MODELLING OF UNEMPLOYMENT DURATION IN THE CZECH REPUBLIC

Adam Čabla, Ivana Malá*

Abstract
The paper examines the duration of unemployment in the Czech Republic in the three selected years (2008, 2010 and 2014). It is based on the Czech Statistical Office data collected from the Labour Force Sample Survey. Lognormal probability distribution (unimodal positively skewed heavy-tailed distribution) is used in the Accelerated Failure Time regression model including the following factor explanatory variables: years (three levels, 2008, 2010, 2014), gender (two levels), education (four levels), five-year age groups (nine levels) and municipality size (five levels). Apart from this parametric model, Turnbull’s nonparametric estimator of the survival function is evaluated for subsamples defined by the year, gender and education. The effects of education and gender – a strong positive one of the former and less significant one of the latter, respectively – are quantified and assessed.

Keywords: unemployment duration, AFT model, Turnbull’s estimator, interval censored data
JEL Classification: E24, J64, C41

1. Introduction
Mass unemployment is a pressing issue in almost all developed countries. Unemployment occurs when a person who is actively searching for employment is unable to find a job. This phenomenon is sometimes used as an indicator of the health of the economy. At present, not only natural unemployment, but also weak demand for workers and structural or regional unemployment pose serious problems. Extended unemployment involves the risk that many of those out of work lose relevant skills and become discouraged from looking for work, thus increasing permanent joblessness. Unemployment affects the whole population, not only the jobless people and their households (Krueger et al., 2011). The global economic crisis strongly influenced the labour market worsening conditions for all participants. The unemployment rate is considered one of the most important indicators of the fitness of the economy, unemployment duration being equally important; the effect of the latter on wages is extensively analysed worldwide (see, e.g. Schmieder et al., 2013). The aim of the paper is to broaden the understanding of the unemployment issue (its duration in particular) in the Czech Republic in the period preceding, during and after the recession. The data from 2008, 2010 and 2014 having been selected, the changes in the unemployment duration (caused by the economic crisis and the following recovery accompanied by job- and labour-market growth) were quantified.

* Adam Čabla, Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics in Prague, Prague (xcaba02@vse.cz);
Ivana Malá, Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics in Prague, Prague (malai@vse.cz).
Unemployment is not spread evenly, groups of people disadvantaged in the labour market are being scattered across the society. Their position is often of interest to government authorities that design ways to create employment opportunities for various population segments divided according to particular economic activities (structural unemployment), place of residence (regional unemployment), education or work experience (e.g. youth unemployment), etc. A problem requiring serious approach is that of single parent households (especially single or divorced mothers) with small children. Among the groups considered at risk of unemployment due to their disadvantage in the labour market are young people without work experience, people over 50 years of age, especially those with low educational attainment and women. Further research in this field can provide deeper understanding of the issue and help develop measures to effectively respond to the situation of the unemployed, increasing their chances to find a new job thus reducing the country’s rate of unemployment. This study represents an attempt to quantify the differences between groups of the unemployed broken down by demographic and socio-economic factors such as gender, age, size of municipality and educational attainment.

Special attention is usually paid to the long-term unemployment of the duration of over one year (recently analysed by Löst and Langhamrová (2011) in terms of the Czech Republic’s development). In the presented paper, unemployment duration for those who are without a job for less than two years is modelled, this limited period covering both short-term and the first year of long-term unemployment. A statistical model is proposed allowing to describe the distribution of unemployment duration and involve subpopulation covariates. Statistical modelling of the probability distribution of unemployment duration data (as an alternative to unemployment rate analysis) provides useful feedback on the issue, identifying its characteristics and distribution heterogeneity across the different subsets of unemployed people.

For the analysis of unemployment duration, there are two major sources of data in the Czech Republic. Labour offices’ databases maintained by the Ministry of Labour and Social Affairs comprise solely the registered unemployed persons, indicating the termination of their previous employment, i.e. the length of unemployment spells. Unlike the above data application in the duration spell description, their usefulness for the probabilistic modelling of unemployment duration remains problematic. The present study obtains the data from the Labour Force Sample Survey conducted quarterly by the Czech Statistical Office (LFSS, CZSO, 2014). According to the data, only 74, 80 and 81 per cent of the unemployed recorded by the survey in 2008, 2010 and 2014, respectively, are registered at labour offices, only about 21 per cent of them receiving unemployment benefits.

The LFS survey provides a thorough examination of employment and unemployment developments in the Czech Republic. All household members involved in the survey were carefully interviewed about their past, present and in some cases also future (e.g. due to the delayed entry into employment. The analysis is based on three datasets containing information about all the unemployed covered by the survey between the fourth quarters of 2007 and 2008 (i.e. five consecutive quarters, designated as 2008), between the first quarters of 2010 and 2011 (referred to as 2010, see Čabla, 2014) and between the fourth quarters of 2013 and 2014.

In Malá (2013), a model of a finite mixture of lognormal distributions is constructed allowing to incorporate the covariates. This study adopts a regression model developed in survival analysis (Lawless, 2002) in order to model unemployment duration distributions.
if the analysed datasets contain censored observation data. A frequently used term “survival analysis” (derived from applications in medical science) is used in the text instead of “duration analysis”, “transition analysis” or “event history analysis” more common in economics and social science, respectively. Nonparametric and semiparametric statistical methods of survival analysis (Turnbull’s estimate and an AFT regression model, respectively) are employed here, Bayesian models (Ibrahim, 2005) being, however, also well applicable in this case.

2. Model Specification and Data

The households in the LFSS survey form a rotating panel data set – a fifth of the sample changing quarterly, monitoring households not exceeding one year. During interviews, no accurate data on the length of unemployment spells were recorded, the unemployed having reported monthly periods of their joblessness (i.e. 0–1, 1–3, 3–6, 6–12, 12–18, 18–24, 24–48 and over 48 months, these interval ranges corresponding to a periodic tabulation of the numbers of unemployed persons published by the Czech Statistical Office, see Figure 3). It follows that all data are censored (i.e. incomplete). For the unemployed who found a job during their studies, the duration analysed is interval-censored in \((l, u]\). For those who did not find a new job, the period of unemployment is obviously longer than the stated interval and the date is treated as right-censored in the left limit of the recorded interval \((l, u]\) (or interval-censored in \((l, \infty)\)). The intervals mentioned above were transformed in order to express the duration of the spell of unemployment as accurately as possible using some additional information from the survey data. Only those unemployed for less than two years were included in the research data set.

The data description indicates that seasonal effects cannot be taken into consideration, despite the fact that chances of finding employment are affected by the season of the year. Moreover, it is difficult to trace an unemployed individual in the database who found a job during the period when the survey was conducted, i.e. within one year. Hence, there are only a few documented cases of the end of the unemployment spell, the estimates of unemployment duration thus being probably exaggerated.

Let \(T\) denote the duration of unemployment in months. This time-to-event random variable with a continuous distribution allows the application of survival analysis tools, location or variability estimation and the quantification of differences between the examined subgroups of unemployed people. Since the distribution of unemployment spell durations is a positively skewed and heavy-tailed probability distribution, quantile characteristics, such as the median (for location) or the quartile deviation (for variability) seem to be more predictive of the location and variability than moment characteristics (the expected value and standard deviation). In the survival analysis, the probability \(S(t)\) of time-to-event longer than a given time \(t\) (survival function, \(S(t) = P(T > t) = 1−F(t), t \in \mathbb{R}\)) is often preferred to the probability of \(T\) which is less than \(t\) (cumulative distribution function \(F(t)\)). A distinctive characteristics of a survival time random variable is also the hazard rate (an instantaneous rate of an event’s occurrence) \(h(t), t > 0\), the function being defined in Equation 4, see the Appendix. For the study of unemployment duration, a decreasing hazard function (the job finding intensity is declining with the length of unemployment) or, alternatively, a function with one maximum (the intensity increasing to the maximum and then falling to zero) are expected to be applied; it is reasonable to assume that the chance of finding work decreases along with
the duration of unemployment, for those with an extremely long unemployment history being almost negligible. In this paper, the two-parameter lognormal distribution \( LN(\mu; \sigma^2) \) is used (for details see Equation 3 in the Appendix), lognormal distribution representing a unimodal positively skewed distribution with the one-maximum hazard function.

The set of unemployed people seems to be rather heterogeneous in terms of the unemployment spell and its probability distribution. The dependence of unemployment duration on factor explanatory variables (gender, size of municipality, educational attainment, age) is revealed and quantified.

For complete (noncensored) data, the survival function can be estimated as a complement of the empirical distribution function (without any assumption about the probability distribution). Interval-censored data modification was introduced by Turnbull (1976); a non-closed-form estimator being based on a numerical iterative procedure. In addition to the nonparametric Turnbull’s estimator, a parametric AFT regression model (Accelerated Failure Time Model or Accelerated Life Model, see Lawless, 2002) is employed. The model allows for the use of explanatory variables, and thus the \( T \) distribution dependence on the covariances (nominal or numerical) can be described. The regression model for the logarithm of \( T \) is expressed as

\[
\ln(T) = \mu + \beta'x + \sigma\varepsilon,
\]

where \( x' = (x_1, ..., x_m) \) is a vector of \( m \) explanatory variables (in the case of factor variables, \( m \) equals the sum of factor levels minus the number of factor variables), \( \mu, \sigma, \beta \) are unknown parameters (\( \beta \) is the vector of \( m \) regression parameters) and \( \varepsilon \) is a random variable with the standard normal distribution.

In the accelerated failure time model (1), the survival function \( S(t; x) \) of an individual with the vector of explanatory variables \( x \) is expressed in the form

\[
S(t; x) = S_0(t \cdot \exp(-\beta'x)),
\]

where \( S_0 \) is the baseline survival function of the distribution \( LN(\mu; \sigma^2) \) for an individual with \( x = 0 \). In the present model, we consider only factor explanatory variables, all the vectors of predictors \( x \) consisting only of dummy variables with the values of 0 and 1. The positive regression coefficient \( \beta_j, j = 1, ..., m \) in the categorical variable means that the unemployment spell tends to be longer than the baseline period (time is decelerated), the negative coefficient implying that the spell tends to be shorter (time is accelerated). The coefficient \( \exp(-\beta) \) quantifies the change in time for two unemployed persons who differ only in the \( j \)-th explanatory variable.

Analysed data were obtained from the LFSS survey (2014) carried out quarterly by the Czech Statistical Office questioning from 50 to 60 thousand respondents in the selected households. Referring to all the unemployed survey respondents who found a job sooner than within 27 months as well as those who are still looking for a job, their period of unemployment not exceeding two years, the present model includes 2,893 (2008), 4,753 (2010) and 2,844 (2014) selected eligible unemployed people, respectively. Instead of the LFSS values of interval limits (0, 1, 3, 6, 12, 18, 24), lower-limit \( l_i \) values of 0, 1, 2, 3, 4, 6, 9, 12, 15 and 18 months and upper-limit values of 1, 3, 4, 6, 7, 9, 12, 15, 18, 21, 24 and 27 months, respectively, were used in the interval \( (l_i, u_i) \) sample (taking into account more details on individual surveyed unemployment spells).
All the explanatory variables (covariates) employed in this analysis are treated as factors – three-, two- and four level ones for years (2008, 2010, 2014), gender (male, female) and educational attainment [basic (ISCED-97 0, 1), secondary without GCSE [ISCED-97 2], secondary with GCSE [ISCED-97 3, 4], tertiary [ISCED-97 5, 6]), respectively. The municipalities were divided into five groups by the population size (under 1,000, 1,000–10,000, 10,000–50,000, 50,000–100,000 and above 100,000 inhabitants) in accordance with the Czech Statistical Office database. The age factor was incorporated as seven five-year period groups from (16,20) to (51,55), the last one being identified as 55-plus. The baseline distribution was specified as a distribution of the unemployment spell for a man aged 30–35 with secondary education living in a municipality with the population between 10,000 and 50,000 inhabitants in the year 2008. In the present model, as many as twenty parameters were to be estimated applying the maximum likelihood method.

Having used the R v. 3.0.3 programme (R core team, 2014) for all the computations performed, the estimates were made employing Interval (Fay, Shaw, 2010) and Survival (Therneau, 2012) packages.

3. **Unemployment in the Czech Republic**

The Czech Republic’s unemployment rate figures are published regularly by the Czech Statistical Office – quarterly for the whole population and separately for men and women as well as the subgroups divided by education, municipality or age. In Figure 1, the development of the unemployment rate is shown for both the entire population and men and women in the Czech Republic (solid lines) and Slovakia (dotted lines). The situation in both countries (originally parts of Czechoslovakia) is not without comparative interest – markedly higher unemployment in Slovakia, similar male and female rates since 2010 and a greater gender gap in the Czech Republic are obvious. In Figure 2, the unemployment rates of the Czech Republic (solid lines) and Germany (dotted lines) are presented suggesting a strong link of the former country’s economy to the latter one. A smaller inverse gender gap in Germany compared with that in the Czech Republic can be seen. For the Czech Republic and Slovakia, the figures show the middle curves for the whole society, the lower and upper ones for men and women, respectively, the unemployment rate being higher among the latter. Germany, on the other hand, reports inverted female-male data, the lower curve showing the female unemployment rate. According to Eurostat data (2014), this is also the case of Belgium, Bulgaria, Estonia or Ireland, while in France, Denmark and Spain the rates are similar for both sexes. Other countries report findings comparable to those from the Czech Republic and Slovakia.

The paper, however, focuses primarily on the duration of unemployment. Figure 3 demonstrates the distribution of the unemployed by the length of unemployment spells in groups of those under three months, between three and six, and six and twelve months, between one and two years, and over two years throughout the period from 1993 to 2013. The long-term unemployment ratio in the Czech Republic is shown in the two upper parts of the diagram. It varies strongly over time, the highest proportion of over 56 per cent being recorded in 2006 (with similar values in 2000–2008); in 2013 and 2014, the long-term unemployment ratio reached 40 and 43 per cent, respectively.
Figure 1 | Unemployment Rate (%) in the Czech Republic and Slovakia (women, whole population, men)

![Figure 1](image1.png)


Figure 2 | Unemployment Rate (%) in the Czech Republic and Germany (women, whole population, men)

![Figure 2](image2.png)


Figure 4 presents the long-term unemployed as a percentage of all unemployed people over the period from 2007. The figure includes data for the Czech Republic, the EU-27 countries, the Eurozone (considered separately) and two selected countries – Germany and the United Kingdom (the former being comparable to the Czech Republic with its steadily decreasing proportion, the latter significantly different due to its low but increasing percentage of the long-term unemployed).
Turnbull’s nonparametric estimates of survival functions were found (without any distribution assumption) and lower quartiles and medians based on piecewise constant functions evaluated (Table 1, left-hand figures). These quantities indicate time periods within which 25 and 50 per cent of the unemployed, respectively, find work. These times are compared to the estimated quantiles from the AFT model (Table 1, right-hand figures). The estimated survival functions for the above subgroups are illustrated in Figures 5 (education) and 6 (gender). The estimated values of quantiles are slightly higher for women than men, the gender gap, however, is not as wide as expected. The impact of educational attainment, on the other hand, is considerable, the unemployed who have received primary or no education being in the most difficult situation, the tertiary educated having the best starting position. The values are comparable for those with secondary education both with
and without a graduation certificate. All these findings remain valid in a more comprehensive and sophisticated AFT model (see Table 2 in the Appendix).

Table 1 | Lower Quartiles and Medians of the Unemployment Spell in Months (Turnbull’s estimate – left, AFT model – right)

<table>
<thead>
<tr>
<th>subgroup</th>
<th>2008 lower quartile</th>
<th>2008 median</th>
<th>2010 lower quartile</th>
<th>2010 median</th>
<th>2014 lower quartile</th>
<th>2014 upper quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>men</td>
<td>7.5</td>
<td>7.0</td>
<td>13.5</td>
<td>12.6</td>
<td>7.5</td>
<td>6.2</td>
</tr>
<tr>
<td>women</td>
<td>7.5</td>
<td>7.6</td>
<td>19.5</td>
<td>13.7</td>
<td>7.5</td>
<td>8.4</td>
</tr>
<tr>
<td>basic</td>
<td>13.5</td>
<td>10.6</td>
<td>19.5</td>
<td>18.9</td>
<td>13.5</td>
<td>12.0</td>
</tr>
<tr>
<td>secondary</td>
<td>7.5</td>
<td>7.1</td>
<td>13.5</td>
<td>12.7</td>
<td>6.5</td>
<td>8.0</td>
</tr>
<tr>
<td>complete</td>
<td>3.5</td>
<td>6.7</td>
<td>13.5</td>
<td>11.8</td>
<td>6.5</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Source: own computations

Figure 5 | Turnbull’s Estimates of Survival Functions of Subsamples Broken Down by Education (first line 2008 (left), 2010 (right), second line 2014)

Source: own computations
In Table 2 (Appendix), the results of the above full AFT model with predictors (year, gender, education, age and the size of municipality) are displayed. The maximum likelihood parameter estimates with the estimated standard deviations and $P$-values for the statistical significance of the parameters are given together with the time change coefficients $e^{-\beta}$.

The baseline duration of unemployment is estimated as $LN (2.61; 0.852)$ with an expected value of 19.4 months, the median 13.6 months, the mode 6.6 months and the quartile deviation of 8.2 months. The hazard rate (job search intensity) reaches a peak within 11 months and then declines to zero with increasing unemployment duration.

The model found a statistically significant difference ($P < 0.05$) between the duration of female and male unemployment. The difference is quantified as the time passing faster for men than women, the coefficient being 0.92 (i.e. time $t$ for a man is 0.92$t$ for a woman). The years 2010 and 2014 are significantly different from 2008. There is a positive parameter implying a longer unemployment spell in 2010 (the time coefficient 0.9 indicates time deceleration of 10 per cent), the negative value for 2014 suggesting time acceleration of 10.6 per cent. The positive effect of education is obvious from Table 2. For an unemployed tertiary-educated person, time passes 1.38 times faster than for someone with lower secondary education, decelerating further by 0.678 for those with basic education. When comparing secondary education with and without GCSE, the graduates of the former show time acceleration of 6.7 percentage points.

The age dependency is rather complex, the estimated coefficients being negative for age groups under 45 and positive for 46-plus. The unemployed people in the 25–55 age interval seemingly experience comparable unemployment duration (non-significant time difference index 0.91–1.16), time passing more slowly for those over 55 with the deceleration of 21 per cent. For the young unemployed (16–25), the results are skewed because of the limited incidence of long-term unemployment due to compulsory schooling and follow-up studies. For these reasons, the current youth unemployment rate is not included in the model.
4. Conclusion

The present article outlines an alternative approach to the modelling of the distribution of unemployment duration in the Czech Republic. The Labour Force Sample Survey not revealing precise, relevant data for the purpose of this paper, the time interval information is available only about those who found a job during the survey, and the lower limit of the unemployment spell is known for those who did not get a new job (along with information on how long the period of unemployment lasts). The AFT regression model allows for the incorporation of explanatory variables (both quantitative and qualitative) into the regression model describing the dependence of a given spell duration on the above variables. Statistical models allow the testing of dependencies and their strength quantification. The interpretation of the AFT model using the concept of the acceleration (or deceleration) of the passage of unemployment time is important for understanding the analysed process. It provides an alternative explanation and quantification of differences between population subgroups defined by combinations of explanatory variables. When applying the AFT model, the basic variable distributions to be made, this decision being crucial for the results, although analogous distributions produce comparable outcomes. As the duration of unemployment is a positively skewed variable (with a really heavy tail even in the case that only the short-term unemployed within the next two years are taken into account), the lognormal distribution was used. This probability distribution proves to be superior to other distributions in the model (according to the Akaike criterion, not referred to in the article) acquiring the properties of its hazard function with a maximum and decreasing with time to zero. This distribution was also successfully used in Malá (2013) as a component distribution in a mixture model. To overcome the problems with the selection of appropriate probability distributions, nonparametric estimates of survival functions were made and compared to the parametric AFT model.

Both approaches enable the evaluation of characteristics of the location and variability as well as testing for differences between groups of the unemployed, statistical analysis allowing a quantification of these differences. Having quantified commonly known facts (i.e. the inferior position of women and unemployed individuals with basic education in the labour market), this paper presents medians and lower quantiles of the unemployment spell, outlining a time acceleration/deceleration interpretation of the AFT model.

The present research results are consistent with the Eurostat analysis (2015) based on the statistics of the long-term unemployment rate in particular, recent labour market trends being well-illustrated by the developments in the selected years. Despite the data and modelling constraints, the key findings of the study provide an alternative view on a widely discussed issue of unemployment.

Appendix

In the paper, the lognormal distribution for \( T \) is used with parameters \( \mu \) and \( \sigma^2 \) if \( \ln T \) is normally distributed as \( (N(\mu; \sigma^2)) \). The distribution is described by the density function

\[
f(t; \mu, \sigma^2) = \begin{cases} 0, & t \leq 0, \\ \frac{1}{t\sigma\sqrt{2\pi}} \exp \left( -\frac{(\ln t - \mu)^2}{2\sigma^2} \right), & t > 0. \end{cases}
\]

(3)
Common formulas for the cumulative distribution function, hazard function and statistical characteristics such as the expected value, variance and quantiles are applied (Malá, 2013). Hazard function is defined as

$$h(t) = \lim_{h \to \infty} \frac{P(t \leq T < t + h)}{h} = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)},$$

where $S'$ is the derivative of the survival function.

### Table 2 | Estimates of Parameters in AFT Model

<table>
<thead>
<tr>
<th>Variable level</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\sigma}$</th>
<th>$P$</th>
<th>$e^{\hat{\beta}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.6065</td>
<td>0.0558</td>
<td>0.0000</td>
<td>0.074</td>
</tr>
<tr>
<td>Year 2010</td>
<td>0.1056</td>
<td>0.0356</td>
<td>0.0030</td>
<td>0.900</td>
</tr>
<tr>
<td>2014</td>
<td>-0.1004</td>
<td>0.0366</td>
<td>0.0060</td>
<td>1.106</td>
</tr>
<tr>
<td>Gender female</td>
<td>0.0837</td>
<td>0.0261</td>
<td>0.0014</td>
<td>0.920</td>
</tr>
<tr>
<td>Isced isced1, 2</td>
<td>0.3893</td>
<td>0.0398</td>
<td>0.0000</td>
<td>0.678</td>
</tr>
<tr>
<td>isced4</td>
<td>-0.0647</td>
<td>0.0302</td>
<td>0.0320</td>
<td>1.067</td>
</tr>
<tr>
<td>isced5</td>
<td>-0.3224</td>
<td>0.0471</td>
<td>0.0000</td>
<td>1.380</td>
</tr>
<tr>
<td>Age 16–20</td>
<td>-0.2277</td>
<td>0.0583</td>
<td>0.0000</td>
<td>1.256</td>
</tr>
<tr>
<td>21–25</td>
<td>-0.1485</td>
<td>0.0493</td>
<td>0.0026</td>
<td>1.160</td>
</tr>
<tr>
<td>25–30</td>
<td>-0.0690</td>
<td>0.0523</td>
<td>0.1870</td>
<td>1.071</td>
</tr>
<tr>
<td>35–40</td>
<td>-0.0596</td>
<td>0.0521</td>
<td>0.2530</td>
<td>1.061</td>
</tr>
<tr>
<td>41–45</td>
<td>-0.0321</td>
<td>0.0551</td>
<td>0.5610</td>
<td>1.033</td>
</tr>
<tr>
<td>46–50</td>
<td>0.0944</td>
<td>0.0567</td>
<td>0.0658</td>
<td>0.910</td>
</tr>
<tr>
<td>51–55</td>
<td>0.0902</td>
<td>0.0549</td>
<td>0.1644</td>
<td>0.914</td>
</tr>
<tr>
<td>55+</td>
<td>0.2248</td>
<td>0.0560</td>
<td>0.0000</td>
<td>0.799</td>
</tr>
<tr>
<td>Municipality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–1,000</td>
<td>-0.1900</td>
<td>0.0380</td>
<td>0.0000</td>
<td>1.209</td>
</tr>
<tr>
<td>1,000–10,000</td>
<td>-0.1008</td>
<td>0.0336</td>
<td>0.0027</td>
<td>1.106</td>
</tr>
<tr>
<td>50,000–100,000</td>
<td>0.0172</td>
<td>0.0470</td>
<td>0.7140</td>
<td>0.983</td>
</tr>
<tr>
<td>100,000+</td>
<td>0.0625</td>
<td>0.0463</td>
<td>0.1771</td>
<td>0.939</td>
</tr>
</tbody>
</table>

Source: own computations

For an unemployed individual with a vector of explanatory variables $\mathbf{x}$, unemployment duration has the (estimated) lognormal distribution with parameters $\hat{\mu} + \hat{\beta}'\mathbf{x}$ and $\hat{\sigma}^2$. The expected value $E(T|\mathbf{x})$ can be expressed as, for example,

$$E(T|\mathbf{x}) = \exp \left( \hat{\mu} + \hat{\beta}'\mathbf{x} + \frac{\hat{\sigma}^2}{2} \right) = \exp \left( \hat{\mu} + \frac{\hat{\sigma}^2}{2} \right) \exp(\hat{\beta}'\mathbf{x}),$$
where \( \exp \left( \mu + \frac{\sigma^2}{2} \right) \) is the baseline expected value (i.e. that of a baseline distribution) and for quantiles it applies

\[
t_{p|x} = \exp \left( \mu + \beta' x + \sigma u_p \right) = t_p \exp \left( \beta' x \right),
\]

where \( u_p \) is a 100\( P\)% quantile of the standard normal distribution and \( t_p \) is a baseline 100\( P\)% quantile of the baseline distribution.

References