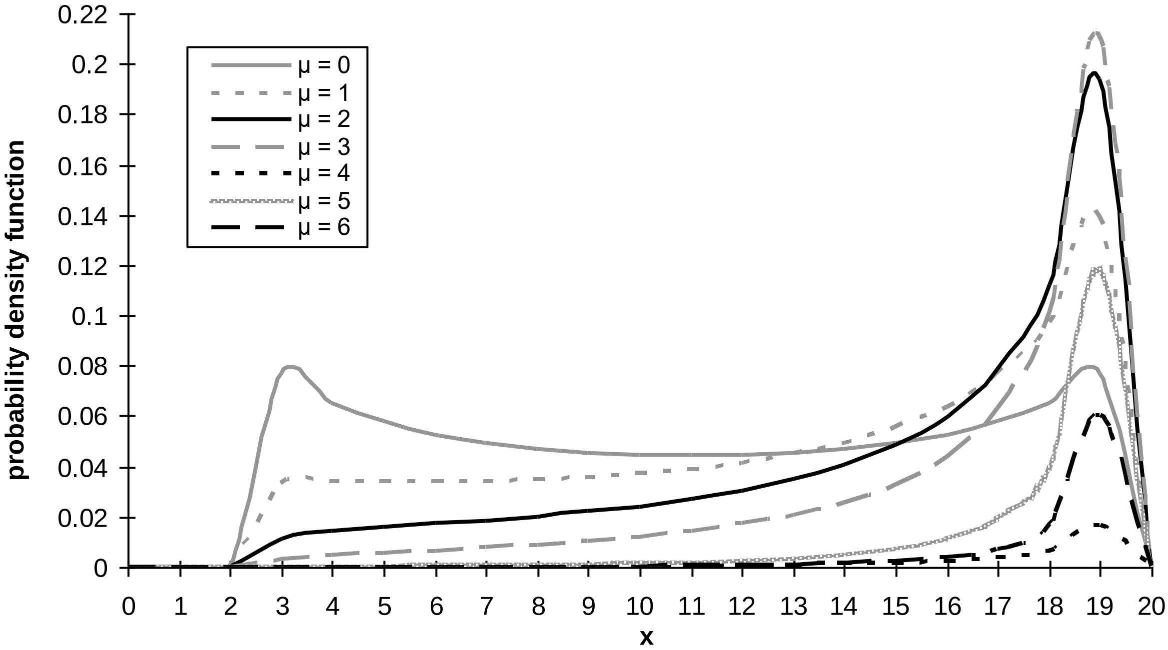
$$Y = \ln\frac{X-\theta}{\tau-X} \tag{2}$$

has normal distribution with parameters μ and σ^2 and random variable

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

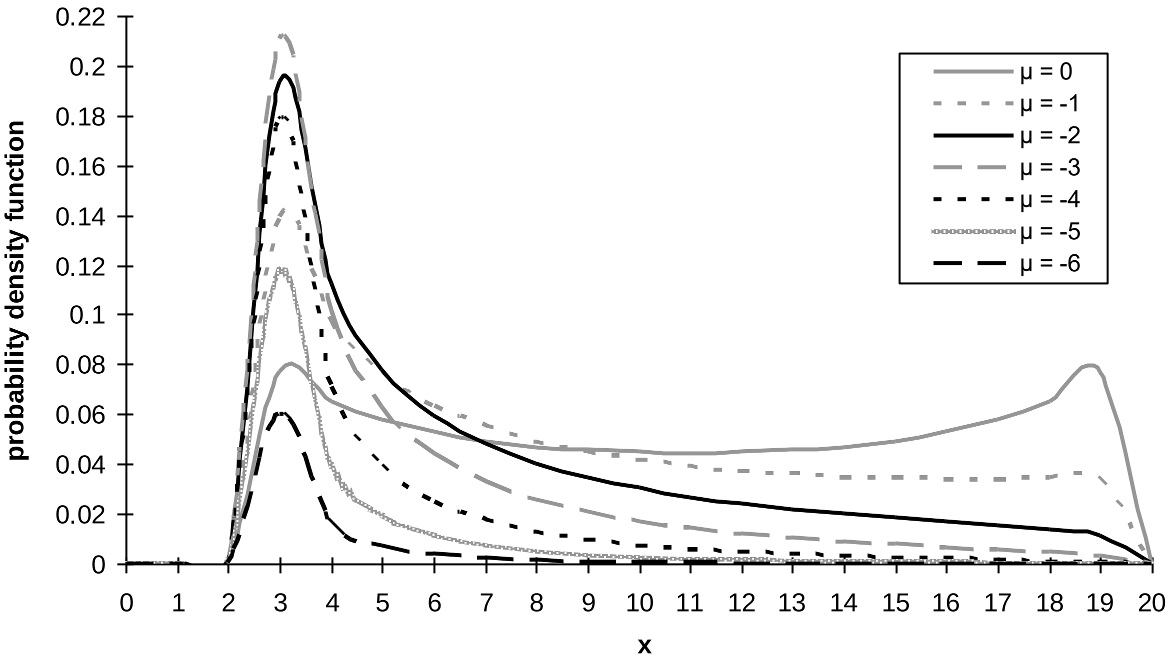
has standardized normal distribution.

Fig. 1: Probability density function of four-parameter lognormal distribution for the values of parameters σ = 2 (σ² = 4); θ = 2; τ = 20



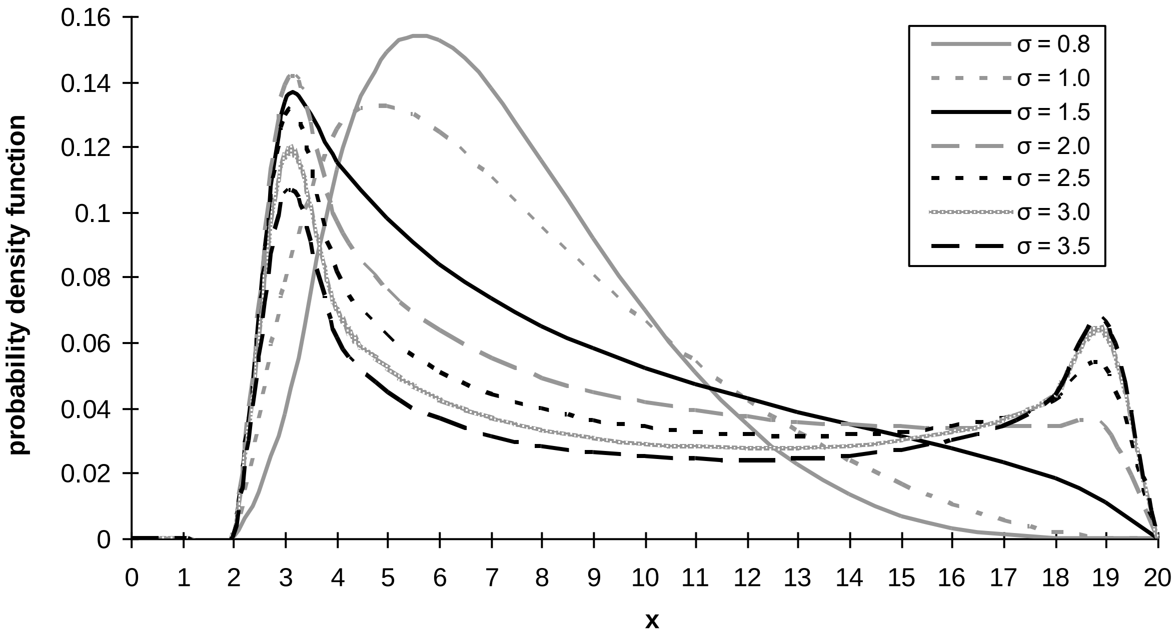
Source: Own research

Fig. 2: Probability density function of four-parameter lognormal distribution for the values of parameters σ = 2 (σ² = 4); θ = 2; τ = 20



Source: Own research

Fig. 3: Probability density function of four-parameter lognormal distribution for the values of parameters μ = –1; θ = 2; τ = 20



Source: Own research

Thus, the parameter μ in the expected value of the random variable (2) and the parameter σ^2 is the variance of this random variable. The parameter θ is the beginning of the distribution (theoretical minimum) of the random variable *X* and the parameter τ represents the terminal point of the distribution (theoretical maximum) of this random variable.

1.2 Three-parameter Lognormal Distribution

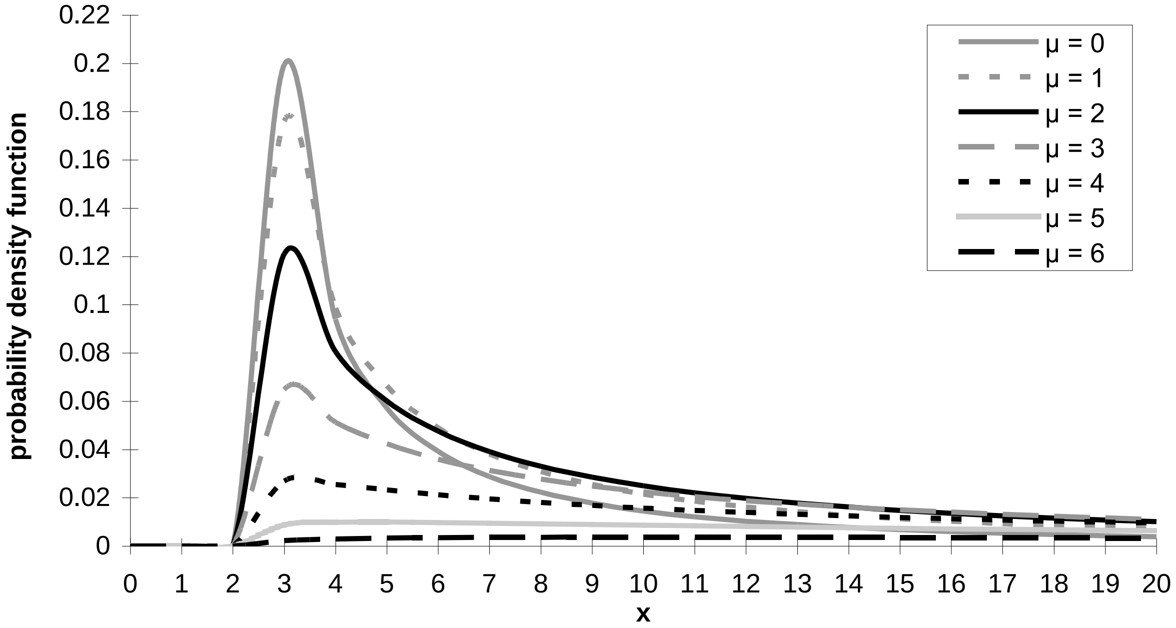
The random variable *X* has a three-parameter lognormal distribution with parameters μ, σ^2 and θ, where $-\infty < \mu < \infty$, $\sigma^2 > 0$, $-\infty < \theta < \infty$, if its probability density function is

$$f(x;\mu,\sigma^2,\theta) = \frac{1}{\sigma\cdot(x-\theta)\cdot\sqrt{2\pi}}\cdot\exp\left[-\frac{[\ln(x-\theta)-\mu]^2}{2\sigma^2}\right], \qquad x>\theta, \tag{4}$$

$$= 0, \qquad \text{jinak.}$$

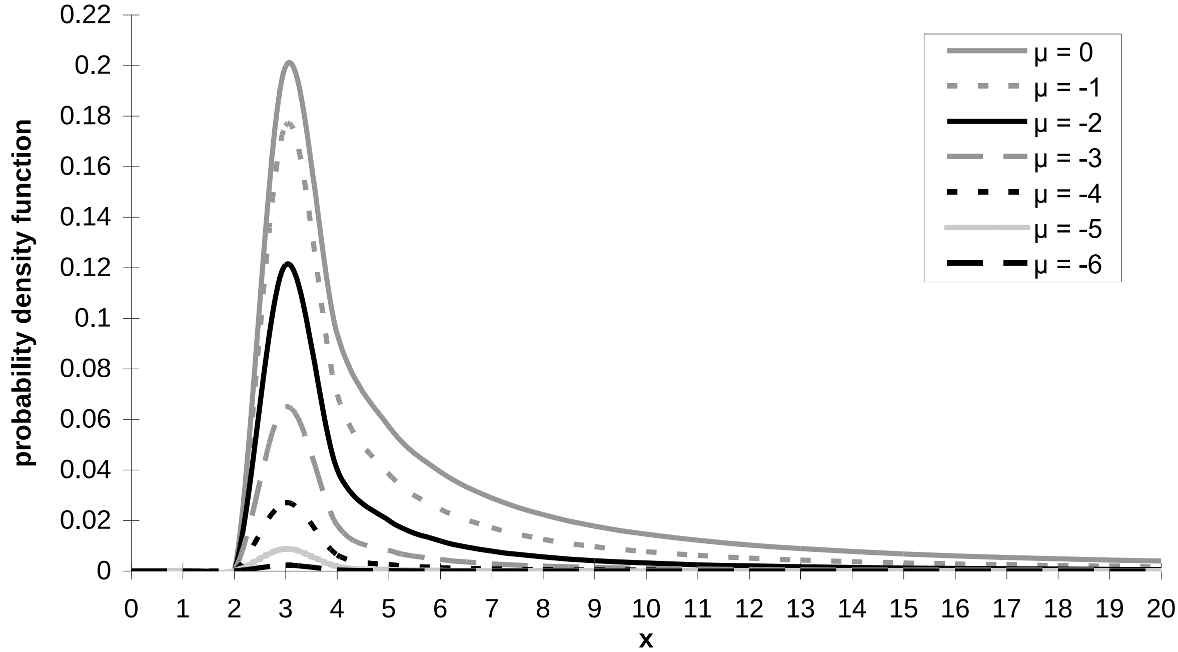
Lognormal distribution with parameters μ, σ^2 and θ will be marked LN (μ, σ2, θ). The probability density function of the three-parameter lognormal distribution has always positive skewness. Figs. 4–6 show the probability density function shapes of three-parameter lognormal distribution depending on parameter values. These figures copiare the parameter values of each corresponding Figure 1–3, but omittering the τ parameter.

Fig. 4: Probability density function of three-parameter lognormal distribution for the values of parameters σ = 2 (σ² = 4); θ = 2



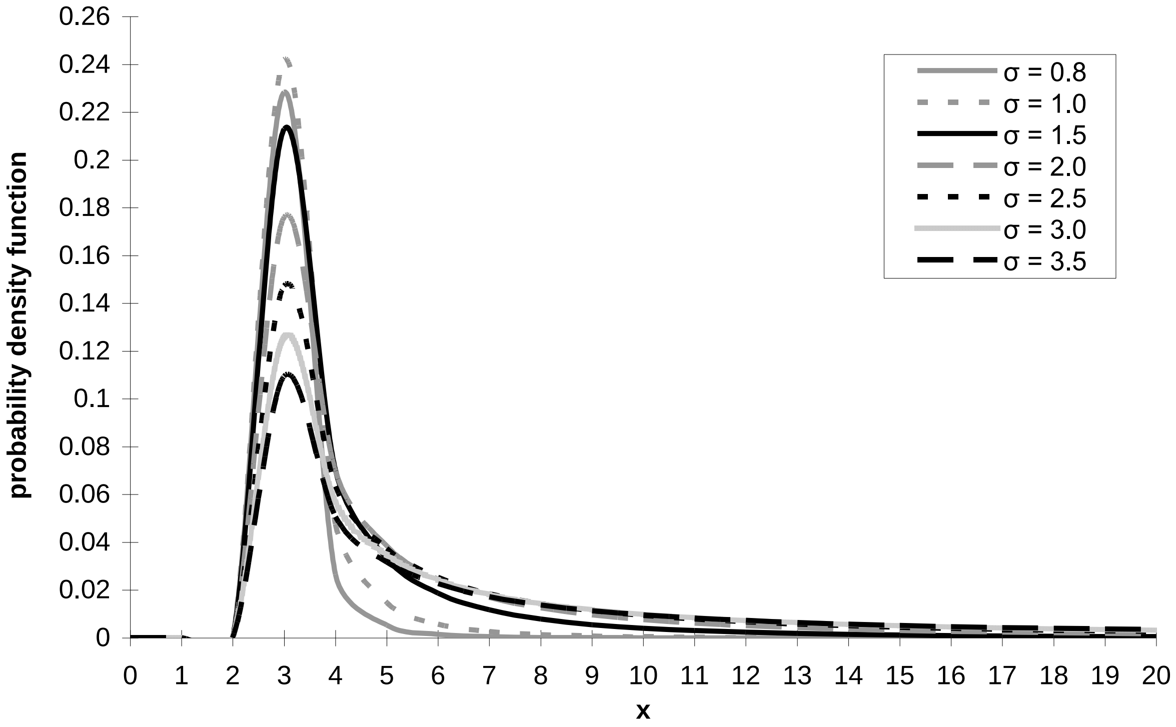
Source: Own research

Fig. 5: Probability density function of three-parameter lognormal distribution for the values of parameters σ = 2 (σ² = 4); θ = 2



Source: Own research

Fig. 6: Probability density function of three-parameter lognormal distribution for the values of parameters μ = –1; θ = 2



Source: Own research

If the random variable *X* has three-parameter lognormal distribution with parameters μ, σ^2 and θ then random variable

$$Y = \ln(X - \theta) \tag{5}$$

has normal distribution with parameters μ and σ^2 and random variable

$$U = \frac{\ln(X - \theta) - \mu}{\sigma} \tag{6}$$

has standardized normal distribution.

Thus, the parameter μ in the expected value of the random variable (5) and the parameter σ^2 is the variance of this random variable. The parameter θ is the beginning of the distribution (theoretical minimum) of the random variable *X*.

1.3 Quantile Method of Parameter Estimation for Four-parameter Lognormal Distribution

We assume that we know the parameter θ because of the existence of the minimum wage institute. As mentioned, the random variable (2) has a normal distribution with parameters μ and σ^2. 100*P% quantile of normal distribution with parameters μ and σ^2 has the form

$$x_P = \mu + \sigma u_P, \tag{7}$$

where u_P is 100*P% quantile of the standardized normal distribution. Thus, $u_{0.50} = 0$ is valid for the median of the standardized normal distribution.

We in total estimate three parameters of four-parameter lognormal distribution (parameter θ is known as minimum wage), so we need a set of quantile equations with three equations. Let $x_{0.25}$ is sample lower quartile, $x_{0.50}$ is sample median and $x_{0.75}$ is sample upper quartile. Analogically, $u_{0.25}$ represents lower quartile of the standardized normal distribution, $u_{0.50}$ represents median of the standardized normal distribution and $u_{0.75}$ represents upper quartile of the standardized normal distribution. Then the relationship $u_{0.75} = -u_{0.25}$ is valid with respect to the symmetry of the probability density function of a standardized normal distribution by zero. We observe a set of three quantile equations

$$y_{0.25} = \ln \frac{x_{0.25} - \theta}{\hat{\tau} - x_{0.25}} = \hat{\mu} + \hat{\sigma} u_{0.25} = \hat{\mu} - \hat{\sigma} u_{0.75}, \tag{8}$$

$$y_{0.50} = \ln \frac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}} = \hat{\mu} + \hat{\sigma} u_{0.50} = \hat{\mu}, \tag{9}$$

$$y_{0.75} = \ln \frac{x_{0.75} - \theta}{\hat{\tau} - x_{0.75}} = \hat{\mu} + \hat{\sigma} u_{0.75}. \tag{10}$$

Equation (9) is inserted into equations (8) and (10), whereby the system of three equations is reduced to two equations. We observe

$$\ln \frac{x_{0.25} - \theta}{\hat{\tau} - x_{0.25}} = \ln \frac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}} - \hat{\sigma} u_{0.75}, \tag{11}$$

$$\ln \frac{x_{0.75} - \theta}{\hat{\tau} - x_{0.75}} = \ln \frac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}} + \hat{\sigma} u_{0.75}. \tag{12}$$

We observe

$$\hat{\sigma} = \frac{\ln \frac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}} - \ln \frac{x_{0.25} - \theta}{\hat{\tau} - x_{0.25}}}{u_{0.75}}, \tag{13}$$

$$\hat{\sigma} = \frac{\ln \dfrac{x_{0.75} - \theta}{\hat{\tau} - x_{0.75}} - \ln \dfrac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}}}{u_{0.75}}. \tag{14}$$

We observe one equation with one unknown parameter τ

$$\hat{\sigma} = \frac{\ln \dfrac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}} - \ln \dfrac{x_{0.25} - \theta}{\hat{\tau} - x_{0.25}}}{u_{0.75}} = \frac{\ln \dfrac{x_{0.75} - \theta}{\hat{\tau} - x_{0.75}} - \ln \dfrac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}}}{u_{0.75}}, \tag{15}$$

so

$$\ln \frac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}} - \ln \frac{x_{0.25} - \theta}{\hat{\tau} - x_{0.25}} = \ln \frac{x_{0.75} - \theta}{\hat{\tau} - x_{0.75}} - \ln \frac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}}. \tag{16}$$

Hence

$$\ln(x_{0.50} - \theta) - \ln(\hat{\tau} - x_{0.50}) - \ln(x_{0.25} - \theta) + \ln(\hat{\tau} - x_{0.25}) =$$
$$= \ln(x_{0.75} - \theta) - \ln(\hat{\tau} - x_{0.75}) - \ln(x_{0.50} - \theta) + \ln(\hat{\tau} - x_{0.50}),$$

$$\ln \frac{(\hat{\tau} - x_{0.25})(\hat{\tau} - x_{0.75})}{(\hat{\tau} - x_{0.50})(\hat{\tau} - x_{0.50})} = \ln \frac{(x_{0.25} - \theta)(x_{0.75} - \theta)}{(x_{0.50} - \theta)(x_{0.50} - \theta)},$$

$$\frac{(\hat{\tau} - x_{0.25})(\hat{\tau} - x_{0.75})}{(\hat{\tau} - x_{0.50})^2} = \frac{(x_{0.25} - \theta)(x_{0.75} - \theta)}{(x_{0.50} - \theta)^2}.$$

We can calculate the constant C

$$C = \frac{(x_{0.25} - \theta)(x_{0.75} - \theta)}{(x_{0.50} - \theta)^2}, \tag{17}$$

so

$$\frac{(\hat{\tau} - x_{0.25})(\hat{\tau} - x_{0.75})}{(\hat{\tau} - x_{0.50})^2} = C. \tag{18}$$

We get a quadratic equation

$$(1-C)\hat{\tau}^2 + (2C\,x_{0.50} - x_{0.25} - x_{0.75})\hat{\tau} - C\,x_{0.50}^2 + x_{0.25} \cdot x_{0.75} = 0. \tag{19}$$

We observe the estimation of parameter τ

$$\hat{\tau} = \frac{-(2C\,x_{0.50} - x_{0.25} - x_{0.75}) + \sqrt{(2C\,x_{0.50} - x_{0.25} - x_{0.75})^2 + 4(1-C)(C\,x_{0.50}^2 - x_{0.25} \cdot x_{0.75})}}{2(1-C)}. \tag{20}$$

We obtain quantile estimations of four-parameter lognormal distribution

$$\hat{\tau} = \frac{-(2C\,x_{0.50} - x_{0.25} - x_{0.75}) + \sqrt{(2C\,x_{0.50} - x_{0.25} - x_{0.75})^2 + 4(1-C)(C\,x_{0.50}^2 - x_{0.25} \cdot x_{0.75})}}{2(1-C)}, \tag{21}$$

$$\hat{\sigma} = \frac{\ln \dfrac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}} - \ln \dfrac{x_{0.25} - \theta}{\hat{\tau} - x_{0.25}}}{u_{0.75}} = \frac{\ln \dfrac{x_{0.75} - \theta}{\hat{\tau} - x_{0.75}} - \ln \dfrac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}}}{u_{0.75}}, \tag{22}$$

$$\hat{\mu} = \ln \frac{x_{0.50} - \theta}{\hat{\tau} - x_{0.50}}, \tag{23}$$

where constant C is determined using equation (17).

1.4 Quantile Method of Parameter Estimation for Three-parameter Lognormal Distribution

The quantile parameter estimation is considerably simpler in the case of a three-parameter lognormal distribution. We again assume that we know the parameter θ because of the existence of the minimum wage institute.

As mentioned, the random variable (5) has a normal distribution with parameters μ and σ^2. We observe a set of two quantile equations

$$y_{0.50} = \ln(x_{0.50} - \theta) = \hat{\mu} + \hat{\sigma} u_{0.50} = \hat{\mu}, \tag{24}$$

$$y_{0.75} = \ln(x_{0.75} - \theta) = \hat{\mu} + \hat{\sigma} u_{0.75}. \tag{25}$$

Equation (24) is inserted into equation (25), whereby the system of two equations is reduced to one equation. We observe equation

$$\ln(x_{0.75} - \theta) = \ln(x_{0.50} - \theta) + \hat{\sigma} u_{0.75}. \tag{26}$$

Hence

$$\hat{\sigma} = \frac{\ln \dfrac{x_{0.75} - \theta}{x_{0.50} - \theta}}{u_{0.75}}. \tag{27}$$

We obtain quantile estimations of four-parameter lognormal distribution

$$\hat{\sigma} = \frac{\ln \dfrac{x_{0.75} - \theta}{x_{0.50} - \theta}}{u_{0.75}}, \tag{28}$$

$$\hat{\mu} = \ln(x_{0.50} - \theta). \tag{29}$$

1.5 Quality Evaluation of Acquired Models

We need to further assess the quality of the constructed models. We can use the known chi-square criterion for this purpose

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n\pi_i)^2}{n\pi_i}, \tag{30}$$

where n_i are the observed frequencies at individual intervals, π_i are the theoretical probabilities of belonging of the statistical unit to the i-th interval, n is the total sample size and $n \cdot \pi_i$ are the theoretical frequencies at individuals intervals, $i = 1, 2, \ldots, k$, k is the number of intervals.

However, the question of the suitability of a given curve for the wage distribution model is not a common mathematical-statistical problem, in which we test the null hypothesis

H_0: The sample comes from the assumed theoretical distribution

against an alternative hypothesis

H_1: non H_0,

because of goodness-of-fit tests, in the case of wage distribution, we often see that we work with large sample sizes. Therefore, the test would almost always lead to the rejection of the null hypothesis. This results not only from the fact that at such a large sample size, the strength of the test at the chosen significance level is such that test reveals all negligible deviations of the sample wage distribution and model, but this also results from the itself principle of the test construction. However, we are not interested is such small deviations, so only an approximate similarity between model and reality is sufficient, and we borrow a model (curve). The test criterion can be used only indicatively in this respect. When evaluating the suitability of the model, we have to proceed to a large extent subjectively and to rely on logical analysis and experience.

2 Results and Discussion

Table 3 presents the sample quartiles of wage distributions needed to estimate the parameters of four-parameter lognormal curves for wage intervals of 5,000 CZK, with the beginning of these curves being presented by the minimum wage in a given year. The sample upper quartiles and medians of Table 3 are also used to estimate the parameters of three-parameter lognormal curves for wage intervals with a width of 5,000 CZK again, with the minimum wage representing the beginning of these curves, too.

Table 3 shows the lowest wage level for the smallest enterprises of less than 10 employees. The wage level then increases with company size up to 1,000–4,999 employees, where it reaches its peak. Further, the wage level does not change fundamentally with the growth of number of the company's employees.

Tab. 3: Sample wage distribution quartiles (in CZK)

Year	Unit size	Quartiles		
		Lower quartile	Median	Upper quartile
2014	less than 10 employees	9,385	15,177	20,478
	from 10 to 49 employees	15,301	21,659	28,297
	from 50 to 249 employees	17,434	23,161	30,411
	from 250 to 999 employees	19,005	24,763	32,972
	from 1,000 to 4,999 employees	21,113	27,647	36,836
	more than 5,000 employees	21,120	27,285	36,246
2015	less than 10 employees	9,683	15,493	21,268
	from 10 to 49 employees	15,803	22,630	29,752
	from 50 to 249 employees	18,043	23,954	31,335
	from 250 to 999 employees	20,025	25,990	34,933
	from 1,000 to 4,999 employees	22,065	28,887	38,227
	more than 5,000 employees	21,693	28,419	37,577
2016	less than 10 employees	10,003	15,949	22,286
	from 10 to 49 employees	16,698	23,645	30,826
	from 50 to 249 employees	19,043	25,235	33,138
	from 250 to 999 employees	21,086	27,112	36,506
	from 1,000 to 4,999 employees	23,325	30,071	39,332
	more than 5,000 employees	22,910	29,636	38,641
2017	less than 10 employees	12,941	16,696	24,019
	from 10 to 49 employees	18,239	25,028	32,481
	from 50 to 249 employees	20,664	27,050	35,639
	from 250 to 999 employees	22,844	29,179	38,473
	from 1,000 to 4,999 employees	25,240	32,118	42,794
	more than 5,000 employees	24,949	32,355	43,814
2018	less than 10 employees	14,412	18,120	24,758
	from 10 to 49 employees	19,983	27,442	36,036
	from 50 to 249 employees	22,729	29,553	38,126
	from 250 to 999 employees	24,847	31,496	41,645
	from 1,000 to 4,999 employees	27,540	35,281	47,265
	more than 5,000 employees	27,430	35,923	48,498

Source: Own calculation

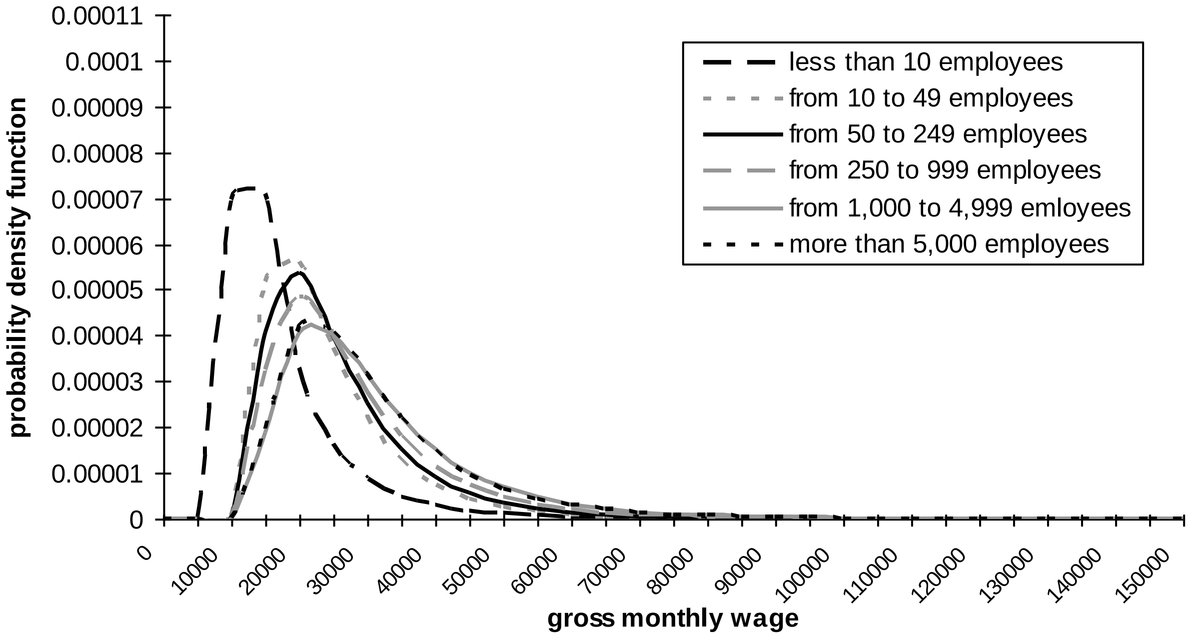
Tab 4: Parameter estimation of four-parameter lognormal curves

Year	Unit size	$\hat{\theta}$	Parameter estimation		
			$\hat{\tau}$	$\hat{\mu}$	$\hat{\sigma}$
2014	less than 10 employees	8,500	252,109,171	-10.538881	0.866471
	from 10 to 49 employees	8,500	1,945,849,366	-11.904105	0.605534
	from 50 to 249 employees	8,500	5,937,914,943	-12.911655	0.595708
	from 250 to 999 employees	8,500	22,372,565,586	-14.134463	0.605886
	from 1,000 to 4,999 employees	8,500	30,923,935,022	-14.294882	0.581146
	more than 5,000 employees	8,500	98,945,504,568	-15.476995	0.578248
2015	less than 10 employees	9,200	241,484,561	-10.555067	0.965431
	from 10 to 49 employees	9,200	1,898,823,320	-11.859251	0.630811
	from 50 to 249 employees	9,200	5,609,449,752	-12.848465	0.601470
	from 250 to 999 employees	9,200	59,014,840,613	-15.072512	0.633039
	from 1,000 to 4,999 employees	9,200	23,129,037,957	-13.976636	0.575665
	more than 5,000 employees	9,200	20,263,998,342	-13.868474	0.577748
2016	less than 10 employees	9,900	230,978,158	-10.550185	1.062661
	from 10 to 49 employees	9,900	2,084,191,626	-11.929221	0.623184
	from 50 to 249 employees	9,900	6,537,123,265	-12.962867	0.616272
	from 250 to 999 employees	9,900	15,486,387,102	-13.709881	0.645742
	from 1,000 to 4,999 employees	9,900	31,820,624,721	-14.271378	0.560186
	more than 5,000 employees	9,900	22,138,140,437	-13.930369	0.557278
2017	less than 10 employees	11,000	1,406,155,571	-12.416549	1.225533
	from 10 to 49 employees	11,000	2,824,251,278	-12.212725	0.631819
	from 50 to 249 employees	11,000	9,733,682,516	-13.315383	0.635494
	from 250 to 999 employees	11,000	57,164,360,790	-14.961164	0.612218
	from 1,000 to 4,999 employees	11,000	70,874,611,882	-15.026309	0.606628
	more than 5,000 employees	11,000	83,974,032,655	-15.184714	0.636859
2018	less than 10 employees	12,200	1,721,507,593	-12.580327	1.114949
	from 10 to 49 employees	12,200	3,573,818,184	-12.365081	0.662954
	from 50 to 249 employees	12,200	9,266,487,217	-13.188130	0.595232
	from 250 to 999 employees	12,200	23,899,457,061	-14.029485	0.626610
	from 1,000 to 4,999 employees	12,200	56,855,677,303	-14.896147	0.626309
	more than 5,000 employees	12,200	75,318,226,791	-14.970803	0.630582

Source: Own calculation

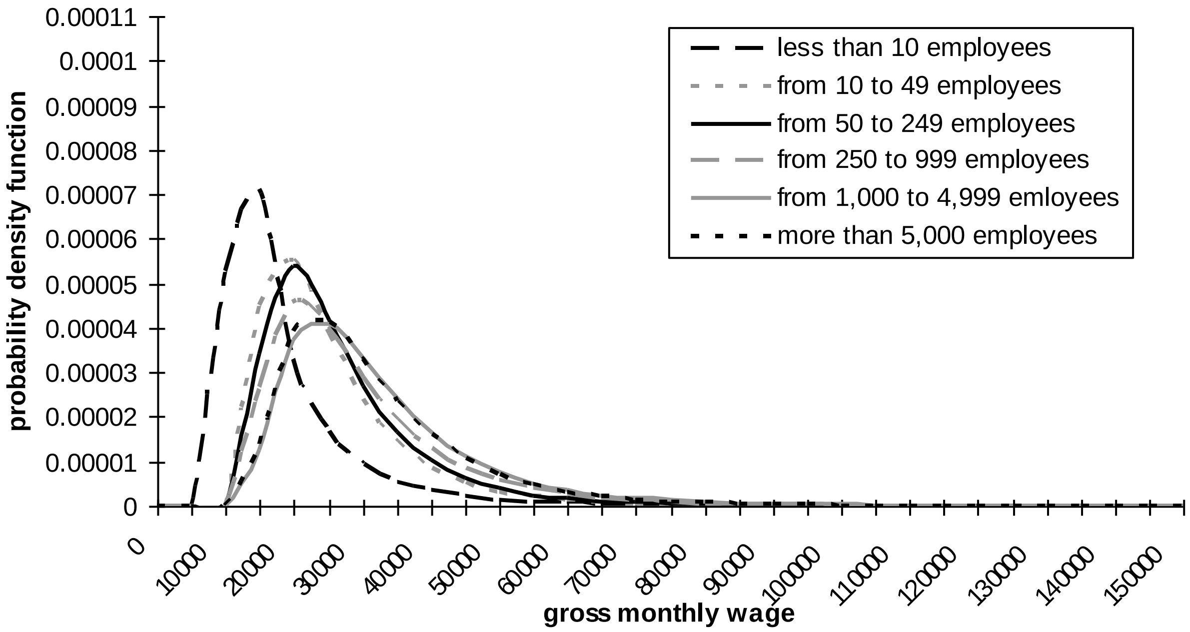
Table 4 shows the parameter estimations for four-parameter lognormal curves representing wage distribution models and Table 5 shows the parameter estimations for three-parameter lognormal curves as wage distribution models.

Fig. 7: Four-parameter lognormal models of gross monthly wage distribution by unit size in 2014 (gross monthly wage in CZK)



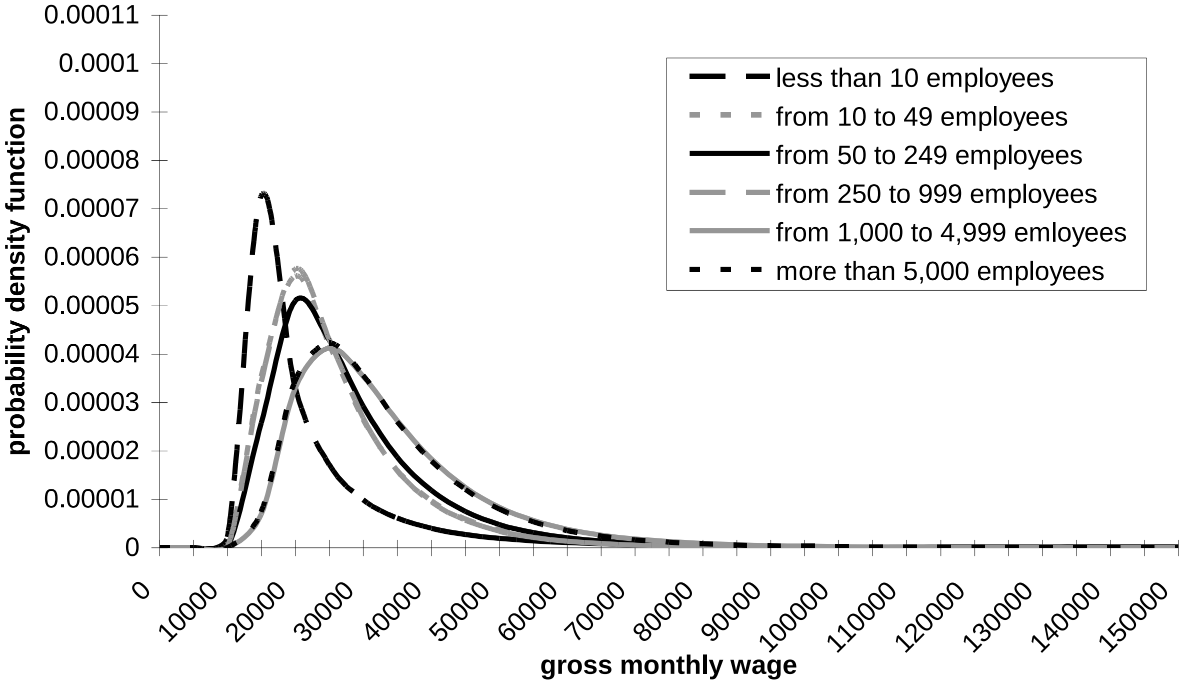
Source: Own calculation

Fig. 8: Four-parameter lognormal models of gross monthly wage distribution by unit size in 2015 (gross monthly wage in CZK)



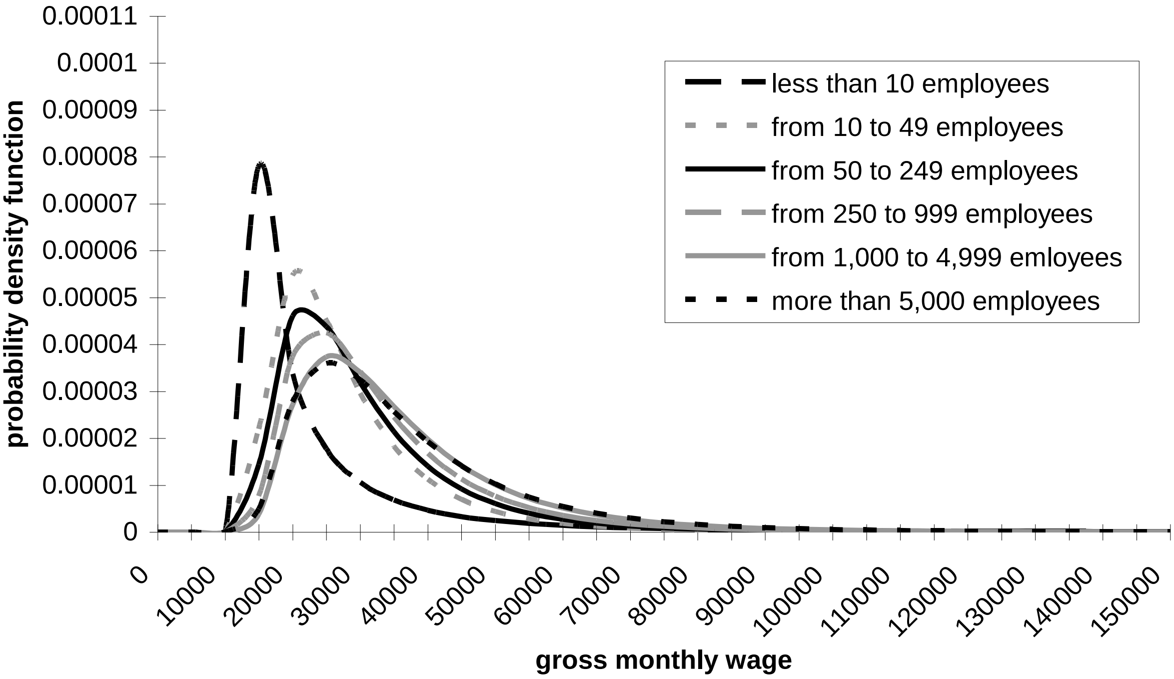
Source: Own calculation

Fig. 9: Four-parameter lognormal models of gross monthly wage distribution by unit size in 2016 (gross monthly wage in CZK)



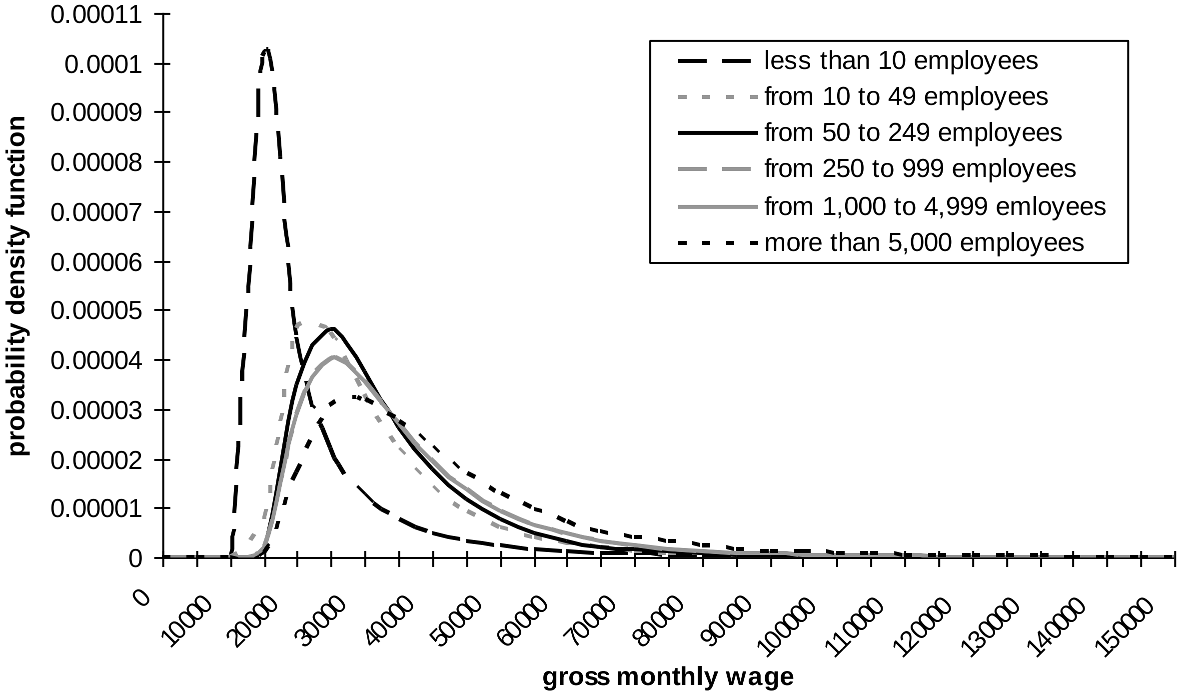
Source: Own calculation

Fig. 10: Four-parameter lognormal models of gross monthly wage distribution by unit size in 2017 (gross monthly wage in CZK)



Source: Own calculation

Fig. 11: Four-parameter lognormal models of gross monthly wage distribution by unit size in 2018 (gross monthly wage in CZK)



Source: Own calculation

Figures 7–11 represent four-parameter models of wage distributions in 2014–2018. These figures show a change in the shape of wage distribution models with the growth of enterprise size, while the shape of wage distribution models is clearly different for the smallest enterprises of less than 10 employees. Figures 12–16 represent three-parameter models of wage distributions in 2014–2018. The clear difference in the shape of wage distribution models for the smallest enterprises of less than 10 employees is apparent from these figures, too.

Table 6 shows the calculated chi-square test criterion values for four-parameter and three-parameter lognormal curves. We can see from this table that the chi-square test criterion is slightly smaller for four-parameter curves than for three-parameter curves in most cases. However, this difference is not essential so much in majority of cases. The chi-square test criterion value is the same for the four-parameter lognormal curves and three-parameter lognormal curves for six from thirty wage models.

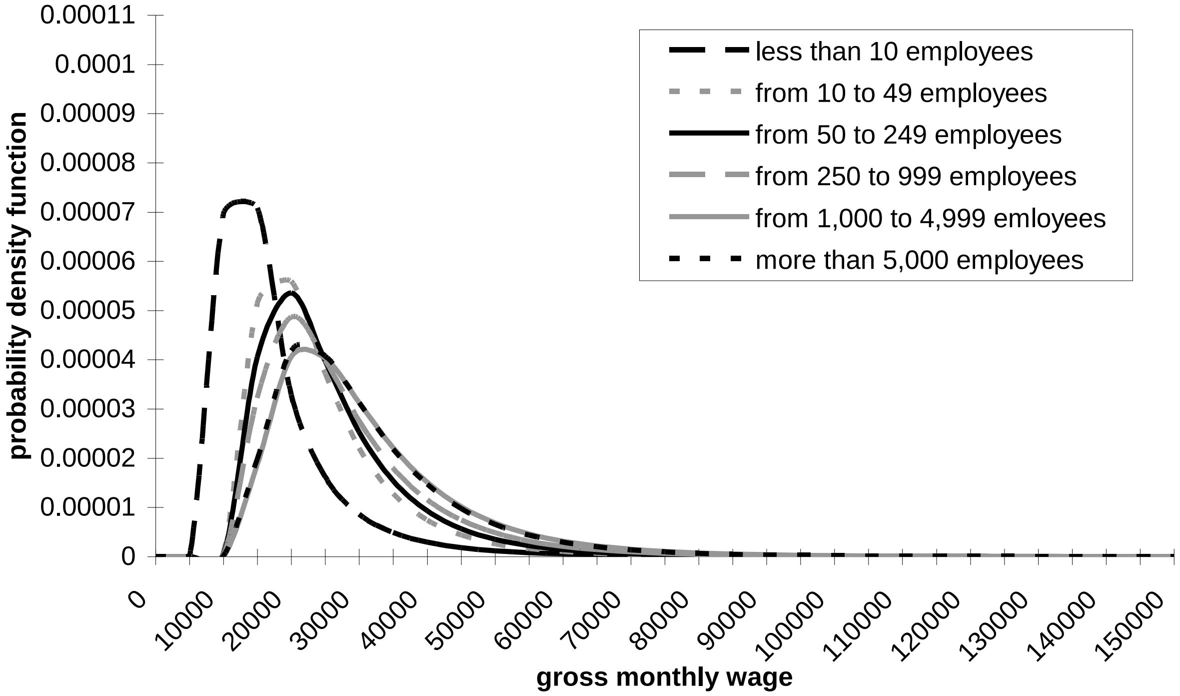
Thus, it can be stated in principle that the four-parameter lognormal curves provided slightly more accurate wage distribution models than three-parameter lognormal curves.

Tab 5: Parameter estimation of three-parameter lognormal curves

Year	Unit size	$\hat{\theta}$	Parameter estimation $\hat{\mu}$	$\hat{\sigma}$
2014	less than 10 employees	8,500	8.806432	0.866440
	from 10 to 49 employees	8,500	9.484848	0.605529
	from 50 to 249 employees	8,500	9.592965	0.595707
	from 250 to 999 employees	8,500	9.696637	0.605885
	from 1,000 to 4,999 employees	8,500	9.859914	0.581146
	more than 5,000 employees	8,500	9.840840	0.578248
2015	less than 10 employees	9,200	8.747185	0.965396
	from 10 to 49 employees	9,200	9.505238	0.630806
	from 50 to 249 employees	9,200	9.599250	0.601468
	from 250 to 999 employees	9,200	9.728543	0.633039
	from 1,000 to 4,999 employees	9,200	9.887718	0.575664
	more than 5,000 employees	9,200	9.863637	0.577748
2016	less than 10 employees	9,900	8.707580	1.062620
	from 10 to 49 employees	9,900	9.528415	0.623179
	from 50 to 249 employees	9,900	9.637892	0.616270
	from 250 to 999 employees	9,900	9.753345	0.645741
	from 1,000 to 4,999 employees	9,900	9.912002	0.560185
	more than 5,000 employees	9,900	9.890198	0.557278
2017	less than 10 employees	11,000	8.647558	1.225524
	from 10 to 49 employees	11,000	9.548775	0.631815
	from 50 to 249 employees	11,000	9.683472	0.635493
	from 250 to 999 employees	11,000	9.808032	0.612218
	from 1,000 to 4,999 employees	11,000	9.957869	0.606628
	more than 5,000 employees	11,000	9.969059	0.636859
2018	less than 10 employees	12,200	8.686112	1.114938
	from 10 to 49 employees	12,200	9.631811	0.662950
	from 50 to 249 employees	12,200	9.761537	0.595231
	from 250 to 999 employees	12,200	9.867635	0.626609
	from 1,000 to 4,999 employees	12,200	10.046775	0.620002
	more than 5,000 employees	12,200	10.074184	0.630581

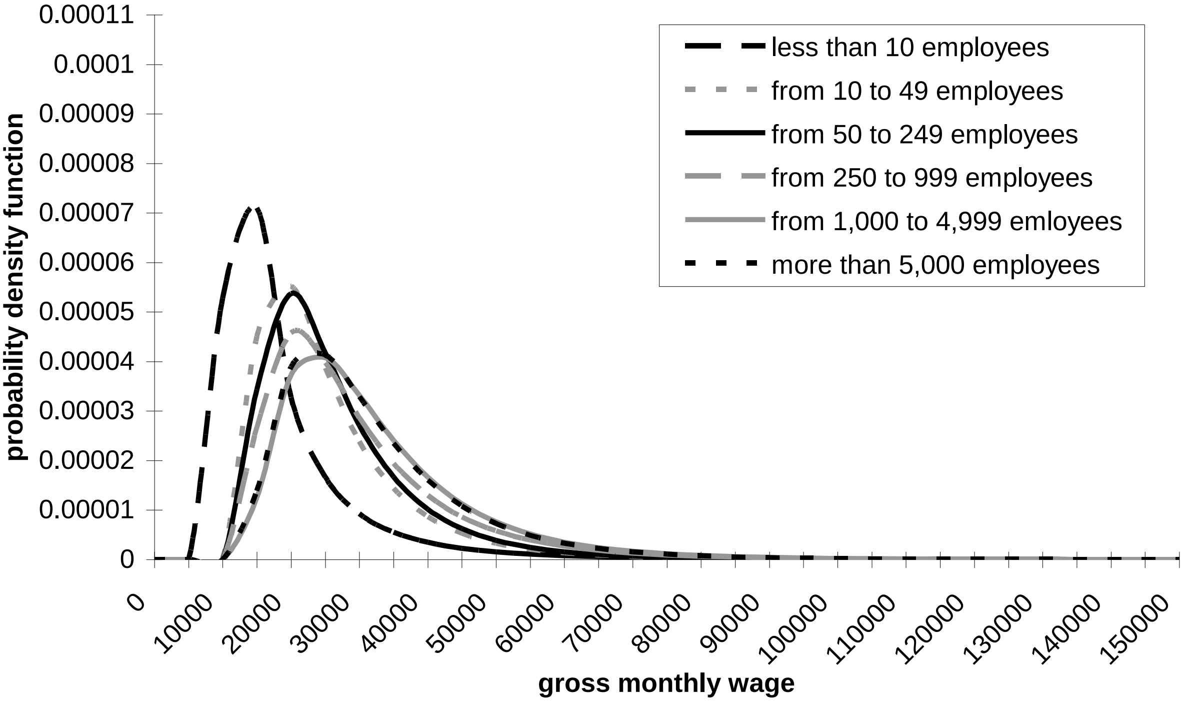
Source: Own calculation

Fig. 12: Three-parameter lognormal models of gross monthly wage distribution by unit size in 2014 (gross monthly wage in CZK)



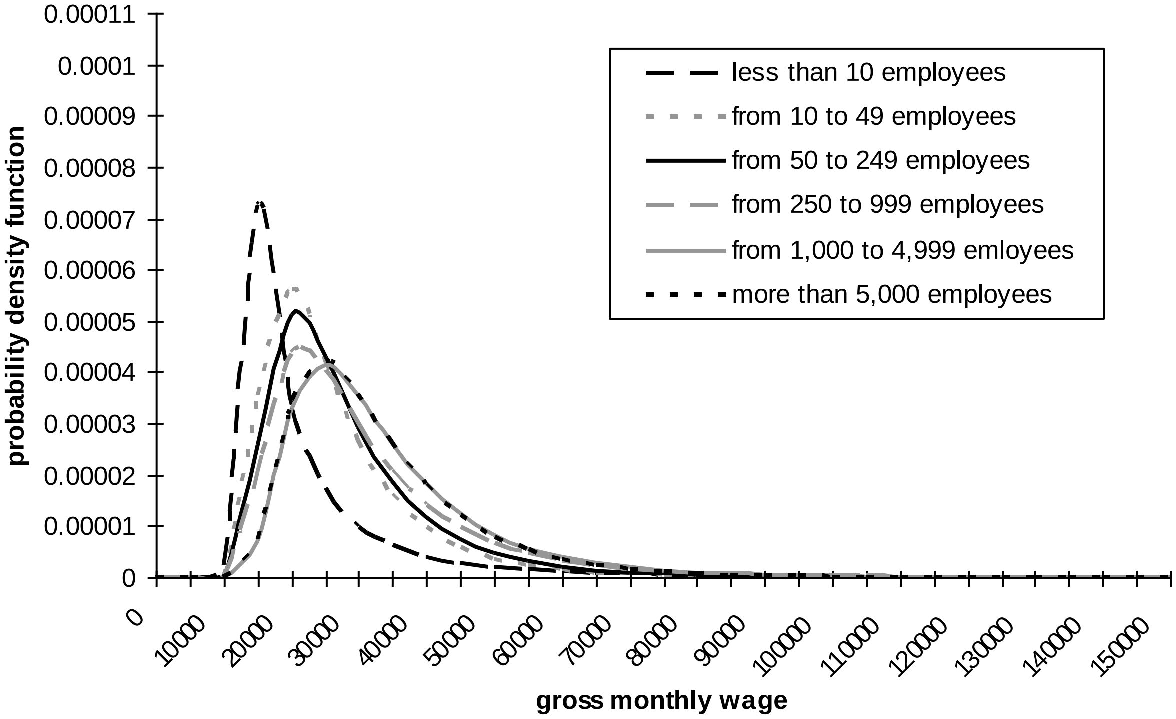
Source: Own calculation

Fig. 13: Three--parameter lognormal models of gross monthly wage distribution by unit size in 2015 (gross monthly wage in CZK)



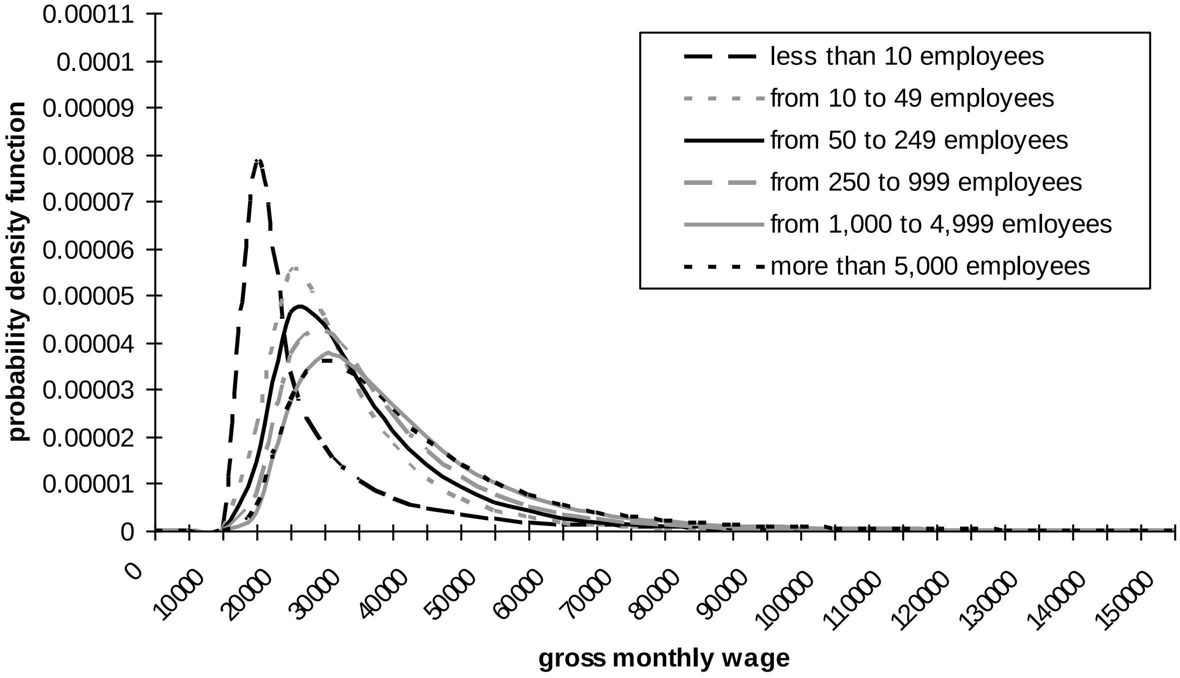
Source: Own calculation

Fig. 14: Three--parameter lognormal models of gross monthly wage distribution by unit size in 2016 (gross monthly wage in CZK)



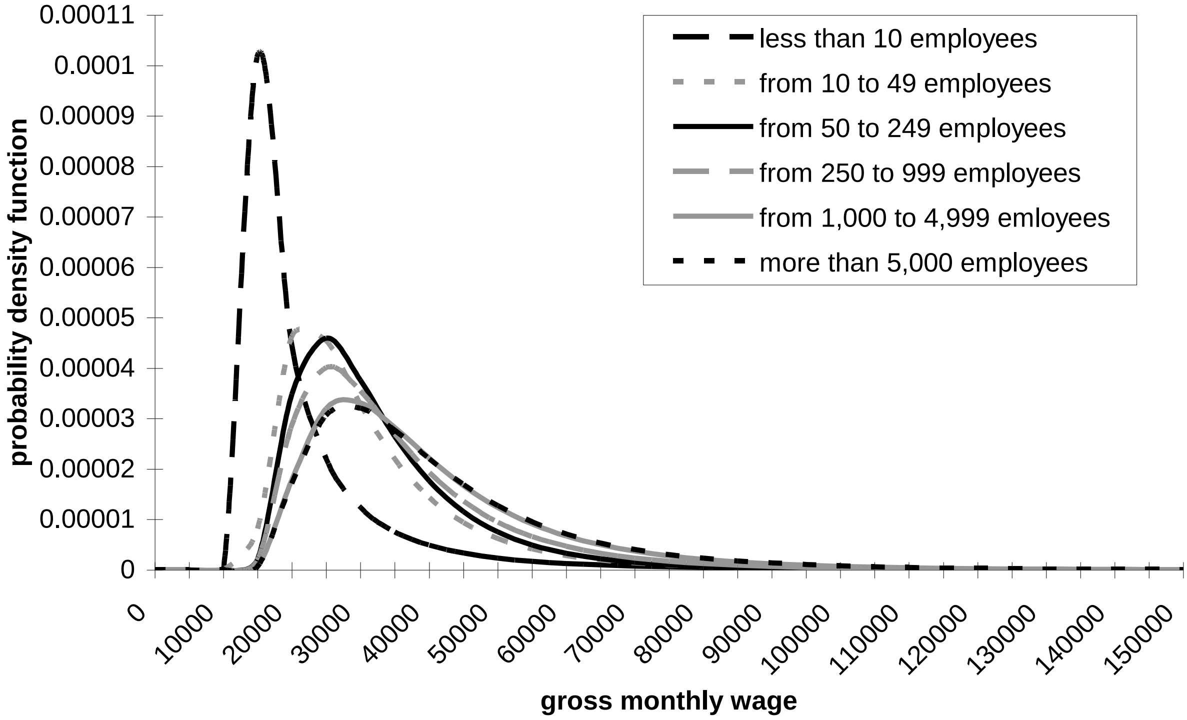
Source: Own calculation

Fig. 15: Three--parameter lognormal models of gross monthly wage distribution by unit size in 2017 (gross monthly wage in CZK)



Source: Own calculation

Fig. 16: Three--parameter lognormal models of gross monthly wage distribution by unit size in 2018 (gross monthly wage in CZK)



Source: Own calculation

We register k = 10 wage intervals in all wage distributions. We estimate p = 3 parameters in the case of four-parameter lognormal distribution and p = 2 parameters in the case of three-parameter lognormal distribution (parameter θ is known in both cases). Assuming the null hypothesis is valid, the test criterion (30) has an asymptotic chi-square distribution of $\nu = k - p - 1$ degrees of freedom. This is $\nu = 6$ in the case of four-parameter lognormal distribution and $\nu = 7$ in the case of three-parameter lognormal distribution. Table 7 contains the critical values at $\alpha = 0.05$ (0.01 or 0.10) significance level.

This is clear that the test leads to the rejection of the null hypothesis assuming a four-parameter or three-parameter lognormal distribution of gross monthly wage in all of thirty cases and at all significance levels considered (5%, 1% and 10%). In all cases, the test leads to the accepting an alternative hypothesis that the gross monthly wage distribution is different than the null hypothesis assumes. As already mentioned, this situation occurs with a regard to the large sample sizes that we encounter in relation to wage distributions, where the test reveals all the slightest differences between the sample distribution and model. However, we are not interested in such small deviations.

Tab 6: Test criterion values

		Test criterion values	
Year	Unit Size	Four-parameter distribution	Three-parameter distribution
2014	less than 10 employees	52,701	52,706
	from 10 to 49 employees	585,678	585,695
	from 50 to 249 employees	323,095	323,095
	from 250 to 999 employees	201,282	201,284
	from 1,000 to 4,999 employees	361,967	361,967
	more than 5,000 employees	7,000	7,000
2015	less than 10 employees	79,541	79,549
	from 10 to 49 employees	1,017,368	1,017,406
	from 50 to 249 employees	698,039	698,056
	from 250 to 999 employees	307,522	307,522
	from 1,000 to 4,999 employees	1,547,992	1,548,024
	more than 5,000 employees	39,217	39,218
2016	less than 10 employees	125,904	125,917
	from 10 to 49 employees	2,874,134	2,874,287
	from 50 to 249 employees	1,626,050	1,626,090
	from 250 to 999 employees	622,184	622,194
	from 1,000 to 4,999 employees	2,818,638	2,818,722
	more than 5,000 employees	388,759	388,759
2017	less than 10 employees	99,243	99,247
	from 10 to 49 employees	64,213,482	64,218,766
	from 50 to 249 employees	20,721,610	20,721,977
	from 250 to 999 employees	62,788, 256	62,788,391
	from 1,000 to 4,999 employees	20,025,793	20,025,801
	more than 5,000 employees	511,089	511,090
2018	less than 10 employees	25,847	25,855
	from 10 to 49 employees	869,035	869,061
	from 50 to 249 employees	1,144,144	1,144,153
	from 250 to 999 employees	401,253	401,258
	from 1,000 to 4,999 employees	83,433	83,440
	more than 5,000 employees	21,705	21,705

Source: Own calculation

Tab 7: Critical ranges at α = 0.05 (0.01 or 0.10) significance level

α	Four-parameter lognormal distribution	Three-parameter lognormal distribution
0.05	$W\alpha = \{\chi^2: \chi^2 \geq 12.592\}$	$W\alpha = \{\chi^2: \chi^2 \geq 14.067\}$
0.01	$W\alpha = \{\chi^2: \chi^2 \geq 16.812\}$	$W\alpha = \{\chi^2: \chi^2 \geq 18.475\}$
0.10	$W\alpha = \{\chi^2: \chi^2 \geq 10.645\}$	$W\alpha = \{\chi^2: \chi^2 \geq 12.017\}$

Source: Own calculation

Conclusion

Within the point parameter estimation using quantile method, four-parameter lognormal curves provided more accurate models of wage distributions than three-parameter lognormal curves in the vast of majority of wage distributions. Four-parameter and three-parameter lognormal curves produced the same accuracy in the results only in six from thirty wage distributions. Three-parameter lognormal curves were not more accurate than four-parameter lognormal curves in either case.

The shape of model wage distributions together with the growing level of wage distributions are changing significantly with the growing size of the firm, up to 1,000 employees. For companies over 1,000 employees, the shape and level of wage distributions do not change strongly. Some other authors deal with the issue if household incomes, employee wages or in total labour market, see for example (Landmesser, 2019; Łukasiewicz, Karpio & Orłowski, 2018; Malkina, 2019; Pernica, 2017, and Megyesiová, & Rozkošová, 2018).

Acknowledgment

This paper was subsidized by the funds of institutional support of a long-term conceptual advancement of science and research number IP400040 at the Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic.

References

Bartošová, J., & Bína, V. (2018). Differences in Income Distribution Given by the Structure of Household. In Váchová, L., Kratochvíl, V. (Eds.), the 36th international conference *Mathematical Methods in Economics (MME 2018)* [flashdisk, online]. Jindřichův Hradec, 12.09.2018–14.09.2018. Praha: MatfyzPress, pp. 19–24. ISBN 978-80-7378-371-6. URL: https://mme2018.fm.vse.cz/wp-content/uploads/2018/09/MME2018-Electronic_proceedings.pdf.

Bartošová, J., & Longford, N. T. (2014). A Study of Income Stability in the Czech Republic by Finite Mixtures. *Prague Economic Papers*, *23*(3), pp. 330–348. ISSN 1210-0455. DOI: 10.18267/j.pep.487.

Landmesser, J. (2019). Differences in Income Distributions for Men and Women in the European Union Countries. *Equilibrium*, *14*(1), pp. 81–98. ISSN 1689-765X. DOI: https://doi.org/10.24136/eq.2019.004.

Langhamrová, J., & Šimpach, O. (2012). Příjmové nerovnosti podnikatelské a nepodnikatelské sféry v České republice (The Income Inequality of Business and Non-business Sphere in the Czech Republic). In Pauhofová, I., Želinský, T. (Eds.), international conference *Nerovnosť a chudoba v Európskej únii a na Slovensku* [online]. Herľany, 26.09.2012–26.09.2012. Košice: Ekonomická fakulta TU, pp. 113–122. ISBN 978-80-553-1225-5. URL:
http://www3.ekf.tuke.sk/NaRE2012/subory/workshop/Herlany_Zbornik_web.pdf.

Łukasiewicz, P., Karpio, K. & Orłowski, A. (2018). Two-Component Structure of Household Income Distributions in Poland. *Equilibrium, 13*(4), pp. 603–622. ISSN 1689-765X. DOI: https://doi.org/10.24136/eq.2018.029.

Malá, I. (2017). L-moments of a Finite Mixture of Probability Distributions. In Löster, T., Pavelka, T. (Eds.), the 11th international conference *International Days of Statistics and Economics (MSED 2017)* [online]. Praha, 14.09.2017–16.09.2017. Slaný: Melandrium, pp. 929–938. ISBN 978-80-87990-12-4.
URL: https://msed.vse.cz/msed_2017/article/65-Mala-Ivana-paper.pdf.

Malá, I. (2016). Properties of Moment Method Estimates Based on L-moments from Right Censored data. In Löster, T., Pavelka, T. (Eds.), the 10th international conference *International Days of Statistics and Economics (MSED 2016)* [online]. Praha, 08.09.2016 – 10.09.2016. Slaný: Melandrium, pp. 1159–1169. ISBN 978-80-87990-10-0.
URL: https://msed.vse.cz/msed_2016/article/245-Mala-Ivana-paper.pdf.

Malkina, M. (2019). Spatial Wage Inequality and Its Sectoral Determinants: the Case of Modern Russia. *Oeconomia Copernicana, 10*(1), pp. 69–87. ISSN 2083-1277. DOI: https://doi.org/10.24136/oc.2019.004.

Marek, L. (2018). Average Wages and Wage Quantiles in the Czech Republic. In Löster, T., Pavelka, T. (Eds.), the 12th international conference *International Days of Statistics and Economics (MSED 2018)* [online]. Prague, 06.09.2018–08.09.2018. Slaný: Melandrium, pp. 1186–1194. ISBN 978-80-87990-14-8.
URL: https://msed.vse.cz/msed_2018/article/19-Marek-Lubos-paper.pdf.

Marek, L., Doucek, P., & Nedomová, L. (2018). Development of ICT Professional Wages in Czech Republic. In Doucek, P., Chroust, G., (Eds.), the 26th international conference *Interdisciplinary Information Management Talks (IDIMT 2018)* [online]. Kutná Hora, 05.09.2018–07.09.2018. Linz: Trauner Verlag Universität, pp. 49–56. ISBN 978-3-99062-339-8.
URL: https://idimt.org/wp-content/uploads/proceedings/IDIMT_proceedings_2018.pdf.

Maryška, M., Nedomová, K., & Doucek, P. (2017). Wages of ICT Professionals in Area of North-East. In the 13th international conference *Liberec Economic Forum (LEF 2017)*.

Liberec, 11.09.2017–13.09.2017. Liberec: Technical University of Liberec, pp. 222–233. ISBN 978-80-7494-349-2.
Megyesiová, S. & Rozkošová, A. (2018). Success of Visegrad Group Countries in the Field of Labour Market. *Journal of Applied Economic Sciences*, *13*(2), pp. 369–377. ISSN 2393-5162. URL: http://cesmaa.org/Docs/JAES%20Spring%20Volume%20XIII%20Issue%202(56)2018%20(1).pdf
Pavelka, T., & Löster, T. (2014). The Development of Unemployment in the European Union's Labor Market Due to the Recent Economic Crisis, Using Cluster Analysis. In the 14th international conference *Eurasia-Business-and-Economics-Society (EBES 2014)* [online]. Barcelona, 23.10.2014 – 25.10.2014. Cham: Springer, pp. 323–337. ISBN 978-3-319-22596-8. DOI: 10.1007/978-3-319-22596-8_23.
Pavelka, T., & Löster, T. (2013). Flexibility of the Czech Labour Market from a Perspective of the Employment Protection Index. In Löster, T., Pavelka, T. (Eds.), the 7th international conference *International Days of Statistics and Economics (MSED 2013)* [online]. Praha, 19.09.2013–21.09.2013. Slaný: Melandrium, pp. 1090–1099. ISBN 978-80-86175-87-4. URL: http://msed.vse.cz/files/2013/7-Pavelka-Tomas-paper.pdf.
Pechrová, M., & Šimpach, O. (2013). The Development Potential of the Regions of the EU. In international conference *Region v rozvoji společnosti 2013* [CD]. Brno, 10.10.2013 – 10.10.2013. Brno: MU v Brně, pp. 257–267. ISBN 978-80-7375-884-4. URL: http://www.icabr.com/respo/RESPO_2013.pdf.
Pernica, M. (2017). Evaluation of the Adequacy of Government Minimum Wage Valorization Policy in the Czech Republic in 2017 in the European Context. *Oeconomia Copernicana*, *8*(1), pp. 21–36. ISSN 2083-1277. DOI: https://doi.org/10.24136/oc.v8i1.2.
Sipková, Ľ., & Sipko, J. (2012). Regionálne porovnanie nerovnosti miezd zamestnancov SR aplikáciou kappa kvantilových modelov (Regional Comparison of the Income Inequality of Employees in the Slovak Republic with the Application of Kappa Quantile Models). In Pauhofová, I., Želinský, T. (Eds.), international conference *Nerovnosť a chudoba v Európskej únii a na Slovensku* [online]. Herľany, 26.09.2012–26.09.2012. Košice: Ekonomická fakulta TU, pp. 87–97. ISBN 978-80-553-1225-5.
URL: http://www3.ekf.tuke.sk/NaRE2012/subory/workshop/Herlany_Zbornik_web.pdf.

Contact

doc. Ing. Diana Bílková, Dr.
University of Economics, Prague
Faculty of Informatics and Statistics
Department of Statistics and Probability
Sq. W. Churchill 1938/4
130 67 Prague 3
Czech Republic
E-mail: bilkova@vse.cz