ANALYSIS OF THE DEVELOPMENT OF WAGE DISTRIBUTION AT THE BEGINNING OF THE 3RD MILLENIUM ACCORDING TO THE GENDER

Diana Bílková

Abstract

Attention of economics is necessary to focus on methods to predict the demand for goods and services in a market economy. Changes in many important indicators occur in the Czech Republic after the transition from a centrally planned economy to a market economy. This text deals with the analysis of the development of description characteristics of wage distribution in the years 2003-2010 in the Czech Republic in total for all employees together and futher broken down by gender. The researched variable was the gross monthly wage in CZK. The author analyses the development of wage distribution in time and a special attention is paid to the differences in the development of wage distribution between males and females. Forecasts of wage distribution of males and females were constructed based on past development of wage distribution and using the three-parameter lognormal distribution.

Key words: Wage distribution, wage distribution according to the gender, characteristics of location and differentiation, development of wage, forecasts of wage distribution

JEL Code: J31, G01, H24

Introduction

Major changes occurred in the Czech Republic during the period since the transition of the Czech economy from a centrally planned economy to a market economy. These changes related not only to the economy, but also to the areas, which are connected with economy. Wages and incomes of population increase and their differentiation increases markedly, too. The group of people with very high wages and incomes was created and it grows stronger gradually. The financial situation of Czech households changes, see for example Bartošová (2009), the purchasing power of Czech households changes, see Bartošová & Forbelská (2010). Unemployment appeared and it develops gradually, see Miskolczi, Langhamrová & Langhamrová (2011) or Löster & Langhamrová (2011). Demographic structure of the population changes, see for example Fiala & Langhamrová (2011), Fiala & Langhamrová (2010) or Malá (2011).

We can use various methods for processing data on wages and incomes, see for example Bartošová & Bína (2007) or Řezanková, Löster & Hušek (2011). L-moments and L-moment method of parameter estimation are described in numerous professional publications and articles, for example Guttman (1993), Hosking (1990), Hosking & Wales (1997) or Kyselý & Picek (2007).

The studied variable for this research was a gross monthly wage (in CZK) and we came out of a set of employees classified by gender. The data come from the official website of the Czech Statistical Office. There are interval frequency distributions with extreme open intervals presented by the CSO website table "Percentages of employees in the bands of gross monthly wages by gender". Table 1 provides information about the sample sizes of the individual sets used. Because the nominal wage is an analysed variable, Table 2 presents data on the average annual inflation for the period. The data were processed using the statistical programs SAS and Statgraphics and using the spreadsheet Microsoft Excel.

Tab. 1: Sample sizes of the wage distribution divided by the gender

Gender	Year							
	2003	2004	2005	2006	2007	2008	2009	2010
Males	559,863	711,551	769,802	813,821	858,656	875,139	846,028	850,788
Females	459,071	692,945	745,725	800,551	814,842	836,672	805,478	812,041
Total	1,018,934	1,404,496	1,515,527	1,614,372	1,673,498	1,711,811	1,651,506	1,662,829

Source: http://www.czso.cz

Tab. 2: Average annual rate of inflation in the years 2003-2010 (in %)

	Year						
	2004	2005	2006	2007	2008	2009	2010
Rate of inflation	2.8	1.9	2.5	2.8	6.3	1.0	1.5

Source: http:// www.czso.cz

1 Development of wage distribution of males and females in the years 2003-2010

Tables 3 and 4 provide an overview on the development of sample characteristics of wage distribution of males and females in the monitored period. Wage distribution of males and females changes in time. The absolute amount of wages grows at a relatively fast pace in the period 2003 − 2008. Wage growth virtually stopped in the Czech Republic in the period immediately after the onset of the economic crisis see Table 5 and Figure 1. Figure 1 presents

Tab. 3: Sample characteristics of the level, variability and scope of the distribution − Males

Characteristics	Year							
	2003	2004	2005	2006	2007	2008	2009	2010
Arithmetic mean (CZK)	20,857	21,992	23,112	24,187	26,715	28,633	28,602	29,064
Median (CZK)	18,241	19,344	20,281	21,199	22,855	24,450	24,164	24,695
Medial (CZK)	21,982	23,078	25,420	25,590	27,888	29,952	30,048	30,655
Standard deviation (CZK)	9,987	11,068	11,465	11,849	15,879	16,736	17,108	17,172
Coefficient of variation (%)	47.88	50.33	49.61	48.99	59.44	58.45	59.81	59.08
Skewness	1.530	1.770	1.612	1.502	2.542	2.362	2.336	2.267

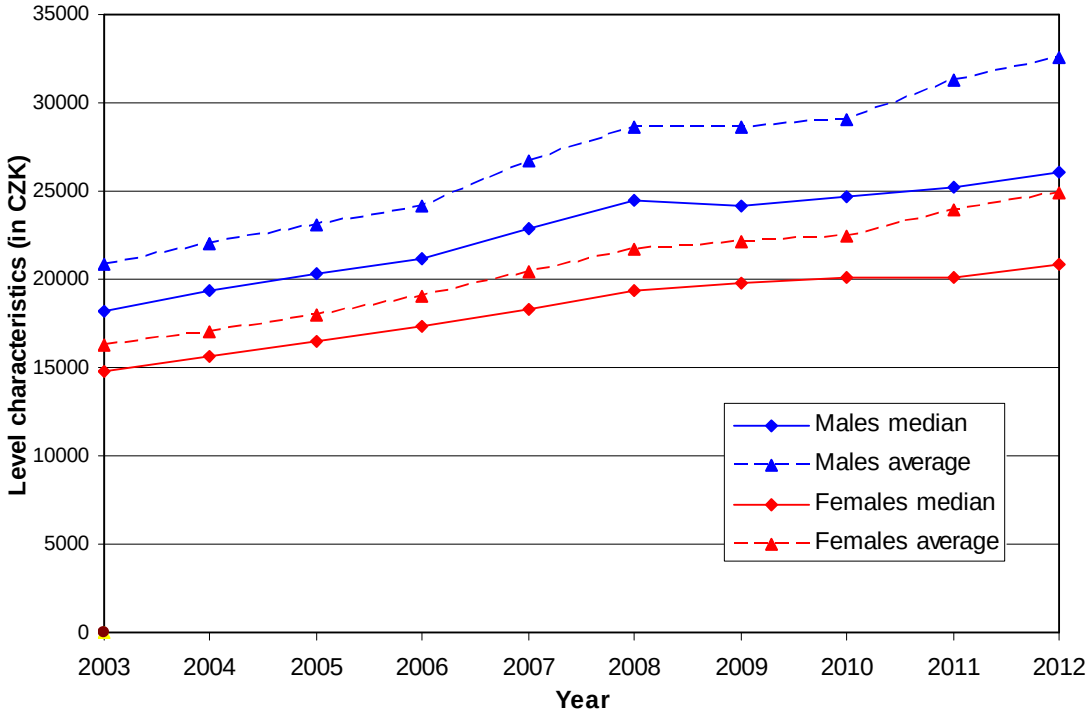
Source: Own calculations

Tab. 4: Sample characteristics of the level, variability and scope of the distribution − Females

Characteristics	Year							
	2003	2004	2005	2006	2007	2008	2009	2010
Arithmetic mean (CZK)	16,313	16,996	17,952	19,042	20,382	21,655	22,124	22,405
Median (CZK)	14,831	15,642	16,454	17,311	18, 328	19,388	19,786	20,085
Medial (CZK)	17,583	18,582	19,598	20,599	22,162	23,373	24,099	24,299
Standard deviation (CZK)	7,752	8,442	8,866	9,202	11,463	12,076	12,328	12,391
Coefficient of variation (%)	47.52	49.67	49.38	48.32	56.25	55.77	55.72	55.30
Skewness	1.930	2.001	1.908	1.926	2.897	2.817	2.675	2.684

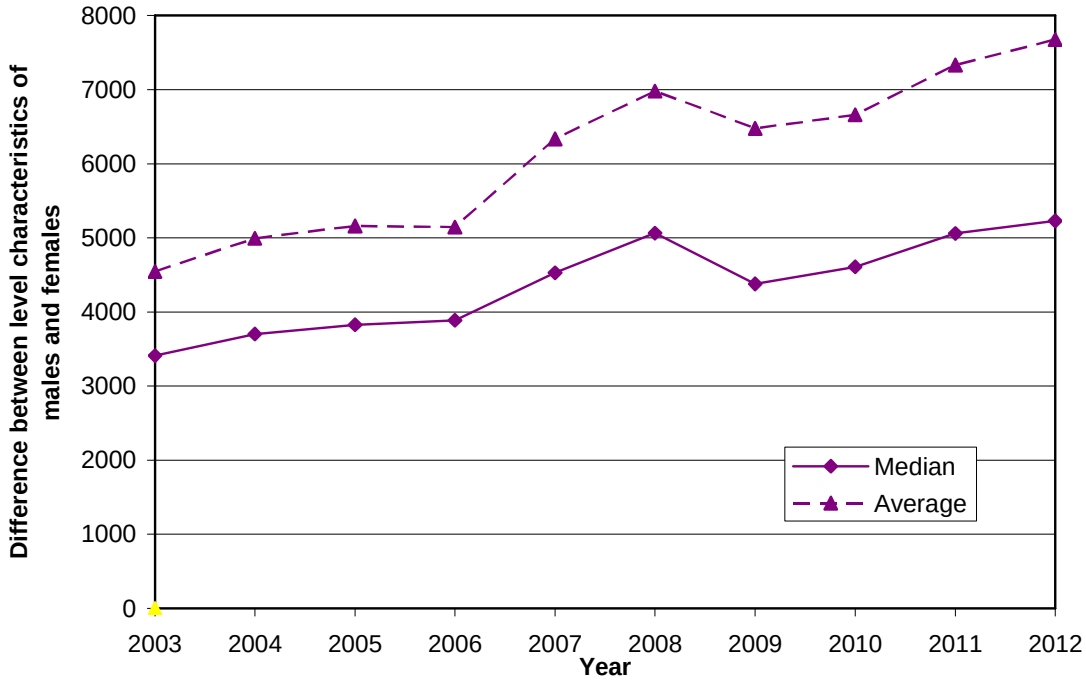
Source: Own calculations

Fig. 1: Development of sample level characteristics of males and females (in CZK)



Source: Own calculations

Fig. 2: Development of the diference between sample level characteristics of males and females (in CZK)



Source: Own calculations

Tab. 5: Annual growth coefficients of middle gross monthly wage in the period 2003 − 2010; annual average growth coefficients of middle gross monthly wage in the period 2003 − 2008 before economic recession and in the period 2008 − 2010 during economic recession

		Year						
	Categhory	2004	2005	2006	2007	2008	2009	2010
Gender	Males	1.0605	1.0484	1.0453	1.0781	1.0698	0.9883	1.0220
		1.0603					1.0050	
	Females	1.0547	1.0519	1.0521	1.0587	1.0578	1.0205	1.0151
		1.0550					1.0178	

Source: Own calculations

the development of middle and average wage of males and females during the monitored period including their forecasts for 2011 and 2012. These forecasts were constructed based on the last development of wage distribution during the period of the years 2003 − 2008, i.e. before the economic crisis, and during the period 2008 − 2010, i.e. the period of economic crisis. In the future, we can expect some recovery of wage growth. Figure 2 presents the differences between sample level characteristics of males and females. We can see from this figure that this difference decreased markedly at the beginning of economical recession in 2008. Then it increased very slowly. We can assume accretion of these differences for 2011 and 2012. Absolute variability of wages increases gradually throughout the researched period. Relative wage variability rather oscillates with increasing tendency.

2 Wage dependence on gender

The dependent variable is the gross monthly wage (in CZK) and the independent variable is the gender. We therefore research the dependence of numerical variable on categorical variable. It is possible to use the test one-factor analysis of variance, which is known as ANOVA for this purpose, provided that certain conditions are met. Table 6 provides an overview of the wage differentiation in terms of intragroup and intergroup variability. We can see from this table that in the process of decomposition of total variability into intragroup and intergroup components, intragroup variability clearly dominates over intergroup variability (the source of the wage dependence on the gender). Total variability represents the variability of wages of individual employees around the total average wage calculated for both genders together. Intragroup variability, on the other hand, is the variability of wages of individual employees around the average wage in a respective category of gender, intergroup variability being the variability of average wages in various categories of gender around the total average wage calculated together for both categories of gender. The sum of intragroup and intergroup

variability yields total variability; i.e. the sum of average variance and the variance of averages is equal to total variance. As already mentioned above, the source of dependence of the wage on the gender is the variability of average wages in various categories of gender around the total average wage for all employees together. This means that the more intergroup variability contributes to total variability, i.e. the less intragroup variability contributes to total variability (the sum of intra and inter-group variability yielding total variability), the stronger the wage dependence on the level of gender and vice versa. Thus we can see from the decomposition of total variability into two individual components in Table 6 that the dependence of the wage on the gender is not too strong in all monitored years.

Table 7 is directly linked to Table 6 giving an overview of the statistical dependence of gross monthly wage on the gender. Having applied a test analysis of variance known as ANOVA (one-factor), the above mentioned dependence was verified for each year of the period. The gross monthly wage dependence upon the gender was demonstrated for virtually any commonly used significance level ($\alpha = 0.05$) with regard to large sample sizes typical for the research of the wage distribution. The critical value for a given number of five decimal places remains consistent in all years of the research period, probably due to large sample sizes used. The same is also valid for the so called P-value, which is the smallest significance level at which we can still reject the tested (null) hypothesis.

We can see from Table 7 that the values of test criterion *F* amply exceed the critical value in all cases. This is because such large sample sizes that are used in the case of wage distributions equate to a very high power of the test. Thus the test leads unambiguously to the rejection of the tested hypothesis, assuming the independence of wages on the gender. The same conclusion has to be drawn from the comparison of P-value and the significance level. We can see from Table 7 that the significance level $\alpha = 0.05$ clearly exceeds the corresponding P-value in all cases. Let us add that the significance level α presents the probability of error of the first type, i.e. the probability that we reject the tested hypothesis (hypothesis of independence), although it is valid. It is evident from Table 7 that the tested hypothesis is rejected, using any significance level in this case ($\alpha = 0.10$ and even $\alpha = 0.01$). A one-way analysis of the variance test (known as ANOVA) clearly leads to the rejection of the tested hypothesis about the independence of the wage on the gender. We can, therefore, conclude that the dependence of wages on the gender is proved with 5% (as well as 1%) risk of error. Or, the dependence of the wage on the gender is statistically significant at a 5% (as well as 1%) significance level.

Tab. 6: Total variance components − total average (in CZK); average variance (in CZK^2); variance of averages (in CZK^2); total variance (in CZK^2); total standard deviation (in CZK); total coefficient of variation (in %) and determination ratio (in %)

Year	Total average	Average variance (intragroup variability)	Variance of averages (intergroup variability)	Total variance	Total standard deviation	Total coefficient of variation	Ratio of determination
2003	18,810	81,875,828	5,111,767	86,987,595	9,327	49.59	5.88
2004	19,527	97,225,978	6,237,001	103,462,978	10,172	52.09	6.03
2005	20,573	105,443,765	6,652,731	112,096,497	10,588	51.47	5.93
2006	21,636	112,768,011	6,615,725	119,383,736	10,926	50.50	5.54
2007	23,631	193,359,436	10,023,726	203,383,162	14,261	60.35	4.93
2008	25,222	214,478,668	12,167,155	226,645,823	15,055	59.69	5.37
2009	25,442	224,057,727	10,486,153	234,543,880	15,315	60.20	4.47
2010	25,812	225,841,683	11,081,641	236,923,324	15,392	59.63	4.68

Source: Own calculations

Tab. 7: Analysis of variance of dependence of gross monthly wage on gender − square means (in CZK^2); test criterion *F*; critical values for 5% and 1% signifikance level and P-value

Year	Square means		Test criterion *F*	Critical value		P-value
	Intragroup variability	Intergroup variability		α = 0, 05	α = 0, 01	
2003	81,875,832	5,208,552,743,087	63,615.265	3.84147	6.63492	0. 0000
2004	97,225,978	8,759,842,842,222	90,097.760	3.84147	6.63492	0. 0000
2005	105,443,766	10,082,394,123,401	95,618.684	3.84147	6.63492	0. 0000
2006	112,768,011	10,680,240,634,696	94,709.843	3.84147	6.63492	0. 0000
2007	193,359,438	16,774,685,610,399	86,753.901	3.84147	6.63492	0. 0000
2008	214,478,669	20,827,870,254,190	97,109.285	3.84147	6.63492	0. 0000
2009	224,057,729	17,317,943,964,875	77,292.330	3.84147	6.63492	0. 0000
2010	225,841,685	18,426,873,915,560	81,591.996	3.84147	6.63492	0. 0000

Source: Own calculations

The ratio of determination then gives the intensity of dependence, i.e. the share of intergroup and total variability. It can take its value from the interval $\langle 0, 1 \rangle$. The closer to one is the value of ratio of determination, the stronger the dependence and vice versa. The determination ratio is presented in percentage terms (when multiplied by a hundred), taking values from the interval $\langle 0\,\%, 100\,\% \rangle$. From Table 6 we can see, however, that it is a considerably weak intensity dependence.

3 Forecasts of wage distribution

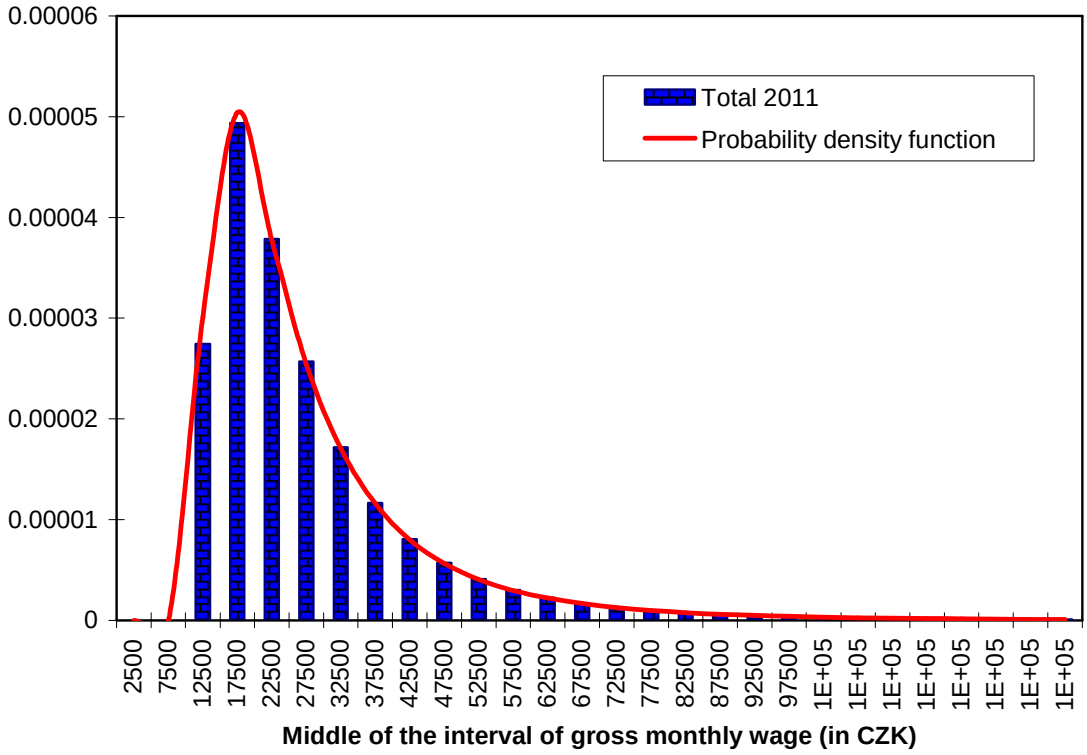
Tab. 8: Forecasts of wage distributions for 2011 and 2012 according to the gender – proportions of employees (in %) in the bands of gross monthly wage (in CZK)

Interval			Gender			
			Males		Females	
		Year	2011	2012	2011	2012
0	−	5,000	0.00	0,00	0.00	0.00
5,001	−	10,000	0.00	0.03	1.15	0.46
10,001	−	15,000	5.05	3.05	22.35	20.07
15,001	−	20,000	23.64	22,46	25.98	26.02
20,001	−	25,000	20.77	21.05	17.97	18.44
25,001	−	30,000	14.53	14.94	11.35	11.83
30,001	−	35,000	9.80	10.29	7.12	7.53
35,001	−	40,000	6.83	7.16	4.54	4.87
40,001	−	45,000	4.80	5.07	2.94	3.21
45,001	−	50,000	3.45	3.67	1.95	2.16
50,001	−	55,000	2.53	2.71	1.32	1.48
55,001	−	60,000	1.87	2.02	0.91	1.04
60,001	−	65,000	1.42	1.55	0.64	0.74
65,001	−	70,000	1.09	1.20	0.46	0.53
70,001	−	75,000	0.85	0.94	0.33	0.39
75,001	−	80,000	0.67	0.75	0.24	0.29
80,001	−	85,000	0.54	0.60	0.18	0.22
85,001	−	90,000	0.43	0.49	0.14	0.17
90,001	−	95,000	0.35	0.40	0.10	0.13
95,001	−	100,000	0.29	0.33	0.08	0.10
100,001	−	105,000	0.24	0.28	0.06	0.08
105,001	−	110,000	0.19	0.23	0.05	0.06
110,001	−	115,000	0.16	0.20	0.04	0.05
115,001	−	120,000	0.14	0.16	0.03	0.04
120,001	−	125,000	0.12	0.14	0.02	0.03
125,001	−	130,000	0.10	0.12	0.02	0.03
130,001	−	135,000	0.09	0.10	0.02	0.02
135,001	−	∞	0.05	0.06	0.01	0.01
Total (in %)			100	100	100	100

Source: Own calculations

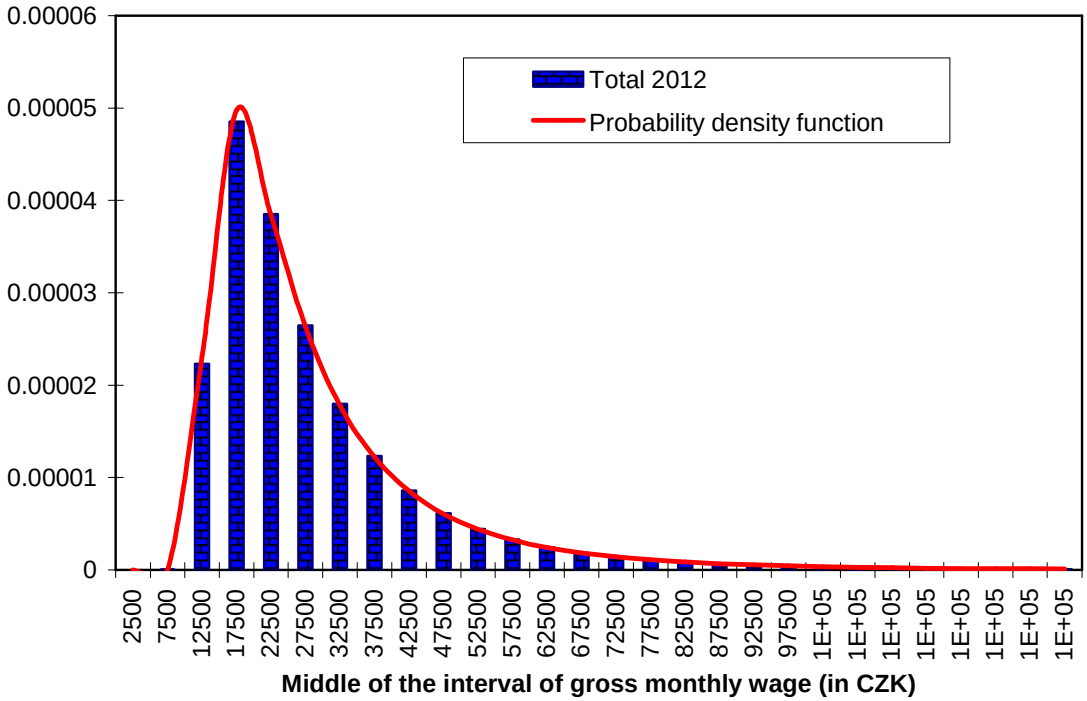
Table 8 presents the forecasts of wage distribution according to the gender for 2011 and 2012. It shows the percentages of employees in the bands of gross monthly wages (in CZK) calculated on the basis of the wage development distribution between 2003 and 2010, including the period of the global economic crisis since autumn 2008. The three-parametric lognormal probability distribution has been used here as a theoretical distribution. It is one of the most widely used probability distributions in wage and income modeling. A lesser-known method, the L-moments method, is employed to estimate parameters of this theoretical distribution. The first three sample L-moments were calculated from sample data. A trend analysis of the development of the first three sample L-moments in the period of 2003-2010 has been performed. Having been based on the trend analysis, the forecasts of the first three sample L-moments development for the years 2011 and 2012

Fig. 3: Forecasts of wage distribution for total set for 2011



Source: Own calculations

Fig. 4: Forecasts of wage distribution for total set for 2012



Source: Own calculations

were calculated. Having been based on the forecasts of the first three sample L-moments with the use of the L-moment method, the values of parameters of three-parametric lognormal distribution for 2011 and 2012 were constructed. Forecasts of wage distributions were made from these three-parametric lognormal curves, see Table 8. Figures 3 and 4 represent the forecasts of wage distribution for 2011 and 2012 for total set.

Conclusion

The paper starts with a development analysis of descriptive characteristics of wage distribution over the last years, monitoring particularly the changes of wage distribution in the context of economic recession at the end of the research period. We can conclude that wage growth has virtually stopped. Wage distributions are classified by the gender. Differences in wage levels were assessed between both genders. The arithmetic mean, median and medial were applied. Since most employees do not reach an average wage, the median was employed as a fundamental characteristic of the level of wage distribution. Wage dependence on the gender being proved by test at any significance level. Both the wage range and distribution are strongly influenced by the amount of the minimum wage. Workers' wages would presumably decline if the minimum wage was reduced or even abolished. The changes are naturally reflected in characteristics of the location, variability and shape of wage distribution. It is noteworthy that the number of extremely well-paid people was increasing gradually over the whole research period 2003–2010. The level of wage distribution was rising until 2008, wage growth having almost stopped in the year when economic recession began. The dual dimension of wage differentiation by the gender – both within and between the groups – had to be taken into account, the latter dimension being already indicated by differences in wage growth rates. It is expected that the deceleration in the growth rate of nominal and real wage level may cause structural changes in household budgets, leading to cuts in money spent on food, clothes and other durable and nondurable goods while energy, housing and transport costs may rise due to relative price changes.

References

Bartošová, J. (2009). Analysis and Modelling of Financial Power of Czech Households. Bratislava 03.02.2009 – 06.02.2009. In: *8th International Conference APLIMAT (Part I).* Bratislava: Slovak University of Technology in Bratislava, pp. 717-722.

Bartošová, J. & Bína, V. (2007). Mixture Models of Household Income Distribution in the Czech Republic. Bratislava 06.02.2007 – 09.02.2007. In: *6th International Conference APLIMAT (Part I).* Bratislava: Slovak University of Technology in Bratislava, pp. 307–316.

Bartošová, J. & Forbelská, M. (2010). Comparison of Regional Monetary Poverty in the Czech and Slovak Republic. Herl`any, Slovakia 13.10.2010 – 15.10.2010. In: *Conference on Social Capital, Human Capital and Poverty in the Regions of Slovakia 2010*. Košice: Technical University Košice, pp. s. 76–84.

Fiala, T. & Langhamrová J. (2011). ICT Experts in the Czech Republic – Development in the Past and Future Prospects. Jindřichův Hradec 07.09.2011 – 09.09.2011. In: *19th Interdisciplinarity Information Management Talks IDIMT-2011*. Linz: Trauner Verlag universitat, pp. 51–58.

Fiala, T. & Langhamrová J. (2010). Population Projection of the Numer and Age Structure of ICT Experts in the Czech republic. Jindřichův Hradec 08.09.2010 – 10.09.2010. In: *18th Interdisciplinarity Information Management Talks*. *IDIMT-2010*. Linz: Trauner Verlag universitat, pp. 115–123.

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

Hosking, J. R. M. (1990). L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society (Series B),* Vol. 52, No. 1, pp. 105 – 124.

Hosking, J. R. M. & Wales, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-moments.* New York, Cambridge University Press. 209 p.

Kyselý, J. & Picek, J. (2007). Regional Growth Curves and Improved design Value Estimates of Extréme Precipitation Events in the Czech Republic. *Climate Research*, Vol. 33, pp. 243 – 255.

Löster, T., Langhamrová, J. (2011). Analysis of Long-Term Unemployment in the Czech Republic. Prague 22.09.2011 – 23.09.2011. In: *International Days of Statistics and Economics at VŠE 2011*. Slaný: Melandrium, pp. 228–234.

Malá, I. (2011). The use of finite mixtures of lognormal distributions in the modeling of incomes in the Czech Republic. *Research Journal of Economics, Business and ICT*, Vol. 4, No. 1, pp. 41–46.

Miskolczi, M., Langhamrová, J. & Langhamrová, J. (2011). Recognition of Differentiation in Unemployment Trends Among Regions in the Czech Republic. Jindřichův Hradec

07.09.2011 – 09.09.2011. In: *19th Interdisciplinarity Information Management Talks IDIMT-2011*. Linz: Trauner Verlag universitat, pp. 387–388.

Řezanková, H., Löster, T. & Hušek D. (2011). Evaluation of Categorical Data Clustering. Fribourg 26.01.2011 – 28.01.2011. In: *7th Atlantic Web Intelligence Conference Advances in Intelligent Web Mastering – 3*. Berlin: Springer Verlag, pp. 173–182.

Contact

Diana Bílková

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
Mail: bilkova@vse.cz

University of Finance and Administration
Faculty of Economic Studies
Department of Informatics and Mathematics
Estonian Street 500/3
101 00 Prague 10
Czech Republic
Mail: diana.bilkova@vsfs.cz