

UNEMPLOYMENT DURATION IN THE CZECH REPUBLIC AFTER THE ECONOMIC CRISIS

ADAM ČABLA

University of Economics in Prague, Faculty of Informatics and Statistics,
Department of statistics and probability
Nám. W. Churchilla 4, Prague, Czech Republic
email: adam.cabla@vse.cz

Abstract

The aim of the paper is to provide deeper look at the probabilistic distribution of the unemployment duration in the Czech Republic after the last economic crisis, specifically during the year 2014. The focus is on the description of the situation.

Methodologically, the paper contains nonparametric and parametric (log-logistic) estimates of the unemployment duration for the whole society and for different strata according to the sex, ISCED, age group and NUTS2 region. The paper contains estimates of the differences, quantiles, mean and variance.

The most surprising finding is the fact that unemployment duration did not differ significantly between different age groups and regions. It is shift from the previous state. Other findings are in line with the previous - women and less educated looks for a job longer.

Key words: *unemployment duration, Turnbull, log-logistic, AFT, crisis.*

1. Introduction

From the statistical point of view the problem of unemployment is often described by the unemployment rate, but little is usually said about unemployment duration. The usual statistic that covers the topic is the rate of long-term unemployment i.e. the rate of those who are unemployed longer than one year.

Some statistics about the average duration of unemployment are provided by the OECD (OECD, 2012), but they seem quite unrealistic. A deeper look at the unemployment duration in the Czech Republic was provided in (Jarošová et al., 2004), (Jarošová and Malá, 2005), (Malá, 2014) and (Čabla, 2014).

The objective and main contribution of the paper to current understanding of the unemployment issue is to provide deeper insight into the probabilistic distribution of the unemployment duration in the Czech Republic during recovery period.

To achieve this aim the author uses two different approaches, namely Turnbull (non-parametric) estimate and accelerated failure time (AFT) model with log-logistic distribution.

To test statistical significance of the differences between distributions the non-parametric asymptotic log rank test and test of the significance of parameters in the AFT model are used.

Based on previous research (Čabla, 2012), the current paper aims at the mapping of the distribution and differences between unemployment durations stratified by the sex, education according to ISCED, region and age group.

The paper is divided into four other parts - first describes how data were collected, which is crucial for understanding choices of methodology; second briefly describes methodology; third shows results and fourth is discussion of them.

2. Data

Data were obtained from the Czech Statistical Office (CZSO) from the Labor Force Surveys (LFS) in the period Q4/2013 – Q4/2014. These data thus represent the up-to-date evaluation of the topic.

The Labor Force Survey is a large sample survey of private households and is an important source for statistics about the labor market (Eurostat, 2013).

The survey is done quarterly and 20 % of the chosen households are changed every quarter. This means that every participant can take part in five consecutive surveys if he is willing to. In one survey approximately 50 – 60 thousand persons are questioned.

One of the questions in the survey asks for the duration of the job search (which can be different from the duration of unemployment *per se*) and another one asks for the duration of the current job. As a person is questioned over five quarters, one can find those, who obtained job in this survey period and compute the search duration.

The status of economic activity is checked at the entry into the survey and exit, which makes it much easier to find the participants who found a job, but it means that there are possible omitted cases – firstly in the situation in which a participant finds a job in between and then loses it, secondly the situation in which participants lose jobs and then find and lastly the situation when they find a job, lose it and find it again. These cases are possible but not very likely, so their omission should not change the overall results.

As the answers to the stated questions are interval censored, so is the consequent computed duration. Finally 677 persons who found a job were identified. Four of these persons were omitted because their interval was nonsensical - stating that the upper limit of the interval was negative etc. 19 observations were actually right censored as their maximum time of looking for a job was unknown.

It is important to recall that the paper deals only with the duration of unemployment of those who were unemployed at the beginning and found a job during the follow-up period, so the unemployment duration is not limited by the length of the follow-up period.

All times are measured in months, which is convenient as original data are in months too.

3. Methodology

The main methodological problem lies in dealing with the interval censored data that are outcome from the LFS data sets. The deeper methodological sources are provided in the author's previous paper on the topic (Čabla, 2012).

The nonparametric Turnbull estimate of survival function (Turnbull, 1976), (Shen, 2012) for the whole population and every stratum in focus was made. Using these estimates the mean duration of the unemployment was estimated. Mean is calculated as restricted (*rmean*) meaning that the highest possible value is firmly set. The mean is calculated using assumption of uniform distribution within the Turnbull intervals.

Note that the Turnbull estimate estimates the probabilities of the observation being in the interval but not the probability distribution within the intervals.

The first test to find those strata for which the survival function is different is asymptotic log-rank two (or *k*) sample test with Sun's scores based on 999 Monte Carlo simulations (Kim, 2006), (Fay and Shaw, 2010). The null hypothesis of the test states that "two survival functions are the same".

The second method to estimate the change of survival function and test it is parametric regression, namely the accelerated failure time model (AFT):

$$S(t|\mathbf{x}) = S_0(\exp(\boldsymbol{\beta}'\mathbf{x})t) \quad (1)$$

The underlying survival function $S_0(t)$ is chosen (see chapter 4) to be log-logistic:

$$S_0(t) = \left(1 + (t/\alpha)^\beta\right)^{-1} \quad (2)$$

In this model it is assumed that the covariate changes time scale, thus $\exp(\beta)$ gives the fold of change of time compared to the selected basis and parameters of distribution are stable, see (1). The estimates of all parameters (of log-logistic distribution and β_i) are done by the maximum likelihood method adapted for interval censored data. Using this method estimates of following quantiles - $x_{0.1}$, $x_{0.25}$, $x_{0.5}$, $x_{0.75}$ and $x_{0.9}$, mean, mode and variance was obtained.

All calculations are done in the R software using packages interval and survival.

All the methodology is described for example in (Klein et al, 1997).

4. Results

For the whole population the plot of nonparametric estimate of the survival function with the parametric estimates is in the Figure 1. Grey areas in the Figure 1 depict the uncertainty of the probability distribution within the intervals of Turnbull estimate. Table 1 shows nonparametric estimate of restricted mean (upper limit 58 months).

Table 1. Turnbull restricted mean estimate for the whole population

| Population | Observations | rmean | s.e. (rmean) |
|------------|--------------|-------|--------------|
| Whole | 673 | 12.9 | 0.475 |

Source: CZSO, own calculations

It is important for the parametric regression to establish proper distribution. Estimates for several easily obtainable distributions were computed and author decided to use log-logistic which has the highest value of the logarithm of likelihood (LogLik) as noted in Table 3. Lognormal distribution has slightly smaller LogLik statistic and is quite similar, as it is shown in Figure 2. Estimated parameters of log-logistic distribution for whole population are $\alpha = 9.19$ and $\beta = 2.16$ and resulting characteristics are in the Table 2.

Table 2. Estimated quantiles and other characteristics via AFT model

| Population | $x_{0.1}$ | $x_{0.25}$ | $x_{0.5}$ | $x_{0.75}$ | $x_{0.9}$ | Mean | Mode | S.D. |
|------------|-----------|------------|-----------|------------|-----------|------|------|------|
| Whole | 3.3 | 5.5 | 9.2 | 15.3 | 25.4 | 13.5 | 5.8 | 29.7 |

Source: CZSO, own calculations

Table 3. Different distributions and LogLik statistic

| Distribution | LogLik |
|--------------|-----------|
| Log-logistic | - 1 003.8 |
| Lognormal | - 1 010.1 |
| Weibull | - 1 046.8 |
| Exponential | - 1 060.1 |
| t | - 1 161.7 |
| Logistic | - 1 204.1 |
| Rayleigh | - 1 207.9 |
| Gaussian | - 1 269.0 |
| Extreme | -1 485.8 |

Source: CZSO, own calculations

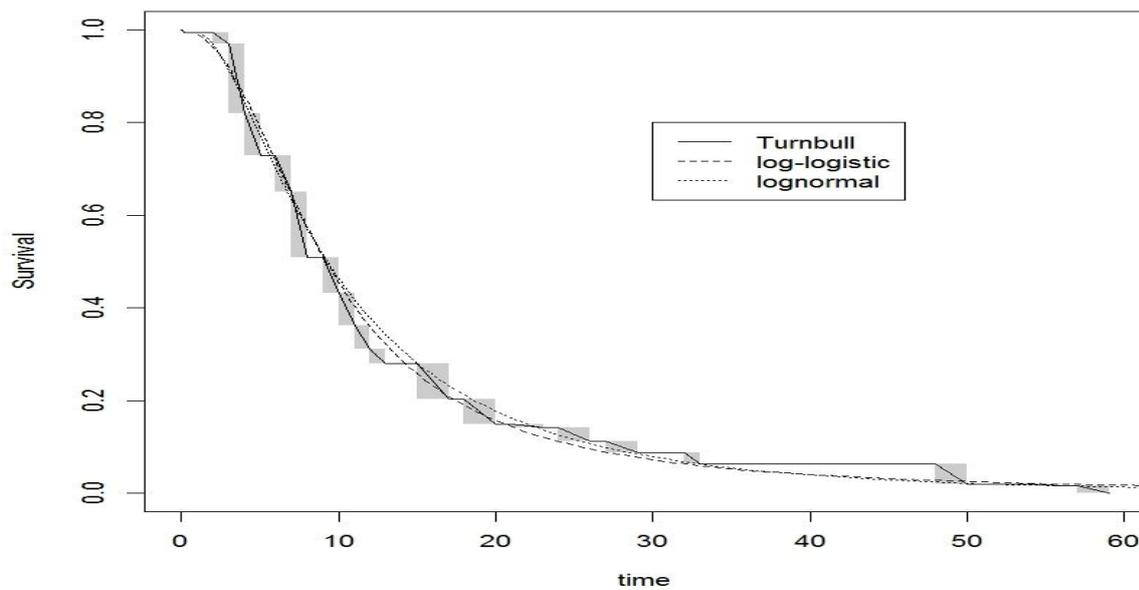


Figure 1. Plot of the Turnbull and two parametric estimates for the whole population

Source: CZSO, own calculations

4.1 Results for Sex

There are statistically significant ($\alpha = 0.05$) differences between the survival functions for men and women. Based upon the Turnbull estimates, it takes on average 2.4 months longer for women to find a job or 20.7 % longer - it shows Table 4 and Figure 2. Table 7 contains results from the standard non-parametric logrank test.

Parametric regression confirms the results stating that the difference is statistically significant and women search a job 17.0 % longer - the results are in Table 5, Table 6 and Figure 6.

Table 4. Turnbull restricted mean estimates for men and women

| Population | Observations | rmean | s.e. (rmean) |
|------------|--------------|-------|--------------|
| Men | 308 | 11.6 | 0.614 |
| Women | 365 | 14.0 | 0.685 |

Source: CZSO, own calculations

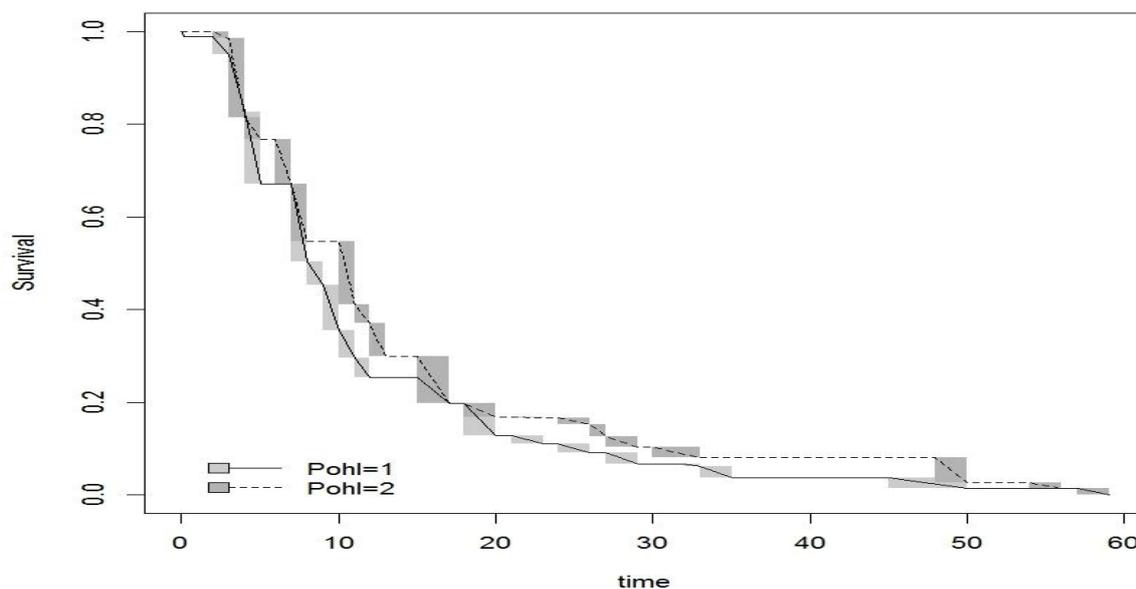


Figure 2. Plot of the Turnbull estimate for men and women separately

Source: CZSO, own calculations

Table 5. Estimated parameters of the AFT model

| Parameter | Estimate | Statistics |
|--------------------|----------|---|
| α | 8.44 | ChiSquare (1 DoF) = 5.76 p-value = 0.016 |
| β | 2.17 | |
| $\exp(\text{Sex})$ | 1.17 | |

Source: CZSO, own calculations

Table 6. Estimated quantiles and other characteristics via AFT model

| Population | $x_{0.1}$ | $x_{0.25}$ | $x_{0.5}$ | $x_{0.75}$ | $x_{0.9}$ | Mean | Mode | S.D. |
|------------|-----------|------------|-----------|------------|-----------|------|------|------|
| Men | 3.1 | 5.1 | 8.4 | 14.0 | 23.2 | 12.3 | 5.3 | 26.4 |
| Women | 3.6 | 5.9 | 9.9 | 16.4 | 27.2 | 14.4 | 6.2 | 30.9 |

Source: CZSO, own calculations

Table 7. Logrank two-sample test

| Population | Observations | Score Statistic | Statistics |
|------------|--------------|-----------------|--------------------------|
| Men | 308 | 25.98 | $z = 2.34$ |
| Women | 365 | -25.98 | $p\text{-value} = 0.020$ |

Source: CZSO, own calculations

4.2 Results for ISCED

Unemployment duration differs significantly ($\alpha = 0.01$) between subgroups constituted by various ISCED levels. Turnbull estimates are in Figure 3 and Table 8. Logrank k-sample test confirms the significance of the difference - it is shown in Table 9.

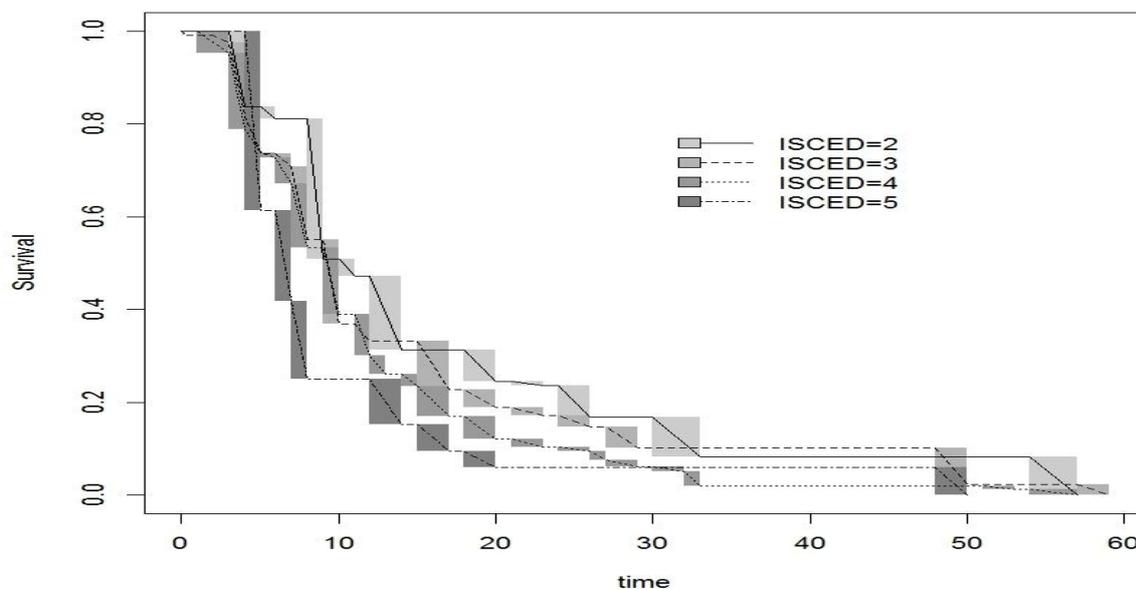


Figure 3. Turnbull estimates for various ISCED levels

Source: CZSO, own calculations

Table 8. Turnbull restricted mean estimates for various ISCED levels.

| Population | Observations | rmean | s.e. (rmean) |
|------------|--------------|-------|--------------|
| ISCED = 2 | 54 | 16.3 | 1.968 |
| ISCED = 3 | 302 | 14.3 | 0.796 |
| ISCED = 4 | 234 | 11.5 | 0.627 |
| ISCED = 5 | 83 | 10.1 | 1.164 |

Source: CZSO, own calculations

Table 10 contains estimated parameters of the parametric regression and Table 11 estimated characteristics via this estimate. It is safe to say, that higher the education, shorter the duration of unemployment.

Table 9. Logrank k-sample test

| Population | Observations | Score Statistic | Statistics |
|-------------------------------------|--------------|-----------------|--|
| ISCED = 2 (Basic) | 54 | -12.47 | ChiSquare (3 DoF) = 16.17 p-value = 0.001 |
| (H.S. without final exam) | 302 | -25.50 | |
| ISCED = 4 (H.S. with final exam) | 234 | 15.93 | |
| ISCED = 5 (University) | 83 | 22.04 | |

Source: CZSO, own calculations

Table 10. Estimated parameters of the AFT model (basis is ISCED = 2)

| Parameter | Estimate | Statistics |
|----------------|----------|--|
| α | 11.28 | ChiSquare (7 DoF) = 16.07 p-value = 0.001 |
| β | 2.19 | |
| exp(ISCED = 3) | 0.87 | |
| exp(ISCED = 4) | 0.79 | |
| exp(ISCED = 5) | 0.61 | |

Source: CZSO, own calculations

Table 11. Estimated quantiles and other characteristics via AFT model

| Population | $x_{0.1}$ | $x_{0.25}$ | $x_{0.5}$ | $x_{0.75}$ | $x_{0.9}$ | Mean | Mode | S.D. |
|------------|-----------|------------|-----------|------------|-----------|------|------|------|
| ISCED = 2 | 4.1 | 6.8 | 11.3 | 18.6 | 30.7 | 16.3 | 7.2 | 32.8 |
| ISCED = 3 | 3.6 | 6.0 | 9.9 | 16.3 | 26.9 | 14.3 | 6.3 | 28.8 |
| ISCED = 4 | 3.3 | 5.4 | 9.0 | 14.8 | 24.4 | 13.0 | 5.7 | 26.1 |
| ISCED = 5 | 2.5 | 4.1 | 6.8 | 11.3 | 18.6 | 9.9 | 4.4 | 19.9 |

Source: CZSO, own calculations

4.3 Results for Age groups

Unemployment duration does not differ significantly ($\alpha = 0.05$) between subgroups constituted by various age groups. Turnbull estimates are in Figure 4 and Table 13. Statistically significant difference between survival curves was not found, as Table 12 shows.

If specific age group 15 - 19 years was not omitted, there would be however significant difference.

Table 12. Logrank k-sample test

| ChiSquare (7 DoF) | p-value |
|-------------------|---------|
| 6.32 | 0.503 |

Source: CZSO, own calculations

Table 13. Turnbull restricted mean estimate for different age groups.

| Population | Observations | rmean | s.e. (rmean) |
|---------------------|--------------|-------|--------------|
| Age Group = 20 - 24 | 123 | 12.2 | 0.996 |
| Age Group = 25 - 29 | 98 | 11.0 | 0.908 |
| Age Group = 30 - 34 | 68 | 14.3 | 1.535 |
| Age Group = 35 - 39 | 95 | 14.5 | 1.392 |
| Age Group = 40 - 44 | 77 | 12.9 | 1.413 |
| Age Group = 45 - 49 | 66 | 13.3 | 1.541 |
| Age Group = 50 - 54 | 57 | 14.2 | 1.606 |
| Age Group = 55 + | 60 | 16.2 | 1.865 |

Source: CZSO, own calculations

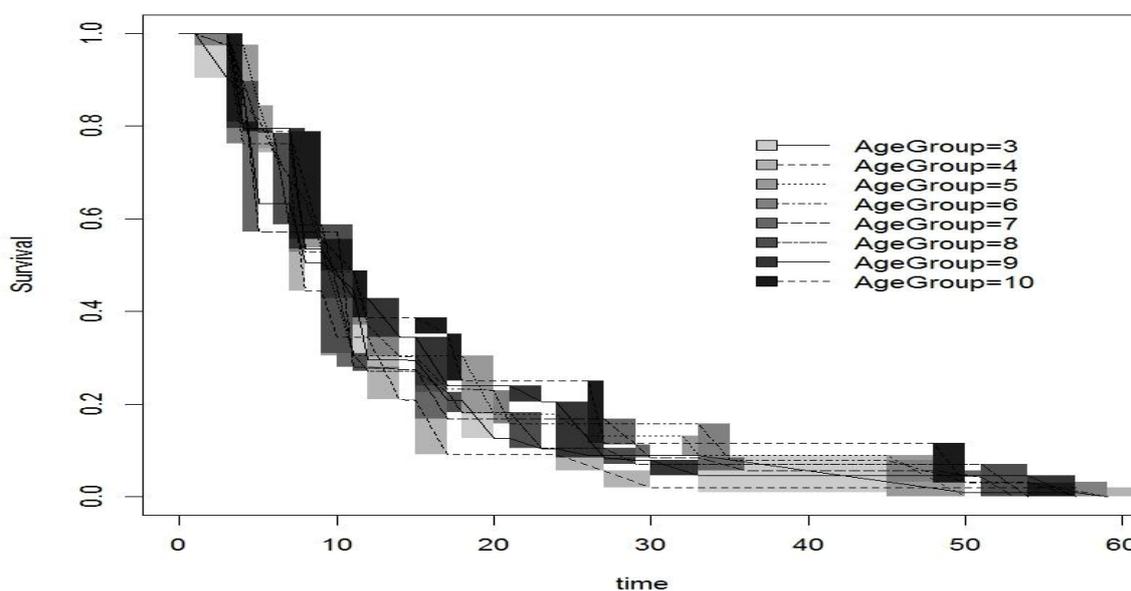


Figure 4. Cox model and AFT model estimates for various ISCED levels

Source: CZSO, own calculations

4.4 Results for NUTS 2 regions

Unemployment duration does not differ significantly ($\alpha = 0.05$) between subgroups constituted by NUTS 2 regions of the habitation. Turnbull estimates are in Figure 5 and Table 15. Statistically significant difference between survival curves was not found, as Table 14 shows.

Table 14. Logrank k-sample test

| ChiSquare (7 DoF) | p-value |
|-------------------|---------|
| 7.42 | 0.387 |

Source: CZSO, own calculations

Table 15. Turnbull restricted mean estimates for various regions

| Population | Observations | rmean | s.e. (rmean) |
|----------------|--------------|-------|--------------|
| NUTS2 = C.B. | 60 | 10.3 | 0.814 |
| NUTS2 = C.M. | 72 | 12.7 | 1.525 |
| NUTS2 = M.S.R. | 76 | 11.4 | 1.262 |
| NUTS2 = NE | 130 | 14.3 | 1.240 |
| NUTS2 = NW | 81 | 13.1 | 1.256 |
| NUTS2 = Prague | 27 | 10.3 | 1.568 |
| NUTS2 = SE | 123 | 13.9 | 1.076 |
| NUTS2 = SW | 104 | 13.6 | 1.268 |

Source: CZSO, own calculations

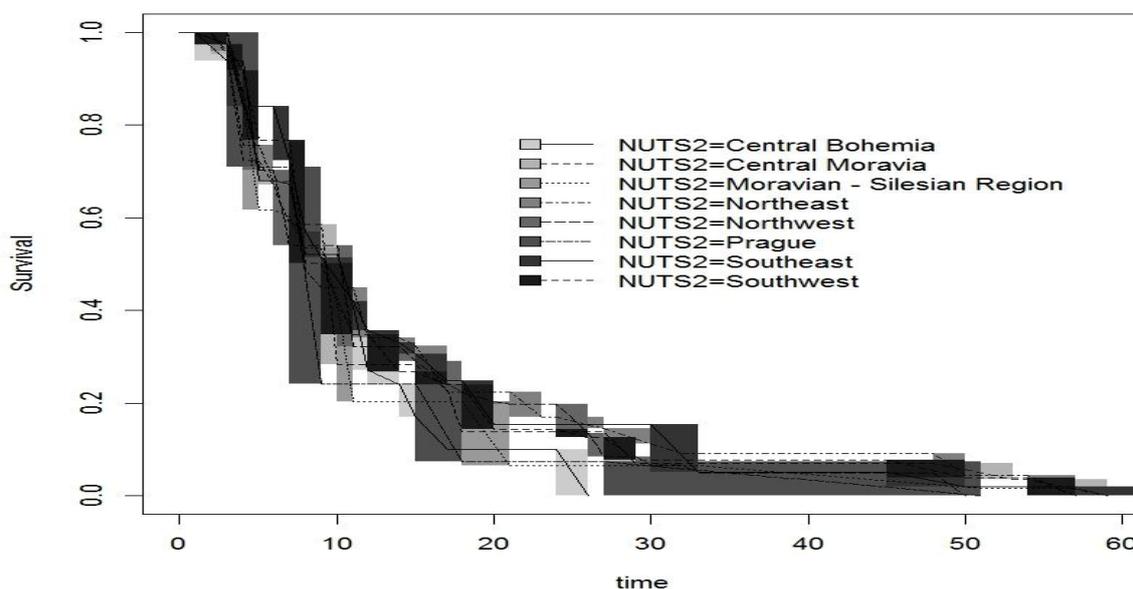


Figure 5. Cox model and AFT model estimates for various regions

Source: CZSO, own calculations

5. Discussion

In the paper author used series of simple models with one independent factor (sex, ISCED region, age group) and one dependent variable - duration of unemployment. This does not take into account possible relations between these variables, e.g. sex and education. On the other hand due to the relatively small number of observations the results would be on shaky grounds especially for some combinations of factors, e.g. regions * age groups.

From the economic perspective there are two surprising findings - the unemployment duration does not differ statistically significantly between various regions and between various age groups. In the previous analysis of the periods before and during the crisis with similar numbers of observations (Čabla, 2014) the differences were statistically significant. It is possible that the differences are not too great or variance is too large to manifest in the sample so small but either way there seem to be a shift that needs deeper examination.

Statistically significant differences are in line with the generally accepted view of the situation on the Czech labor market - women and less educated have worse position on it, which is demonstrated e.g. by the higher unemployment rates and lower earnings. How the unemployment duration of these groups changed from the crisis years is out of the scope of the article and needs further investigation.

It would be also possible to enlarge the sample with the right censored observations, i.e. with those who were looking for a job but not found it during the follow-up period. The sample would be heavily right censored and it is not predictable, if it would lead to better informational value.

6. Conclusion

The main conclusions are simple - mean unemployment duration among those, who found a job in at the end of the year 2013 or in the year 2014 is around 13 months and differs among different sexes and educational levels. Women looks for a job about 15 - 20 % longer than men and more educated shorter than less educated for all educational levels. On the other hand it was not proven, that it differs among different age groups or regions. If later research confirms the findings, it would mean the shift from the crisis and pre-crisis state of affairs.

The paper deepened current understanding of the situation on the labor market after economic crisis by quantification of the duration of unemployment for the whole population of economically active and for different subgroups of this population.

It also shows that, two best fitting distributions of the unemployment duration among usually used are log-logistic and log-normal, former being slightly better.

Acknowledgements

The support of the grant scheme IG410045 by the University of Economics IGA is gladly acknowledged.

References

1. ČABLA, A. 2012. Unemployment duration in the Czech Republic. Prague 13.09.2012 – 15.09.2012. In The 6th International Days of Statistics and Economics, Conference Proceedings. Prague, 2012, pp. 257 – 267. ISBN 978-80-86175-86-7.
2. ČABLA, A. 2014. Unemployment Duration before and during The Economic Crisis in The Czech Republic. Acta Aerarii Publici, 2014, vol. 11, iss.. special, pp. 19–26. ISSN 1336-8818.
3. EUROSTAT. 2013. "EU labour force survey" - Statistics Explained (2013/9/0), http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/EU_labour_force_survey# [cit. 14-09-2013].
4. FAY, M. P., SHAW P. A., 2010. "Exact and Asymptotic Weighted Logrank Tests for Interval Censored Data: The interval R Package." Journal of Statistical Software, vol. 36, iss. 2, Aug 2010.

5. FAY, M. P. 2013. Package "interval". R Project. [Online] 06. 05 2013. [Citation: 17. 11 2013.] <http://cran.r-project.org/web/packages/interval/interval.pdf>.
6. HENSCHL et al. 2013. Package "intcox". R Project. [Online] 18. 02 2013. [Citation: 17. 11 2013.] <http://cran.r-project.org/web/packages/intcox/intcox.pdf>.
7. JAROŠOVÁ E., MALÁ I., ESSER M., POPELKA J. 2004. Modelling time of Unemployment via Loglocation-scale Model. COMPSTAT 2004 Symposium: pp. 1-8.
8. JAROŠOVÁ, E., MALÁ I. 2005. Modelling time of unemployment in the Czech Republic. In: KOVÁČOVÁ, M. Proceedings of the 4th International Conference APLIMAT 2005. Bratislava: Slovak University of Technology, 2005. ISBN 978-80-969264-2-8.
9. KIM, J., KANG, D. R., NAM C. M. 2006. Logrank-type tests for comparing survival curves with interval-censored data. In Computational Statistics & Data Analysis [online]. 2006, vol. 50, iss. 11, 3165-3178 [cit. 2015-06-08]. DOI: 10.1016/j.csda.2005.06.014.
10. KLEIN, J. P., MOESCHBERGER, MELVIN L. 1997. Survival Analysis: Techniques for Censored and Truncated Data. New York: Springer-Verlag New Your, Inc., 1997.
11. MALÁ, I. 2014. Modelling of the distribution of the unemployment duration in the Czech Republic. In: 29th International Workshop on Statistical Modelling. [online] Göttingen, 14.07.2014 – 18.07.2014. Bremen : Georg-August-University, 2014, pp. 97–100. <http://www.uni-goettingen.de/de/programme/443670.html>.
12. OECD. 2012. "Average duration of unemployment." OECD. StatExtracts. 31 July 2012. <<http://stats.oecd.org/Index.aspx?QueryId=25006>>.[cit. 15-09-2013].
13. SHEN, P. Nonparametric analysis of interval censored and doubly truncated data. Journal of Statistical Computation and Simulation [online]. 2012, vol. 82, iss. 12, 1845-1854 [cit. 2015-06-08]. DOI: 10.1080/00949655.2011.599326.
14. THERNEAU, T. 2013. Package "survival". R Project. [Online] 26. 02 2013. [Citation: 17. 11 2013.] <http://cran.r-project.org/web/packages/survival/survival.pdf>.
15. TURNBULL, B. W. 1976. The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. In Journal of the Royal Statistical Society B38, 1976, 290 – 295002E