Modelling of the Unemployment Duration in the Czech Republic Based on Aggregated Complete and Individual Censored Data[1]

*Ivana MALÁ – Adam ČABLA**

Abstract

The unemployment rate is considered to be one of the essential characteristics of the state of the economy. Unemployment duration can also describe the situation in the labour market. There are two sources of data on the duration of unemployment in the Czech Republic – data from the Labour Force Sample Survey provided by the Czech Statistical Office (aggregated or individual data) and aggregated data from the database of registered unemployed people held by labour offices under the Ministry of Labour and Social Affairs. Two parametric lognormal distribution is used to model the distribution of durations quarterly from 1Q 2000 to 2Q 2019. The maximum likelihood estimates of parameters are found from individual data taking into account censored (incomplete observations) when observing unemployment duration; the minimum chi-squared method is used to estimate parameters from aggregated data. Time series of estimated parameters from different data sources, estimation procedures and data types are presented and compared. The relationship between the rate of unemployment and the duration of unemployment is shown.

Keywords: *unemployment duration, Labour Force Sample Survey, survival analysis, unemployment rate*

JEL Classification: J64, C34, C41, E24

DOI: https://doi.org/10.31577/ekoncas.2022.02.04

* Ivana MALÁ – Adam ČABLA, Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Statistics and Probability, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czech Republic; e-mail: malai@vse.cz; cabla@vse.cz

[1] This paper is supported by the grant F4/80/2018 which has been provided by the Internal Grant Agency of the Prague University of Economics and Business and a long term institutional support of research activities by the Faculty of Informatics and Statistics, Prague University of Economics and Business. LFSS data access in the SafeCentre of the Czech Statistical Office is gratefully acknowledged.

Introduction

Unemployment is a severe problem in almost all countries; this phenomenon is sometimes used to measure the health of the economy. The unemployment rate is usually considered but equally important is also unemployment duration. Since 2000 the Czech Republic had experienced a period of growth, economic crisis, recovery from the crisis to the pre-COVID state when unemployment was very low, and the number of vacancies higher than the number of applicants. Before the coronavirus outbreak, the labour market suffered only from natural unemployment as demand for employees exceeded the supply strongly. The impact of this outbreak and its reaction to it is yet to be seen in the data, but based on the presented paper with historical data; we can imagine possible future development in reaction to an incoming economic crisis.

The indicators such as unemployment rates and the length of unemployment are quite different in nature. The unemployment rate is generally rising in the case of economic problems when there is a shortage of labour supply on the labour market and thus an excess of demand oversupply. Unemployment time decreases when many employees lose their jobs and, although people are not finding work, the unemployed with short periods of unemployment prevails. For this reason, the distribution of unemployment spells shifts to a lower level with the increasing percentage of people with a short period of unemployment. Then, due to the recession or the crisis, people are unable to find a job, the duration of the unemployment spell will increase. On the other hand, time will shorten if more people find work with the prolonged recovery of the labour market and long-term unemployed can find a new job.

Unemployment has a severe impact on the economics of countries and all the unemployed and their families. It causes economic and social problems as a direct and immediate consequence of unemployment and implications for a future position in the labour market. With the increasing unemployment spell, the unemployed lose their work habituates. In Pollmann-Schult and Büchel (2005) the impact of unemployment duration on the quality of the future job is analysed for Germany. The text by Winkelmann and Winkelmann (1998) deals with the life satisfaction of the unemployed, and Hanappi and Lipps (2019) analysed the relationship between parenthood and family factors and job insecurity and unemployment. The impact of unemployment on households is a topic of Krueger et al. (2011). There is a considerable number of studies dealing with the relationship between unemployment benefits, unemployment duration and economic conditions (Bover et al., 2002).

Different data sets are applied in the literature to model unemployment duration and search for positive or negative determinants affecting it. If incomplete observations for those unemployed are included in the data, the survival analysis provides powerful tools for the analysis. Lim and Lee (2019) applied survival analysis (Cox regression model) to analyse unemployment of Korean graduates based on panel data from the Korean Education and Employment Panel and find out determinants to access youth unemployment. Terracol (2009) used data from the European Community Household Panel to evaluate the impact of a French guaranteed income program on the hazard out of unemployment (a multivariate causation duration model is used). Arslan and Senturk (2018) performed a questionnaire survey of the unemployed persons in 15 regions of Turkey and studied individual factors affecting the unemployment duration in Turkey, survival analysis using the econometric approach has been applied. Grogan and van den Berg (2001) use the Russian Longitudinal Monitoring Survey to construct Kaplan-Meier and Weibull parametric model to identify factors influencing the unemployment spell. The use of Weibull distribution is in line with the conclusions of (Čabla, 2017) for the Czech Republic. Uysal and Pohlmeier (2001) used individual unemployment data from the German Socio-Economic for the panel econometric duration analysis of the impact of personal traits determining individual unemployment duration. Rosholm (2001) studies individual unemployment durations using unemployment spells of a sample of Danish workers. McCall and Chi (2008) applied a discrete-time hazard model for unemployment durations based on the National Longitudinal Survey of Youth (1979) to analyse the relationship between unemployment support, unemployment duration spell and reemployment wages. The study of Setyadi et al. (2019) describes the characteristics of unemployed workers in Central Java Province and determines the multiple linear equations model of educated unemployment duration (without application of survival analysis). It uses data of individuals sampled from the National Labor Survey 2015. Kupets in (Kupets, 2006) uses individual data from the Ukrainian Longitudinal Monitoring Survey and estimates a discrete time-independent competing risks model with gamma distribution to describe random errors. In the survey SHARE, there is available a Job episodes panel dataset (Brugiavini et al., 2019) for the working history of respondents from the European Union including year time series of employee working status.

The paper aims to compare models of unemployment spell in the Czech Republic based on two different sources of data (Labour Force Sample Survey and the Ministry of Labour and Social Affairs). We show similar development in time for data defined by the different definitions and various types of data – aggregated or individual censored. The lognormal distribution was selected as a reasonable model for the analysed highly positively skewed variable (based on Čabla, 2017 or Čabla and Malá, 2017).

The paper is organised as follows: first, we present four analysed datasets (Section 1). The statistical models and their estimate that enables the description of the distribution of unemployment duration in the Czech Republic from 2000 to 2019 for aggregated and censored individual data are described (Sections 2.1 for aggregated data and 2.2 for censored data). In part 3 we present our results from all estimated models. We discuss various estimates of parameters, expected values, probability densities, hazard functions, and their development in time. We also show the impact of gender comparing in figures estimated parameters of lognormal distribution for the whole population and men and women.

1. Data

Duration of unemployment refers to the length of the period during which the person recorded as unemployed was seeking or available for work. For the analysis of unemployment duration in the Czech Republic, there are mainly two possible credible data sources. The source of data on the numbers of available job applicants out of work, who are registered by the labour offices, is the Ministry of Labour and Social Affairs (MoLSA). In these datasets, all registered unemployed people are included with the exact date of the end of previous work (exact unemployment spell), but the database is limited to registered ones. In the presented study, we use aggregated quarterly data on registered unemployed people (to March, 31st, June 30th, September, 30th, December 31st of each year), the whole database, and those eligible for the unemployment benefits. Absolute frequencies of intervals 0 – 3, 3 – 6, 6 – 9, 9 – 12, 12 – 24, 24+ months are presented. The unemployment benefits are paid to the unemployed (based on requirements on the working history) during 5 months for people under 50, 8 months for 50 – 55 and 11 months for those over 55. By extending the period of unemployment benefits, the authorities try to improve the position of older workers in the labour market and give them a longer time to find a new job. At the same time, people undergoing requalification are also eligible for unemployment benefits for their whole period.

The Labour Force Sample Survey (LFSS) is performed by the Czech Statistical Office (CZSO, 2014; LFSS, 2019) and provides thorough information about employment and unemployment in the Czech Republic. The survey is harmonised in all the European Union member countries, candidate countries and EFTA countries (Eurostat, 2022). For this reason, it can be used mainly for international comparisons, as the definition of unemployment differs from the definition used by MoLSA. The survey forms a rotating panel with sampled households included in the panel for five consecutive quarters. All members of households included in the survey are interviewed about their working history, present and in some cases also future (for example, delayed entry into employment). No exact values of the unemployment spell are recorded; the respondents state their unemployment only in intervals 0 – 1, 1 – 3, 3 – 6, 6 – 9, 9 – 12, 12 – 24, 24 – 18, 48+ months. Czech Statistical Office publishes the number of the unemployed in these intervals, but no information about people leaving unemployment for economic non-activity or employment is given. For this reason, we also used individual data provided by the SafeCentre of the CZSO. This approach is preferable from the point of view of more detailed information included in individual observations than in aggregated values resulting in a wider spectrum of possible methods of analysis as well to more precise estimates of unknown characteristics.

For our modelling, we have four datasets:

1. aggregated data from labour offices (MoLSA)
   a1. all registered applicants; referred to as MoLSA a1
   a2. registered applicants eligible for benefits; referred to as MoLSA a2
2. aggregated data from LFSS; referred to as LFSS b
3. individual data from LFSS; referred to as LFSS c.

The definition of unemployment differs according to methodology in MoLSA (MoLSA a1, MoLSA a2) in comparison to unemployment defined for LFSS datasets b and c. The target population of "unemployed" people is considered by the methodology of the International Labour Organization (ILO, 2019) as all persons above a specified age who during the reference period were without a job did not work an hour for pay, and was in an active manner seeking job they would be able to join within two weeks at the latest.

Our study uses quarterly data from Q1 2000 to Q2 2019. These time series include 78 observations, 19 years and two quarters from 2019.

2. Statistical Models

We denote T duration of unemployment in the Czech Republic (in months). This variable is a time-to-event random variable; we will suppose its distribution to be continuous and positively skewed with heavy tails. For such distributions, usually, the quantile characteristics as median (for location) or quartile deviation (for variability) seem to be more predictive characteristics of central tendency and variability than moment characteristics (expected value and standard deviation). The moment characteristics are too sensitive to outliers and rare, but too high values. In this paper, we use two-parametric lognormal distribution ($LN(\mu;\sigma^2)$), where $\mu = E(\ln T), \sigma^2 = Var(\ln T)$. This income distribution fulfils intuitive requirements, it is a long-tailed, unimodal distribution (at the time $\exp(\mu-\sigma^2)$).

For the unemployment duration in the Czech Republic, the distribution is applied in Čabla and Malá (2017). Survival distributions as gamma, loglogistic and Weibull distributions were also fitted, the lognormal distribution seems to be the most effective (based on AIC criterion and a chi-square goodness-of-fit statistics). Results from loglogistic distribution are comparable to the lognormal distribution.

In survival analysis, an important characteristic of the distribution of a time-to-event variable is a hazard function, it characterises the intensity occurrences of events in time. If f is a density function and F a cumulative distribution function of the distribution of random variable T, the hazard function h (Lawless, 2002) is defined as

$$h(t) = \lim_{\varepsilon \to 0} \frac{P(t \le T < t + \varepsilon)}{\varepsilon} = \frac{f(t)}{1 - F(t)}.$$

The function describes the intensity of finding a new job depending on the length of unemployment. The function quantifies the probability, that the unemployed at time t finds a job in a small time interval $(t, t + dt]$. For lognormal distributions with parameters corresponding to the analysed variable, it is increasing to a maximum, where the intensity of leaving unemployment is the highest (upside-down bathtub shape of intensity) and the hazard function is decreasing for longer unemployment spells. It decreases, but very slowly – the intensity is almost constant; it is in line with the conclusion in Čabla (2017). This property supports the meaningfulness of application of lognormal distribution; from Weibull distribution in some quarters, a function decreasing to a minimum and then increasing was obtained as the distribution is more complex and enables hazard functions with both maximum and minimum (this distribution is applied in Grogan and van den Berg (2001) for Russia).

We construct the quarterly time series of estimated medians, modes and expected values. For the lognormal distribution, we have characteristics depending only on location parameter μ (median) or on scale parameter σ (coefficients of variation, skewness, or kurtosis, different type ranges), or those that are functions of both parameters (expected value, mode, quantiles excluding the median, variance). To smooth the time series we apply the moving average filter (a window with four values to account for quarterly seasonality).

2.1. Estimation for Aggregated Complete Data

For aggregated data, the estimates of unknown parameters based on minimum chi-square (Harris and Kanji, 1983) were applied. The function is defined as

$$\chi^2(\mu,\sigma)=\sum_{j=1}^{k}\frac{\left(n_j-n\pi_j(\mu;\sigma)\right)^2}{n\pi_j(\mu;\sigma)} \tag{1}$$

where k is a number of time intervals, n is a sample size, n_j are frequencies of time intervals, $\pi_j(\mu;\sigma),\ j=1,2,...,k$ are probabilities of intervals based on the lognormal distribution. The χ^2 characteristics was numerically minimised with respect to parameters $\mu\in R$ a $\sigma>0$.

Data from LFSS (LFSS b) are given in intervals 0 – 1, 1 – 3, 3 – 6, 6 – 12, 12 – 18, 18 – 24, 24 – 48, and above 48, we obtain $k=8$. Numbers of registered applicants in MoLSA dataset are given in intervals 0 – 3, 3 – 6, 6 – 9, 9 – 12, 12 – 24, 24+, then $k=6$. The asymptotic distribution of $\chi^2(\hat{\mu},\hat{\sigma})$ is the chi-squared distribution with 5 (8 – 3) and 3 degrees of freedom (6 – 3), respectively. Because of large samples, the chi-square test is too strict and even small deviations from the lognormal distribution are identified as statistically significant. For this reason, we take this value only as a measure of the quality of fit.

2.2. Estimation for Censored Individual Data

The individual data from LFSS in SafeCentre of the CZSO (LFSS c) includes not only the unemployed but also all unemployed in the sample who found a job. For this reason, based on these data, we can model time to reemployment, not only an unemployment duration.

Unemployed respondents state their unemployment spell duration during each survey interview only in intervals (0 – 1, 1 – 3, 3 – 6, 6 – 12, 12 – 18, 18 – 24, 24 – 48 and above 48). Moreover, the rotating panel type of the survey was used. We take all respondents who found a job as interval censored and those who remain in unemployment as right censored. Therefore, all data are censored (incomplete). In the survey, the week of the interview is recorded. For an analysed quarter, we use all unemployed respondents with the unemployment spell in the semi-closed interval in months $\left(ll,ul\right\rangle$. Moreover, we try to utilise information from the previous and the consecutive quarters.

If a respondent was unemployed in the previous quarter in the interval $\left(ll,ul\right\rangle$, and employed in the present one, we obtain interval censored observation. We set a time interval of finding a job (in months) as

$$\left(ll,ul+\left(week\ reference-week\ previous\ \right)/\,4.33\right\rangle \tag{2}$$

where we use a mean month length of 52/12 = 4.33 weeks. This procedure enables us to utilise information on the weeks of the survey, and from the point of view of statistical quality of estimates, it improves the quality of fit, as we have more time intervals (not only 8). Then we check, whether a person unemployed in the reference quarter is unemployed in the next quarter (or observation is missing due to panel data). In this case, we set the unemployment duration to the lower limit of interval in the reference quarter. In the end, we look at those unemployed in the reference quarter to see if they are employed in the next one. In such a case, we will use the unemployment spell duration interval censored with value

$$(ll,\ ul + (week\ consecutive - week\ reference)\ /\ 4.33).$$

Estimation of models based on the data with a relatively high proportion of respondents without an event result in an overestimation of lognormal parameters. In our data, we obtain only 8 – 34% of successful applicants in particular quarters who find a new job.

For the estimation of unknown parameters, the maximum likelihood method reflecting interval censoring (for those who found a job) and right censoring (for those who stayed unemployed) was applied separately for each quarter (Lawless, 2002). To take into account the gender of the unemployed, the gender covariate was used in the model.

For all computations the program R version 3.6.0 (R Core Team, 2017) was used, the estimation of survival models was performed with package *Survival* (Therneau, 2015).

3. Results and Discussion

3.1. Comparison of Unemployment Rates and Median Times of Unemployment Duration

The choice of the definition of the unemployed and the population to which the number of unemployed refers enables the construction of different unemployment rates suitable for different analyses and decisions. The general unemployment rate in the Czech Republic is published regularly by the Czech Statistical Office (CZSO, 2014) quarterly for the whole population and separately for men, women and subgroups are given by education, municipality or age groups. The ratio of unemployed persons is (from 2013) defined as the ratio of registered available job seekers aged 15 to 64 years in the same age population. This ratio differs from the previously used measure of registered unemployment, defined as the ratio of registered available job seekers to economically active persons. The ratio of unemployed persons is calculated and provided back to the year 2005 by CZSO.

In Figures 1 and 2 no models are used and empirical aggregated frequencies of time intervals are shown. In Figure 1, both rates are shown (%, in black) for the general unemployment rate (2000 – 2019) and the ratio of unemployed persons (2005 – 2019). To describe the relationship between unemployment rates and the duration of the unemployment spell, the estimated medians from aggregated data are plotted (month, in grey). Estimated values for all registered unemployed (data MoLSA a1) and aggregated data from LFSS (data LFSS b) are given. A similar development of both medians and the same for both unemployment rates is observed. Differences in values of medians are due to different definitions of unemployment in the LFS survey and the Ministry of Labour and Social Affairs. Moreover, there are unregistered unemployed included in the LFS survey, especially those with a long unemployment spell not eligible for benefits.

The unemployment rate was quite stable from 2001 to the beginning of 2006 and was accompanied by a slowly increasing median unemployment duration. The situation possibly means that the persons with a longer unemployment duration started to accumulate compared to persons with a shorter duration of unemployment who were obtaining new jobs with a higher probability. From the beginning of 2006, a sharp decline of unemployment rate began with the then absolute low being hit in the second quarter of the year 2008 for both unemployment rate and the ratio of unemployed persons (4.2% and 3.8%, respectively). The median unemployment duration was stable or rising at the beginning of this period depending on the data source and then began declining around the beginning of the year 2007, with a lag of around one year. These numbers characterise a situation where there is a shortage of available workers with a short duration of unemployment, and companies start to hire more employees with a longer duration of unemployment due to the rising demand for the labour force. With the hit of the crisis in the middle of the year 2008, the unemployment ratio started to increase, and median unemployment duration accelerated downwards as the consequence of a wave of layoffs. The lowest values of median unemployment duration were achieved in the second and the third quarters of the year 2009 when unemployment ratios were near their peaks which occurred in the first quarter of 2010 with values similar to those before the crisis. Since then, the situation reminded the one observed at the beginning of the presented time-series with stable unemployment rates and slowly increasing the median unemployment duration. Interestingly from the fourth quarter of the year 2013, the ratio of unemployed persons and general unemployment rate started to diverge, with the first one increasing to the peak at the first quarter of the year 2014 and the second one slowly declining to achieve thus steadily lower values from then since. Median unemployment duration again reacted with the lag; the peak was achieved for the data from MoLSA in the first quarter of 2015 and for the data from LFSS half a year later. The observed lags of the peaks of median unemployment duration and ratio of unemployed persons were one year, and a similar lag for the data from LFSS was two and half years when taking the first quarter of the year 2013 as the peak of the unemployment rate. The situation in the middle of 2019 was unique from those previously observed. Both the unemployment ratio and unemployment of unemployed persons were at their historic lows, and the median duration of unemployment was almost minimal for data from MoLSA and minimal for the data from LFSS. It underlines how unprecedented the economic situation in the Czech Republic was and was possibly the consequence of the long-lasting and more importantly, stable economic growth.

F i g u r e 1

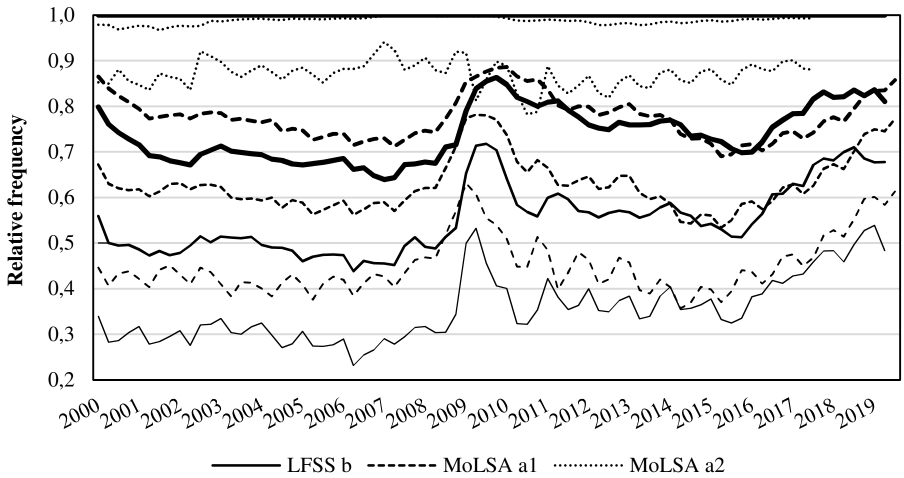
Comparison of Unemployment Measures

Source: Own computations.

Long-term unemployment, referring to people who have been unemployed for more than one year, is addressed in Figure 2. The relative frequencies of the unemployed with the unemployment spell under six months (thin line), one year (thicker line) and two years (thick line) are shown. Solid lines describe the situation based on the LFSS database (LFSS b), dashed lines for registered (MoLSA a1) and points for registered people eligible for unemployment benefits (MoLSA a2). The frequencies of the unemployed under one year are lower for data from LFSS than for MoLSA to 2014. Then both curves are comparable in the period of favourable conditions on the labour market. A similar situation is for all three curves. The empirical median is close to one year for data from LFSS to 2008; from this year, it is higher; the empirical median for registered applicants (MoLSA a1) is higher than one year for the whole analysed period. From the rules for obtaining unemployment benefits, the curve for the group MoLSA a2 is closed to one even for 6 months.

Figure 2

Relative Frequencies of the Unemployed with Unemployment Duration
under Six Months (thin line), One Year (thicker line) and Two Years (thick line)



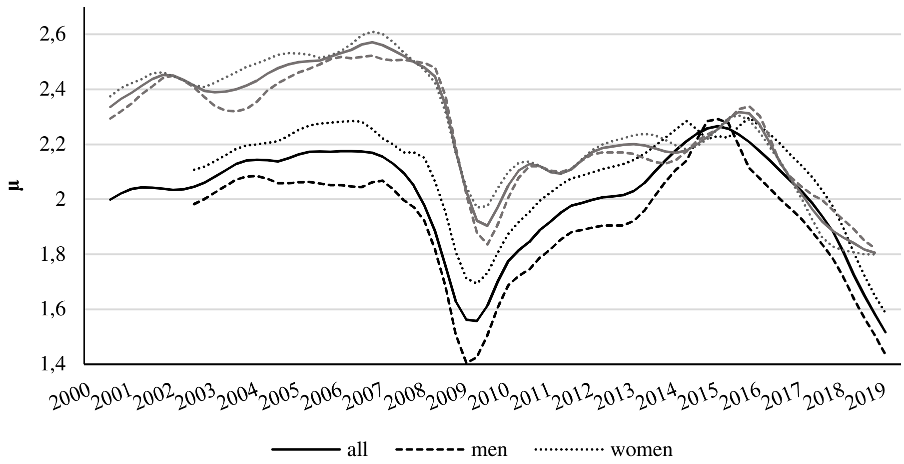
Source: Own computations.

3.2. Results from the Modelling

The quarterly time series of fitted parameters are given for the whole population of unemployed people and men and women separately. In Figures 3 and 4, the estimated values of parameters are presented for aggregated data from the database of MoLSA (data MoLSA a1) and LFS survey (LFSS b). We can observe a higher value of the parameter μ and lower for the parameter σ for LFSS and the opposite is true for data from MoLSA. It means that median times of unemployment spell from LFSS are higher and the estimated probability distribution is less skewed.

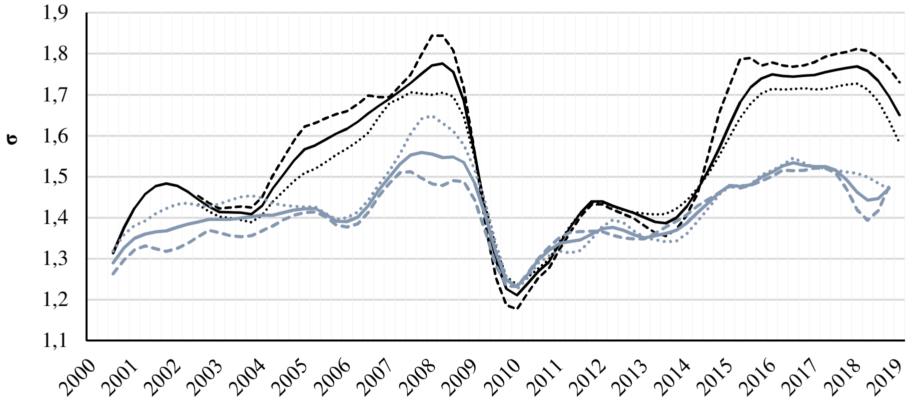
For women, values of the parameter μ are higher than for men, as the median unemployment spell is higher for them than for men, the difference being higher in MoLSA data. In Figure 4, smaller values of σ for MoLSA data and higher for LFSS data for women than for men are shown. It suggests that in MoLSA data, the men had a higher probability of longer unemployment spells and contrary to LFSS data, the men had a lower probability of higher unemployment spells up to 2009. From that year on, the estimates of σ have no clear relation to gender.

F i g u r e 3
Estimates of the Parameter μ for Data LFSS b (grey) **and MoLSA a1** (black)



Source: Own computations.

F i g u r e 4
Estimates of the Parameter σ for Data from LFSS b (grey) **and MoLSA a1** (black)

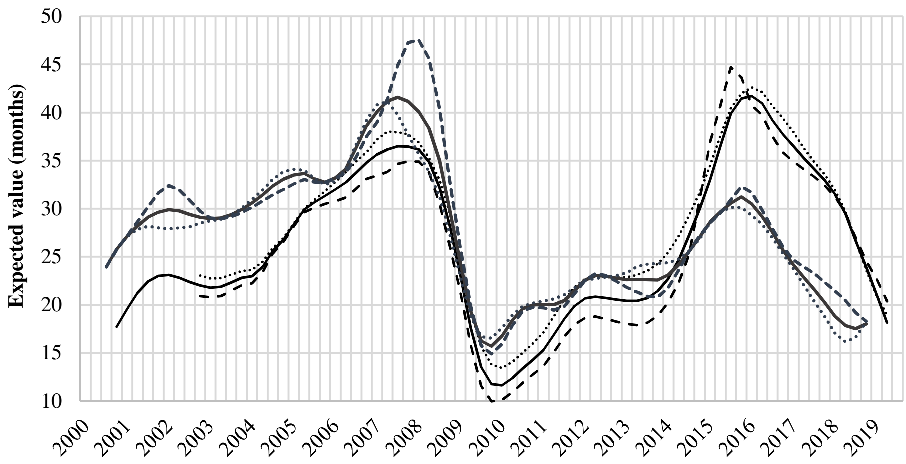


Source: Own computations.

In Figure 5, the estimated expected values are presented. The expected value depends on both parameters $(EX = \exp(\mu + \sigma^2/2))$, but it is strongly influenced by large values and usually is not applied as a representative of the distribution in the analysis of highly skewed data or survival analysis. The sample medians are included in Figure 1. We recommend interpreting only the development, not absolute values. Two peaks are included, one immediately before the crisis and one in 2016. During the crisis, many people were losing a job; for this reason, the expected duration decreased quickly because of many people with a short unemployment spell. We see almost the same decline for all curves for two years. From 2010 the duration was slowly getting longer to 2015 when the excellent opportunities to find a new job have occurred. Economic problems caused the decline in 2008; in 2015, on the contrary by favourable labour market conditions. The gender gap is not visible in the expected values for longer times. It is the result of the different values of the parameters described previously.

F i g u r e 5

Estimated Expected Values for Data from LFSS b (grey) **and MoLSA a1** (black)



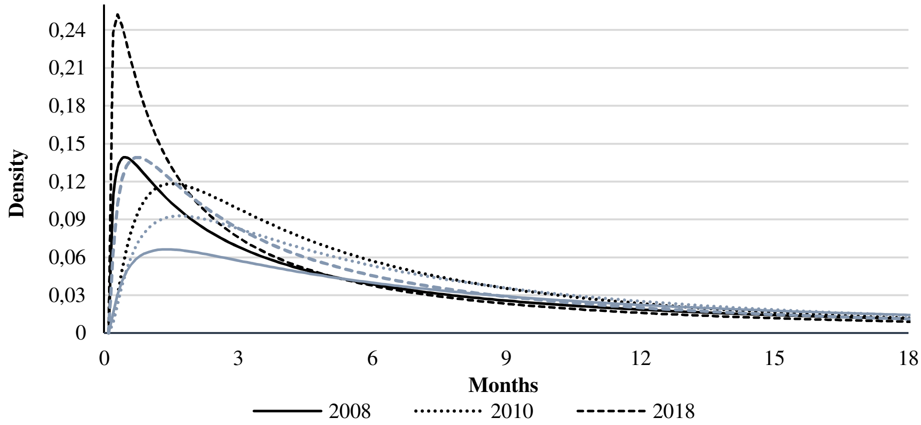
Source: Own computations.

In Figure 6, we choose three distributions from pre-crisis Q2 2008, peak crisis Q1 2010 and economic boom Q4 2018. The estimated probability densities are shown. For MoLSA a1 data there is a clear shift rightwards during the crisis and then back leftwards with a much more pronounced mode for the last economic boom. This again suggests that much more unemployment spell is in the lower part of the distribution than in pre-crisis levels. For LFSS b data we can observe somewhat similar development, alas with increasing value of mode.

In Figure 7, the estimated parameters from all four analysed datasets are given. The values from censored data are the highest, estimated parameters from LFSS aggregated data and data from MoLSA are comparable and values for registered unemployed (MoLSA a2) are the lowest. Higher values from the survival model reflect the definition of a variable as they refer to the time to reemployment, not to the distribution of the unemployment spells. The unemployment spells discussed so far can be viewed as a minimum reemployment duration – the time to reemployment if all the unemployed obtained job at the time of data collection.

F i g u r e 6

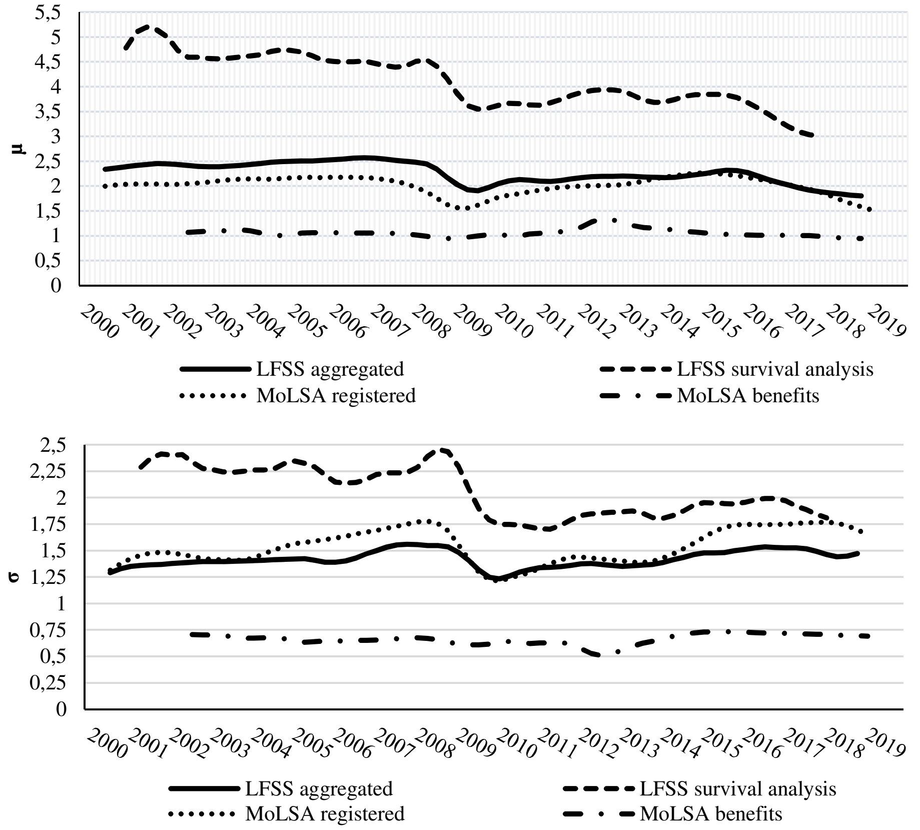
Estimated Densities for Data from LFSS B (grey) **and Molsa A1** (black)



Source: Own computations.

F i g u r e 7

Estimated Parameters from Individual LFSS Data in Comparison

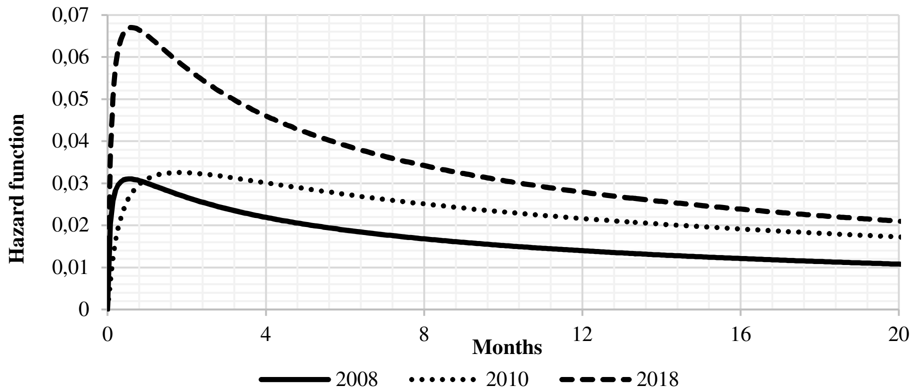


Source: Own computations.

Figure 8 depicts the estimated hazard functions for quarters Q2 2008, Q1 2010 and Q4 2018, the same quarters as shown in Figure 6. In the first quarter of the year 2010, which is at the peak of the crisis, the intensity of finding a new job was increasing slowly at the first month compared to pre-crisis values, but was higher since then. The last quarter shows a much higher intensity of reemployment than the other two with a much more significant peak. All of the hazards approaches the constant rate of change for higher values, which suggests the close-to-exponential distribution of the right tail. The interpretation could be as follows: from some sufficiently long time of unemployment (near two years), the probability of reemployment starts being independent of the length of unemployment.

F i g u r e 8

Estimated Hazard Function for Selected Quarters from Figure 6
(based on individual LFSS c data)



Source: Own computations.

Conclusion

In the text, four available data sources for the modelling of the length of the unemployment spell in the Czech Republic. The data include aggregated published data based on the database of the unemployed of the Ministry of Labour and Social Affairs and datasets from the regular sample survey organized by the Czech statistical office Labour Force Sample Survey. The data are published in an aggregated version and the individual data are available for researchers in the SaveCentre. These datasets can be used to analyse the same phenomenon but are of different types from the point of view of origin, individual or aggregated form as well as there are two definitions of the unemployed. The unemployment rate is usually applied to describe the phenomenon and we try to describe different answers to the same situation on the labour market and the shift between time series of the unemployment rate and median unemployment spell length.

Lognormal distribution is used to model the probability distribution of the unemployment spell length in the Czech Republic from all datasets and provide information on the development of this distribution from 2000 through 2019. From the study it seems to be a generally acceptable model for the empirical distributions with theoretical properties corresponding to the requirements however the goodness-of-fit test does not support the hypothesis. Moreover, the straightforward interpretation of the parameters (as the mean and the standard error of the log of the analysed spell length) is useful. But unlike the maximum likelihood estimation, the estimates are not independent and applied univariate presentation could be used to take it into account. For this reason, we use the estimated distributions to describe characteristics and to compare results. We show that the mean and median unemployment duration decreases at the beginning of the economic crisis, which is caused by the inflow of newly unemployed, then slowly rises even at the beginning of economic recovery and boom, as the layoffs are limited. Probably the persons with lower unemployment spell are preferred for reemployment. If the economic boom is long and stable enough, the mean and median unemployment duration starts to decrease, which can be viewed as a sign of reemployment of persons with long unemployment spells.

From the aggregated data, we can model only the time of unemployment (as observed times of reemployment are not observed); in case of incomplete individual values, the time to reemployment is analysed and modelled. We show similar development of duration modelled based on different data and the impact of data to the estimated parameters, usually, it is not reasonable to compare absolute values of estimated characteristics or parameters.

In our model, we obtain highly skewed estimated distributions because of the relatively large ratio of unemployment spells longer than two years for aggregated data (Figure 2) or unemployment spells longer than 4 or 10 years in individual data. Unlike some analyses (for example, Čabla and Malá, 2017), we considered all unemployed people, not only those with a period shorter than two years. It results in larger characteristics of a location in our study and even longer times based on the individual data. This data better describes the tails using respondents with a very long spell. We do not estimate the proportion of those, who will never reemploy. From the parametric point of view, the decrease during an economic boom is driven mainly by a decrease of location parameter μ and not by the parameter σ, which means that skewness and relative variability of the distribution remains stable.

The gender gap is visible in terms of higher μ (and so the median) for women than for men and higher σ (and so skewness and relative variability) for men in MoLSA dataset with no clear persistent difference in the LFSS dataset. Gender is not supposed to be highly significant in comparing labour market position but we show differences in estimated parameters.

References

ARSLAN, H. – SENTURK, I. (2018): Individual Determinants of Unemployment Duration in Turkey. Eskisehir Osmangazi University Journal of Economics and Administrative Sciences, *13*, No. 1, pp. 113 – 127.

BOVER, O. – ARELLANO, M. – BENTOLILA, S. (2002): Unemployment Duration, Benefit Duration and the Business Cycle. The Economic Journal, *112*, No. 479, pp. 223 – 265.

CZSO (2014): Czech Statistical Office. Available at: <http://www.czso/eng>.

BRUGIAVINI, A. – ORSO, C. E. – GENIE, M. G. – NACI, R. – PASINI, G. (2019): Combining the Retrospective Interviews of Wave 3 and Wave 7: The Third Release of the SHARE Job Episodes Panel. [SHARE Working Paper Series 36-2019.] Munich: SHARE-ERIC. Available at: <http://www.shareproject.org/uploads/tx_sharepublications/WP_Series_36_2019_Brugiavini.pdf>.

ČABLA, A. (2017): Right Tail of the Unemployment Duration in the Czech Republic. In: Applications of Mathematics and Statistics in Economics. [20th International Scientific Conference AMSE.] Szklarska Poreba, Wroclaw: Publishing House of Wrocław University of Economics, pp. 59 – 74.

ČABLA, A. – MALÁ, I. (2017): Modelling of Unemployment Duration in the Czech Republic. Prague Economic Papers, *26*, No. 4, pp. 438 – 449.

EUROSTAT (2022): Eurostat Statistics Explained. Available at: <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Unemployment_statistics>.

GROGAN, L. – van den BERG, G. J. (2001): The Duration of Unemployment in Russia. Journal of Population Economics, *14*, No. 3, pp. 549 – 568.

ILO (2019): International Labour Organization. Available at: <https://www.ilo.org/global/lang--en/index.htm>.

HANAPPI, D. – LIPPS, O. (2019): Job Insecurity and Parental Well-being the Role of Parenthood and Family Factors. Demographic Research, *40*, Article 31, pp. 897 – 932.

HARRIS, R. R. – KANJI, G. K. (1983): On the Use of Minimum Chi-Square Estimation. Journal of the Royal Statistical Society, Series D (The Statistician), *32*, No. 4, pp. 379 – 394.

KRUEGER, A. B. – MUELLER, A. – DAVIS, S. J. – SAHIN, A. (2011): Job Search, Emotional Well-being, and Job Finding in a Period of Mass Unemployment: Evidence from High Frequency Longitudinal Data (with Comments and Discussion). Brookings Papers on Economic Activity, pp. 1 – 81.

KUPETS, O. (2006): Determinants of Unemployment Duration in Ukraine. Journal of Comparative Economics, *34*, No. 2, pp. 228 – 247.

LAWLESS, J. F. (2002): Statistical Models and Methods for Lifetime Data. 2nd ed. Wiley: Series in Probability and Statistics. LFSS (2019): Labour Force Sample Survey. Available at: <http://www.czso.cz/eng/redakce.nsf/i/lfs_analyses_in_news_releases>.

LIM, J. Y. – LEE, Y. M. (2019): Exit Duration and Unemployment Determinants for Korean Graduates. Journal for Labour Market Research, *53*, No. 5.

McCALL, B. – CHI, W. (2008): Unemployment Insurance, Unemployment Durations and Re-employment Wages. Economics Letters, *99*, No. 1, pp. 115 – 118.

POLLMANN-SCHULT, M. – BÜCHEL, F. (2005): Unemployment Benefits, Unemployment Duration and Subsequent Job Quality: Evidence from West Germany. Acta Sociologica, *48*, No. 1, pp. 21 – 39.

R CORE TEAM (2017): R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Available at: <http://www.r-project.org/>.

ROSHOLM, M. (2001): Cyclical Variations in Unemployment Duration. Journal of Population Economics, *14*, No. 1, pp. 173 – 191.

SETYADI, D. – SUDARSO, Y. – NAHAR, M. – SUGIYANTA. (2019): Duration of Educated Unemployment. Economic Journal of Emerging Markets, *11*, No. 2, pp. 213 – 222.

TERRACOL, A. (2009): Guaranteed Minimum Income and Unemployment Duration in France. Labour Economics, *16,* No. 2, pp. 171 – 182.

THERNEAU, T. (2015): A Package for Survival Analysis in S. R Package Version 2.38. Available at: <https://CRAN.R-project.org/package=survival>.

UYSAL, S. D. – POHLMEIER, W. (2001): Unemployment Duration and Personality. Journal of Economic Psychology, *32*, No. 6, pp. 980 – 992.

WINKELMANN, L. – WINKELMANN, R. (1998): Why Are the Unemployed So Unhappy? Evidence from Panel Data. Economica, New Series, *65*, No. 257, pp. 1 – 15.