A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 286, 2013

Beata Bieszk-Stolorz[*], *Iwona Markowicz*[**]

INFLUENCE OF UNEMPLOYED PERSON AGE ON THE HAZARD AND ITS CHANGES IN TIME

Abstract. The aim of the article is to analyse the influence of unemployed person's age on job seeking time. This influence can changes in time. That analysis is possible by using Cox non-proportional hazard model. We get the marks of constant in time parameters by using proportional hazard model. Hazard ratio for two units with different value of analyzed feature is going to be averaged for all research period. The constancy of hazard ratios can be examined with graphic methods or by researching significance of parameters of Cox non-proportional hazard model.

Key words: Cox proportional hazard model, Cox non-proportional hazard model, hazard ratios, unemployment.

I. INTRODUCTION

In the analysis of unemployment, the initiating incident is registration of the person who is seeking a job in a Local Labour Office and the ending incident is signing this person off, and the time between these incidents is a random variable. The aim of the article is to analyze the influence of unemployed person's age on the job seeking time and the analysis of changes of this influence in time[1]. The Cox regression models of proportional and non-proportional hazard were used as the research tools. The first of them assumes stability of hazard ratio in time, but the second one the absence of this stability. Cox proportional hazard model can be used, if the influence of explanatory variable X on hazard does not depend on time. However, in practice this influence often changes in time. In that situation we have to deal with the extended Cox model (Kleinbaum, Klein, 2005, p. 95), it means with non-proportional hazard model. Variable X_i affects hazard changes in time and changes in time of hazard ratios (Schemper, 1992). Choosing the right model could be a problem. Making the test of hazard proportionality can help in making the decision but we also have to consider the aim of research and interpretation of results. To examine if influence of analyzing explanatory

[*] Ph.D. , Department of Econometrics and Statistics, University of Szczecin.
[**] Ph.D. , Department of Econometrics and Statistics, University of Szczecin.
[1] Stage the research project (2010-2012, MNiSW, N N111 273538).

variable on hazard changes in time we have to use graphic and analytical methods. They were used in this article.

Researches were carried out on the basis of individual data taken from Local Labour Office in Szczecin (resources of computer system Syriusz). Data of persons who signed off in 2009 were used in empiric example (19,398 persons). The reasons for signing off were different, but because of the range of research, we focused on those related with finding a job. Unemployed persons were grouped by age: 18–24, 25–34, 35–44, 45–54, 55–59, 60–64.

II. MODEL OF COX PROPORTIONAL HAZARD REGRESSION

Analysis of duration of the phenomenon can be made with the use of the Cox regression model. It can be a proportional hazard model and then it is assumed that the influence of individual endogenous variables on hazard is constant in time. We can talk then about proportional hazard. Exactly, the hazard ratio for two unemployed persons with different variant of given feature is constant in time. In this model it is assumed that the hazard function is the function of independent variables and it can be written in that way (Bednarski, 2005):

$$h(t, X_1, ..., X_n) = h_0(t)\exp\left(\sum_{i=1}^{n}\beta_i X_i\right) \tag{1}$$

where: X_1, ..., X_n – independent variables, $h_0(t)$ – reference hazard or hazard`s line zero, β_1, ..., β_n – model coefficients, t – time of observation.

This is the semi-parametric model, because function of hazard is a product of parametric unspecified base hazard function and explanatory variable function for which the parameters are estimated (by maximizing the partial likelihood function (Cox, 1972, pp. 187-220; Cox, 1975, pp. 269-276; Cox, Oakes, 1984). Base hazard $h_0(t)$ is a hazard function when all explanatory variables are zero. The second part of the model is exponential function with variables affecting the duration of the phenomena (factors of risk and chance) and parameters defining so-called relative risk (chance). The method used for defining the variables (0–1 for transformed dichotomous variables), allows to designate n – 1 parameters for n variables (avoid collinearity). One, chosen variable becomes a point of reference. In the conducted analysis as a reference group were taken unemployed persons in age group of 18–24. Value $\exp(\beta_i)$ is interpreted as a relative (hazard ratio – HR) risk (or chance). The results of estimated parameters of Cox proportional hazard model are shown in Table 1. The age of unemployed persons was divided in to 6 sections, making in this way 6 endogenous variables (W_1–W_6).

Table 1. The results of the estimation of Cox proportional hazard model for age groups of unemployed persons registered out of PUP in Szczecin in 2009

Variable	Estimated coefficient	Standard error	Wald statistic	*p*-value	Hazard ratio (*HR*)
W_2	0.1253	0.0342	13.4509	0.0003	1.1335
W_3	–0.0115	0.0409	0.0791	0.7785	0.9886
W_4	–0.2036	0.0403	25.5292	0.0000	0.8158
W_5	–0.7026	0.0593	140.2092	0.0000	0.4953
W_6	–1.6849	0.1677	100.9869	0.0000	0.1855

Source: own study.

Hypothesis that the model parameters are insignificant (H_0: $\beta_i = 0$) was verified with the test of likelihood ratio. As alternative hypothesis says, at least one explanatory variable is related with explaining variable. Statistic has $\chi 2$ distribution, and received value $p = 0{,}0000$ shows on rejecting the null hypothesis. Significance test of individual parameters is based on Wald`s statistic. The youngest unemployed persons, who are aged 25 or below, are the reference group. Hazard ratio greater than one was received only for variable W_2, which shows that persons aged 25–34 have more chances of finding a job in a short period of time. Hazard ratios for other age groups are smaller, which means that when the age is higher, the chance of finding a job by unemployed persons is smaller. Significance of all these estimated parameters was confirmed except one. Value p is higher than accepted 0,05 for variable W_3. It means that although there is an absence of significance, the estimated parameters are important. It has to be underlined, that received hazard ratios were accepted at the same stage of research as time-independent. However, it can turn out that the hazard ratio is not constant in time and its value changes in some points or time periods. Then, more precise results could be received using Cox non-proportional hazard model. Verification of hazard stability in time can be made with graphic or analytical methods.

III. GRAPHIC METHOD OF RESEARCHING PROPORTIONALITY OF HAZARD

There are a few graphic methods of researching proportionality of hazard (Bieszk-Stolorz, Markowicz, 2012). One of them was used in the study (generally used by researchers). It relies on the examination if transformed survival curves (designated with for example Kaplan-Meier`s estimator) for individual categories of researching independent variable are parallel to each other. To value „parallelism" the value of survival function $S(t,X)$ should be transformed into form[2] $-\ln(-\ln S(t,X))$. If the assumption of proportionality is satisfied, on the graph with horizontal axis t and vertical axis $-\ln(-\ln S(t,X))$ survival curves describing appropriate categories should be equally distant from each other[3]. This „parallelism" provides that the influence of analyzed explanatory variable on hazard does not depend on time and designated hazard ratio in every time unit is the same.

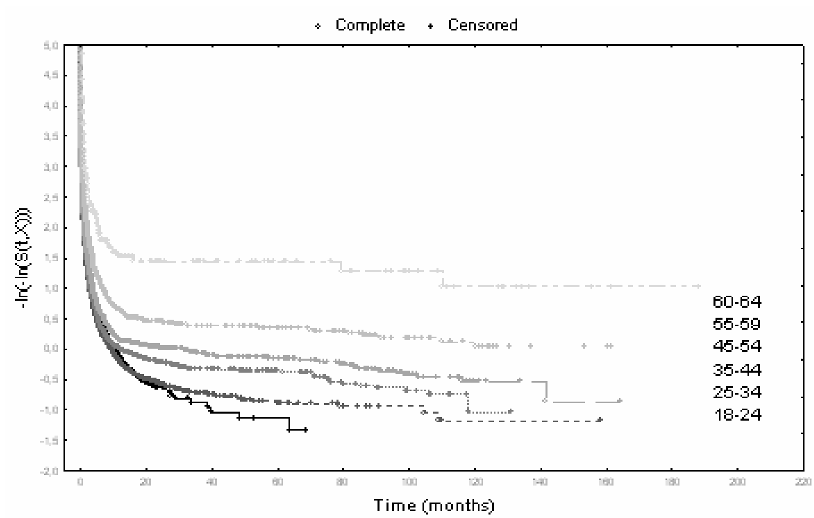

Fig. 1. Transformed Kaplan-Meier`s estimators for age groups of unemployed persons registered out from PUP in Szczecin in 2009.

Source: own study.

[2] This transformation was first proposed by Kalbfleisch, Prentice (1980).

[3] This distance is equal to the absolute value of the difference of ordinates for a given abscissa.

By analyzing transformed Kaplan-Meier`s estimators in Figure 1 we notice that not all of them are parallel to each other. So there is a conjecture about absence of hazard proportionality in case of categorized variable *Age*.

IV. COX NON-PROPORTIONAL HAZARD MODELS

In addition to graphic methods there are various types of analytical testing of non-proportionality in Cox model[4]. To research the influence of explanatory variables X_i on hazard depending on time we can use hazard model in the following form:

$$h(t, X_1, ..., X_n) = h_0(t)\exp\left(\sum_{i=1}^{n}\beta_i X_i + \sum_{i=1}^{n}\delta_i X_i \times g_i(t)\right). \quad (2)$$

In case of significance of parameter δ_i (we reject hypothesis $H_0 : \delta_i = 0$) the influence of variable X_i on hazard changes in time. Function of hazard ratio is not constant. Function g occurring in formula (2) can have various forms. If variable X_i is dichotomous and takes the value 0 and 1, the function g and functions of hazard ratio have the form:

$$g(t) = t\,,\ HR(t) = \exp(\beta_i + \delta_i t) \quad (3)$$

$$g(t) = \ln t\,,\ HR(t) = \exp(\beta_i + \delta_i \ln t) \quad (4)$$

$$g(t) = \begin{cases} 0 & \text{dla t} < \text{t}_0 \\ 1 & \text{dla t} \geq \text{t}_0 \end{cases}, HR(t) = \begin{cases} \exp\beta_i & \text{dla t} < \text{t}_0 \\ \exp(\beta_i + \delta_i) & \text{dla t} \geq \text{t}_0 \end{cases}. \quad (5)$$

Function g in formula (5) is called Heavyside`s function (Ata, Sözer, 2007, p. 161). We should remember about appropriate choice of the critical moment t_0 which should be a point of crossing the survival curves or a point from which the curves are approaching or receding from each other. Model (2) with functions (3)-(5) was marked as Models A, B and C. The results of the estimation of those models are in Table 2.

---

[4] Selected methods in Keele (2010).

Table 2. The results of the estimation of Cox non-proportional hazard models for age groups

Model	Variable	Estimated coefficient	Standard error	Wald statistic	p–value	Hazard ratio (HR)
Model A	W_2	0.2441	0.0415	34.5611	0.0000	1.2765
	W_3	0.1536	0.0482	10.1694	0.0014	1.1660
	W_4	–0.0134	0.0477	0.0784	0.7795	0.9867
	W_5	–0.4628	0.0680	46.3754	0.0000	0.6295
	W_6	–1.3602	0.1854	53.8120	0.0000	0.2566
	W_2xt	–0.0388	0.0066	34.4405	0.0000	
	W_3xt	–0.0511	0.0069	54.5283	0.0000	
	W_4xt	–0.0551	0.0068	66.4062	0.0000	
	W_5xt	–0.0606	0.0075	64.7757	0.0000	
	W_6xt	–0.0647	0.0109	34.9703	0.0000	
Model B	W_2	0.1366	0.0381	12.8616	0.0003	1.1464
	W_3	0.0501	0.0451	1.2359	0.2663	1.0514
	W_4	–0.0899	0.0443	4.1182	0.0424	0.9140
	W_5	–0.5319	0.0674	62.2476	0.0000	0.5875
	W_6	–1.3942	0.2000	48.6015	0.0000	0.2480
	W_2xlnt	–0.0380	0.0279	1.8445	0.1744	
	W_3xlnt	–0.1204	0.0321	14.0283	0.0002	
	W_4xlnt	–0.1884	0.0313	36.3456	0.0000	
	W_5xlnt	–0.2279	0.0429	28.2905	0.0000	
	W_6xlnt	–0.2969	0.1044	8.0975	0.0044	
Model C	W_2	0.1334	0.0350	14.4958	0.0001	1.1427
	W_3	0.0263	0.0418	0.3949	0.5297	1.0266
	W_4	–0.1392	0.0411	11.4605	0.0007	0.8701
	W_5	–0.6240	0.0613	103.7732	0.0000	0.5358
	W_6	–1.4852	0.1741	72.7738	0.0000	0.2265
	$W_2(t \geq 12)$	–0.4670	0.1593	8.5898	0.0034	
	$W_3(t \geq 12)$	–1.1383	0.1953	33.9850	0.0000	
	$W_4(t \geq 12)$	–1.5295	0.1946	61.7987	0.0000	
	$W_5(t \geq 12)$	–1.4992	0.2419	38.4162	0.0000	
	$W_6(t \geq 12)$	–2.2505	0.6234	13.0341	0.0003	

Source: own study.

Values of the parameters indicate that the function g in form (3)-(5) can be used for modeling of time varying hazard for explanatory variables W_i (significance of parameters δ_i). Absence of parameter significance in W_2xln$t$ (Model B) informed us that with the time passage, the influence of variable W_2 (unemployed persons aged 25–34) on hazard does not change significantly. Also parameters β_i in estimated models were statistically significant except three cases. Level $p > 0{,}05$ for variables: W_4 (unemployed persons aged 45–55) in model A, W_3 (unemployed persons aged 35–45) in models B and C informed that in moment $t = 0$ (Models A and B) and in first year (Model C) the level of hazard of these variables is similar to the level of hazard for the reference group – unemployed persons aged 18–24 (hazard ratio for this variables is close to 1). Changes of hazard ratios with passage of time are presented in Table 3 (Models A and B; example months) and Table 4 (Model C).

Table 3. Values of hazard ratios for Models A and B

t	Hazard ratios									
	Model A					Model B				
	W_2/W_1	W_3/W_1	W_4/W_1	W_5/W_1	W_6/W_1	W_2/W_1	W_3/W_1	W_4/W_1	W_5/W_1	W_6/W_1
0	1.2765	1.1660	0.9867	0.6295	0.2566					
1	1.2279	1.1079	0.9338	0.5925	0.2405	1.1464	1.0514	0.9140	0.5875	0.2480
2	1.1812	1.0528	0.8838	0.5577	0.2255	1.1167	0.9672	0.8021	0.5017	0.2019
3	1.1362	1.0003	0.8364	0.5249	0.2113	1.0996	0.9211	0.7432	0.4574	0.1790
4	1.0930	0.9505	0.7915	0.4940	0.1981	1.0877	0.8898	0.7040	0.4284	0.1643
5	1.0513	0.9032	0.7491	0.4650	0.1857	1.0785	0.8662	0.6750	0.4071	0.1538
6	1.0113	0.8582	0.7089	0.4377	0.1740	1.0711	0.8474	0.6522	0.3906	0.1457
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
43	0.2406	0.1297	0.0923	0.0465	0.0159	0.9939	0.6685	0.4501	0.2493	0.0812
44	0.2314	0.1233	0.0873	0.0438	0.0149	0.9931	0.6666	0.4481	0.2480	0.0806
45	0.2226	0.1171	0.0826	0.0412	0.0140	0.9922	0.6648	0.4462	0.2467	0.0801
46	0.2141	0.1113	0.0782	0.0388	0.0131	0.9914	0.6631	0.4444	0.2455	0.0796
47	0.2060	0.1057	0.0740	0.0365	0.0123	0.9906	0.6614	0.4426	0.2443	0.0791
48	0.1981	0.1003	0.0701	0.0343	0.0115	0.9895	0.6597	0.4408	0.2431	0.0786
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Source: own study.

Table 4. Values of hazard ratios for Model C

Hazard ratios	$t < 12$	$t \geq 12$
W_2/W_1	1.1427	0.7163
W_3/W_1	1.0266	0.3289
W_4/W_1	0.8701	0.1885
W_5/W_1	0.5358	0.1197
W_6/W_1	0.2265	0.0239

Source: own study.

The analysis of results contained in Tables 3 and 4 leads to the following conclusions:

– unemployed persons aged 25–34 and 35–44 in first months of job seeking, have bigger chances of finding employment than the reference group (aged 18–25); HR > 1,

– chances of finding a job by unemployed persons in all age groups relative to the youngest persons are lower together with time passage and with age of unemployed person.

V. CONCLUSIONS

The two types of model of Cox regression: proportional and non-proportional hazard are more and more often used in socio-economic researches. In most of the analysis Cox proportional hazard model is used to modeling, so it is assumed that the influence of explanatory variables on hazard does not depend on time. Estimated in this way model parameters made it possible to compute the hazard ratios, which are constant in research period of time. This is some kind of averaging. Researching hazard ratios can be changed in time. Graphic methods can be used for preliminary estimation of proportionality. In case of any doubts it is worth using the analytical methods. They allow checking, if assumptions of proportionality of hazard are satisfied and, at the same time, allow to estimate the hazard ratios changing in time.

The interpretation of parameters of Cox regression model with proportional hazard indicated which groups of the unemployed have bigger chances of finding a job in all researched periods. While the interpretation of parameters of Cox regression model with non-proportional hazard made it possible to study changes in chances of research groups in time. By proper defining of function *g* it can be established in what way the hazard ratios change in time. Based on the analysis we can draw the following conclusions:

– hypothesis that the influence of variable *Age* on hazard changes in time was confirmed,

– choosing the model with proportional or non-proportional hazard is up to the researcher; but in the first case we can get the averaged (in time) values of hazard ratios.

REFERENCES

Ata N., Sözer M.T. (2007), Cox regression models with nonproportional hazard applied to lung cancer survival data, *Hacettepe Journal of Mathematics and Statistics*, Vol. 36 (2), pp. 157-167.

Bednarski T. (2005), Ocena przydatności danych Bael dla charakterystyki rozkładu czasu poszukiwania pracy na przykładzie danych z lat 2001-2002, *Studia Ekonomiczne* nr 4, Instytut Nauk Ekonomicznych PAN, Warszawa.

Bieszk-Stolorz B., Markowicz I., *Modele regresji Coxa w analizie bezrobocia*, CeDeWu, Warszawa, 2012.

Cox D. R. (1972), Regression Models and Life-Tables, *Journal of the Royal Statistical Society*, Series B (Methodological), Vol. 34, No. 2, pp. 187-220.

Cox D. R. (1975), Partial Likelihood, *Biometrika*, Vol. 62, No.2, pp. 269-276.

Cox D.R., Oakes D. (1984), *Analysis of Survival Data*. Chapman and Hall. New York, London.

Kalbfleisch, J.D., Prentice, R. L. (1980), *The statistical analysis of failure time data*, Wiley, New York.

Keele L. (2010), Testing for Nonproportional Hazards in Cox Models, *Political Analysis* 18, pp. 189-205.

Kleinbaum D.G., Klein M. (2005), *Logistic Regression. A Self-Lerning Text*, Springer-Verlag, New York.

Schemper M. (1992), Cox analysis of survival data with non-proportional hazard function, *The Statistician* 41, pp. 455-465.

Beata Bieszk-Stolorz, *Iwona Markowicz*

WPŁYW WIEKU OSÓB BEZROBOTNYCH NA HAZARD I JEGO ZMIANY W CZASIE

Celem artykułu jest analiza wpływu wieku osób bezrobotnych na czas znalezienia zatrudnienia. Wpływ ten może zmieniać się wraz z czasem. Analizę taką umożliwia zastosowanie modelu nieproporcjonalnego hazardu Coxa. Stosując model proporcjonalnych hazardów uzyskujemy oceny parametrów stałe w czasie. Iloraz hazardów dla dwóch jednostek różniących się wartością analizowanej cechy jest wówczas uśrednieniem dla całego badanego okresu. Stałość ilorazów hazardu można zbadać metodami graficznymi lub badając istotność parametrów modelu nieproporcjonalnego hazardu Coxa.