Model wage distribution - mixture density functions

Luboš Marek and Michal Vrabec

Abstract**—**The authors try to contribute to the discussion of the possibility to predict the trend of the wage distribution in the article. For this purpose data from Czech Republic are used. But computed model is useable for all similar data types. Classical models use the probability distribution such as lognormal, Pareto, etc., but their results are not very good. Authors suggest using a mixture of normal probability distribution (normal mixture) in our model. Authors focus mainly on the possibility of constructing a mixture of normal distributions based on parameter estimation. Parameters are estimated these parameters on the basis of their evolution in time. The data cover last 15 years. The data are divided into groups with respect to sex, age and regions.

Keywords**—**mixture normal density functions, probability models, wage distribution.

I. Introduction

We want to contribute to the discussion on suitability of the arithmetic mean as a characteristic for the wage level in the Czech Republic. There is recurring expression of surprise with the fact that the income of more than fifty percent of the population is lower than the average wage“. If the intended effect is to have "more" wage recipients above the officially announced level, a simple solution would be to use different characteristics of this level. For example, the median (50% quantile) is defined by the condition that exactly 50% wage recipients are below this value, while the remaining 50% are above it. Choosing a suitable quantile, we can always get the required percentage of wage recipients above the quantile level. E.g., 60% of wage recipients are above and 40% below, the second pentile. Whichever characteristic is chosen, we have to keep in mind that it is a simplification. Another possible approach comprises monitoring a higher number of characteristics (of not only the location). In addition to location, we can also pay attention to variability, skewness, kurtosis, etc.

Another approach is to describe the frequency distribution of individual income groups. Apart from other advantages, this approach enables us to derive any of the above-mentioned characteristics at the required level of accuracy. We can also predict the future distribution on the basis of the time evolution of the parameters in the model. To build a good wage model is very important in crises time [10], [11].

Lubos Marek. Author works at University of Economics, Prague, W. Churchill sq. 4, 130 67 Prague, Czech Republic (corresponding author to provide phone: +420 224 095 401; e-mail: marek@vse.cz).

Michal Vrabec. Author, works at University of Economics, Prague, W. Churchill sq. 4, 130 67 Prague, Czech Republic (corresponding author to provide phone: +420 224 095 481; e-mail: vrabec@vse.cz).

II. Data

A. Data description

We work with wages in the Czech Republic in the years from 1995 to 2012. So, we have data over 18 years. The annual data is reported in quarterly units. The scope of the data set on which the analyses were carried out was gradually increased from more than 300,000 observations in 1995 to approximately two million in 2012. This data is structured in a very detailed way. The wage values are divided into intervals whose widths are 500 CZK. Such a detailed structure enables us to achieve quite accurate results. The structure of data is shown in Table I.

TABLE I. DATA INTERVALS

lower bound	-	upper bound	absolute frequency
15,000	-	15,500	41,541
15,500	-	16,000	42,942
16,000	-	16,500	44,960

Source: Own calculations

For information, the rate of Czech Crown vs. USD is as follows: 1 USD = 20 CZK.

B. Data characteristics

We have at our disposal basic characteristics of wages

- arithmetic mean,
- standard deviation,
- median,
- upper and lower quartiles,
- 10% and 90% quantiles.

In particular, the mean and standard deviation are very important for us – without those we would not have been able to estimate the density of probability distribution for wages. Characteristics were calculated for

- the entire Czech Republic,
- by gender,

- by age – 3 groups:
  - up to 30 years,
  - 30 to 50 years,
  - over 50 years of age,
- by regions - 14 groups.

The overall number of characteristic categories hence was $2 \times 3 \times 14 = 84$.

II. Distribution of Wages

A. Common description the frequency distribution

If the wage distribution is more or less "smooth", it can be adequately modeled with the aid of a suitable theoretic (continuous) distribution, such as a lognormal one [1], [3], [4] or [9].

Fig. 1 below shows that the wage distribution could be modeled by lognormal distribution in the first years. It also indicates, however, that the wage distribution has been becoming multimodal in the recent years and the use of the lognormal model is thus problematic.

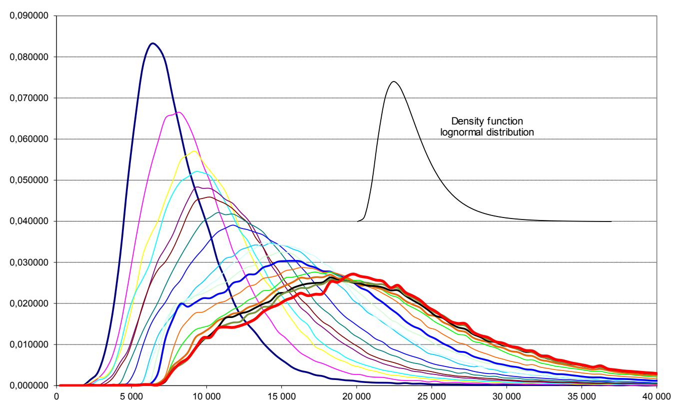

Fig. 1. Empirical wage distribution
Source: Own calculations

On the other hand, the multimodal character might be well explained if the population is suitable subdivided. A secondary effect of a subdivision is that skewness values of the component distributions are smaller. All these reasons led us to modelling the wage distribution with the aid of a mixture of normal distributions – see [2], [8] or [12]. So, we will not work with one distribution, but with a mixture of several distributions.

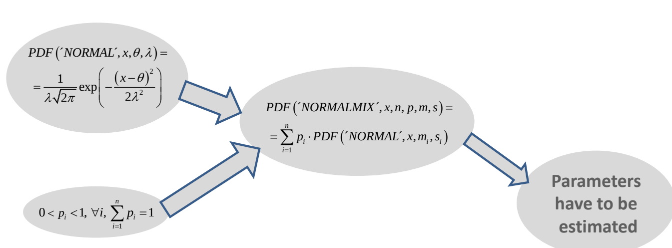

Fig. 2. Schema of normal mixture densities
Source: Own schema

B. Empirical frequency distribution by gender

Fig. 3 and 4 show the empirical density function for men and women over all years. Fig. 5 shows the empirical density function for men and women in year 2012 in the same picture. We can see that these two groups are different in characteristics of location, variability, skewness and kurtosis, too.

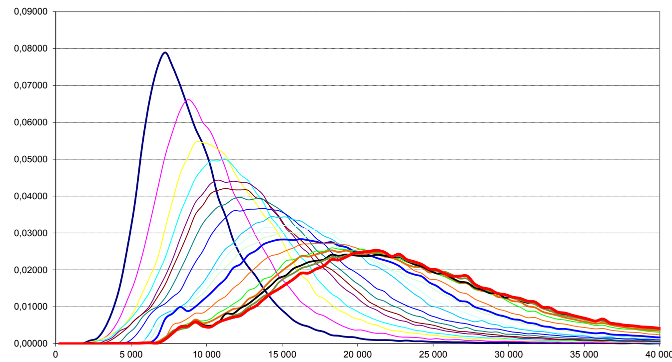

Fig. 3. Empirical wage distribution - men
Source: Own calculations

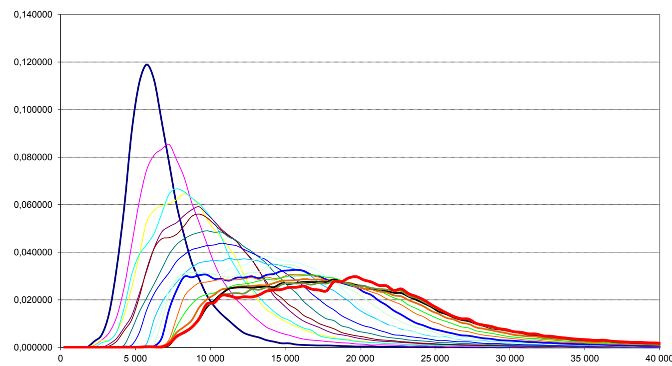

Fig. 4. Empirical wage distribution - women
Source: Own calculations

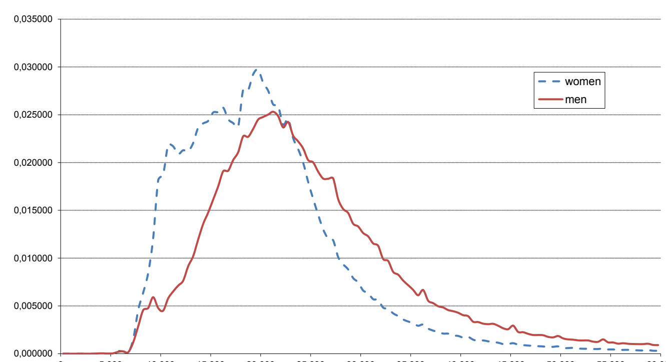

Fig. 5. Example of empirical wage distribution – men and women
Source: Own calculations

Fig. 5 compares empirical distribution of relative frequency for men and women for year 2012 only. It is quite clear that this is a two-peaks distribution. So, final model we will build as a mixture of two distributions.

C. Theoretical frequency distribution by gender (Model 1)

The probability density for a general model of a normal mixture can be written as follows

$$PDF\left(\acute{}NORMALMIX\acute{}, x, n, p, \mu, \sigma\right) = \sum_{i=1}^{n} p_i \cdot PDF\left(\acute{}NORMAL\acute{}, x, \mu_i, \sigma_i\right)$$

Here PDF stands for a probability density of a mixture of normal distributions (´NORMALMIX´) or a normal distribution as such (´NORMAL´), x for the argument, n for the number of components in the mixture, and p is the vector of weights, for which holds

$$0 < p_i < 1, \ \forall i, \ \sum_{i=1}^{n} p_i = 1, \qquad (1)$$

μ and σ are vectors of mean values and standard deviations of individual components in this mixture.

The density of normal distribution (of individual components in this mixture) is expressed by the following formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty \qquad (2)$$

where $-\infty < \mu < +\infty, \ \sigma^2 > 0$ are parameters.

The standard approach (parameter estimation on the basis of selected optimization criteria) is rather good for describing the history (even though interpretation is not easy) but it cannot be used for useful prediction of the future development. Several methods for estimating such parameters have been described in the literature (Expectation Maximisation (EM), Markov Chain Monte Carlo, Moment Matching, EF3M algorithm, etc.). The EM algorithm is most frequently used for practical applications – it is an iterative method for establishing the estimate with the aid of the Maximum Likelihood or MAP - Maximum Aposteriori Probability [4]. This algorithm is included in SAS [5], [13]. In the general case, 3n + 1 parameters have to be estimated (among them n itself - see [9] for details). Hence we decided for another method, namely, that of factual determination of parameters and a construction of the mixture on the basis of standard prediction of parameters within the mixture.

This approach brings about considerable advantages. The first such advantage is the factual interpretation. E.g., the simplest model (division of the population by sex, to men and women) we get n=2, are the expected 2013 wage values for men and women (respectively), and are the corresponding standard deviation values. Another advantage is a simple construction of the prediction for the future period (2013). The Figures below illustrate the linear evolution of these parameters in time.

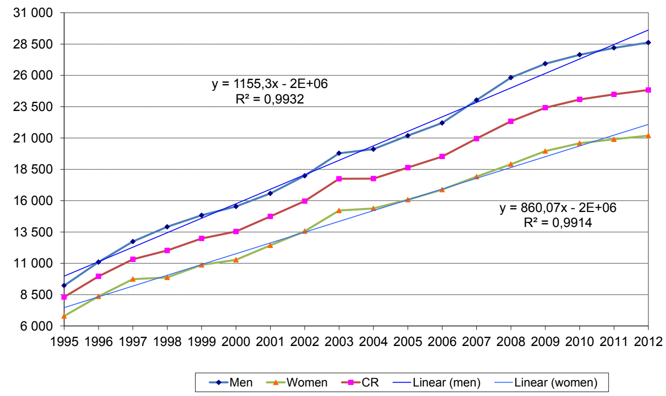

Fig. 6. Average wage – men and women
Source: Own calculations

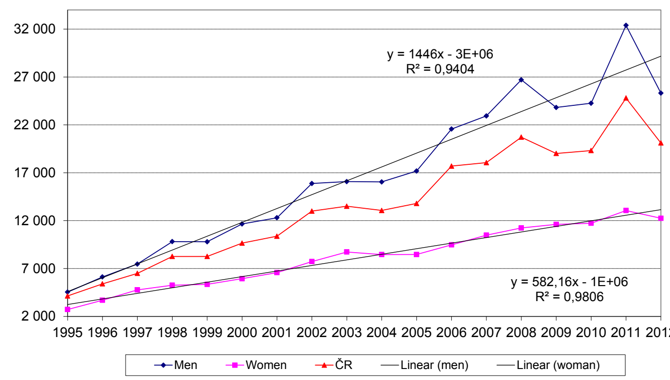

Fig. 7. Standard deviation of wage – men and women
Source: Own calculations

TABLE II. EMPIRICAL PARAMETERS – GROUPS BY GENDER

Year	Men			Women		
	Count	Mean	StdDev	Count	Mean	StdDev
1995	0.646	9,221	4,538	0.354	6,794	2,720
1996	0.599	11,100	6,118	0.401	8,363	3,683
1997	0.532	12,737	7,462	0.468	9,740	4,766
1998	0.538	13,914	9,808	0.462	9,872	5,255
1999	0.535	14,835	9,790	0.465	10,878	5,345
2000	0.531	15,537	11,654	0.469	11,281	5,936
2001	0.557	16,580	12,299	0.443	12,435	6,569
2002	0.542	17,987	15,876	0.458	13,565	7,722
2003	0.554	19,784	16,078	0.446	15,217	8,726
2004	0.503	20,109	16,042	0.497	15,380	8,459
2005	0.502	21,188	17,183	0.498	16,076	8,463
2006	0.497	22,203	21,565	0.503	16,882	9,472
2007	0.497	24,026	22,933	0.503	17,916	10,480
2008	0.496	25,821	26,701	0.504	18,912	11,233
2009	0.496	26,929	23,814	0.504	19,957	11,605
2010	0.495	27,644	24,261	0.505	20,585	11,726
2011	0.491	28,196	32,390	0.509	20,903	13,056
2012	0.490	28,617	25,318	0.510	21,189	12,245

Source: Own calculations

Hence we can estimate the mixture parameters for 2013 by a linear trend (cf. the Table II).

The resulting mixture (its parameters) is given by this formula (the last row in table)

$$PDF\left(\acute{}NORMALMIX\acute{}, x, 2, (0.49; 0.51), (28617; 21189), (25318; 12245)\right)$$

The quality of the model was checked according to various criteria. We tested the null hypothesis

H_0: *difference between theoretical and empirical densities is equal zero*

again alternative hypothesis

H_1: *non H_0*

The results of this test are in Table II. We cannot reject null hypothesis H_0 at level $\alpha = 0.05$.

In the next step we applied Wilcoxon Signed Rank test and Sign test. The results of these tests are at the end of article.

TABLE III. TEST OF QUALITY – MODEL FOR GENDER

Sex	0.00001	t-Ratio	1.316089
Count	0.00001	DF	100
Mean Difference	2.9e-22	Prob > |t|	0.1912
Std Error	2.2e-22	Prob > t	0.0956
Upper 95%	7.1e-22	Prob < t	0.9044
Lower 95%	-1e-22		
N	101		
Correlation	1		

Source: Own calculations

The accordance between empirical and theoretical distribution is shown at Fig. 9 and is it very good. The correlation between the theoretical and empirical frequencies nearly equals one, which is another indication of a good fit between these frequencies.

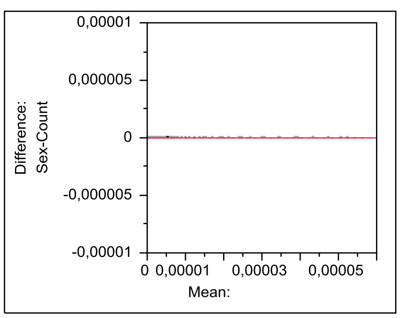

Fig.8. Test for Model 1
Source: SAS output

There is the corresponding estimated empirical density of the wage. The following Figure illustrates the estimated wage distribution in the Czech Republic for 2012 for Model 1 - mixture 2 gender groups (men, women)

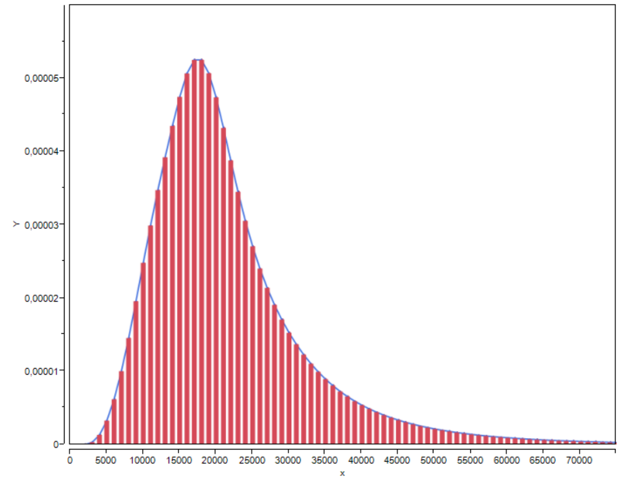

Fig. 9. Wage distribution – Model 1
Source: Own calculations, SAS output

The result of Chi-square test confirms the good quality of the model. It is clear that we cannot reject null hypothesis H_0 at level $\alpha = 0.05$.

TABLE IV. CHI-SQUARE TEST – MODEL 1

Chi-Square Test for Specified Proportions	
Chi-Square	188.5258
Df	199
P-value	0.6921

Source: SAS output

D. *Empirical frequency distribution by age*

Empirical frequency distributions by age have the form

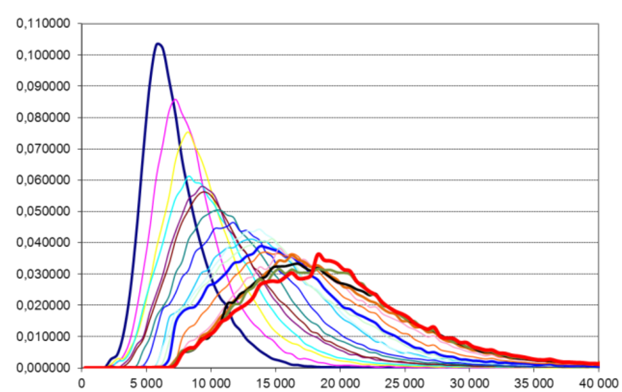

Fig. 10. Wage distribution by age – under 30
Source: Own calculations

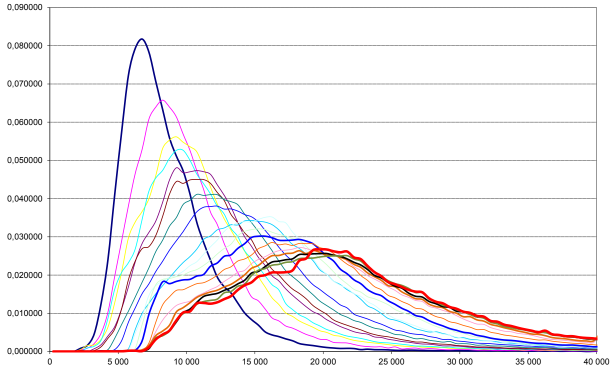

Fig. 11. Wage distribution by age –30-50
Source: Own calculations

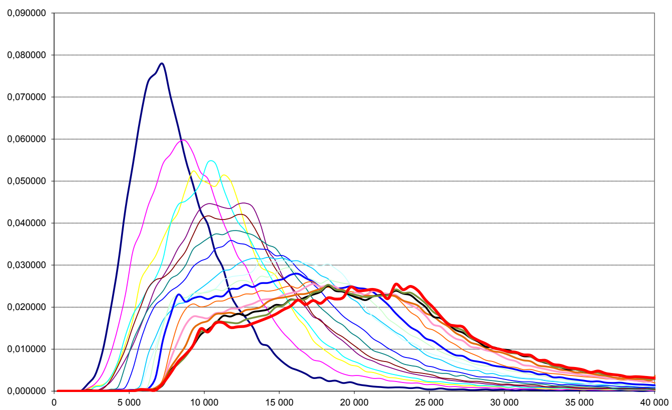

Fig. 12. Wage distribution by age – over 50
Source: Own calculations

Fig. 13 shows the empirical density function for men and women in year 2012 in the same picture. We can see that these two groups are different in characteristics of location, variability, skewness and kurtosis, too.

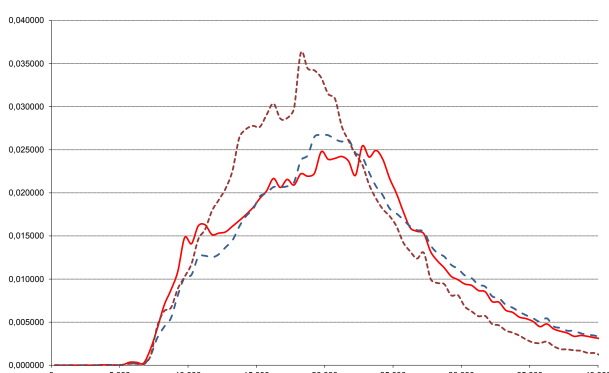

Fig. 13. Empirical wage distribution –3 age groups
Source: Own calculations

E. *Theoretical frequency distribution by age (Model 2)*

TABLE V. EMPIRICAL PARAMETERS – GROUPS BY AGE

Year	Under 30			30 - 50			Over 50		
	Weight	Average	StdDev	Weight	Average	StdDev	Weight	Average	StdDev
1995	0.215	7,266	2,597	0.551	8,654	4,150	0.234	8,779	4,919
1996	0.209	8,623	3,210	0.547	10,337	5,504	0.244	10,679	6,394
1997	0.207	9,807	4,297	0.558	11,635	6,700	0.235	12,181	6,955
1998	0.223	10,674	6,026	0.548	12,311	8,681	0.230	12,821	8,816
1999	0.224	11,671	6,317	0.532	13,310	8,435	0.244	13,796	9,243
2000	0.210	11,922	6,378	0.516	13,906	10,146	0.274	14,091	10,594
2001	0.218	13,014	7,036	0.505	15,260	10,992	0.276	15,164	11,238
2002	0.206	13,934	8,691	0.492	16,594	14,019	0.303	16,319	13,552
2003	0.197	15,523	8,336	0.492	18,472	14,979	0.311	18,015	13,546
2004	0.182	15,457	7,469	0.505	18,346	14,411	0.313	18,146	13,176
2005	0.180	16,144	7,632	0.505	19,357	15,228	0.315	18,918	13,954
2006	0.180	16,838	8,179	0.504	20,346	18,825	0.316	19,752	16,850
2007	0.183	17,862	8,725	0.507	21,976	19,977	0.310	21,111	18,627
2008	0.180	19,076	9,079	0.511	23,562	23,825	0.309	22,214	19,831
2009	0.164	19,867	8,940	0.527	24,632	21,149	0.308	23,235	18,859
2010	0.161	20,004	8,503	0.535	25,356	21,451	0.304	23,981	19,171
2011	0.160	20,173	9,446	0.538	25,861	29,650	0.302	24,311	20,246
2012	0.156	20,420	8,512	0.544	26,180	22,094	0.300	24,673	20,286

Source: Own calculations

Parameters for the groups by age (year 2012 only) were estimated in a similar way. We obtained the results in Table III. The resulting mixture (its parameters) is given by this formula:

$$PDF\begin{pmatrix} ´NORMALMIX´, x, 3, (0.156; 0.544; 0,300), (20420; 26180; 24673) \\ ,(8512; 22094; 20286) \end{pmatrix}$$

The quality of the model is not good. We reject null hypothesis H_0 at level $\alpha = 0.05$.

TABLE VI. TEST OF QUALITY – MODEL FOR AGE

Age	0.00001	t-Ratio	4.716771
Count	0.00001	DF	100
Mean Difference	4.78e-7	Prob > |t|	<.0001*
Std Error	1.01e-7	Prob > t	<.0001*
Upper 95%	6.8e-7	Prob < t	1.0000
Lower 95%	2.77e-7		
N	101		
Correlation	0.99792		

Source: Own calculations

The accordance between empirical and theoretical distribution is shown at Fig. 15.

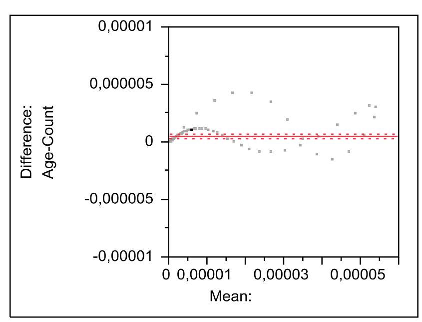

Fig.14. Test for Model 2
Source: SAS output

There is the corresponding estimated empirical density of the wage. The following Figure illustrates the estimated wage distribution in the Czech Republic for 2012 for Model 2 - mixture 3 age groups (under 30, 30-50, over 50).

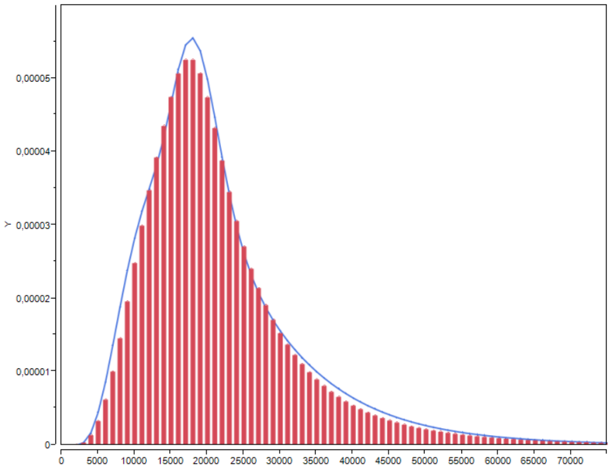

Fig. 15. Wage distribution – Model 2
Source: Own calculations

The result of Chi-square test do not confirms the good quality of the model for usual value $\alpha = 0.05$. We reject null hypothesis H_0 at level $\alpha = 0.05$ for Model 2, but we cannot reject this null hypothesis H_0 at level $\alpha = 0.01$.

TABLE VII. CHI-SQUARE TEST – MODEL 2

Chi-Square Test for Specified Proportions	
Chi-Square	247.101
Df	199
P-value	0.0115

Source: SAS output

F. *Empirical frequency distribution by regions*

There are 14 empirical frequency distributions by regions. Therefore, we publish as an example, the only distribution for 4 selected regions in 2012.

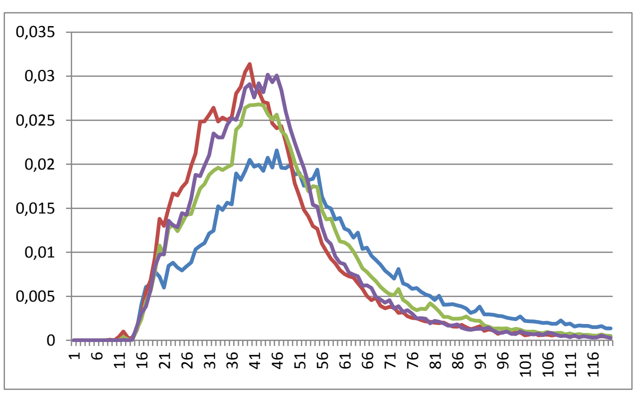

Fig. 16. Empirical wage distribution – 4 selected regions
Source: Own calculations

The table of empirical parameters for the groups by regions is very large. Therefore, we published only last year 2012. The parameters were calculated by the same way as in the previous cases.

TABLE VIII. EMPIRICAL PARAMETERS – GROUPS BY REGIONS

Region	n	Count	Average	StdDev
Praha	14	0.18	33,842	35,774
Středočeský		0.10	24,995	17,913
Jihočeský		0.06	22,625	19,735
Plzeňský		0.05	23,501	13,948
Karlovarský		0.02	22,330	16,023
Ústecký		0.07	23,080	15,617
Liberecký		0.04	23,432	14,573
Královehradecký		0.05	22,752	13,140
Pardubický		0.05	22,220	16,574
Vysočina		0.05	22,941	14,147
Jihomoravský		0.11	24,040	17,177
Olomoucký		0.06	22,341	14,966
Zlínský		0.05	21,842	14,222
Moravskoslezský		0.12	23,448	17,226

Source: Own calculations

G. *Theoretical frequency distribution by age (Model 3)*

The resulting mixture (its parameters) is given by this formula

$$PDF\left(\begin{array}{l}'NORMALMIX', x, 14,\\ (0.1943;0.09493;0.0563;0.05493;0.0239;0.0674;0.0368;\\ 0.0494;0.0469;0.04510;0.1054;0.0562;0.0477;0.1208),(32301;\\ 25288;22215;23241;21306;22764;23193;22683;21604;22239;\\ 23977;21797;21349;23366),(31456;19347;13805;14383;13632;\\ 15241;14700;13262;13959;13204;18334;14050;15218;15414)\end{array}\right)$$

The quality of the model is not good. We reject null hypothesis H_0 at level $\alpha = 0.05$.

TABLE IX. TEST OF QUALITY – MODEL FOR REGIONS

Region	8.93e-6	t-Ratio	-4.71677
Count	0.00001	DF	100
Mean Difference	-4.8e-7	Prob > |t|	<.0001*
Std Error	1.01e-7	Prob > t	1.0000
Upper 95%	-2.8e-7	Prob < t	<.0001*
Lower 95%	-6.8e-7		
N	101		
Correlation	0.99776		

Source: Own calculations

The accordance between empirical and theoretical distribution is shown at Fig. 18.

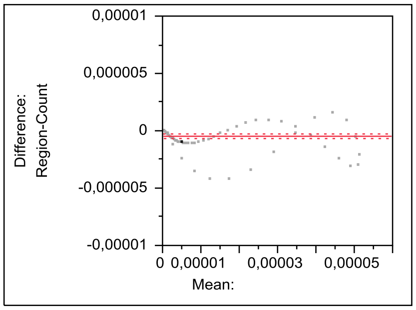

Fig.17. Test for Model 3
Source: SAS output

There is the corresponding estimated empirical density of the wage. The following Figure illustrates the estimated wage distribution in the Czech Republic for 2012 for Model 3 - mixture 14 groups by regions.

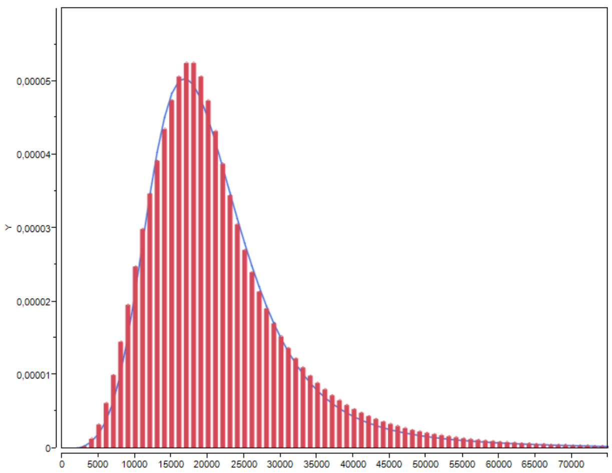

Fig. 18. Wage distribution – Model 3
Source: Own calculations

The result of Chi-square test do not confirms the good quality of the model for usual value of $\alpha = 0.05$. We reject null hypothesis H_0 at level $\alpha = 0.05$ for Model 3, but we cannot reject null hypothesis H_0 at level $\alpha = 0.01$.

TABLE X. CHI-SQUARE TEST – MODEL 3

Chi-Square Test for Specified Proportions	
Chi-Square	241.4527
Df	199
P-value	0.0224

Source: SAS output

H. *Other tests of model quality*

We applied other tests to evaluate the quality of the model. We used Wilcoxon Signed Rank test and Sign test. The results of these tests are in the next tables.

Wilcoxon Signed Rank test is a nonparametric version of the paired t-test that compares the sizes of the positive differences to the sizes of the negative differences.

Sign test is a nonparametric version of the paired t-test that uses only the sign (positive or negative) of the difference for the test.

TABLE XI. WILCOXON SIGNED RANK TEST

	Sex-Count	Age-Count	Region-Count
Test Statistic S	31.500	1376.50	-1376.5
Prob>|S|	0.6452	<.0001*	<.0001*
Prob>S	0.3226	<.0001*	1.0000
Prob<S	0.6774	1.0000	<.0001*

Source: SAS output

TABLE XII. SIGN TEST

	Sex-Count	Age-Count	Region-Count
Test Statistic M	1.500	12.500	-12.500
Prob ≥ |M|	0.7552	0.0165*	0.0165*
Prob ≥ M	0.3776	0.0083*	0.9953
Prob ≤ M	0.7336	0.9953	0.0083*

Source: SAS output

The results of both tests confirm the previous results of testing. Only Model 1 is good, quality of other two models is not so good.

I. *Forecast of wage density function*

Parameters for our models are estimated from known data (usually from the current or from last year). None forecasts are possible in classical approach. Because we know long history of parameters (mean and standard deviation), we can apply trend analysis for estimation future values of parameters. So, we can forecast the model of wage distribution for the next years. Details can be found in [6], [7] or [9].

Because the best model is Model 1 (mixture by gender) we use this model to calculation the future shape of wage density function. For estimation mean and standard deviation we use the formula of trend at Fig. 7 and Fig. 8. The trend formulas and index of determination for mean have the exact form

$$Y_t = 1155.3\,t - 2E + 06\,, \quad R^2 = 0.9932 \quad \text{for men}$$

$$Y_t = 860.07\,t - 2E + 06\,, \quad R^2 = 0.9914 \quad \text{for women}$$

The trend formulas and index of determination for standard deviation have the form

$$Y_t = 1446\,t - 2{,}172{,}273\,, \quad R^2 = 0.9404 \quad \text{for men}$$

$$Y_t = 562.16\,t - 1{,}158{,}164\,, \quad R^2 = 0.9806 \quad \text{for women}$$

There is very high index of determination for all trend models. This index is very near to 1 and it signs that there is very strong linear dependency. This means that the quality of the predictions will be very good. After calculations we obtained the values

TABLE XIII. FORECASTS OF PARAMETERS

2013	mean	std. dev.
men	30,776.7	30,616,3
women	22,945.4	13,722.8

Source: Own calculations

We use these forecasts as an input in mixture equation. The resulting mixture (its parameters) is given by this formula

$$PDF\,(\acute{}NORMALMIX\acute{}, x, 2, (0.49; 0.51), (30776.7; 30616.3), (22945.4; 13722.8)$$

Because we have no empirical data for year 2013, we show the shape of theoretical mixture density function.

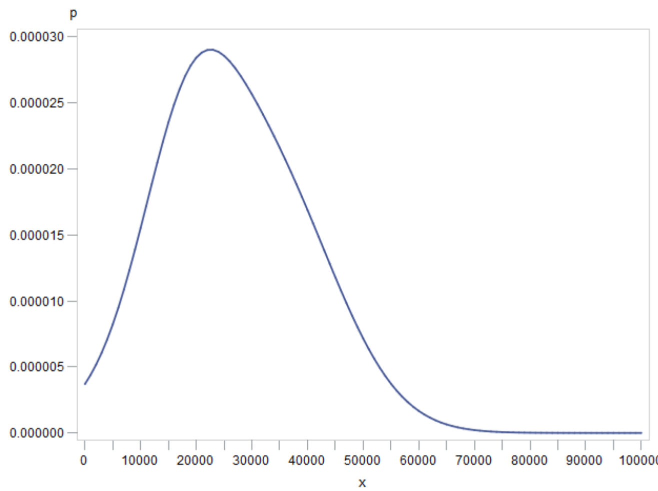

Fig. 19. Forecast of wage distribution for year 2013
Source: Own calculations

A. Conclusions

Neither tables nor estimate charts for empiric densities are shown for our models. Differences in frequencies implied by individual subdivisions of the basic population are not very large. We can provide these results, together with the SAS code, to interested parties. We used 3 mixtures (gender, age, regions) in the article. We note that

- Classical model for wage distribution is based on one density function (usually lognormal distribution).
- Quality of such model is not so good because there are often significant differences between model and empirical data.
- We suggest using of new approach – we work with mixture of normal density functions.
- The achieved results are very good - the results are better then for classical approach.
- The best model is for mixture of two densities (groups by gender).
- Results are confirmed by appropriate t-tests.
- The correlation between the theoretical and empirical frequencies nearly equals one, which is another indication of a good fit between these frequencies.
- Wilcoxon Ranked Rank test and Sign test also confirm the fit between the theoretical and empirical frequencies.
- Our approach allows the construction of forecasts for wage distribution.
- The forecast of parameters is based on time series analysis – quality of trend model is very good.

Knowledge of the theoretical model allows us to perform all probability calculations such as

- Construction of confidence intervals.
- Estimation of the relative and absolute frequency of employees in wage intervals.
- Estimations of the future values many of characteristics – mean, standard deviation, median, percenttiles measures etc.

Achieved estimates can also be consequently used in estimating tax revenues, in social politics (construction of minimum wage) and in many other areas of state administration. The authors will continue this study and will publish practical applications of their research in other articles.

REFERENCES

[1] J. Arlt, M. Arltová, J. Klůfa, Empirical Analysis for Testing the Validity of the Phillips Curve in the Czech Republic. Rhodes Island July 16-19, 2013. Recent Advances in Economics and Business Administration, pp. 184-188. EBA 2013 - Proceedings of the 2013 International Conference on Economics and Business Administration. ISBN 978–1–61804–198–2

[2] J. Bartošová, V. Bína, "Mixture Models of Household Income Distribution in the Czech Republic", Bratislava 06.2.2007– 09.02.2007. In: 6th International Conference APLIMAT 2007, Part I. Bratislava: Slovak University of Technology, 2007, s. 307-316, ISBN 978-80-969562-4-1.

[3] D. Bílková, "Lognormal distribution and using L-moment method for estimating its parameters", In International Journal of Mathematical models and methods in Applied Sciences, Volume 6, 2012, ISSN: 1998-0140

[4] A.P. Dempster, N.M. Laird; D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society, 1977. Series B (Methodological) 39 (1): 1–38. JSTOR 2984875. MR0501537.

[5] E. P. Hughes, T. D. Kearney, "Optimization with the SAS® System". SAS Institute Inc. 2012.

[6] L. Marek, "Analýza vývoje mezd v ČR v letech 1995-2008". In Politická ekonomie, 2010, roč. 58, č. 2, s. 186–206. ISSN 0032-3233.

[7] L. Marek, M. Vrabec, "Wage Distribution Models", Rhodes Island July 16-19, 2013. Recent Advances in Economics and Business Administration, pp. 184-188. EBA 2013 - Proceedings of the 2013 International Conference on Economics and Business Administration. ISBN 978–1–61804–198–2

[8] L. Marek, M. Vrabec, "K moznostem modelovani mzdovych rozdeleni" Praha 13.12.2010 – 14.12.2010. In: Reprodukce lidskeho kapitalu – Vzajemne vazby a souvislosti. [online] Praha : KDEM VSE, 2010, s. 1–9. ISBN 978-80-245-1697-4.
URL: http://kdem.vse.cz/resources/relik10/Index.htm.

[9] L. Marek, M. Vrabec, "Forecast of the Income Distribution in the Czech" Republic in 2011. Ras Al Khaimah 29.11.2010 – 03.12.2010. In: ICABR 2010 – VI. International Conference on Applied Business

Research. Brno: Mendel University, 2010, s. 142. ISBN 978-80-7375-462-4.

[10] L. Smrčka, M. Arltová, "Increasing Threat of a "Total Financial Crisis" in the Upcoming Years" In International Journal of Mathematical models and methods in Applied Sciences, Volume 6, 2012, ISSN: 1998-0140

[11] J. Strouhal. Testing Accounting Harmonization upon the Globalization of Financial Reporting Systems. Cambridge 22.02.2012 – 27.02.2012. In: RUDAS, Imre J., ZAHARIM, Azami, SOPIAN, Kamaruzzaman, STROUHAL, Jiří (ed.). *Recent Research in Engineering Education and Software Engineering.* Athens : WSEAS Press, 2012, s. 163–168.

[12] M. Vrabec, L. Marek, "The Using of Normal Mixture Distribution for Wages Models", Johor Bahru 28.11.2011 – 02.12.2011. In: ICCDA 2011 – IC Computer Desing and Applications [CD-ROM]. Brno : Mendel University in Brno, 2011, s. 159. ISBN 978-80-7375-557-7.

[13] SAS Institute. B"ase SAS(R) 9.2 Procedures Guide: Statistical Procedures", Third Edition.

Luboš Marek graduated from the Charles University in Prague in 1984, majoring in Probability and Mathematical Statistic. He currently works as an associate professor in the Department of Statistics and Probability at the University of Economics Prague. He focuses on time series analyses, probability and computational statistics.

Michal Vrabec graduated from the Charles University in Prague in 1979, majoring in Probability and Mathematical Statistic. He currently works as an assistent in the Department of Statistics and Probability at the University of Economics Prague. He focuses on multidimensional methods, theory of sampling methods and computational statistics.