



CONFERENCE PROCEEDINGS

FULL TEXT PAPERS

edited by
Zofia Rusnak and Beata Zmysłona



APPLICATIONS OF
MATHEMATICS AND STATISTICS
IN ECONOMICS

International Scientific Conference | Poland • 27-31 August 2014

Scientific Committee

*Richard Hindls, Stanislava Hronová, Rudolf Zimka, Walenty Ostasiewicz,
Emília Zimková, Zofia Rusnak, Martin Bod'a*

Organizing Committee

Beata Zmysłona, Cyprian Kozyra, Grzegorz Rogoziński, Kristýna Vltavská

Reviewers

*Milan Bašta, Diana Bílková, Martin Bod'a, Joanna Dębicka, Tomáš Fiala, Jakub Fischer,
Stanisław Heilpern, Karel Helman, Lenka Hudrlíková, Miroslav Hužvár, Nikola Kaspříková,
Alena Kaščáková, Kamil Kladívko, Jindřich Klůfa, Pavol Král, Katarzyna Kuziak,
Jana Langhamrová, Ivana Malá, Tomáš Marcinko, Luboš Marek, Miloš Maryška, Petr Mazouch,
Zofia Mielecka-Kubień, Witold Miszczak, Petr Musil, Gabriela Nedelová, Walenty Ostasiewicz,
Iva Pecáková, Viera Roháčová, Zofia Rusnak, Mária Stachová, Jana Špírková, Šárka Šustová,
Jana Tepperová, Vladimír Úradníček, Kristýna Vltavská, Michal Vrabec, Dariusz Wawrzyniak,
Henryk Zawadzki, Jaroslav Zbranek, Tomáš Zeithamer, Martin Zelený, Jan Zeman, Rudolf Zimka,
Emília Zimková, Pavel Zimmermann, David Žižka*

Layout

Martin Bod'a, Beata Zmysłona, Grzegorz Rogoziński

Front page design

Grzegorz Rogoziński

CD cover design

Beata Dębska

Articles published in the form submitted by the authors

All rights reserved. No part of this book may be reproduced in any form
or in any means without the prior permission in writing of the Publisher

© Copyright by Wrocław University of Economics
Wrocław 2014

ISBN 978-83-7695-421-9

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
53-345 Wrocław, ul. Komandorska 118/120

www.ue.wroc.pl

Sprzedaż książek tel./fax 71 36-80-602

e-mail: econbook@ue.wroc.pl www.ksiegarnia.ue.wroc.pl

Contents

Foreword.....	5
Diana Bílková: TL-Moments: Analogy of Classical L-Moments.....	7
Dagmar Blatná: Application of Robust Regression in the Analysis of Internet Access in European Countries.....	17
Martin Bod'a, Mária Kanderová: Rebalancing Issues in Tracking Error Variance Minimization.....	26
Martin Bod'a, Viera Roháčová: Application of Six Sigma Ideas to Timing Decisions at Financial Markets	36
Anton Dekrét, Rudolf Zimka: On the Price Hartwick's Task and Its Inverse in a Dynamic Model of an Economy with Exhaustible Resources	46
Joanna Dębicka, Agnieszka Marciniuk: Comparison of Reverse Annuity Contract and Reverse Mortgage on the Polish Market.....	55
Petra Dotlačilová, Jitka Langhamrová: The Influence of Mortality Models for the Expected Future Life-time of Older People	65
Marek Ďurica, Lucia Švábová: Delta and Gamma for Chooser Options	75
Vlastimil Farkašovský: New Concept of Pension Funds Performance Evaluation	85
Albert Gardoň: The Normality of Weekly Relative Changes of the Freight Rate in Container Shipping.....	95
Mária Grausová, Miroslav Hužvár, Jana Štrangfeldová: Healthcare Systems Efficiency in the Visegrád Group	104
Stanisław Heilpern: Multiple Life Insurance - Pension Calculation	114
Alena Kaščáková, Gabriela Nedelová: Changes in Slovak Households' Economy	122
Igor Kollár, Pavol Král', Peter Laco: Methodology for Assessing Website Improvement in Corporate Environment.....	131
Maciej Kostrzewski: Some Method of Detecting the Jump Clustering Phenomenon in Financial Time Series.....	141
Cyprian Kozyra, Beata Zmysłona, Katarzyna Madziarska: Complementary Objective and Subjective Measures of Hospital Services Quality.....	150
Pavol Král', Mária Stachová, Lukáš Sobíšek: Utilization of Repeatedly Measured Financial Ratios in Corporate Financial Distress Prediction in Slovakia	156
Ivana Malá: The Use of Finite Mixture Model for Describing Differences in Unemployment Duration	164
Lukáš Malec: Studying Economics and Tourism Industry Relations by Smooth Partial Least Squares Method Depending on Parameter.....	173

Tomáš Marcinko: Consequences of Assumption Violations Regarding Classical Location Tests.....	180
Edyta Mazurek: The Income Tax Progression Depending on Social Insurance Contribution in Poland.....	190
Petr Musil, Jana Kramulová, Jan Zeman: Regional Consumption Expenditures: An Important Starting Point for Regional Input-output Tables.....	200
Katarzyna Ostasiewicz, Walenty Ostasiewicz: Good Life: From Political to Human Economy	208
Anna Sączewska-Piotrowska: Analysis of Poverty Transitions in Poland Using Multilevel Discrete-Time Event History Models	219
Martina Šimková, Petra Švarcová: Disadvantaged University Students in the Czech Republic	229
Michal Široký: The Use of Short-term Business Statistics for Quarterly GDP Flash Estimates in the Czech Republic.....	239
Zdeněk Šulc, Hana Řezanková: Evaluation of Recent Similarity Measures for Categorical Data.....	249
Lucia Švábová, Marek Ďurica: The Relationship Between the Finite Difference Method and Trinomial Trees	259
Kristýna Vltavská, Jaroslav Sixta: The Estimation of Final Consumption Expenditures	270
Lenka Vraná: Business Cycle Analysis: Tracking Turning Points	277
Janusz Wywiół: On Bayesian Testing in Auditing.....	284
Emília Zimková: Window Analysis of Supper-efficiency Change: Case of the Slovak Banking System	294
Beata Zmyślona: Statistical Modelling of the Impact of Diabetes on the Risk of Hospitalization	301

THE USE OF FINITE MIXTURE MODEL FOR DESCRIBING DIFFERENCES IN UNEMPLOYMENT DURATION

IVANA MALÁ

University of Economics, Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, W. Churchill Sq. 4, 130 67, Prague 3
email: malai@vse.cz

Abstract

The unemployment is a serious problem of all developed economies. There is a wide range of analyses of this phenomenon, especially of its rate and duration, from different points of view in economic or statistical literature. In this contribution the probability distribution of unemployment duration in the Czech Republic is analysed. The model of finite mixtures of probability distributions is used to model the distribution of unemployment duration and to describe (and quantify) differences of its distribution in subgroups of the unemployed given by factor explanatory variables gender (with two levels) and education (with three levels). The impact of the choice of component probability distributions on the estimates of survival and hazard functions and the estimates of population characteristics of location and variability based on quantiles is treated. Data from Labour Force Sample Survey performed quarterly by the Czech Statistical Office are used in order to find maximum likelihood estimates of component parameters in subpopulations and mixing proportions. Frequently used positively skewed distributions (lognormal, loglogistic, gamma and Weibull) are used and quality of different fits is compared. The positive impact of education on the unemployment spell as well as gender gap will be quantified and shown in figures. The lognormal distribution appears to be the best of analysed distributions.

Key words: *interval censored data, finite mixture of distributions, unemployment, ML estimates.*

DOI: 10.15611/amse.2014.17.18

1. Introduction

In this text we try to discuss the possibility to use finite mixture models in order to describe the probability distribution of the duration of unemployment in the Czech Republic in 2010. Suitable models for this phenomenon can be useful for all who are interested in this serious problem affecting strongly the society. Data dealing with the unemployment from different points of view are used by national or international authorities, business or media, and other public institutions. Also the general public is interested in changes in unemployment. When describing unemployment, the unemployment rate is frequently used. This characteristic (there are more definitions, Czech Statistical Office, 2014) is used to characterize the development of unemployment, regional differences or differences between countries. For recent analyses of the unemployment in the Czech Republic we refer to Čabla, 2012, Löster, Langhamrová, 2011 or Pavelka, Löster, 2013. There is a huge literature dealing with the problem of unemployment duration modelling from economic, econometric

or statistical point of view. The serious negative consequences of unemployment are discussed for example in Jürges, 2007.

In this contribution the parametric approach is used and four two-parametric probability distributions (lognormal, loglogistic, Weibull and gamma) are fitted into data and they are also used as the component distributions in mixture models. The Turnbull nonparametric estimate of survival function (Lawless, 2003) is also used to describe the distribution of the length of unemployment spell. Wichert and Wilke (2008) construct nonparametric model for unemployment in Germany up to 2001 and dependencies on age or previous wage are shown. In this text we try to quantify differences between groups in the Czech labour market given by gender (we can be interested whether the unemployment for women tends to be longer than for men) or education (the impact of high education on the unemployment duration is of interest).

A mixture of probability distributions might be a useful tool to describe very non-homogenous distribution of the length of unemployment spell and include influence of above mentioned factors.

2. Methodology

In this part a finite mixture of probability density is defined and its properties that are used in this article are given (McLachlan, Peel, 2000). Suppose now, that T is positive time to event random variable with continuous distribution (the duration of unemployment, time to finding a job). We will denote f density, F cumulative distribution function, S survival function ($S(t) = 1 - F(t) = P(T > t), t \in R$) and t_p 100P % quantile of T . As the distribution of T should be positively skewed and heavy tailed probability distribution, quantile characteristics median $t_{0.5}$ and quartile deviation $q = 0.5(t_{0.75} - t_{0.25})$ will be more predicative characteristics of the location and variability than moment characteristics as expected value and standard deviation.

A mixture of K probability distributions (with component densities $f_j, j = 1, \dots, K$) can be described with a density

$$f(t; \boldsymbol{\psi}) = \sum_{j=1}^K \pi_j f_j(t; \boldsymbol{\theta}_j), \quad (1)$$

where $\boldsymbol{\theta}_j, j = 1, \dots, K$ are vectors of unknown component parameters, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$ is a vector of mixing proportions (weights of the components) satisfying $\sum_{j=1}^K \pi_j = 1$ and $\boldsymbol{\psi}$ is a vector of all unknown parameters of the mixture model (1). From (1) we obtain (F_j (S_j) is the distribution function (resp. survival function) of the j^{th} component)

$$F(t; \boldsymbol{\psi}) = \sum_{j=1}^K \pi_j F_j(t; \boldsymbol{\theta}_j), \quad (2)$$

and

$$S(t; \boldsymbol{\psi}) = 1 - \sum_{j=1}^K \pi_j F_j(t; \boldsymbol{\theta}_j) = \sum_{j=1}^K \pi_j S_j(t; \boldsymbol{\theta}_j). \quad (3)$$

Quantiles of mixture distributions are evaluated numerically from the definition as a root of equation $F(t_p) = P$ or $S(t_p) = 1 - P$, $0 < P < 1$, these functions are given in (2) and (3).

In our data we have only right censored observations (for those who did not find a job) and interval censored observations (for those who have got a job). The observable explanatory variables gender ($K = 2$; 5 parameters) and education ($K = 3$; 8 parameters) were used, and the problem of estimation of $\boldsymbol{\psi}$ can be split into estimation of component parameters in subsets defined by these factors (McLachlan, Peel, 2000). Maximum likelihood estimates of mixing proportions are given by relative frequencies of these subgroups in the whole dataset.

The unemployment duration is in the survey observed in intervals (in months) under one month, 1 - 3, 3 - 6, 6 - 12, 12 - 18, 18 - 24, 24 - 48 and over 48 months and no exact durations are recorded. Moreover we know whether the unemployed found or did not find a job. Intervals were transformed in order to describe the duration in the spell as exactly as possible with the use of additional information obtained from the survey data. For the i -th unemployed denote the above mentioned interval (l_i, u_i) .

There are no explicit formulas for maximum likelihood estimates of component distributions in the presence of censored data. Right censored observation t from the j -th component censored in the interval (l, u) contributes to the component likelihood function with $\pi_j S_j(l; \boldsymbol{\theta}_j)$, interval censored observation with $\pi_j (F_j(u; \boldsymbol{\theta}_j) - F_j(l; \boldsymbol{\theta}_j))$.

In the analysis we use probability distributions in scale and shape parameterization in the form:

lognormal distribution (scale parameter σ^2 , shape parameter μ), $\boldsymbol{\psi} = (\boldsymbol{\pi}, \mu_j, \sigma_j^2, j = 1, \dots, K)$

$$f_{LN}(t; \boldsymbol{\psi}) = \sum_{j=1}^K \frac{\pi_j}{\sqrt{2\pi}\sigma_j t} \exp\left(-\frac{(\ln t - \mu_j)^2}{2\sigma_j^2}\right), \quad (4)$$

gamma distribution (scale parameter δ , shape parameter m), $\boldsymbol{\psi} = (\boldsymbol{\pi}, m_j, \delta_j, j = 1, \dots, K)$

$$f_{gamma}(t; \boldsymbol{\psi}) = \sum_{j=1}^K \frac{\pi_j}{\delta_j^{m_j} \Gamma(m_j)} t^{m_j-1} \exp\left(-\frac{t}{\delta_j}\right), \quad (5)$$

loglogistic distribution (scale parameter δ , shape parameter m), $\boldsymbol{\psi} = (\boldsymbol{\pi}, m_j, \delta_j, j = 1, \dots, K)$

$$f_{LL}(t; \boldsymbol{\psi}) = \sum_{j=1}^K \pi_j \left(1 + \left(\frac{t}{\delta_j}\right)^{m_j}\right)^{-2} \frac{m_j}{\delta_j} \left(\frac{t}{\delta_j}\right)^{m_j-1}, \quad (6)$$

Weibull distribution (scale parameter δ , shape parameter m), $\Psi = (\pi, m_j, \delta_j, j = 1, \dots, K)$

$$f_{Weibull}(t; \Psi) = \sum_{j=1}^K \pi_j \frac{m_j}{\delta_j} \left(\frac{t}{\delta_j} \right)^{m_j-1} \exp \left[\left(- \frac{t}{\delta_j} \right)^{m_j} \right]. \quad (7)$$

For the hazard rate of a mixture we obtain (Lawless, 2003)

$$h(t; \Psi) = \lim_{h \rightarrow 0} \frac{P(t \leq T < t+h)}{h} = \frac{\sum_{j=1}^K \pi_j f_j(t; \theta_j)}{\sum_{j=1}^K \pi_j S_j(t; \theta_j)} = \frac{\sum_{j=1}^K \pi_j S_j(t; \theta_j) h_j(t; \theta_j)}{\sum_{j=1}^K \pi_j S_j(t; \theta_j)}. \quad (8)$$

It follows from this formula (8) that the hazard rate of a mixture is the weighted average of component hazards, but the weights depend on t and component survival functions.

Maximum likelihood estimates of unknown parameters (in (4) - (7) for $K=1$ (one distribution), $K=2$ (gender) a $K=3$ (education)) were used in parametric models.

All computations were performed in the program R (R CORE TEAM, 2014), package FITDISTRPLUS (Delignette-Muller et al., 2014) was used to estimate parameters. This package the estimates for gamma distribution a shape parameter and a rate parameter, the scale parameter was evaluated as 1/rate and standard deviation was estimated with the use estimated standard deviation of rate parameter and second order Taylor approximation. Standard deviations of estimated parameters are provided by the procedure *fitdistcens*, 90% confidence limits for estimated quantiles were obtained by the procedure *bootdistcens*, where bootstrap is used for the construction of limits. Number of resamples was set to 1,000.

Package SURVIVAL (Therneau, Grambsch, 2000, Therneau, 2014) includes procedures for evaluating nonparametric estimates for interval censored data and Turnbull estimates were evaluated and plotted in this package. The package FAdist (Aucoin, 2012) was used for calculations with loglogistic distribution.

3. Results

In the analysis data from the Labour Force Sample Survey (LFSS) organised quarterly by the Czech Statistical Office were used (Czech Statistical Office, 2014). Declared unemployment spell was used from the survey that provides thorough information about employment and unemployment in the Czech households. All unemployed respondents with unemployment spell under two years in five consecutive surveys (one year, surveys from the first quarter of 2010 to the first quarter of 2011) were included together with information about gender and education. This dataset consists of 4,482 respondents (2,245 men and 2,237 women). From these people only 886 (20%) were registered in local Labour Office and obtained unemployment benefits, 1,842 were registered without obtaining benefits and 1,750 were even unregistered. More detailed description of the data is given in Čabla, 2012.

In the Table 1 the point estimates of parameters accompanied by an estimate of accuracy (estimated standard deviation, SD) are given for all distributions. Estimated component

weights are 0.501 for men and 0.499 for women and 0.155 for basic, 0.774 secondary and 0.071 for tertiary education.

Table 1. Estimates of parameters for all analysed models.

model	par.	lognormal		gamma		Weibull		loglogistic	
		est.	SD	est.	SD	est.	SD	est.	SD
all	shape	2.7096	0.0251	1.9756	0.0754	1.6177	0.0416	0.5318	0.0141
	scale	0.9202	0.0237	9.0456	0.5143	18.8716	0.4124	2.7107	0.0233
men	shape	2.6598	0.0339	0.5107	0.1037	1.5978	0.0568	0.5333	0.0195
	scale	0.9166	0.0325	8.7573	0.6823	18.1719	0.5408	2.6625	0.0318
women	shape	2.7623	0.0372	0.5003	0.1101	1.6436	0.0613	0.5292	0.0203
	scale	0.9226	0.0346	9.3374	0.7756	19.6023	0.6284	2.7601	0.0343
basic	shape	3.1681	0.0915	0.4442	0.2638	1.9091	0.1598	0.4909	0.0416
	scale	0.9321	0.0765	11.2752	2.1132	25.1701	1.8438	3.1065	0.0761
secondary	shape	2.6473	0.0267	0.4991	0.0853	1.6200	0.0459	0.5238	0.0153
	scale	0.8989	0.0256	8.4462	0.5249	17.9845	0.4217	2.6518	0.0253
tertiary	shape	2.3972	0.0821	0.5265	0.2590	1.5091	0.1324	0.5270	0.0492
	scale	0.8901	0.0807	7.2087	1.4084	14.8346	1.1195	2.3998	0.0805

Source: Own calculations.

In the Table 2 the quantile characteristics of the level and variability are provided. Median as a characteristic of the location and quantile deviation (both characteristics are given in months) as a characteristic of variability are given, medians for the whole population and subgroups were evaluated with the use of bootstrap, for the mixture these values were evaluated numerically from the estimated mixture distribution.

Table 2. Quantile characteristics of location (median) and variability (quantile deviation) for all analysed models.

model		lognormal		gamma		Weibull		loglogistic	
		median	q	median	q	median	q	median	q
all		15.02	9.93	14.96	7.77	15.05	7.18	15.04	9.30
gender	men	14.29	9.41	14.33	7.48	14.45	6.98	14.33	8.89
	women	15.84	10.50	15.66	8.08	15.68	7.36	15.80	9.71
	mixture	15.04	9.96	14.98	7.79	15.05	7.18	15.05	9.31
education	basic	23.76	15.94	21.74	10.48	20.77	8.38	22.34	12.64
	secondary	14.12	9.09	14.20	7.32	14.34	6.83	14.18	8.62
	tertiary	10.99	7.00	11.38	6.05	11.64	5.96	11.02	6.74
	mixture	14.95	9.97	14.90	7.82	14.99	7.18	14.96	9.29

Source: Own calculations.

The positive impact of the education is visible for all distributions; the only median less than one year is that in the group with tertiary education (10.99-11.63 months). We can see comparable estimates in all probability distributions, usually varying within one month. The highest variability is in the group of the unemployed with basic education.

All models were compared with Bayesian information criterion BIC, the lowest value of BIC the better model. From the Table 3 the superiority of lognormal distribution as a distribution for components is obvious. Moreover the estimation in this distribution seems to be more numerically stable with no problems even in 1,000 bootstrap replications. The model is constructed based on given explanatory variables; the results do not lead to better model (with respect to the BIC criterion). The model with two components (gender, $K = 2$) has 5 parameters to be estimated, in case of 3 components (education, $K = 3$) it is 8 parameters that are reflected in the BIC value. Comparing overall model with two mixture models graphically, the differences between mixtures and one component model are negligible.

Table 3. Values of BIC for the analysed models.

model		lognormal	gamma	Weibull	loglogistic
all		5 196	5 206	5 213	5 231
gender	men	2 707	2 717	2 722	2 728
	women	2 497	2 497	2 346	2 511
	mixture	5 221	5 231	5 237	5 256
education	basic	618	613	610	616
	secondary	4 126	4 137	4 145	4 157
	tertiary	392	396	398	395
	mixture	5 214	5 220	5 301	5 323

Source: Own calculations.

In the Figure 1 estimated cumulative distribution function are shown (one distribution, $K=1$) for the whole dataset together with Turnbull nonparametric estimate (step function with 90% confidence limits). We can see very similar functions in the time limits where we have observations (under 24 months). Moreover we can see long steps of Turnbull estimate due to the type of data with long censoring intervals. In our data (due to the limitations set on the unemployment duration) we cannot estimate survival function after 20 months of unemployment. On the other hand it is possible to extend the estimated (parametric) cumulative density function or survival function to longer unemployment spells.

In the Figure 2 the hazard rate functions are illustrated for both mixture models (together with component hazard rates). These functions have maximal values (the higher intensity of finding a new job) at 10.8 months for men, 11.8 months for women (the mixture at 11.5 months), 17.3 months for basic, 11.2 months for secondary and 8.8 months for tertiary education (mixture at 11.0 months). Hazard rates for both mixtures are very similar.

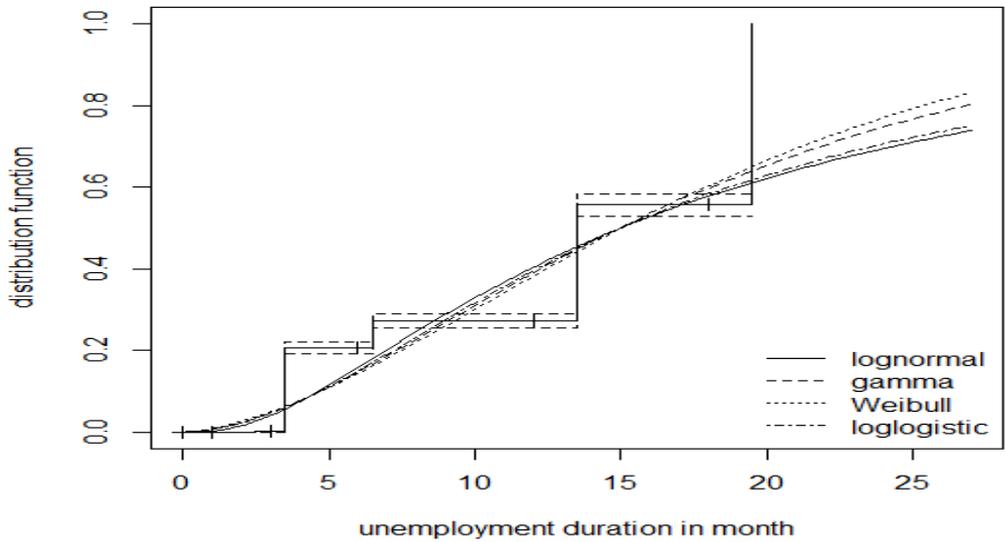


Figure 1: Estimated distribution functions of unemployment duration (months) and Turnbull estimate with 90% confidence limits
 Source: own calculations.

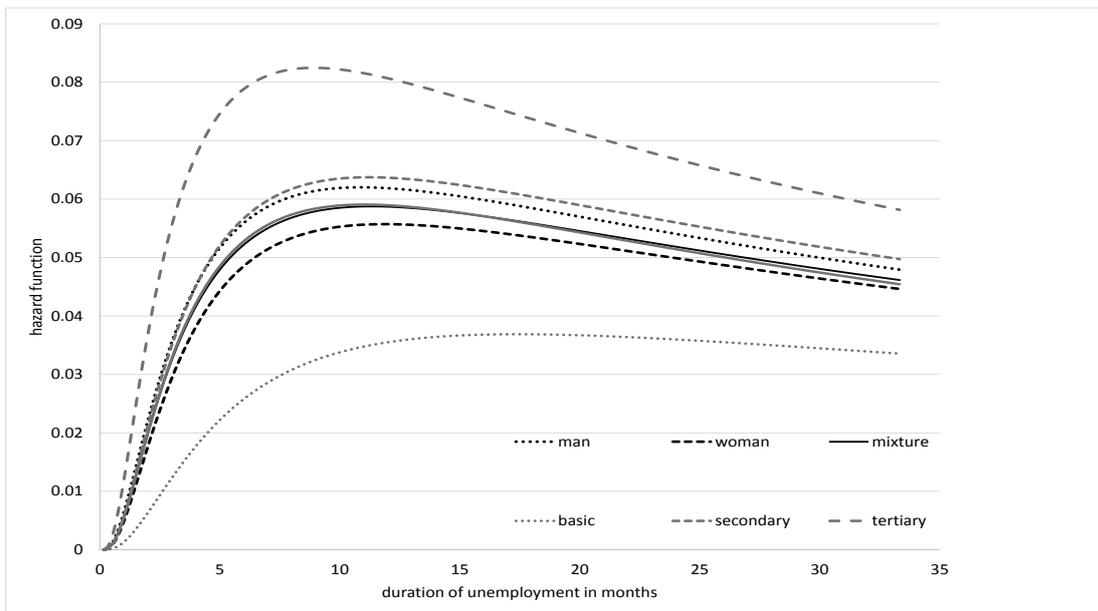


Figure 2: Estimated hazard functions for subgroups (men, women (black lines), basic, secondary and tertiary education (grey lines)) and mixtures (solid lines)
 Source: own calculations.

In the Figure 3 the Turnbull estimate of overall unemployment duration is compared with the mixture of Turnbull estimates in subgroups (with mixing proportions estimated by relative frequencies). This procedure combines nonparametric and parametric approach.

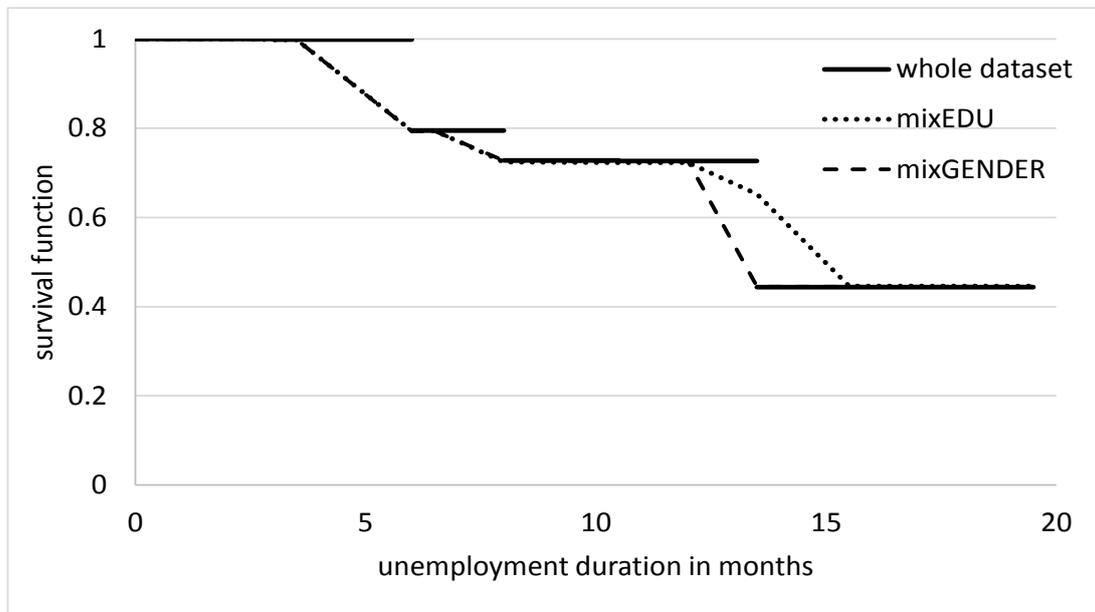


Figure 3: Survival functions, combination of Turnbull nonparametric estimate and parametric estimate of mixing proportions

Source: own calculations.

4. Conclusion

In the text various models for the unemployment duration in the Czech Republic are constructed, unknown parameters are estimated and characteristics of location and variability are given.

In the analysed period the results from four used distributions are similar, however the lognormal distribution seems to be superior. Moreover the mixture model does not improve the fit. We obtain information about components, but the resulting model is not better (measuring by BIC or graphically). We may conclude (with caution) that the duration of unemployment in the Czech Republic in 2010 can be modelled with the use of single lognormal distribution, mixture models with explanatory variables gender and education are comparable.

Mixture models give quantification of differences between subgroups. In this text we used median, quantile deviation and maximum of hazard rate. If we use lognormal distribution of components and median duration of unemployment, women have median duration of unemployment 15.8 months comparing to 14.3 month for men (difference of 1.5 months). Positive impact of education can be described by median durations of 23.7 months for basic education, 14.1 for secondary education and 11 months for tertiary education (more than one year long interval). Variability is greater for women than for men, for basic education than for secondary and tertiary education. All hazard functions have one maximum.

The program R was used for all computations, the packages Survival and Fitdistrplus provides procedures that are necessary for computations needed for the presented analysis (for results shown in the text or not incorporated).

References

- 1 AUCCOIN, F. 2012. Distributions that are sometimes used in hydrology. R package version 2.0. <http://cran.r-project.org/web/packages/FAdist/index.html>.
- 2 CZECH STATISTICAL OFFICE. www.czso.cz. 12.3.2014
- 3 ČABLA, A. 2012 Unemployment duration in the Czech Republic, pp. 257-267, Prague 13.09.2012 – 15.09.2012. In: *International Days of Statistics and Economics at VŠE, Prague*. Praha: VŠE.
- 4 DELIGNETTE-MULLER, M. L., POUILLOT, R., DENIS, J.-B., DUTANG, C. Fitdistrplus: help to fit of a parametric distribution to non-censored or censored data, R package version 1.0-2. 2014. <http://cran.r-project.org/web/packages/fitdistrplus/index.html>.
- 5 JÜRGES, H. 2007. Unemployment, Life Satisfaction and Retrospective Error. *Journal of the Royal Statistical Society. Series A, Statistics in Society*, Vol. 170, No. 1, pp. 43-61.
- 6 LAWLESS, J. F. 2003. *Statistical models and methods for lifetime data*. 2nd ed. Hoboken: John Wiley & Sons.
- 7 LÖSTER, T., LANGHAMROVÁ, J. 2011. Analysis of Long-term Unemployment in the Czech Republic. Praha 22.12.2011 – 23.12.2011. In: LÖSTER, T., PAVELKA, (ed.). *International Days of Statistics and Economics*. Slaný: Melandrium, pp. 228–234.
- 8 PAVELKA, T., LÖSTER, T. 2013. Flexibility of the Czech Labour Market from a Perspective of the Employment Protection Index. In: *The 7th International Days of Statistics and Economics. Conference Proceedings*. [online] Praha, 19.09.2013 – 21.09.2013. Slaný: Melandrium, 2013, pp. 1090–1099.
URL: <http://msed.vse.cz/files/2013/7-Pavelka-Tomas-paper.pdf>.
- 9 MCLACHLAN, G. J., PEEL, D. 2000. *Finite Mixture Models*. Wiley series in Probability and Mathematical Statistics: Applied Probability and Statistics Section, New York.
- 10 R CORE TEAM 2014: R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, <http://www.r-project.org/>.
- 11 WICHERT, L., WILKE, R. A. 2008. Simple Non-Parametric Estimators for Unemployment Duration Analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 57, No. 1, pp. 117-126.
- 12 THERNEAU, T. 2014. *A Package for Survival Analysis in S*. R package version 2.37-7. <http://cran.r-project.org/web/packages/survival/survival.pdf>
- 13 THERNEAU, T. M., GRAMBSCH, P. M. *Modeling Survival Data: Extending the Cox Model*. Springer, New York. 2000.