

## Evaluation of the Duration of Unemployment Spells Using Kaplan-Meier Estimator. A study on Botoșani County's Labor Market

Sandu Christiana Brigitte

Darabă Gabriel

Jaba Elisabeta

*"Alexandru Ioan Cuza" University of Iasi,  
Faculty of Economics and Business Administration*

[christiana.balan@uaic.ro](mailto:christiana.balan@uaic.ro)

[gdaraba@yahoo.com](mailto:gdaraba@yahoo.com)

[ejaba@uaic.ro](mailto:ejaba@uaic.ro)

### Abstract

*The aim of this paper is analysing unemployment spells for Botoșani County's labour market. The database includes individual information on a sample of 200 unemployment spells registered at the Employment Agency of Botoșani County during January 2012 - December 2015.*

*We compare Kaplan-Meier estimates of the probability of staying unemployed in function of several characteristics, namely cohort, gender, and age. The method is applied using STATA 13.0 software. The results of our estimations show that probabilities of staying unemployed are influenced by the variable cohort, when controlling for gender and age.*

**Key words:** unemployment, Kaplan- Meier estimator, survival probabilities, cohort

**J.E.L. classification:** P33, P34

### 1. Introduction

Unemployment duration is an important variable that explains the changes in labour market and is generally used in modelling jobs creation, especially when analysing the relationship among the employment, unemployment, and labor market exits flows (Ciucă and Matei, 2010, p. 354). A labour market where few people become unemployed, but they tend to remain unemployed for a very long time, is more insecure than one in which there are many more persons that become unemployed but remain in that position for only a short period of time (Bajram, 2013, p. 36).

High unemployment rate and high long term unemployment duration have important economic, social and individual consequences. At the macroeconomic level, unemployment generates an increase in the social costs that an economy have to manage by the unemployment benefits (Dănăicică and Paliu-Popa, 2017, p. 964). These effects become more important when exit rates fall and the current spell of unemployment lengthens (Carroll, 2004, p. 1). At the individual level, it is generally agreed that unemployment, and especially long term unemployment, deprives people not only from their livelihood but also from meaningful activity in their lives (Tansel and Taşçi, 2010, p. 502).

For the identification of the factors that explain unemployment spells and the exit rates (unemployment exit rates by reason of getting employed or of ending the legal period of unemployment) there are used survival analysis methods applied in labour market research (Mussida, 2007, p.49).

Explaining the unemployment in Botoșani County according to different factors requires understanding the dynamics of the unemployed population. During the period 2012-2016, the monthly evolution of the number of unemployed registered at the Employment Agency of Botoșani County presents an increasing trend with a maximum value observed in January 2014 (Darabă, Sandu and Jaba, 2017, p. 155).

This study aims analysing the unemployment duration on the labour market in Botoşani County using a statistical tool useful in estimating the probability of remaining unemployed (the Kaplan-Meier estimator) in relation to several characteristics (cohort, age, and gender). The results show a significant influence of the cohort on the unemployment duration when controlling for gender and age.

The paper is organized as follows: in Section 2 we give a description of the Kaplan - Meier estimator and we present the data used in this study; in Section 3 we analyse the unemployment duration for the Botoşani County's labour market; Section 4 concludes the analysis.

## 2. Method and data

Duration data are not normally distributed and they often contain incomplete observations (censored data) (Nonyana, 2015, p. 17). The data collected on the unemployment spells from the Employment Agency of Botoşani County include data that are right censored (unemployment duration for those persons that exit unemployment after the end of the study).

Under the assumption that the distribution of unemployment duration is not known, we use the Kaplan - Meier method to estimate the probability of staying unemployed. The survival functions are graphically represented by the Kaplan - Meier survival curves and they are used to highlight the potential influence of the explanatory variables and to test the difference among the categories of factor variables.

### 2.1. Kaplan - Meier estimator

The Kaplan Meier estimator is a non-parametric estimator of survival function  $S(t)$ , the cumulative probability of not leaving unemployment at time  $t$  (Kaplan and Meier, 1958). The method is well-known for its capability of handling censored data. The Kaplan - Meier survival function has the following expression:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

where  $t_i$  represents the survival time;  $d_i$  represents the number of unemployed persons that exit unemployment at time  $t_i$ , and  $n_i$  represents the number of persons at risk of leaving unemployment at time  $t_i$ .

It is recommended to apply this method when the sample is not very large or when the time unit considered for expressing the durations is small (days or months for socio - demographic events).

The Kaplan-Meier estimator of the survival function allows preliminary analysis of duration data and the exploration of the univariate association between the categories of factors and the duration of unemployment (Nonyana, p. 20).

### 2.2. Data

The empirical analysis is based on a sample of 200 unemployment spells registered at the Employment Agency of Botoşani County from January 2012 to December 2015.

The available information for each individual concerns: the beginning and the end date of the unemployment spell, gender, age, and the reason for leaving unemployment.

The variables used in this study for the analysis of unemployment duration are:

- **Unemployment duration** is expressed as the number of unemployment days and is calculated as the difference between the exit date and the entry date in unemployment. The minimum duration value is equal to 0 days and the maximum duration value is equal to 1020 days;
- **Cohort** is represented by the four years from the period between 2012 and 2015. There were randomly selected 50 cases by year for the observed period;
- **Gender** is defined with the two categories: 1 - Male and 2 - Female;

- **Age** is grouped into two categories: 1 – unemployed persons less than 25 years old (24.5% of the overall sample) and 2 – unemployed persons aged 25 years or over (75.5% of the overall sample). By these two age groups the Ministry of Labor defines the long term unemployment: 6 months of unemployment for the persons aged 16-24 years, and more than 12 months of unemployment for the persons aged 25 years or over (Ministry of Labor, 2014, p.2).
- **Status** is defined by the reason of leaving unemployment during the period 2012-2015. We grouped the 16 reasons of leaving unemployment observed for the 200 persons registered at the Employment Agency (in accordance with Law 76/2002) into two categories: 1 – employed (first employment or re-employment) and 0 – unemployed. This variable allows to differentiate between the persons who had registered the event of interest (becoming employed) and the persons who hadn't registered the event of interest (stayed unemployed).

### 3. Results of the Kaplan - Meier method

By applying the nonparametric method Kaplan – Meier, we estimated the survival probabilities at time  $t_i$  for the overall sample, and by cohorts, age and gender groups; we represented the Kaplan - Meier survival curves and we tested the differences among the survival probabilities.

The results for the descriptive statistics on unemployment duration are presented in Table 1. The indicators calculated are the mean, the quartiles and the 95% confidence interval for mean.

*Tabel no. 1. Descriptive statistics for the unemployment duration (days)*

Group	N	Mean	95% Confidence Interval	Survival time		
				25%	50%	75%
Overall sample	192	416.8496	[344.492 ; 489.207]	35	361	894
<b>Cohort</b>						
2012	50	534.9108	[398.674 ; 671.148]	147	539	918
2013	48	483.3813	[347.818 ; 618.944]	35	444	894
2014	47	232.3466	[169.613 ; 295.08]	25	341	365
2015	47	202.8427	[155.593 ; 250.092]	17	271	366
<b>Gender</b>						
Male	128	431.3575	[346.801 ; 515.914]	56	362	979
Female	64	414.0469	[263.006 ; 565.088]	9	365	894
<b>Age group</b>						
< 25 years	45	128.4517	[104.065 ; 152.838]	35	182	–
≥ 25 years	147	417.9025	[342.019 ; 493.786]	37	361	894

*Source:* Authors' calculation with Stata 13.0 software

For the overall sample, the mean unemployment duration is almost 417 days (14 months) and the median unemployment duration is 361 days (12 months). Therefore, 50% of the unemployed persons have found a job only after 12 months of unemployment.

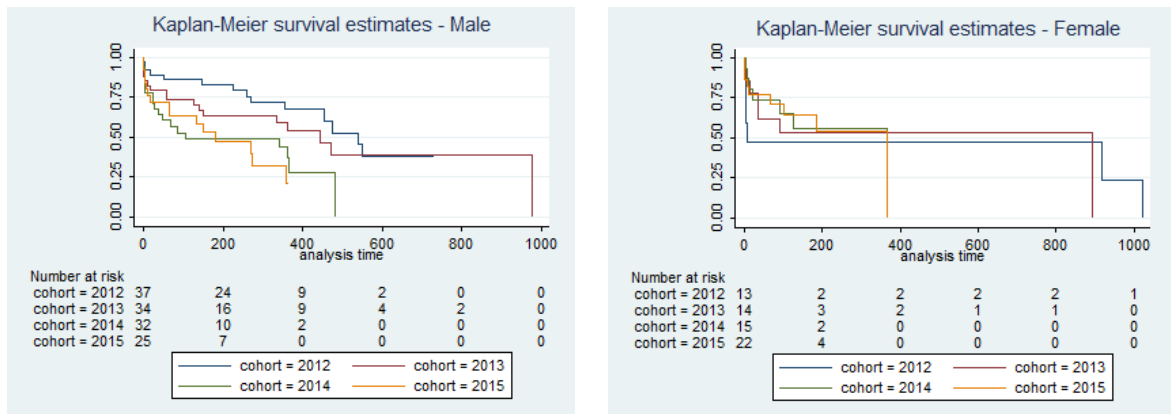
The cohorts that entered unemployment in 2012 and 2013 have the highest mean unemployment duration (534.9 and 483.38 days respectively) while the cohorts that entered unemployment in 2014 and 2015 have the lowest mean unemployment duration (232.35 and 202.84 days respectively). Also, 50% of the persons of the 2012 cohort stayed unemployed for more than 18 months.

By gender, the mean values of the unemployment duration of the two groups are comparable: the male unemployed group has a mean unemployment duration of 431.36 days and the female unemployed group has a mean unemployment duration of 414.05 days.

The mean unemployment duration of young persons (aged less than 25 years) is, on average, smaller than for persons aged 25 years and over. Moreover, 50% of the young persons registered long term unemployment (stayed unemployed for more than 6 months), while 50% of the persons aged 25 years and over stayed unemployed for more than 12 months.

The probabilities of staying unemployed (survival probabilities) are presented by cohort and by groups defined by gender and age using the Kaplan – Meier survival curves in Figures 1 and 2.

Figure no. 1 Kaplan – Meier Survival functions by cohort and gender groups

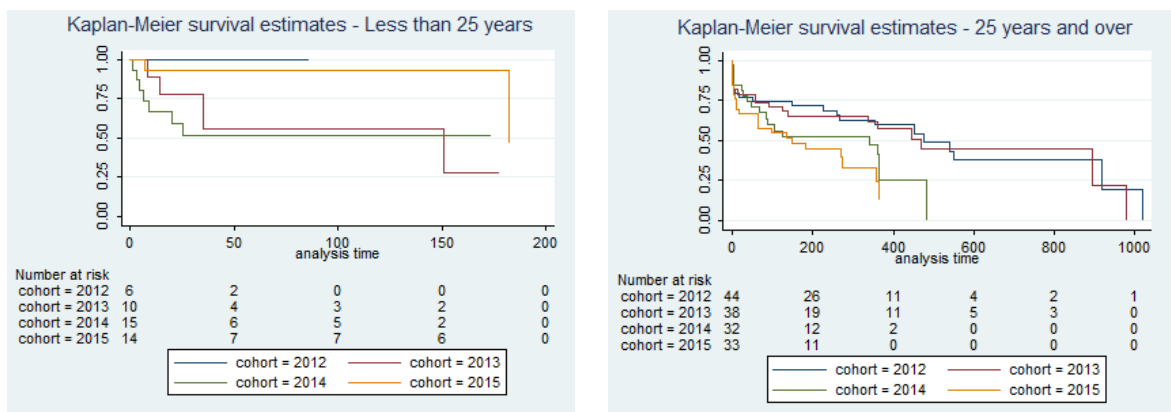


Source: Authors' calculation with Stata 13.0 software

The tables under the survival curves show the numbers of persons at risk of leaving unemployment at time  $t_i$  for each cohort and by groups defined by the variables gender or age.

In Figure 1 – left, it can be seen that, for male persons, the unemployment survival probabilities are highest for the unemployed persons of the 2012 cohort, while the 2014 and 2015 cohorts have the lowest unemployment survival probabilities. For female persons (Figure 1 - right), the unemployment survival curves are not very different among the four cohorts. Moreover, the probabilities of staying unemployment in the first 6 months of unemployment are smaller for the 2012 and 2013 cohorts than from the 2014 and 2015 cohorts.

Figure no. 2 Kaplan – Meier Survival functions by cohort and age groups



Source: Authors' calculation with Stata 13.0 software

In Figure 2 – left, we can notice that probabilities of staying unemployed for the young persons are smaller for the 2013 and 2014 cohorts. Moreover, the unemployment duration is smaller for the young unemployed persons than for persons aged over 25 years.

The results for the Log-rank test for equality of survival functions are presented in Table 2. There are significant differences in the probabilities of staying unemployed among the four cohorts for the male sample. The Chi-Square shows significant differences among the male cohorts (probability level smaller than 1%), while the female cohorts are not statistically significant.

Tabel no. 2. Log-rank test for equality of survival functions

Variable	Chi - Square	Degrees of freedom	Pr > chi2
Cohort	6.63	3	0.0847
Cohort for Gender = Male	11.37	3	0.0099**
Cohort for Gender = Female	0.46	3	0.9272
Cohort for Age < 25 years	8.05	3	0.0449*
Cohort for Age ≥ 25 years	10.32	3	0.0161*

Source: Authors' calculation with Stata 13.0 software;

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

By age groups, we can see that there are significant differences among the four cohorts both for the young unemployed persons and for the unemployed persons over 25 years. For both test, the probability of the Chi-Square statistic is smaller than 5% risk.

Therefore, the factors cohort, age and gender can be considered as influence factors of the unemployment survival probabilities, and consequently of the unemployment duration. However, we cannot say that there are significant differences among the survival curves for female cohorts.

#### 4. Conclusions

The aim of this paper was to analyze the duration of unemployment spells in the Botoșani County. The unemployment in Botoșani County had registered an increasing trend during the period 2012-2015. Nonparametric methods such as Kaplan-Meier estimator are an important tool for estimating the average unemployment duration and the influence of the factors with impact on the unemployment duration. The analysis of the duration of unemployment spells was based on the data obtained from the Employment Agency of Botoșani County, which consists of the unemployment spells between January 2012 and December 2015. For a sample of 200 individuals experiencing unemployment, we have estimated the survival probabilities using the Kaplan – Meier estimator and we have tested the differences in survival probabilities among the groups of unemployed persons.

The results have shown that there are significant differences among the observed cohorts. The differences by cohorts are significant for the male cohorts, whereas the differences for the female cohorts are not statistically significant. Furthermore, the unemployment duration is different among the four cohorts both for young persons and for the persons aged 25 years and over.

The probability of staying unemployed for persons aged 25 years and over is higher for the persons that belong to the 2012 and 2013 cohorts (50% of the unemployed persons from these two cohorts stayed unemployed for more than 16 months) than for the persons belonging to 2014 and 2015 cohorts (50% of the unemployed persons from the 2015 cohort stayed unemployed for more than 4.5 months and for more than 6 months in the case of the 2014 cohort). The more disadvantaged persons are the one that remained unemployed for a long time and had experienced a depreciation of their skills on the labour market.

We can conclude that the differences in the probabilities of staying unemployed among the four cohorts under the impact of gender and age show the necessity of a managerial strategy that should consider the differences in age and gender of the unemployed persons.

#### 5. References

- Bajram, N., 2013. "Distribution of Unemployment Spells and Survival Analysis: The Case of Bosnia and Herzegovina". *Southeast Europe Journal of Soft Computing*, 2(2), pp.36-39.
- Carroll, N., 2014. "Explaining Unemployment Duration in Australia". *Discussion paper no. 483*, pp. 1-28.
- Ciucă, V., Matei, M., 2010. "Survival analysis for the unemployment duration". *Proceedings of the 5th WSEAS International Conference on Economy and Management Transformation (EMT'10)*, vol. I, pp. 354-360;
- Darabă, G., Sandu, C.B., Jaba, E., 2017. « Estimation of Unemployment Duration in Botoșani County Using Survival Analysis". *Ovidius University Annals, Economic Sciences Series 17 (1)*, pp. 155-161.

- Dănăicică, D.E., Babucea, A.G., 2010. "Using survival analysis in economics". *Scientific Annals of the Alexandru Ioan Cuza University of Iași*, pp. 439-450.
- Dănăicică, D.E., Paliu-Popa, L., 2017. « Determinants of unemployment spells and exit destinations in Romania in a competing-risks approach". *Economic Research-Ekonomska Istraživanja*, Vol. 30, No. 1, pp. 964-984.
- Kaplan, E.L., Meier, P., 1958."Nonparametric Estimation from Incomplete Observations". *Journal of the American Statistical Association*, Vol. 53, No. 282, pp. 457-481.
- Mussida, C., 2007. "Unemployment duration and competing risks: A regional investigation". *Serie Rossa: Economia – Quaderno N. 49*, pp.1-57.
- Nonyana, J.Z., 2015. "Statistical modeling of unemployment duration in South Africa", *University of South Africa, Pretoria*, [online]. Available at: <http://uir.unisa.ac.za/handle/10500/20982> [Accessed 10 November 2017].
- Tansel, A, Taşçi, H.M., 2010. "Hazard analysis of unemployment duration by gender in a developing country: The case of Turkey". *Labour 24 (4)*, pp. 501-530.
- [http://www.mmuncii.ro/j33/images/Documente/Transparenta/2014/2014-02-25\\_DO-IntrebariFrecvente\\_DOES.pdf](http://www.mmuncii.ro/j33/images/Documente/Transparenta/2014/2014-02-25_DO-IntrebariFrecvente_DOES.pdf) [Accessed 5 May 2018].