Quantitative Methods in Economics
Vol. XI, No. 1, 2010, pp. 212–222

A DYNAMIC APPROACH TO THE STUDY OF UNEMPLOYMENT DURATION

Joanna Landmesser

Chair of Econometrics and Statistics, Warsaw University of Life Sciences
Nowoursynowska 159, 02-776 Warszawa, Poland
e-mail: joanna_landmesser@sggw.pl

Abstract: In this research work we investigate which factors influence the probability of leaving the unemployment state among people registered in the District Labor Office in Słupsk. The multiepisode hazard models with time-varying variables are suitable tools for this analysis. We introduced the changing labor market structure into the risk model. The main results achieved show that the job finding process depends on the historical time of the entry into the unemployment state and the actual historical time.
Also, the specific individual characteristics of people unemployed, such as gender, age, marital status, place of residence, education level, influence the probability of exiting the unemployment state. There is a greater tendency to leave the unemployment state when the person doesn't receive the unemployment benefit. The participation in the vocational training doesn't increase the transition rate into employment.

Key words: unemployment duration, hazard models, factor analysis

Introduction

Unemployment is a problem frequently connected with specific regions. In our research work we try to analyze in detail the situation in district Słupsk in north Poland (voivodeship Pomorskie).

The widely used unemployment measure – the unemployment rate – presents the proportion of the unemployed workforce which is seeking employment. However, such a ratio masks the dynamic nature of the labor market by failing to cover the length of time individuals are unemployed. It will be more useful to understand how the probability of exiting unemployment varies with demographic and economic characteristics. Therefore, to analyze the duration of unemployment among residents of district Słupsk, we use econometric risk models.

The main goal of our study is to introduce the changing labor market structure into the risk model in order to treat the time-dependent nature of the unemployment duration process in an adequate manner. There are two main ways in which the changing labor market affects finding job opportunities. First, people start their unemployment episodes in different labor market contexts (cohort effect, see also [B], [BGR]), and second, the labor market structure influences the opportunities of all people unemployed within the labor market at every moment (period effect).

We attempt to find measures for the macro effects on the labor market and suggest using 12 time series from official statistics, indicating the development of the labor market structure in Słupsk region. But such variables often measure similar features of a process and are highly correlated, which implies an identification problem. When we choose only uncorrelated series, it may only capture specific features of the labor market development. If time series represent aspects of an underlying regularity, it is more appropriate to look for these latent dimensions. To do this, we apply the statistical method – the exploratory factor analysis. The application of this method allows to isolate the basic principal factors from the given set of variables describing the development of the labor market in Słupsk region. The factors obtained are introduced into the multiepisode hazard model as explanatory variables.

The second goal of our analysis is to estimate the effect of the unemployment compensation system on the individual's unemployment duration. Duration models with time-varying covariates serve as proper tools for the analysis of the influence of the unemployment benefit received at the risk of leaving unemployment. We would also like to look at the means being used by the labor office to combat unemployment. Vocational activation of people belonging to risk groups in the local job market is very important for the labor office. Therefore, additionally, we intend to investigate the impact of vocational training on the unemployment duration.

Our research work is based on the data obtained from the District Labor Office in Słupsk from 1999 to 2007.

Description of the analysis method

Hazard models

The dependent variable we are interested in is the duration of time an individual spends in the state of being unemployed. Empirical data for the duration variable can take only positive values the negative duration periods do not exist. Moreover, duration of the phenomenon can be observed only temporarily (censoring problem). All this makes it impossible to apply traditional models of regression. An appropriate approach, which considers right censoring of unemployment spells, and which controls for observable personal characteristics of individuals that influence the unemployment duration, is the application of hazard models.

In case of continuous hazard models the duration variable T is a continuous non-negative random variable describing duration in any state, where t is realization. The distribution function of T is denoted F and is defined as $F(t) = Pr[T \leq t]$. The density function of the duration variable T is $f(t) = dF(t)/dt$. The probability of survival to t is given by the survivor function $S(t)$:

$$S(t) = Pr[T > t] = 1 - F(t)$$

The survival function $S(t)$, gives the probability of the surviving of the process over a certain moment t. The survival function lies between zero and one; it is equal to one at the beginning of the spell ($t = 0$); and its slope is non-positive.

Hazard models are concerned with observation of the instantaneous rate of leaving a certain state (e.g. unemployment) per unit time period at t:

$$h(t) = \frac{f(t)}{S(t)} = \lim_{dt \to 0} \frac{Pr[t \leq T < t + dt | T \geq t]}{dt}$$

The hazard function $h(t)$ is the limit of probability that the spell is completed during the interval $[t, t + dt]$, given that it has not been completed before time t, for $dt \to 0$. The hazard rates describe the intensity of transition from one state to another. A higher value of hazard function means that the transition from state A to state B follows faster.

For general surveys of hazard rate models – also called survival models or duration models – see, e.g. [KP], [CO], [Ki], [HL], [CT]. The first intensive application of duration models is the analysis of individual unemployment duration data by Lancaster [L]. A forecasting duration period of the unemployed people, over which they are without jobs, until taking up a job is a typical example of the use of hazard models (see, for example, [DK], [L]).

As far as the use of hazard models for unemployment duration in Poland is concerned, literature is modest. The first hazard models to analyze the time spent in the unemployment state have been applied by Frątczak, Jóźwiak, Paszek [FJP] and later by Malarska [M]. The other researchers performed microeconometric analysis of job market based on the logit or probit model. However, such a methodology fails to cover the individual duration of unemployment period. These researchers concentrated only on the detecting of determinants, which influence the probability of finding jobs.

Hazard models are constructed whenever there is the purpose of forecasting the moment, in which a certain event will occur. These models differ in assumptions concerning distribution of individual time in which the event T occurs. Among duration models, which allow to estimate the influence of different determinants, the following parametric hazard models can be noted: proportional hazard models (PH) and accelerated failure-time models (AFT).

In the PH models, the conditional hazard rate $h(t|X)$ can be factored into separate functions: $h(t|X) = h_0(t)\, g_0(X\beta) = h_0(t)\, exp(X\beta)$, where $h_0(t)$ is called the baseline hazard and $exp(X\beta)$ is a function of explanatory variables vector X. The characteristics of hazard function change proportionally to the influence of explanatory variables. This category of models comprises the whole range of models which show differences when it comes to assumptions concerning distribution of baseline hazard.

The most widely applied semiparametric method of analyzing the effect of covariates on the hazard rate is the proportional hazard model proposed by Cox [C]. The Cox model states that the hazard rate for the j-th subject in the data is $h(t|X_j) = h_0(t)\, exp(X_j\beta)$. Compared with the parametric approaches, the advantage of the semiparametric Cox model is that we have no need to make assumptions about baseline hazard; $h_0(t)$ is left unestimated. This model is particularly attractive when the researcher has only a weak theory supporting a specific parametric model and is only interested in the magnitude and direction of the effects of observed covariates.

Factor Analysis

Explanatory factor analysis (EFA) is a statistical technique for data reduction. The term factor analysis was first introduced by Thurstone [T]. EFA is used to uncover the latent structure of a set of variables. The main aim of this method is to get a small set of variables from a large set of variables (most of which are correlated to each other). The variability among observed variables will be described in terms of fewer unobserved variables called factors. All factors will be orthogonal to one another, meaning that they will be uncorrelated.

In the EFA the observed variables $X_i, i = 1, ..., p$ are modeled as linear combinations of the common factors $F_j, j = 1, ..., k\,(k < p)$, plus the unique factors u_i ("error" terms):

$$
\begin{aligned}
X_1 &= a_{11}F_1 + a_{12}F_2 + ... + a_{1k}F_k + u_1 \\
X_2 &= a_{21}F_1 + a_{22}F_2 + ... + a_{2k}F_k + u_2 \\
&\dots \\
X_p &= a_{p1}F_1 + a_{p2}F_2 + ... + a_{pk}F_k + u_p
\end{aligned}
$$

where a_{ij} is the linear coefficient called the factor loading. Everything except the left-hand-side variables is to be estimated, which implies that the model has an infinite number of solutions.

EFA assumes that the variance in the measured variables can be decomposed into that accounted for by common factors and that accounted for by unique factors. EFA estimates how much of the variability is due to common factors ("communality").

We use EFA when we are interested in making statements about the factors that are responsible for a set of observed variables and when the goal of the analysis is to detect structure.

Advantages of EFA are the following: the reduction of the number of variables by combining two or more variables into a single factor and the identification of groups of inter-related variables. However, there is a disadvantage, too. Interpreting factor analysis is based on using a "heuristic". More than one interpretation of the same data can be made.

Subject of the research

In our analysis we use data taken from the District Labor Office in Słupsk in Poland concerning registered unemployed people in the period from 1999 to 2007.

Our selected sample consists of 1690 persons, who were registered unemployed in the labor office during the survey time at least for one day. They are residents of district Słupsk and the city with district of Słupsk status.

On the basis of the registered history of events in the labor office Puls computer system we can find out for how long was a person looking for a job every time or for how long unemployed is actually looking for a job (in days). The time spent in the unemployment state is called an episode. The episode finishes when the event occurs (finding a job etc.). The duration of a single episode is marked by the neighboring days, during which a given person has been in a given state. Unemployment spells are completed if they end up with transition from unemployment state. Otherwise, unemployment spells are treated as right censored. The last observed exit from unemployment was noted at the 3235 nd day. While our data basis contains multiple spells for 1690 persons we have got 3614 episodes.

Method of episode splitting

For survival data, however, individuals may be observed at several stages during an episode. and relevant time-varying regressors may take different values over an episode. Time-dependent covariates can be included in parametric and semiparametric transition rate models by applying the method of episode splitting. The idea of this method can be described as following: time-dependent qualitative covariates change their values only at discrete points in time. At all points in time, when at least one of the covariates changes its value, the original episode is split into subepisodes (splits).

The Table 1 shows, for example, that the individual with "id"=19 has had two unemployment episodes (numbers 26 and 27). The variable "des" serves as the censoring indicator ("des=0" for right censoring). The episode 26, for example, is divided into two subepisodes: the first one with starting time 2004-06-01 and the second one with tstart=2005-06-02. The symbol "R B" marks the registration of unemployed received unemployment benefit, "T U" means the change in the type of unemployment, "N PU" – new type: the unemployed without unemployment benefit, "W PC" means finding a job. The episode 27 starts with the registration of unemployed, who receives no unemployment benefit; contains three subepisodes; the second one concerns the vocational training.

Table 1. Records of data after episode splitting for one individual.

id	newid	des	type	tstart	type	tfin	tf	t1	benefit	training
19	26	0	R B	2004-06-01	T U	2005-06-01	365	365	1	0
19	26	1	N PU	2005-06-02	W PC	2005-06-22	20	385	0	0
19	27	0	R P	2006-04-12		2006-04-30	28	28	0	0
19	27	0	Z X	2006-05-01	O O	2006-05-15	14	42	0	1
19	27	1		2006-05-16	W PC	2006-05-26	10	52	0	0

For the whole data set with 3614 episodes we created 6751 subepisodes, 3445 of them were right censored.

Covariates in hazard models

Estimated hazard models will not only comprise present duration as a determinant for the probability of leaving the state of unemployment, but also other observable characteristics of individuals such as gender, age, marital status, place of residence, education level. The Table 2 defines the first part of covariates used to explain the joblessness duration with hazard models.

Table 2. Definitions of Variables.

Variable	Description	Change in time
sex	1 if individual is male	no
age	age of the individual at the beginning of a subepisode in years	yes
edu1	1 if individual has incomplete primary, primary, lower secondary or basic vocational education level	no
edu2	1 if individual has general secondary, vocational secondary or post-secondary education level	no
edu3	1 if individual has tertiary education level	no
marr	1 if individual is married	no
town	1 if the place of residence is town	no
language	1 if individual declares any foreign language skills	no
benefit	1 if individual receives unemployment benefit	yes
training	1 if individual takes part in the vocational training	yes

Below, we present the results of the empirical calculation in which the method of factor analysis was applied. The application of this method to the 12 time series from official statistics (Regional Data Bank) allowed to isolate the basic factors describing the development of labor market in Słupsk region in the time period 1999-2007. The suggested measures for the macro effects are shown in the Table 3.

Table 3. Suggested measures for the macro effects in Słupsk region.

Variable	Description	Data source
workpop	the proportion of the population at working age in % of total population	district słupski and city with district of Słupsk status
service	the proportion of the employed in the service sector in % of all persons employed	district słupski and city with district of Słupsk status
industry	the proportion of the employed in the industrial sector in % of all persons employed	district słupski and city with district of Słupsk status
offers	job offers per 1000 unemployed registered	district słupski and city with district of Słupsk status
entities	entities of the national economy recorded in the REGON register per 1000 of the population	district słupski and city with district of Słupsk status
dwellings	dwellings completed per 1000 of the population	district słupski and city with district of Słupsk status
budgreven	revenue of districts and cities with district status budgets per 1 inhabitant	district słupski and city with district of Słupsk status
unempl	the unemployment rate registered	subregion Słupsk
wages	average monthly gross wages and salaries	voivodeship Pomorskie
gdp	gross domestic product per capita	voivodeship Pomorskie
invest	investment outlays per capita	voivodeship Pomorskie
rd	expenditures on R&D per 1 inhabitant	voivodeship Pomorskie

We perform the factor analysis with principal factoring and the Equimax rotation. The variance accounted for by factors is summarized in the Table 4. In the second column (Eigenvalue), we find the variance on the new factors that were extracted. In the third column, these values are expressed as a percentage the total variance. The next column contains the cumulative variance extracted.

The Table 4 attempts to determine the number of orthogonal factors to be retained for further analysis. The Kaiser criterion for determining the number of factors, is the "eigenvalue greater than 1" criterion [Ka]. This entails that, unless a factor extracts at least as much as the equivalent of one original variable, we drop it. Since the first two factors were the only ones that had eigenvalues greater than 1, the final factor solution will represent only 87.649% of the variance in the 12 time series.

Table 4. Extraction of factors.

	Eigenvalues			Sum of square of loadings after extraction			Sum of square of loadings after rotation		
Factors	Eigen-value	% of variance	Cum. % of variance	Eigen-value	% of variance	Cum. % of variance	Eigen-value	% of variance	Cum. % of variance
1	6.838	56.984	56.984	6.838	56.984	56.984	6.648	55.397	55.397
2	3.680	30.665	87.649	3.680	30.665	87.649	3.870	32.251	87.649
3	0.736	6.133	93.781						
4	0.496	4.133	97.914						
5	0.141	1.179	99.093						
6	0.080	0.668	99.761						
7	0.022	0.186	99.947						
8	0.006	0.053	100.000						
9	0.000	0.000	100.000						
10	0.000	0.000	100.000						
11	0.000	0.000	100.000						
12	0.000	0.000	100.000						

A graphical method for determining the number of factors is the scree test. We can plot the eigenvalues shown above in a simple line plot (Figure 1).

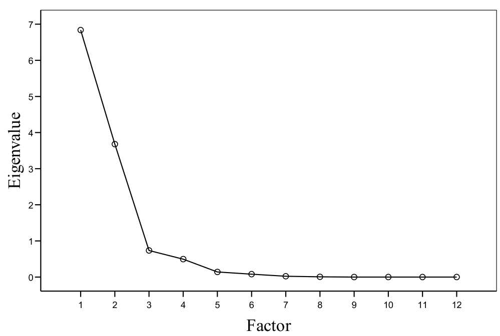

Figure 1. Plot of eigenvalues.

We have to find the place where the smooth decrease of eigenvalues appears to level off to the right of the plot. According to this criterion, we would retain two factors (principal components).

The Table 5 shows the loadings representing a correlation between the items and the factors. The loadings are distributed between the two factors. In order to obtain interpretable results the basic solution was rotated (we used Equimax rotation).

Table 5. Factor loading matrix after Equimax rotation.

	Factors	
Variables	1	2
gdp	**0.971**	0.234
rd	**0.959**	0.107
budgreven	**0.943**	0.278
invest	**0.927**	-0.329
wages	**0.926**	0.294
offers	**0.924**	0.088
unempl	**-0.875**	0.429
service	0.087	**0.968**
entities	0.285	**0.881**
workpop	0.566	**0.798**
industry	0.387	**-0.779**
dwellings	0.003	**0.619**

Table 6. Table of factor scores.

	Factors	
Variables	1	2
gdp	0.142	0.033
rd	0.144	0.000
budgreven	0.137	0.045
invest	0.152	-0.115
wages	0.134	0.050
offers	0.139	-0.004
unempl	-0.147	0.139
service	-0.016	0.253
entities	0.018	0.224
workpop	0.063	0.194
industry	0.083	-0.217
dwellings	-0.018	0.163

The Table 6 shows the factor scores for each row of the data file; we use these to obtain the values for each factor.

The first factor could be interpreted as representing the changing "level of economy development" in the whole region (voivodeship) and, the second one as a measure of changes in the "local economic activity".

As can be seen from the plots of the scores of the two factors in Figure 2, the first factor shows a trend with an increasing slope, while the "local economic activity" factor shows contrary development with downturn around 2004.

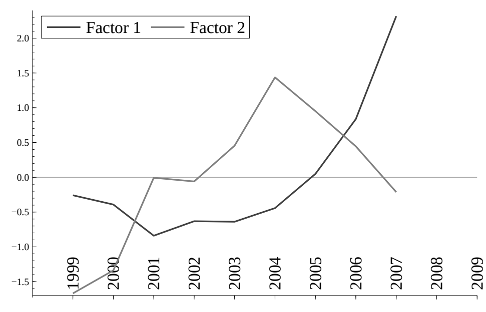

Figure 2. Plots of factor values.

Because the factors obtained are orthogonally constructed, it is possible to introduce both measures into the multiepisode hazard model as time-varying explanatory variables.

To represent the changing conditions under which cohorts enter the unemployment state, we use the factor values at the day, on which the person was registered in the labor office (variables fac1rd and fac2rd). To introduce the period effects, we use the earlier splitted episodes so that the factor values are updated in each job subepisode (variables fac1 and fac2).

Results of models estimation

Below, you can find estimated Cox proportional hazards model describing the unemployment duration. The results of the estimation of cohort and period effects and the effects of the other variables on the probability of leaving the unemployment state in the Cox model are shown in the Table 7.

Table 7. Results of Cox PH models estimation for the risk of leaving the unemployment state.

	Cox regression		
Variable	Coef.		Haz. Ratio
sex	0.490	***	1.632
age	-0.009	***	0.991
edu1	-0.194	***	0.824
edu2	-0.096		0.909
marr	0.169	***	1.184
town	0.136	***	1.146
language	0.126	***	1.134
benefit	-6.504	***	0.001
training	-0.463	**	0.629
fac1rd	-0.646	***	0.524
fac2rd	-0.504	***	0.604
fac1	0.788	***	2.199
fac2	0.566	***	1.760
	Log likelihood = -22,913.006		
	LR chi2(13) = 2,412.20		
	Prob > chi2 = 0.0000		

***,**,* - significant at 1%, 5%, 10% level respectively.
Source: own computations using Stata Statistical Software.

The semiparametric Cox model was estimated by the partial likelihood method. By interpreting the results of models parameters it can be stated that:

- the hazard of leaving the unemployment state in the case of a man is 63.2% greater than in the case of a woman,
- the age coefficient implies that older people are at a disadvantage; the one-year-older age of the individual at the beginning of a jobless subepisode leads to 0.9% decrease of chance for exiting unemployment,
- the primary, lower secondary or basic vocational education levels, in comparison with the tertiary education level, lead to the significant decrease of opportunities to break unemployment,
- the chance of leaving unemployment is greater in urban than in rural areas,
- the hazard of breaking unemployment is greater in the case of people with any foreign language skills (by 13.4%),

- there persists a lower tendency to leave the unemployment state if the person registered receives the unemployment benefit (hazard rate decreases by 99.9%),
- the participation in the vocational training significantly decreases the transition rate from unemployment.

Now we concentrate on the interpretation of the effects of changes in markets development. The parameters, which we are interested in, are statistically significant.

The "level of economy development" at the time of entry into the unemployment state has a negative effect on the chance of finding a job. The higher the "level of economy development", the smaller the attractiveness of the people newly registered in the labor office. It is less likely that these people will be further moved from unemployment. It is also true for the negative effect of the "local economic activity" level at the entry into jobless state.

Conversely, the period effect of the "economy development" is positive. The continuous developing of the regional economy leads to the increasing opportunities for all people to move up to the job. The same is true for the period effect of the "local economic activity" level; the better this level is, the more current opportunities for unemployed people to find a job.

Conclusions

Using the multiepisode hazard models with time-varying variables, we investigated which factors influence the probability of leaving the unemployment state among the people registered in the District Labor Office in Słupsk.

We introduced the changing labor market structure into the risk model in order to treat the time-dependent nature of the unemployment duration process in an adequate manner. The main results achieved show that the job finding process depends on the historical time of the entry into the unemployment state and the actual time. We found that the "level of economy development" in the whole region and the "local economic activity" level at the time of entry into the unemployment state have a negative effect on the chance of finding a job, while the period effects are positive.

Having estimated the Cox PH model for the risk of leaving the unemployment state, we can take into account three future scenarios for the regional and local economy development. The optimistic scenario A assumes a continuous development of the regional and local economy, the worst-case scenario B reflects the decline in development at the regional and local level and the base scenario C assumes a continuation of existing tendencies. The continuous development (scenario A) leads to the increasing opportunities for all registered unemployed to move up to the job. The reduction of development (scenario B) leads to fewer opportunities to find a job. In scenario C, the effects of increasing development at the voivodeship level and decline in local economic activity may be reduced.

Apart from that, the specific individual characteristics of people unemployed, such as gender, age, marital status, place of residence, education level, influence the probability of exiting the unemployment state.

The examination of the impact of the unemployment benefit received exhibits, that there is a greater tendency to leave the unemployment state when the person doesn't receive the unemployment benefit. What is more, the actual participation in the vocational training doesn't increase the transition rate into employment.

Acknowledgments

This work was supported by grant No. N N111 209436 from Polish Ministry of Science and Higher Education.

References

[B] H. P. Blossfeld, Career opportunities in the Federal Republic of Germany: A dynamic approach to the study of life-course, cohort, and period effects, European Sociological Review 2 (1986), 208–225.

[BGR] H. P. Blossfeld, K. Golsch K. and G. Rohwer, Event History Analysis with Stata, Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey, 2007.

[CT] A. C. Cameron and P.K. Trivedi, Microeconometrics: Methods and Applications, Cambridge University Press, New York, 2005.

[C] D. R. Cox, Regression Models and Life Tables (with Discussion), Journal of the Royal Statistical Society, Series B 34 (1972).

[CO] D. R. Cox and D. Oakes, Analysis of Survival Data, Chapman and Hall, London, 1972.

[DK] D. Devine and N. Kiefer, Empirical Labor Economics, New York, Oxford, 1991.

[FJP] E. Frątczak, J. Jóźwiak and B. Paszek, Zastosowanie analizy historii zdarzeń w demografii, Szkoła Główna Handlowa, Warszawa, 1996.

[HL] D. Hosmer and S. Lemeshow, Applied Survival Analysis: Regression Modeling of Time to Event Data, John Wiley and Sons, New York, 1999.

[Ka] H. F. Kaiser, The application of electronic computers to factor analysis, Educational and Psychological Measurement 20 (1960), 141–151.

[KP] J. Kalbfleisch and R. Prentice, The Statistical Analysis of Failure Time Data, John Wiley and Sons, New York, 1980.

[Ki] N. Kiefer, Economic Duration Data and Hazard Functions, Journal of Economic Literature 26/2 (1988), 646–679

[L] T. Lancaster, Econometric Methods for the Duration of Unemployment, Econometrica 47/4 (1979), 939–956 .

[M] A. Malarska, Diagnozowanie determinantów bezrobocia w Polsce nieklasycznymi metodami statystycznymi, Wydawnictwo Biblioteka, Łódź, 2007.

[T] L. L. Thurstone, Multiple factor analysis, Psychological Review 38 (1931), 406–427.