Estimation of Unemployment Duration in Botoşani County Using Survival Analysis

Darabă Gabriel
Sandu Christiana Brigitte
Jaba Elisabeta
"Alexandru Ioan Cuza" University of Iasi,
Faculty of Economics and BusinessAdministration
gdaraba@yahoo.com
christiana.balan@uaic.ro
ejaba@uaic.ro

Abstract

In this paper we aim at estimating the unemployment duration in Botosani County in order to study the impact of individual characteristics (gender, age, place of residence, unemployment benefit, etc.) on the length of unemployment spells. We use Cox regression model to measure the effects of gender, age, residential environment, etc. on the hazard rate of leaving unemployment andKaplan-Meier estimator to compare survival probabilities among different categories of unemployed persons. The study is carried out on a sample of 200 unemployment spellsregistered with the Employment Agency of Botoşani County from January 2012 to December 2015. The results reveal that place of residence, unemployment benefit and unemployed category have a significant impact on unemployment spells.

Key words: unemployment spells, Kaplan - Meier estimator, Cox model, censored data, hazard rate
J.E.L. Classification: P33, P34

1. Introduction

The North-East Region is the least developed region of Romania. Within this region, the poorest areas are: the southeast of Neamţ County, the south of Iasi County, the eastern of Bacău County, and the counties of Botoşani and Vaslui. The low level of development of Botoşani County is determined by the following factors: reduced economic alternatives for the rural population (almost 75% of the population at risk of poverty or social exclusion lives in rural areas), low level of competitiveness of the economic environment and low level of job attractiveness, especially in rural areas (in rural area there is a high level of youth unemployment, practically one of every three young people aged 19-25 is unemployed) (Botoşani County Council, 2014, pp. 12-13). During the period 2012-2016, the monthly evolution of the number of unemployed registered at the Employment Agency of Botoşani Countypresents an increasing trend and a maximum value observed in January 2014.

The analysis of labour market indicators such as employment rate and the number of unemployed, over a certain period of time, is not completely relevant for supporting labour market policies in a region. In order to implement adequate policies on the labour market, it is essentially to study the entrance / exit of individuals on / off thelabour market, the individuals' ability to quickly find alternative jobs, and the degree to which different groups of individuals are more affected than others by changes in the labour market (Bajram, 2013, p. 36).

The issue of unemployement duration implies the use of particular statistical methods, such as survival analysis. Over the last decades, duration models have been applied in social sciences for analysing the length of unemployment spells and the duration of labour strikes (Boršić and Kavkler, 2009, p. 124).

In this paper we aim at estimating the unemployment duration in Botoşani County using survival analysis techniques. We use Cox regression model to measure the effects of individual characteristics (gender, age, residential environment, unemployment benefit, etc.) on the risk of leaving unemployment and Kaplan-Meier estimator to compare survival probabilities among different categories of unemployed persons.

2. Method and data

The analysis of duration data requires different statistical analysis compared to quantitative data due to their particularities (Dănăcică and Babucea, 2010, pp. 439-440). Data on the length of unemployment spell are not normally distributed and they often contain incomplete information (censored data). For example, an individual for which the end of the unemployment spell had not occurred by the end of the study represents a right censoring.

The methods proposed in the literature for the statistical analysis of duration data are nonparametric methods (Kaplan – Meier estimator, Nelson – Aalen estimator, etc), parametric methods (Weibull, exponential, log-normal, and log-logistic models) and regression models (Cox model).

In this study, we apply the Cox regression model and the Kaplan-Meier survival function estimator.

2.1. Cox regression model

The Cox model is a regression model proposed by Cox in 1972 to model the relationship between the hazard rate and the explanatory factors (Nonyana, 2015, p. 39). Using the Cox regression allows to identify the individual characteristics that have a significant impact on the hazard function (Kavkler, 2009, p.87). The hazard function, $h(t)$, describes the risk of occurrence of an event (leaving unemployment through finding a job) and has the following expression:

$$h(t, X) = h_0(t) * e^{\beta X} \qquad (2)$$

where $h_0(t)$ represents the baseline hazard, and βX represents the linear combination of the explanatory variables vectors.

The Cox model assumes the proportional hazards hypothesis (the hazard ratio of two individuals is independent of time, thus the effect of the explanatory variables does not vary over time).

2.2. Kaplan – Meier estimator

The Kaplan-Meier method is a nonparametric method used for estimating the survival function *S(t)*, the cumulative probability of not leaving unemployment since time *t* (Dănăcică and Babucea, 2007, p.35). The Kaplan - Meier estimator allows the estimation of the survival function in the presence of right censoring (Ciucă, 2010, p. 355). This method is also called the *Product Limit Estimator* (PLE). The Kaplan - Meier survival function has the following expression:

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

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

2.3. Data

The study is carried out on a sample of 200 unemployment spells registered at the Employment Agency of Botoşani County from January 2012 to December 2015. There were randomly selected 50 cases for each year during the period 2012 – 2015.

The available information for each individual is: the beginning date of unemployment spell, gender, age, education, place of residence (urban or rural), unemployment benefit (yes / no), unemployed category (new graduates, fired or persons without any income), the reason for leaving unemployment and the end date of the unemployment spell.

The variables used in the analysis of unemployment duration are presented in Table 1.

Table no. 1 The list of the variables used in the analysis of unemployment duration

Variables	Expression	Observations
Duration and event variables		
Length of unemployment spell (duration of unemployment):	The number of unemployment days (the difference between the date of entry and the date of leaving unemployment).	The minimum value is 0 days and the maximum value is 1020 days (34 months). The mean value is 184.21 days (6.14 months), while the median duration is 113.5 days (3.78 months).
The status of the person (or the census index)	We grouped all the different reasons for leaving unemployment into 2 main outgoing destinations: 1 - employment (first employment or re-employment); 0 - economically inactive	There are 16 different reasons for leaving unemployment for the 200 persons registered with the local agency, according to Employment Agency and Law 76/2002. The status variable allows for a distinction between the individuals who registered the event (employment) and the individuals who did not register the event.
Explanatory variables		
Gender	Male (0); Female (1)	Out of f the total of 200 unemployed persons registered between January 2012 and December 2015, 68 (34%) are women and 132 (66%) are men.
Age	We grouped the ages into 5 categories (according to the methodology of the Romanian National Institute of Statistics): Age_1 → 15-24 years Age_2 → 25-34 years Age_3 → 35-44 years Age_4 → 45-54 years Age_5 → 55-64 years	The average age of the persons registered in the database is 38.4 years, while the median is 38.5 years. The youngest unemployed person is 17 years old and the oldest is 63 years old
Education level	The level of education is grouped in 3 categories: Educ_1 → Primary and secondary or vocational education Educ_2 → High school and post-secondary education Educ_3 → Tertiary education	Out of the sample, 39% of total persons have primary and secondary or vocational education, 48% of total persons have high school and post-secondary education, and 13% of total persons have tertiary education
Place of residence	Urban (0); Rural (1)	73 people (36.5%) live in rural area and 127 (63.5%) live in urban area.
Unemployment benefit	No (0); Yes (1)	Out of the sample, 148 persons (74%) receive unemployment benefits, while 52 persons (26%) do not receive unemployment benefit.
Unemployed category	Recent graduates (1); Fired (0)	Out of the sample, 48 persons (24%) are recent graduates and 152 persons (76%) are registered as fired or persons without income.

Source:Authors' presentation of the variables used in the paper

3. Results and discussions

In this section, we firstly present the results of theCox model used to estimate the effect of the explanatory variables (gender, place of residence, unemployment benefit, etc.) on the risk of leaving unemployment. Secondly, we discuss the results of the Kaplan - Meier nonparametric method used to estimate the survival probabilities and to compare survival curves by categories of factor variables.

3.1Results of the Cox model

We estimated the Cox regression model including simultaneously all the explanatory variables. The reference categories are: *male* for Gender, *urban* for Place of residence, *No benefits* for Unemployment benefits, *Fired or without income* for unemployment category, *55-64 years* for Age and *Tertiary education* for Education level. The results are presented in Table no. 2.

Table no. 2 The estimations of the coefficients of the Cox regression model

Variables	Parameter estimation (b)	Standard Error	Wald Test	df	Sig.	Hazard ratio (e^b)
Gender	0.068	0.240	0.080	1	0.778	1.070
Place of residence	0.408	0.225	3.285	1	**0.070**	1.503
Unemployment benefit	-3.785	0.354	114.237	1	**0.000**	0.023
Unemployed category	-0.766	0.411	3.477	1	**0.062**	0.465
Education level			0.646	2	0.724	
Educ_1	0.310	0.406	0.580	1	0.446	1.363
Educ_2	0.312	0.402	0.600	1	0.439	1.366
Age			1.669	4	0.796	
Age_1	0.360	0.432	0.695	1	0.404	1.434
Age_2	0.336	0.372	0.817	1	0.366	1.400
Age_3	0.310	0.340	0.832	1	0.362	1.363
Age_4	0.030	0.341	0.008	1	0.930	1.030

Source: Authors' calculation with SPSS software

The hazard rate of leaving unemployment through a job is 7% higher for a woman than for a man, but this difference is not statistically significant (Sig. = 0.778).

We can notice that the hazard rate of leaving unemployment for unemployed persons living in rural area is 50.3% higher than for the unemployed persons living in the urban area. The variable place of residence has a significant impact on the hazard rate and on the unemployment duration for a significance level of 10%.

Moreover, the hazard rate of leaving unemployment for unemployment beneficiaries is lower by 97.7% than for people that have no benefits. This result is consistent with the results obtained at national level,where the lack of unemployment benefits increase the exit to a job hazard rate almost four times (Dănăcică and Cîrnu, 2014, p. 1178).The descriptive statistics apllied on our data show that persons that receive unemployment benefit have the average unemployment duration higher by 18 months than persons that do not receive unemployment benefits.

For new graduates, the hazard rate of leaving unemploymentis smaller by 53.5% than for the fired persons. We can assume that there are significant differences among the two categories of unemployed persons for a 10% significance level.The descriptive statistics on our sample show that among the 48 new graduates, 20.8% left unemployment for a job, while among the 152 persons fired or without any income, 59.2% have got a job.

For persons with primary, secondary and post-secondary education, the hazard rate of leaving unemployment is higher than persons with tertiary education by approximately 36%. However, education level does not have a significant impact on hazard rate.

The hazard rate of leaving unemployment does not differ significantly for the considered aged groups of unemployed persons and the reference group, 55-64 years. The average length of unemployment spells for persons aged 15 - 24 years is 125 days, while for unemployed persons aged 45 - 54 years and 55 - 64 years, the average unemployment duration is equal to 445 days and 426 days, respectively. If we had considered another age group as reference category, we may have had identified significant differences.

3.2. Kaplan – Meier survival probabilities by groups of unemployed persons

Figure no.1 presents the survival probabilities (the probabilities of being unemployed) by gender, place of residence, unemployment benefit, unemployed category, age, and education levels. If the survival curves overlap, we can say that survival probabilities do not differ among the groups of the considered variable, and if the survival curves do not overlap, then, the survival probabilities are different by groups.

Figure no. 1 Kaplan – Meier Survival functions by groups of unemployed persons
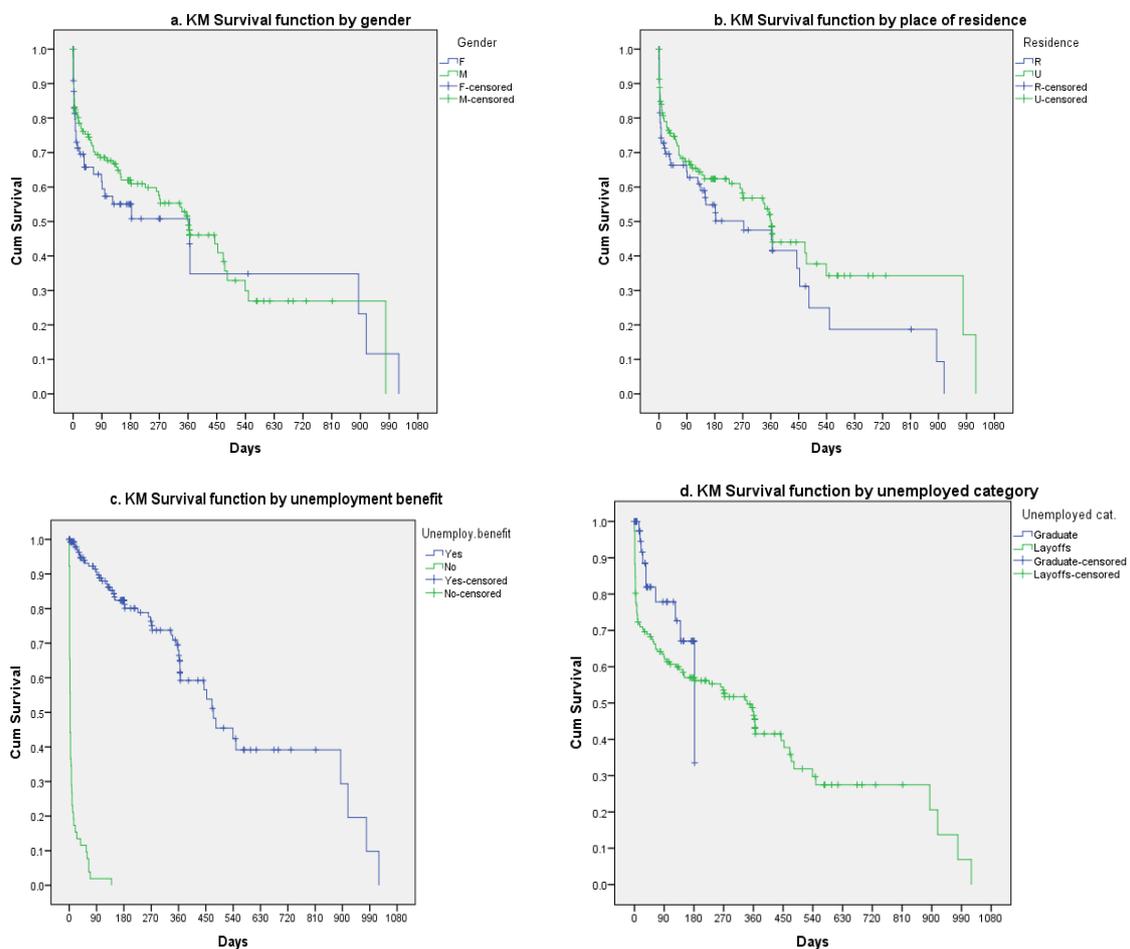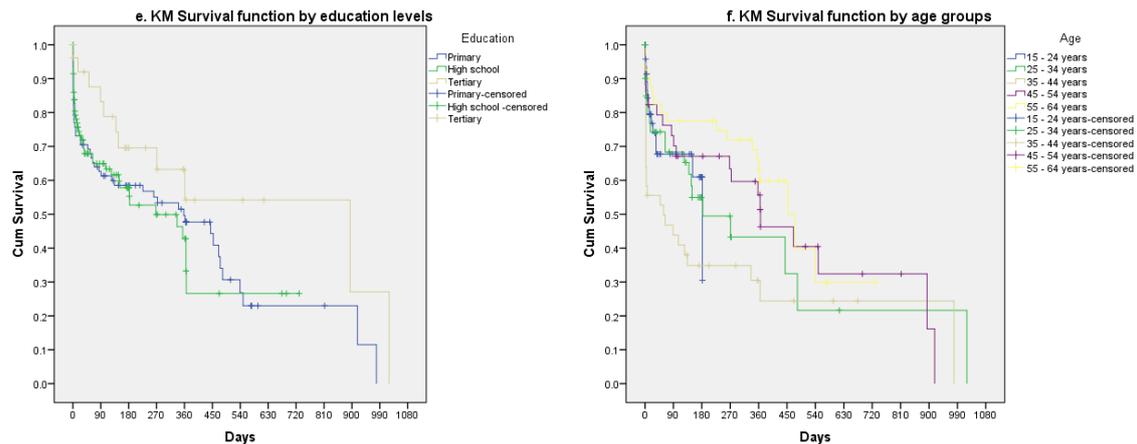
Source: Authors' calculation with SPSS software

We can notice that the probabilities of being unemployed are different by age, place of residence,unemployment benefit, and unemployed category groups. The differences among the considered groups have been tested with the Log-Rank test (Mantel and Haenszel, 1959) and the results show that survival probabilities are significantly different by age,unemployment benefit and unemployed category groups.

4. Conclusions

In this paper we focused on the analysis of the duration of unemployment in the Botoşani County. The data used consisted in a sample of 200 unemployment spells observed during the period 2012-2015 at the Employment Agency of Botoşani County. In order to quantify the effects of the individual characteristics (gender, education, age, etc.) on the hazard rate of leaving unemployment, we have used the Cox regression model. We have estimated the survival probabilities using the Kaplan – Meier estimator and we have compared the survival curves among the groups of unemployed persons.

The results obtained for the Cox model show that, among the considered factors, the explanatory variables with the highest impact on hazard rate of leaving unemoployment are: the place of residence, unemployment benefit, and unemployed category. The most disadvanteged groups of unemployed persons are people living in urban areas, receiving unemployment benefits and having lost their jobs due to layoffs.

The probabilities of being unemployed are significantly different between the unemployment beneficiaries and the unemployed persons that do not receive benefits; between the recent graduates and the fired persons, and also between unemployed belonging to different age groups. Aged persons have more difficulties in leaving unemployment because it is more complicated for them to adapt to the new challenges of the labour market.

The results of this paper can help improve the implementation of adequate policies on the labour market of Botoşani County by identifying potential vulnerable groups of persons that look for employment. A potentially important area for future research is modeling the hazard rate of leaving unemployment by different groups of unemployed persons.

5. Acknowledgement

6. 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.
- Boršić, D., Kavkler, A.,2009. Duration of regional unemployment spells in Slovenia.*Managing Global Transitions*, nr. 2/vol. 7,pp.123-146.
- Botoşani County Council, 2014. Strategia de dezvoltare socio - economică a judeţului Botoşani 2014-2020, [online]. Available at: <http://www.cjbotosani.ro/files/h043_27.03.2014.pdf> [Accessed 4 April 2017].
- 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;
- Dănăcică, D.E., Babucea, A.G., 2007. Using Kaplan - Meier curves for preliminary evaluation the duration of unemployment spells.*Annals of the University "Constantin Brancusi" of Targu Jiu,*pp. 33-38.
- Dănăcică, D.E., Babucea, A.G., 2010. Using survival analysis in economics.*Scientific Annals of the AlexandruIoanCuza University of Iaşi*, pp. 439-450.
- Dănăcică, D.E., Cîrnu, D.,2014. Unemployment duration in rural Romania,*Procedia Economics and Finance*, 15, pp. 1173-1184;
- Nonyana, J.Z.,2015.Statisticalmodeling of unemployment duration in South Africa,*University of South Africa, Pretoria,*[online]. Available at: <http://uir.unisa.ac.za/handle/10500/20982>[Accessed 27 March 2017].