Survival rates in unemployment

Vasilica Ciucă, Monica Matei

Abstract— In this paper we use survival analysis tools for the examination of some aspects of the labor market in 16 counties of Romania. Our analysis is developed on a database which includes individual information about the subjects registered at the National Agency for Employment of Romania during three years, starting with 2007. We analyzed the unemployment duration and we used the Kaplan Meier estimator of the survival functions in unemployment in order to estimate the probabilities of leaving unemployment.The results of our estimation show that the survival rates in unemployment are influenced by age, education, gender and also that each county has a different survival function.

Keywords— Hazard, Kaplan Meier, Survival, Unemployment Duration.

I. Introduction

It is not enough to study the labor market only by analyzing the static variables like rates of employment and unemployment. For the decision making on the labor market it is essential to see the movement of people into and out of jobs, the extent to which they can or cannot quickly find alternative employment and to which extent different groups of the labor force are more affected than others.

Unemployment duration refers to the amount of time that an individual remains unemployed. What makes the unemployment duration an important variable in modeling the labor market? The unemployment duration is an important variable which can explain the changes in labor markets and it is widely used in the job destruction and job creation models when analyzing the flows between employment, unemployment and out of the labor force. Blanchard and Diamond [1] examined the unemployment duration dependence and suggested that this affects both the matching function and the wage function. Also it is interesting to study the unemployment duration over the economic cycle and afterwards to see which are the implications on other economic indicators are and how these affects other labor market variables. During recession the unemployment rate increases but also the unemployment durations. This has consequences for household spending and financial solvency which requires specific labor market policies.

In the "Employment in Europe 2009" we find an example which illustrates the importance of estimating the unemployment duration: "a 10% unemployment rate can represent two entirely different realities in terms of their implications for the welfare of those affected: one where every individual in the labor force experiences unemployment during 5 weeks per year and another where 10% of the population are unemployed during the whole year. " A labor market in which few people become unemployed, but where those who do are likely to remain unemployed for a very long time is likely to be more damaging than one in which are many more who become unemployed but remain in that position for only a short period of time.

The welfare of unemployment depends on the probability of leaving unemployment and obtaining a job (hazard rate). Duration of unemployment is sensitive to the level of unemployment benefits and to the entitlement period to benefits. In this paper we try to determine other factors (age, education) which may produce variations of the unemployment duration, using the survival analysis.

Often the unemployment duration is computed as the average duration of unemployment spells currently in progress (or interrupted/ incomplete spells) but there are also some nonparametric and parametric methods available for estimating it. There are some studies [7] which highlight the main differences existing between the results obtained from the two methods. The nonparametric methods are those used in medical studies, known as survival analysis.

Survival analysis was used for analyzing the Romanian labour market in 2009. In that paper a semi-parametric Cox regression approach was employed to model the impact of the variables age, gender, education level, and region on the duration of unemployment spells in five Central and Eastern European countries [5]. Their results have shown that unemployed persons with higher levels of education are in a better position in the labour market. Another conclusion drawn in the paper was that the longer the unemployment spell lasts, the less pronounced the differences between different age groups are.

II. Methodology. Essential concepts and notation

Nonparametric methods based on actuarial methods such as those of Kaplan and Meier provide important tools for estimating the average unemployment duration.

The Kaplan-Meier or product limit estimator was developed by Kaplan and Meier [4]. Their procedure is widely used in medical studies where "death" and "alive" are used frequently. In the context of our paper "alive" is remaining unemployed and "death" is finding a job. Thus, the survival function (S[t]) is the probability of an individual surviving (remaining unemployed) t units of time from the beginning of the study.

The survival function shows what proportion of a cohort of people who become unemployed remains unemployed as time passes. In this paper we present a few survival functions for various breakdowns- counties, gender, etc.

In general**, the survival function** $S(t) = \Pr(T > t)$, gives the population probability of surviving beyond *t*, where T is a random variable called the lifetime or the survival time(1958). The cumulative distribution of T is $P(t) = \Pr(T \le t)$ and the probability density function is $p(t) = dP(t)/dt$. This means the survival function is $S(t) = 1 - P(t)$.

Also, very important in the survival analysis is the **hazard function.** This function assesses the instantaneous risk of demise at time t, conditional on survival to that time**:**

$$h(t) = \frac{p(t)}{S(t)}. \tag{1}$$

The Kaplan- Meier estimator allows the computation of an estimated survival function in the presence of right censoring [2].

The usual sample survival function for uncensored data is the following [8]:

$$\hat{S}(t) = \prod_{t_{(i)} < t} \left\{ 1 - \frac{m_{(i)}}{r_{(i)}} \right\} \tag{2}$$

Where:

- The distinct failure times are denoted by: $t_{(1)} < t_{(2)} < ... < t_{(k)}$
- The number of failure times equal to $t_{(i)}$ is denoted by $m_{(i)}$
- The number of those individuals whose failure or censoring time is at least $t = t_{(i)}$ is denoted by $r_{(i)}$.

This estimator is also called the product-limit estimator because one way of describing the procedure is that it multiplies together conditional survival curves for intervals in which there either no censored observations or no deaths. This becomes a step function where the estimated survival is reduced by a factor $(1 - 1/r_{(i)})$ if there is a death at time $t_{(i)}$ and a population of $r_{(i)}$ is still alive and uncensored at that time [2].

In order to see whether two or more survival curves are identical we will use the **log rank test** which is based on looking at the population at each death time and computing the expected number of deaths in proportion to the number of individuals at risk in each group. This is then summed over all death times and compared with the observed number of deaths by a procedure similar to the χ^2 test. [2]

The null hypothesis of this test states that there is no difference between the population survival curves. The test statistic is calculated as follows:

$$\chi^2(\text{log rank}) = \frac{(\mathrm{O}_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \tag{3}$$

Where the O_1 and O_2 are the total numbers of observed events in groups 1 and 2, respectively, and E_1 and E_2 the total number of expected events. In the context of our paper the event is given by leaving the unemployment. The total expected number of events for a group is the sum of the expected number of events at the time of each event. The expected number of events at the time of an event can be calculated as the risk for "death" at that time multiplied by the number alive in that group. [11]

Log rank test and Kaplan Meier procedure have been generalized in order to allow for addition during the study (gradual entry into the study or staggered entry).

The survival data can be analyzed by regression models, known as **Cox proportional hazards** models, and we will use it in order to examine the relationship of the survival distribution to covariates. The scale on which linearity is assumed is the log-hazard scale. A linear model for the log-hazard may be written as:

$$\log h_i(t) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}, \tag{4}$$

Where *i* is a subscript for observation and *x* are the covariates. The constant α in this model represents a kind of log-baseline hazard, since $\log h_i(t) = \alpha$, when all of the *x*'s (explanatory variables are zero. The Cox model leaves the baseline hazard function $\alpha(t) = \log h_0(t)$ unspecified:

$$h_i(t) = \alpha(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}) \tag{5}$$

This model is semi-parametric because while the baseline hazard can take any form, the covariates enter the model linearly. Consider, now, two observations *i* and i' that differ in their *x*- values, with the corresponding linear predictors

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{ik} + ... + \beta_k x_{ik} \tag{6}$$

$$\eta_{i'} = \beta_1 x_{i'1} + \beta_2 x_{i'2} + ... + \beta_k x_{i'k} \tag{7}$$

The hazard ratio for these two observations,

$$\frac{h_i(t)}{h_{i'}(t)} = \frac{h_0(t)e^{\eta_i}}{h_0(t)e^{\eta_{i'}}} = \frac{e^{\eta_i}}{e^{\eta_{i'}}} \tag{8}$$

is independent of time t. Consequently, the Cox model is a proportional-hazards model. So we don't have to make assumptions about the form of the baseline hazard. [3]

The coefficient β_j will be interpreted in terms of the relative risk when the covariate x_{ij} is increased by 1:

$$\frac{\alpha(t)\exp(\beta_1(x_{i1} + \beta_2 x_{i2} + ... + \beta_j(x_{ij} + 1) + ... + \beta_k x_{ik})}{\alpha(t)\exp(\beta_1(x_{i1} + \beta_2 x_{i2} + ... + \beta_j x_{ij} + ... + \beta_k x_{ik})} = \exp(\beta_j)$$

III. ESTIMATION

A. Data

In this paper we use survival analysis tools for the examination of some aspects of the labor market in 16 counties of Romania: Bistrita Nasaud, Bacau, Bihor, Caras Severin, Calarasi, Cluj, Galati, Giurgiu, Harghita, Iasi, Maramures, Mehedinti, Satu Mare, Salaj, Sibiu, Ilfov.

Our analysis is developed on a database which includes individual information about the subjects registered at the National Agency for Employment of Romania during three years, starting with 2007. The information available for every individual is: the month and the year of registration to the employment agency known as the start date, the month and the year when the individual stopped being registered at the agency, known as the end date, gender, age, educational level and county. We can not say that the end date is the date of leaving the unemployment because there is no information available about the reason why individuals stopped coming to the agency. It is obvious that only some of them changed their status from unemployed to employed.

We are dealing with right censored data, meaning some individuals are still "alive" at the end of the study so the event of interest (death= leaving unemployment) has not occurred. Also length of follow up varies due to staggered entry. So we can not observe the event for those individuals with insufficient follow up time. In addition to censoring because of insufficient follow up other reason for censoring includes loss to follow up: individuals stop coming to the agency for unknown reasons.

Also, we can see from our data that during the period of observation individuals experience not only one spell of unemployment. So multiple unemployment spells are a common experience in Romania.

Another problem in this database is that for many individuals there is no information about their end date. This happens because the contact with some individuals is lost after few months after registering. Unfortunately we don't know when this happened. If we had this information we could have used this data as censored data (death from other reasons). Given this problem we had to eliminate all this individuals from our study.

The initial database contains *1059 985* registrations. After the elimination of the individuals over *65* years old, of those under *15* years old, and of those with no information on the end date, the size of our sample was *851 164*. Given some errors in the data series, and given the fact that there are individuals with no information available for the "end date" the final sample size became *596 772*. In conclusion in our data we have *596 772* completed spells of unemployment.

Figure 1 shows the numbers of persons registered at the Public Employment Services every month. At the beginning of each year there are more entries than in other months of the year. 27% of the individuals registered in 2007 were still unemployed in 2008 and 3.2% of them were still unemployed in 2009. From all the unemployed registered at the National Employment agency in 2008, 30% are still unemployed in 2009.Thus we can not say that the unemployment rate has declined in 2008 or 2009 compared to 2007. Also the situation presented in the figure is an aggregation of the registrations in the 16 counties, each of these having a different pattern in terms of monthly registrations distribution.

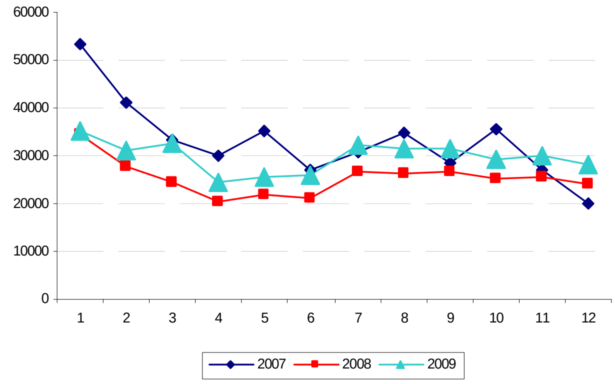

Fig.1 Number of monthly registrations per year

Approximately 60% of the unemployed persons registered each year are men. In terms of percentages of entries by age groups, there are small variations. Thus in 2007, 20% of the persons registered was in the age group 15- 25 years, 25 % was in the 26-35 years group, 26% was in the 36-45 years group, 22% was in the 46-55 years group and only 8% in the 56-65 years group. In 2008 the share of those registered from the age group 15-25 years increased to 23%, at the expense of other three age groups: 26-35 years group, 36-45 years group, 46-55 years group, their shares decreasing by 1 percentage point. In 2009, compared to 2008 the distribution have changed so that the age group 15-25 years share fell to 21%, and the share of those from the 56-65 years group fell to 7%.

In terms of percentages of registrations by education level has been a significant change given that the share of those with tertiary education increased from 3.5% in 2007, to 5.4% in 2008 and 7.5% in 2009.

In the year 2007, 16% of total inflows took place in Iasi county which is the county with the largest share. Its share declines to 12% in 2009 but it still remains the maximum share. There are also counties like Bacau, Maramures, Ilfov, Sibiu, Salaj, Satu Mare where there was an increase in the inflow share in 2009.

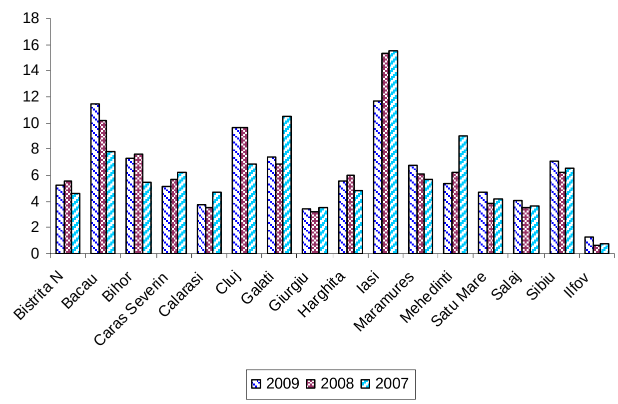

Fig.2 Share of counties registrations per year

The most important variable in our study is the duration of unemployment: the number of months between start date and end date. The average of this variable, computed for the 16 counties is *5.65* months and its standard deviation is *4.40*. The average duration has its maximum in Ilfov County (6.64 months) with a standard deviation equal to *3.95*. The minimum average duration was in Cluj County (4.68 months) with a standard deviation of *4.14*.

Table I Average duration of unemployment by county

County	**Average duration (months)**	**Std. dev.**	**County**	**Average duration (months)**	**Std. dev.**
Bistrita N	4.95	3.28	Harghita	5.57	4.17
Bacau	5.3	3.01	Iasi	6.26	4.94
Bihor	5.31	4.01	Maramures	5.5	3.93
Caras Severin	6.22	4.36	Mehedinti	6.14	5.57
Calarasi	5.37	4.23	Satu Mare	6.01	3.3
Cluj	4.68	4.14	Salaj	5.53	4.15
Galati	6.07	5.43	Sibiu	5.45	3.9
Giurgiu	6.22	5.67	Ilfov	6.64	3.96

The histogram build on the 596 772 completed spells, is represented in the figure below.

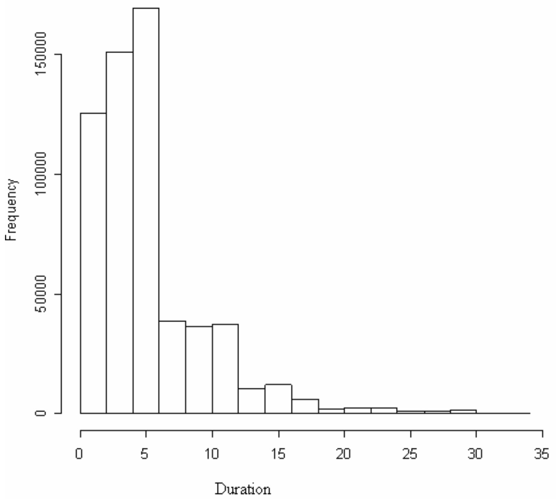

Fig. 3 Histogram of the unemployment duration

The average duration of unemployment for women is 5.8 months and it is larger than the men average duration which is 5.54 months. Unemployed individuals with higher education remain unemployed, on average 4.88 months, less than the unemployed who have completed only high school. Their average unemployment duration is 5.68 months. The persons who have completed only elementary school experience the event in question (finding a job) after 5.81 months.

There are also differences by age group in terms of average duration of unemployment as it can be seen in Table II.

Table II Average durations by age group

Age Group	<=25 years	26-35 years	36- 45 years	46-55 years	56-65 years
Duration (months)	5.13	5.48	5.84	5.96	6.07

The highest value of the unemployment duration was obtained for the 56-65 years age group and the lowest value was that calculated for the age group 15-25 years. As shown in Table I the number of months spent in unemployment increases as people age.

When computing all these average durations we did not use those individuals with no information about the end date. Approximately 90% of those registrations were made in 2009 thus we can conclude that they are still unemployed at the end of our study. In terms of survival analysis, they are still alive, so they might be considered as censored data.

B. Estimation of survival functions –staggered entry

First we have analyzed a staggered entry situation, meaning we have considered as inputs all the registrations in our database, whatever being the registration date.

The Kaplan Meier estimator of the survival function is presented in fig. 4

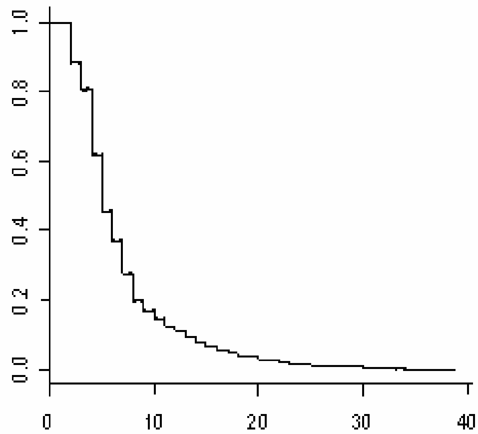

Fig. 4 Survival function for staggered entry

The first column of Table III shows the month of our study, the second one contains the surviving probability for the persons included in the study and the last two columns represent the lower and the upper bound of a 95% confidence interval of this probability.

Table III Surviving probabilities for staggered entry

Month	Survival	Lower 95% CI	Upper 95%CI
1	0.882	0.879	0.885
2	0.805	0.802	0.808
3	0.616	0.613	0.620
4	0.455	0.452	0.458
5	0.371	0.368	0.374
6	0.277	0.274	0.279
7	0.204	0.202	0.205
8	0.172	0.170	0.174
9	0.149	0.148	0.151
10	0.130	0.129	0.132
11	0.117	0.116	0.119
12	0.099	0.098	0.100
13	0.084	0.083	0.085
14	0.073	0.072	0.074
15	0.0617	0.0610	0.0624
16	0.05272	0.05212	0.05333

The probability of survival is 0.882 in the first month of our study period. The 95 percent confidence interval of this value is (0.879, 0.885). In the third month this probability declines to 0.616 and its confidence interval is (0.613, 0.620). After 12 months the survival probability gets under 10%. The probability of survival gets close to zero in the 28 month of the study. In the 31 month the survival rate is 0.0096.

C. Estimation of surviving functions- entry at the same time

The results presented in the following paragraphs are obtained from an analysis developed on a cohort entering unemployment at the same time: first month of our study. Thus we considered only the inflows occurred on 1 January 2007. This means our sample size is 36 746 and we want to test on it if the survival rates are different for different age groups, educational levels, gender and counties.

Considering only these observations we obtained a survival function which shows that the probability of survival is 0.88 in the first month of our study period. The 95 percent confidence interval of this value is (0.85, 0.89). Obvious this confidence interval is larger than in the staggered entry situation. In the third month this probability declines to 0.384 and its confidence interval is (0.379, 0.389). After 6 months the survival probability gets under 10%. The probability of remaining unemployed gets close to zero after 13 months from the beginning of the study.

We wanted to see if the survival function is the same for women and men and Fig.5 shows that there are minor differences with an advantage for men over women.

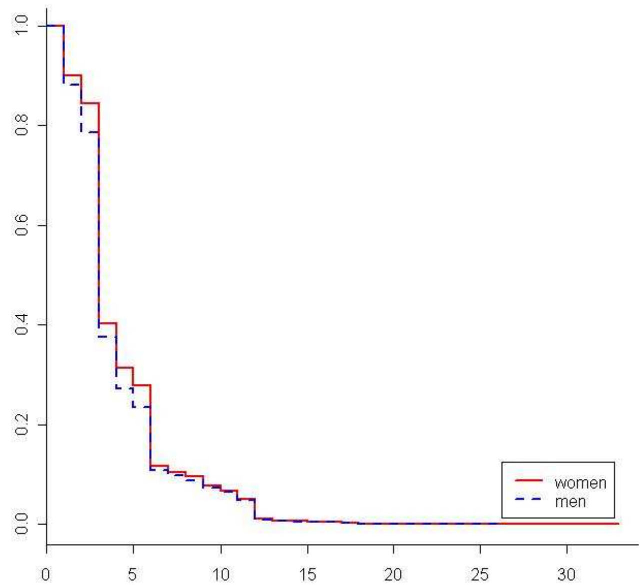

Fig.5 Survival functions per gender

There are significant differences between women and men only in the first six months of the study. As can be seen in table IV, in the 10^{th} month there is very little difference between the survival rate for women which is 0.067 and men's rate which is 0.064.

Table IV Surviving rates on gender

Month	Gender	Surv.	Lower 95% CI	Upper 95% CI
1	Fem.	0.900	0.895	0.905
	Masc.	0.881	0.877	0.885
2	Fem.	0.844	0.838	0.856
	Masc.	0.787	0.781	0.792
3	Fem.	0.402	0.394	0.411
	Masc.	0.375	0.369	0.381
4	Fem.	0.311	0.304	0.320
	Masc.	0.270	0.265	0.276
…	…	…	…	…
10	Fem.	0.067	0.063	0.072
	Masc.	0.064	0.061	0.067

The Log rank test (Table V) indicates that the survival functions are not the same for women group and men group.

Table V Log rank test for survival function on gender

	Observed	Expected	(O-E)^2/E
women	12193	12800	28,80
men	24553	23946	15,40

Chisq= 72.9 on 1 degrees of freedom, p=0

In Fig. 7 are presented the Kaplan Meier estimators of the survival curves on educational levels. We have considered four educational levels: level 1 refers to elementary school, level 2 refers to vocational education, level 3 refers to high school, and level 4 refers to higher education. The figure shows that major differences in the survival curves appear after the forth month of unemployment.

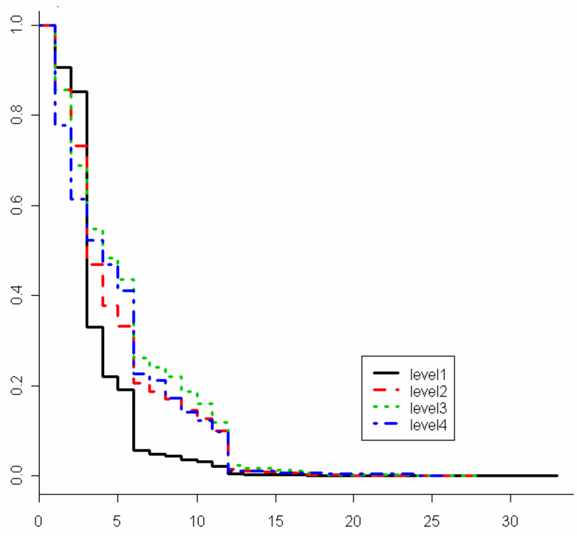

Fig.7 *Survival curves in unemployment on education levels*

Not only the figure but also the Log Rank test (Table VI) suggests that education is a major factor affecting the duration of unemployment (the p-value= 0.0). The individuals within the low level of education group have the highest probability of employment. Those individuals with high school diplomal have the lowest probability of employment. The educational level has no influence on the survival rates after 12 months of unemployment

Table VI Log Rank test for survival functions on education levels

	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Level 1	24830	22654	210.8	965.1
Level 2	7226	8247	126.4	273.9
Level 3	3923	4975	222.6	434.9
Level 4	767	879	14.2	23.8

Chisq=1009 on 3 degrees of freedom, p=0

Also the age influences the probability of employment. This fact is proved by fig. 8 presenting the Kaplan Meier estimators of the survival curves in unemployment by age. Only after 12 months of unemployment the curves coincide. We notice a strong difference among these curves starting with month 3 of unemployment and an inversion of them in month 6. It seems that for the individuals in the age groups 36-45 years and 46-55 years it is the most difficult to find a job and that it is easier for those in the age group 15-25 years.

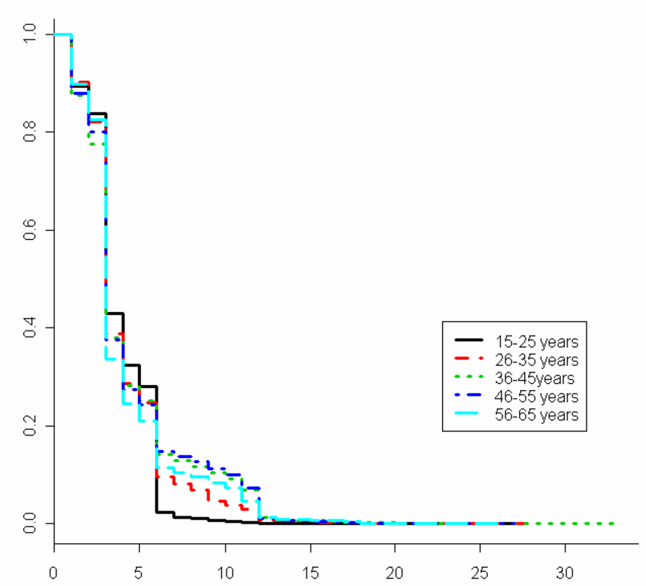

Fig. 8 *Survival curves in unemployment on age groups*

The age influence is also indicated by the log rank test (Table VII). Given that the value of *p* is *1.51e-12*, the null hypothesis has been rejected.

Table VII Log Rank test for survival functions on age groups

	Observed	Expected	(O-E)^2/E	(O-E)^2/V
15- 25 years	5297	5050	12.03	23.29
26-35 years	9618	9370	6.54	14.45
36-45 years	10510	10749	5.31	12.42
46-55 years	8564	8880	11.26	24.62
56-65 years	2757	2696	1.39	2.46

Chisq=61.4 on 4 degrees of freedom, p=1.51e-12

The analysis presented in the next paragraph is developed on 8 counties, one for every development region of Romania. Considering only these 8 counties and only the individuals registered in January 2007 we obtained a survival function which shows that the survival rate in the first month is 0.9 and in the third month is 0.41. The probability of remaining unemployed gets close to zero in the twelve month of the study.

Next, we test if the survival rates are the same for all the eight counties. We plotted in Fig. 6 only four of the eight survival functions. The figure shows that each county has another survival function with a different behavior. For example, in the first months Caras Severin county has the lowest unemployment duration meanwhile in the interval 6-12 has the highest unemployment duration. Thus in Caras Severin county the probability of remaining in unemployment after the first month is under 0.8 whereas in Satu Mare or Mehedinti this probability is close to 0.9. In October 2007 the probability of remaining unemployed gets close to zero in Calarasi county, meantime in Caras Severin is 0.2. After 15 unemployment months the survival curves coincide.

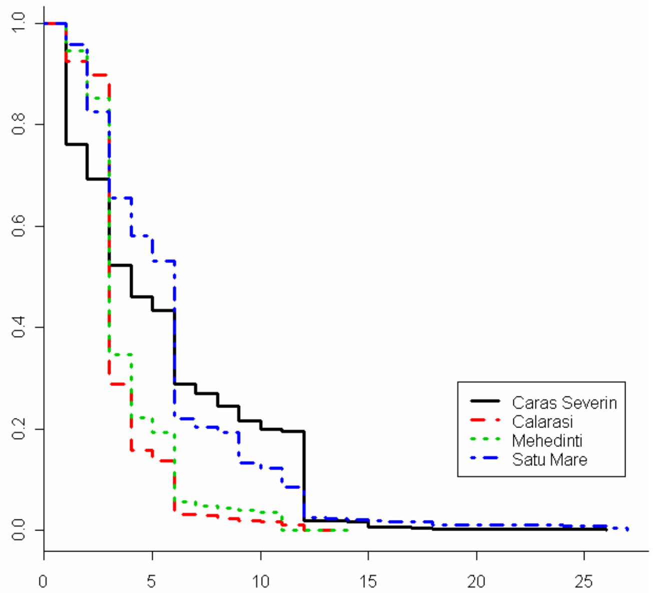

Fig.6 Survival functions per county

The Log rank test indicates that the survival times are indeed different: given that the chi-square statistic is *1160* on *7* degrees of freedom and the value of *p* is *0*, the null hypothesis has been rejected.

We have estimated the survival curves for every age group on the eight counties in order to see if there is a different behavior for every group. In the age groups 56-65 years Satu Mare county seems to have the biggest problem with the employment. For this age groups, low survival rates in unemployment have counties like: Calarasi, Mehedinti, Galati. In the age groups 26-35 years, 36-45 years, and 46-55 years we have the same situation as described in Fig. 6, with higher survival rates in unemployment for Caras Severin county. For the age group 36-45 years we find a significant increase of the survival probability in unemployment for Ilfov County.

Estimating the survival functions for each educational level for all the eight counties we found that for level 1, Satu Mare County has the highest survival rate in unemployment. Mehedinti and Calarasi remain the counties with the lowest survival rates independent of the educational level.

For the other educational levels Caras Severin remains the county with high probability of survival.

Finally we estimated the survival functions separately for all 16 counties and we summarized the results in the Table VIII.

These results demonstrate once again that the counties are very heterogeneous. Thus, there are three counties where the survival rate for the unemployed men is higher than the women rate: Caras Severin, Satu Mare and Ilfov.

Table VIII a Probability of surviving in the 6th month by gender

County	Men	Women
Bacau	0.176	0.188
Bihor	0.084	0.097
Bistrita Nasaud	0.113	0.243
Caras Severin	0.343	0.174
Calarasi	0.029	0.034
Cluj	0.179	0.182
Galati	0.052	0.091
Giurgiu	0.062	0.0569
Harghita	0.165	0.216
Iasi	0.052	0.065
Maramures	0.041	0.098
Mehedinti	0.037	0.064
Satu Mare	0.227	0.218
Salaj	0.068	0.122
Sibiu	0.113	0.168
Ilfov	0.273	0.18

In the previous sections we concluded that the unemployed persons with high levels of education are not advantaged on the labor market. They experience high levels of unemployment duration. The results presented in Table VIII c show that there are two exceptions to this rule: Satu Mare and Maramures. In these counties the individuals with higher education have the lowest survival rate.

Table VIII b Probability of surviving in the 6th month by education level

County	Level 1	Level 2	Level 3	Level 4
Bacau	0.124	0.265	0.345	0.333
Bihor	0.022	0.193	0.437	0.322
Bistrita Nasaud	0.13	0.163	0.224	0.2
Caras Severin	0.109	0.494	0.424	0.529
Calarasi	0.017	0.056	0.1	0.2
Cluj	0.092	0.296	0.325	0.314
Galati	0.038	0.169	0.25	0.161
Giurgiu	0.029	0.097	0.216	0.285
Harghita	0.076	0.351	0.471	0.217
Iasi	0.024	0.103	0.242	0.18
Maramures	0.035	0.082	0.093	0.027
Mehedinti	0.034	0.098	0.113	0.171
Satu Mare	0.249	0.176	0.195	0.142
Salaj	0.034	0.226	0.301	0.157
Sibiu	0.072	0.223	0.32	0.191
Ilfov	0.201	0.238	0.24	0.33

When estimating the survival functions on age groups we saw that the unemployed from the 36-45 years and 46-55 years age groups are facing the biggest problems on the labor market. We can draw the same conclusion from the Table VIII c. But we must mention that in Bihor and Iasi it is even more difficult for those in 56-65 years age group to find a job.

Table VIII c Probability of surviving in the 6th month by age

County	15- 25 years	26-35 years	36- 45 years	46-55 years	56-65 years
Bacau	0.022	0.164	0.239	0.223	0.194
Bihor	0.012	0.068	0.124	0.13	0.165
Bistrita Nasaud	0.008	0.096	0.178	0.242	0.177
Caras Severin	0.326	0.298	0.37	0.315	0.126
Calarasi	0.015	0.019	0.04	0.0501	0.024
Cluj	0.01	0.133	0.207	0.237	0.225
Galati	0.009	0.05	0.063	0.105	0.073
Giurgiu	0	0.055	0.097	0.072	0.063
Harghita	0.229	0.148	0.248	0.264	0.185
Iasi	0.011	0.042	0.0689	0.076	0.088
Maramures	0.12	0.044	0.058	0.083	0.064
Mehedinti	0.012	0.045	0.077	0.059	0.057
Satu Mare	0.13	0.201	0.255	0.258	0.244
Salaj	0.011	0.064	0.117	0.16	0.118
Sibiu	0.06	0.122	0.159	0.161	0.14
Ilfov	0.071	0.229	0.279	0.239	0.125

D. Survival functions and Cox Models

In this paragraph we examine the relationship of the survival distribution to an explanatory variable, using a Cox regression of time to employment on age. The results are presented in the table below.

Table IX Cox model results

Explanatory variable	coef.	exp(coef)	se(coef)	p value
Age	-0.00367	0.996	0.000483	3.1e-14

The results in Table IX show that the age has a statistically significant coefficient. The exponentiated coefficient in the third column is interpreted as multiplicative effect on the hazard. Thus, increasing with one year the age reduces the monthly hazard of employment by 0.9996 on average.

IV. Conclusion

Given the problems in our data, the results of our estimations should be interpreted with caution. They may be considered as a picture of the factors shaping unemployment duration, which may be helping policy makers.

From the results of the survival analysis developed on a database that consists of information about the individuals registered at the employment agency during 2007-2009 in 16 counties of Romania, we may conclude that:

- Unfortunately our analysis, suggests that the unemployed persons with higher education are not advantaged on the labor market; may be it is more difficult for them to re-orientate than it is for the unemployed with low levels of education, and also we must investigate their specialization in order to find out the reasons of their long spells;
- There are not strong differences between women and men, concerning the unemployment duration;
- The unemployed persons from the age groups 36-45 and 46-55 years have a higher probability of remaining unemployed; we think that it may be more difficult for them to adapt to the new labor market conditions;
- Given the fact that the survival curves are different in the analyzed counties, in our future research we must investigate which factors are causing these differences.

We stress that this is a limited analysis since we have studied only some of the Romanian counties so these must be considered as very preliminary results

References

[1] O., J., Blanchard, P., Diamond, "The flow approach to labor markets", *The American Economic review*, Vol.82, No.2,1992 pp 354-359

[2] P. Dalgaard , "*Introductory Statistics with R"*, Springer, 2008, ISBN 978-0-387-79053-4

[3] J.,Fox, "Cox Proportional Hazards Regression for Survival Data, Appendix to an R and S-Plus Companion to Applied Regression", 2002

[4] E. L. Kaplan, P. Meier ,Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, Vol. 53, No 282, 1958, pp 457-481

[5] A. Kavkler, D. E.Danacica, A., G. Babucea, I.Bicanic, B. Bohm, D. Tevdovski, K. Tosevska, D. Borsic,"Cox regression models for unemployment duration in Romania, Austria, Slovenia, Croatia and Macedonia", Romanian *Journal of Economic Forecasting*, 2009

[6] K., Pollock, S., R., Winterstein, C., M., Bunck, P., D., Curtis, Survival analysis in telemetry studies: The staggeres entry design, *The Journal of Wildlife Management*, 1989

[7] Employment in Europe 2009, European Commission, Directorate-General for Employment, Social Affairs and Equal Opportunities, Belgium, 2009

[8] D.R. Cox, "Regression models and life tables", Journal of the Royal Statistical Society. Series B (Methodological), Volume 34, Issue 2 , pp.187-220, 1972

[9] G. M. Baker,., P. K. Trivedi, „*Estimation of unemployment duration from Groued Data: A Comparative Stud"y*, Journal of Labor Economics, vol. 3, no 2, 1985

[10] S. Davis, J. Haltiwanger, '*Gross Job Creation, Gross Job Destruction and Employment Reallocation*', Quarterly Journal of Economics 107 (3), 819-863, 1992

[11] V. Bewick, L. Cheek, J. ,Ball, "Statistics review 12: Survival Analysis", Criticial Care 8, 389-394, 2004, available online **http://ccforum.com/content/8/5/389**

[12] Jiang H, Fine JP., Survival Analysis, Methods Mol Biol., 2007, 404: 303-18

[13] P. Klein, M. L. Moeschberger, Survival Analysis: Techniques for Censored and Truncated Data, New York , Springer, 2003

[14] D. Diez., "Survival analysis in R", available online, http://www.stat.ucdavis.edu/~hiwang/teaching/10fall/R_tutorial%201.pdf

[15] V. Ciuca., M. Matei , „Survival analysis for the unemployment duration", Proceedings of the 5th WSEAS International Conference on Economy and Management Transformation (EMT'10), West University of Timişoara, Romania, October 24-26, 2010, Volume I, ISBN 1792-5983, pp. 354-360

[16] I. Vasadan., „The relevance of the unemployment rate for Romania's labor market", Proceedings of the 5th WSEAS International Conference on Economy and Management Transformation (EMT'10), West University of Timişoara, Romania, October 24-26, 2010, Volume I, ISBN 1792-5983, pp.423-428

[17] A. Alexandru., I. Dobre , C. Ghinararu, "Revisiting the relationship between unemployment rate and the size of the shadow economy for United States using JOHANSEN approach for cointegration" Proceedings of 11 th WSEAS International Conference on Mathematics and Computers in Business and Economics, pp.199-204, 2010, ISSN: 1790-2769.

[18] A. Alexandru., I. Dobre , C. Ghinararu." The relationship between unemployment rate and the size of the shadow economy. The case of

United States", Wseas Transactions on Business and Economics, issue 4, vol.7, pp.359-369, october 2010, ISSN: 1109-9526.

[19] A. Alexandru., I. Dobre , C. Ghinararu "The relationship between shadow economy and unemployment rate: a SVAR approach" Proceedings of 5th WSEAS International Conference on Economy and Management Transformation, pp.242-249, 2010, ISSN: 1792-5983.