Stratified Cox Model with Interactions in Analysis of Recurrent Events

Abstract: The purpose of this paper is the assessment of relative intensity of exit from registered unemployment by means of the analysis of recurrent survival episodes and the comparison of these results with the results obtained for an individual episode. The stratified Cox model with interactions was used. Statistical data collected by labour offices indicate that a large fraction of the unemployed persons is registered multiple times. However, many of them resign from the mediation of labour offices and are subsequently removed from the register. In the article, the intensities of de-registration due to various causes for men and women were compared. The study data came from the database of personal details of people registered by the Poviat Labour Office in Szczecin in 2013. The observation covered the records of their registration until the end of 2014. Gender of the unemployed persons influenced the intensity of de-registrations in the first episodes, partially in the second and third ones, due to various causes, such as finding a job or removal from the register, whereas it did not influence the intensity of de-registrations in the fourth and subsequent episodes. As for the other causes in the subsequent episodes, the differences were also not statistically significant. The proposed analysis may be important for implementing a good policy in the labour market. The identification of persons that resign from the mediation of the labour office is as interesting as the identification of these persons who find a job.

Keywords: survival analysis, recurrent events, stratified Cox model, discontinuous risk intervals, unemployment

JEL: C41, J64
1. Introduction

The survival analysis focuses on how long an individual survives in a given state until the moment when a specific event occurs. It is often the case that we examine processes in the course of which an individual enters the pre-defined state several times. The recurrent event processes are defined as processes that generate specific events more than once (Cook, Lawless, 2007). They can be analysed by means of selected survival analysis methods, which are used in technical sciences to examine assembly line stoppage or to detect and remediate software bugs. In medicine, the duration to relapse is analysed (Sagara et al., 2014), while in socio-economic sciences we can study the length of subsequent episodes of entering and leaving poverty (Sączewska-Piotrowska, 2015), the time between recurrent warranty or insurance claims or the duration of unemployment (Galecka-Burdziak, 2016; Galecka-Burdziak, Góra, 2017). In such studies the analysed random variable is the time to event. In earlier works on textile industry, the random variable was the yarn length, while the event – the occurrence of a yarn defect. Other examples include vehicle mileage until warranty repair or a number of production cycles (Cook, Lawless, 2007).

The purpose of this paper is the assessment of relative intensity of exit from registered unemployment by means of the analysis of recurrent survival episodes and the comparison of these results with the results obtained for an individual episode. Using the stratified Cox regression model with interactions, the relative intensity of ceasing the use of services provided by the labour office in particular episodes by women with respect to men was compared. The time from the moment of registration to de-registration was observed. The de-registration from the labour office could have happened due to various causes: taking up a job, removal from the register or other causes. The study is an important contribution to the process of creating a labour market policy. It allows policy-makers to select from the pool of the registered unemployed a group of individuals who should be covered by activation programmes.

2. Methodology of research

Selected survival analysis methods for recurrent events and discontinuous risk intervals were used in the research. In medicine, when we analyse the time to a relapse of a chronic disease (e.g.: asthma, epileptic seizures or osteoporotic fractures), the term of continuous risk intervals is used (Twisk, Smidt, de Vente, 2005). During the investigation of the duration of registered unemployment, the fact that recurrent episodes are separated by periods of de-registration from
the labour office (Figure 1), thus creating discontinuous risk intervals (Guo, Gill, Allore, 2008; Tan, 2014: 60–77), should be considered. On that account, in this study, the conditional approach was used (Hosmer, Lemeshow, 1999: 308–311; Machin, Cheung, Parmar, 2006: 247; Aalen, Borgan, Gjessing, 2008: 473). The time to the next event was determined using the notion of a time gap (Prentice, Williams, Peterson, 1981; Jiang, Landers, Rhoads, 2006; Bijwaard, Franses, Paap, 2006), i.e. – the beginning of each event was restarted to zero (Figure 2).

The actual study was preceded by the estimation of the median duration. It is the survival time at which the survival function is equal to 0.5. The survival function is defined as follows:

\[
S(t) = \prod_{t_i < t} \left(1 - \frac{d_i}{m_i} \right)
\]

where \(d_i\) is the number of events and \(m_i\) is the number of at-risk individuals at time \(t_i\).
\[ S(t) = P(t > T) = 1 - F(t), \]  

where:
\( T \) – the event duration,
\( F(t) \) – the cumulative distribution function of the random variable \( T \).

The most frequently used estimator of the survival function is the Kaplan-Meier estimator (Kaplan, Meier, 1958):
\[
\hat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right),
\]

where:
\( d_j \) – the number of events at the moment \( t_j \),
\( n_j \) – the number of individuals at risk by the moment \( t_j \).

In the first stage of the analysis, the author estimated the intensity of de-registration from labour office lists due to any cause and the intensity of de-registration due to three main causes for a single episode. To this end, the Cox hazards model was applied, where (Bieszk-Stolorz, Markowicz, 2012):
\[
h(t, X) = h_0(t) \exp \left( \sum_{i=1}^{n} \beta_i X_i \right),
\]

where:
\( t \) – time,
\( X = [X_1, X_2, ..., X_n] \) – the vector of explanatory variables,
\( h_0(t) \) – the baseline hazard.

The second stage of the study included the analysis of recurrent episodes. Here the stratified Cox proportional hazards model with interactions could be used (Kleinbaum, Klein, 2005: 352):
\[
h_g(t, X) = h_{0g}(t) \exp \left( \sum_{i=1}^{n} \left( \beta_i X_i + \sum_{j=1}^{k-1} \delta_{ij} D_j X_i \right) \right),
\]

where:
\( t \) – time,
\( k \) – the number of strata,
\( X = [X_1, X_2, ..., X_n] \) – the vector of explanatory variables,
\( g \) – the number of stratum,
\( h_{0g}(t) \) – the baseline hazard for the stratum \( g \),
\( D_j \) – the dummy dichotomous variable determining the episode number.

The alternative version of the model (4):
Stratified Cox Model with Interactions in Analysis of Recurrent Events

\[ h_g^*(t, X) = h_{0g}^*(t) \exp \left( \sum_{i=1}^{n} \sum_{j=1}^{k} \delta_{ij} D_j X_i \right), \]  \hspace{1cm} (5)

where:
- \( t \) – time,
- \( k \) – the number of strata,
- \( X = [X_1, X_2, \ldots, X_n] \) – the vector of explanatory variables,
- \( g \) – the number of stratum,
- \( h_{0g}^*(t) \) – the baseline hazard for the stratum \( g \),
- \( D_j \) – the dummy dichotomous variable determining the episode number.

The intensity of unemployment leaving with respect to the de-registered person’s gender was analysed. The dichotomous variable \( X \) was 1 for women and 0 for men. The strata (\( g \)) were defined as follows: 1 – the first episode, 2 – the second episode, 3 – the third episode, 4 – the fourth and subsequent episode. Therefore the model (4) took the following form:

\[ h_g^*(t, X) = h_{0g}^*(t) \exp(\beta X + \delta_2 D_2 X + \delta_3 D_3 X + \delta_4 D_4 X) \] for \( g = 1, 2, 3, 4 \), \hspace{1cm} (6)

where:
\[ D_j = \begin{cases} 1 & \text{j-th episode} \\ 0 & \text{not j-th episode} \end{cases} \] for \( j = 2, 3, 4 \). \hspace{1cm} (7)

Using the estimated parameters of the model (5), we could determine the hazard ratio (\( HR \)) for the \( j \)-th episode:

\[ HR_1 = \exp \beta, \] \hspace{1cm} (8)

\[ HR_j = \exp(\beta + \delta_j) \] for \( j = 2, 3, 4 \). \hspace{1cm} (9)

While the alternative model (5) took the following form:

\[ h_g^*(t, X) = h_{0g}^*(t) \exp(\delta_1 D_1 X + \delta_2 D_2 X + \delta_3 D_3 X + \delta_4 D_4 X) \] for \( g = 1, 2, 3, 4 \), \hspace{1cm} (10)

where:
\[ D_j = \begin{cases} 1 & \text{j-th episode} \\ 0 & \text{not j-th episode} \end{cases} \] for \( j = 1, 2, 3, 4 \). \hspace{1cm} (11)

In this case, the hazard ratio (\( HR \)) for the \( j \)-th episode, which represents the intensity of women’s unemployment leaving against men’s, is given by:
3. The data used in the research

The study used individual data of unemployed people registered for the first time by the Poviat Labour Office in Szczecin in 2013. The observation covered the records of their registration until the end of 2014. Each record consists of episodes of registration. The duration of subsequent episodes was analysed. Labour offices collect several dozens of causes for the unemployed person’s de-registering from their lists. Based on the existing research into competing risks in unemployment (Bieszk-Stolorz, 2017a; 2017b), in this study, the causes were categorised into three groups: taking up a job (Job), removal from the lists due to causes related to the unemployed person (Removal), and the remaining causes (Other). The detailed description of these categories is shown in Table 1. The category Job contains such causes for de-registering as taking up a subsidised and non-subsidised job or starting an entrepreneurial activity. Removal is understood as the de-registration of an unemployed person from the register because of their absence in the office within a specified time as well as their refusal to accept a job offer. The category Other includes de-registering due to the granting of disability pension, allowance or old-age-pension, the individual’s moving abroad or death.

Every individual registered by the labour office has their own record of registration. Such records include several episodes, i.e. the subsequent periods of registered unemployment. After a preliminary analysis of the number of episodes in the registration records, a decision was made to divide them into four groups: the records with one, two, three, four or more episodes. The selection of the last category resulted from a small number of people with at least four unemployment episodes. The analysis covered 5,418 records, including 2,644 records of unemployed women. A vast majority (4,100) contained one episode of unemployment, while 1,078 contained two episodes (Table 2). The first episodes usually ended with taking up a job followed by de-registration (for both men and women). In the case of the second episodes, the pattern was similar in the group of females, while in the group of males de-registration predominantly preceded taking up a job (Table 3).
Table 1. Groups of causes of de-registration from the labour office

| Major causes | Sub-causes                          |
|--------------|-------------------------------------|
| Job          | Non-subsidised job                  |
|              | Subsidised job                      |
|              | Entrepreneurial activity            |
| Removal      | Removal                             |
| Other        | Disability pension, allowance or old-age-pension |
|              | The individual's moving abroad      |
|              | Other                               |

Source: own study

Table 2. Registration records by number of episodes

| Number of episodes | Number of records (women) |
|--------------------|---------------------------|
| 1                  | 4100 (2073)               |
| 2                  | 1078 (476)                |
| 3                  | 193 (73)                  |
| 4 and more         | 47 (22)                   |
| Total              | 5418 (2644)               |

Source: own study

Table 3. The number of subsequent episodes

| Episode number | Total (women) | Type of observation | Any cause (women) | Job (women) | Removal (women) | Other (women) |
|----------------|---------------|---------------------|-------------------|-------------|-----------------|---------------|
| 1st            | 5418 (2644)   | Uncensored          | 5048 (2448)       | 2790 (1359) | 1862 (710)      | 696 (379)     |
|                |               | Censored            | 370 (196)         | 2628 (1285) | 3556 (1934)     | 4722 (2265)   |
| 2nd            | 1318 (571)    | Uncensored          | 1035 (428)        | 466 (232)   | 472 (156)       | 97 (40)       |
|                |               | Censored            | 283 (143)         | 852 (339)   | 846 (415)       | 1221 (531)    |
| 3rd            | 240 (95)      | Uncensored          | 162 (64)          | 82 (39)     | 65 (17)         | 15 (8)        |
|                |               | Censored            | 78 (31)           | 158 (56)    | 175 (78)        | 225 (87)      |

www.czasopisma.uni.lodz.pl/foe/ FOE 3(35) 2018
There were two types of right-censored observations in the research. The first type of censoring was when at the end of the study the individual was still in the register. In the analysis of episodes due to the de-registration causes, there was one more censoring type – de-registering because of causes other than the analysed ones.

4. Analysis of relative intensity of de-registering from the labour office

The actual study was preceded by the analysis of median duration of episodes (Table 4). Note that the 50th percentile (median) for the cumulative survival function is usually not the same as the point in time up to which 50% of the sample survived. This would only be the case if there were no censored observations prior to this time (Sokołowski, 2010). In some cases, the estimation of median is impossible. Such a situation takes place when the intensity of exit from unemployment is low. In the case of other causes of de-registration in the analysed period (24 months), less than 50% of persons left the unemployment. It referred to all unemployed persons – both men and women. That is why this cause was not presented in Table 4. It also happened in several subgroups of finding a job and removal (appropriate situations in Table 4 were marked by the hyphen). In these cases, we can only state that the median duration is higher than 24 months. In the case of any cause of de-registration, the median duration was longer for women than for men (4.47 vs 3.35 months, respectively). The median time to finding a job was lower for women. It results from the fact that women found jobs sooner than men. Men were more intensively removed from the register because the median duration for men was lower than for women with the exception of the fourth episode.

In the first stage of the study, the parameters for the Cox hazards models (1) were estimated for single episodes (Table 5). Women were de-registered less intensively than men due to any cause (by 16%) and due to removal (by 45%), but...
they took up a job with higher intensity (by 10%). Because the $\beta$ parameter is not significant, the intensities of de-registration due to other causes for women and men were not significantly different.

Table 4. The median duration of episodes by the cause of de-registration and gender

| Groups          | Any cause of de-registration | Any episode | 1st episode | 2nd episode | 3rd episode | 4th and next episodes |
|-----------------|-----------------------------|-------------|-------------|-------------|-------------|----------------------|
| Total           | 3.95                        | 4.24        | 2.99        | 2.88        | 1.41        |
| Women           | 4.47                        | 4.77        | 3.24        | 2.96        | 2.04        |
| Men             | 3.55                        | 3.72        | 2.79        | 2.79        | 1.13        |
| Job             |                             |             |             |             |             |
| Total           | 8.98                        | 9.00        | 9.05        | 7.49        | 2.64        |
| Women           | 8.71                        | 8.78        | 8.58        | 5.52        | 1.62        |
| Men             | 9.47                        | 9.46        | 9.88        | 8.50        | –           |
| Removal         |                             |             |             |             |             |
| Total           | 11.78                       | 12.74       | 8.48        | –           | 5.13        |
| Women           | 17.13                       | 17.10       | –           | –           | 3.26        |
| Men             | 8.05                        | 8.46        | 5.96        | 8.57        | –           |

Source: own study

The second stage of the analysis, i.e. the estimation of parameters of the stratified Cox model with interactions (8), allowed the assessment of the recurrent episodes (Table 6). During the first and second episode, women were leaving the register with lower intensity than men (by 15% and 19% respectively). They were also taking up a job with higher intensity during the first episode (by 9%) and were removed from the register with lower intensity during the first (by 44%), second (by 43%) and third (by 49%) episode. In other cases, de-registration intensity between men and women was not significantly different.

Table 5. The results of Cox regression parameters estimation for single episodes

| De-registration cause | Estimator of parameter $\beta$ | Standard error | Wald statistics | $p$-value | Hazard ratio |
|----------------------|--------------------------------|----------------|-----------------|-----------|--------------|
| Any cause            | $\chi^2 = 46.4918$, $p = 0.0000$ | –0.1725        | 46.3904         | 0.0000    | 0.8415       |
| Job                  | $\chi^2 = 6.5721$, $p = 0.0104$ | 0.0931         | 6.5587          | 0.0104    | 1.0975       |
| Removal              | $\chi^2 = 200.9263$, $p = 0.0000$ | –0.5882        | 193.4751        | 0.0000    | 0.5553       |
| Other                | $\chi^2 = 0.1438$, $p = 0.7045$ | 0.0267         | 0.1437          | 0.7046    | 1.0271       |

Source: own study
Table 6. The results of the stratified Cox model parameters estimation with interactions for multiple episodes

| De-registration cause | Parameter | Estimator of parameter | Standard error | Wald statistics | p-value | Hazard ratio |
|-----------------------|-----------|------------------------|----------------|----------------|---------|-------------|
| Any cause             | \( \delta_1 \) | -0.1620                | 0.0282         | 32.9958        | 0.0000  | 0.8505      |
|                       | \( \delta_2 \) | -0.2100                | 0.0634         | 10.9747        | 0.0009  | 0.8106      |
|                       | \( \delta_3 \) | -0.0543                | 0.1612         | 0.1137         | 0.7360  | 0.9471      |
|                       | \( \delta_4 \) | 0.6466                 | 0.3624         | 3.1824         | 0.0744  | 1.9090      |
| Job                   | \( \delta_1 \) | 0.0858                 | 0.0403         | 4.5353         | 0.0332  | 1.0896      |
|                       | \( \delta_2 \) | 0.1333                 | 0.0930         | 2.0517         | 0.1520  | 1.1426      |
|                       | \( \delta_3 \) | 0.2897                 | 0.2226         | 1.6937         | 0.1931  | 1.3360      |
|                       | \( \delta_4 \) | 0.7627                 | 0.4671         | 2.6656         | 0.1025  | 2.1440      |
| Removal               | \( \delta_1 \) | -0.5816                | 0.0478         | 148.2220       | 0.0000  | 0.5590      |
|                       | \( \delta_2 \) | -0.5635                | 0.0982         | 32.9287        | 0.0000  | 0.5692      |
|                       | \( \delta_3 \) | -0.6706                | 0.2824         | 5.6409         | 0.0176  | 0.5114      |
|                       | \( \delta_4 \) | 0.4853                 | 0.6341         | 0.5857         | 0.4441  | 1.6246      |
| Other                 | \( \delta_1 \) | 0.0518                 | 0.0762         | 0.4625         | 0.4965  | 1.0532      |
|                       | \( \delta_2 \) | -0.2391                | 0.2073         | 1.3309         | 0.2486  | 0.7873      |
|                       | \( \delta_3 \) | 0.4583                 | 0.5178         | 0.7832         | 0.3762  | 1.5813      |
|                       | \( \delta_4 \) | 0.3722                 | 1.4159         | 0.0691         | 0.7927  | 1.4509      |

Source: own study

5. Conclusions

Applied methods of survival analysis allowed to assess the intensity of unemployment exit with respect to gender of the unemployed person. The examination of the effect of gender on the intensity of unemployment exit in the case of a single episode was possible because of using the Cox regression model. In this case, the intensities of de-registrations from the labour office of men and women due to any cause, finding a job and removal were significantly different. The stratified Cox regression model with interactions was used to evaluate the relative intensity of de-registrations from the labour office during the subsequent episodes, i.e. the subsequent periods of registration in the labour office. In the case of the first episodes, the intensities of de-registrations of men and women were significantly different for any cause, finding a job and removal. These differences were significantly different for the second episodes in the case of any cause and removal, while for the third episodes – only for removal from the register. The intensities...
for women and men were not significantly different for the fourth and subsequent episodes. As for the other causes in the subsequent episodes, the differences were also not statistically significant. The proposed analysis may be important for implementing a good policy in the labour market. Other researches very often focus only on unemployed persons finding a job. The data published by the Central Statistical Office indicate that there is a considerable group of the unemployed persons that resign from the mediation of labour offices without giving a reason. It would be interesting to extend the study and examine the impact of the unemployed Poles’ other reasons for de-registration. Unfortunately, the restricted access to individual data is a serious research limitation.

References

Aalen O.O., Borgan O., Gjessing H.K. (2008), Survival and Event History Analysis. A Process Point of View, Springer, New York.

Bieszk-Stolorz B. (2017a), Cumulative Incidence Function in Studies on the Duration of the Unemployment Exit Process, “Folia Oeconomica Stetinensia”, vol. 17, issue 1, pp. 138–150.

Bieszk-Stolorz B. (2017b), Funkcja skumulowanej częstości i modele hazardu w ocenie konkurencyjnych form wyjścia z bezrobocia, “Taksonomia”, no. 29, “Prace Naukowe UE we Wrocławiu”, no. 469, pp. 21–31, doi: 10.15611/pn.2017.469.02.

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

Bijwaard G.E., Franses P.H., Paap R. (2006), Modeling Purchases as Repeated Events, “Journal of Business & Economic Statistics”, vol. 24, no. 4, pp. 487–502, doi.org

Cook R.J., Lawless J.F. (2007), The Statistical Analysis of Recurrent Events, Springer, New York.

Gałecka-Burdziak E. (2016), Multiple unemployment spells duration in Poland, Szkoła Główna Handlowa, Kolegium Analiz Ekonomicznych Working Papers Series, 2016/019, http://kolegia.sgh.waw.pl/pl/KAE/Documents/WorkingPapersKAE/WPKAE_2016_019.pdf [accessed: 15.01.2018].

Galecka-Burdziak E., Góra M. (2017), How Do Unemployed Workers Behave Prior to Retirement? A Multi-State Multiple-Spell Approach, Discussion Paper Series, IZA DP no. 10680, ftp.iza.org [accessed: 15.01.2018].

Guo Z., Gill T.M., Allore H.G. (2008), Modeling repeated time-to-event health conditions with discontinuous risk intervals: an example of a longitudinal study of functional disability among older persons, “Methods of Information in Medicine”, vol. 47, issue 2, pp. 107–116.

Hosmer D.W., Lemeshow S. (1999), Applied Survival Analysis. Regression Modeling of Time to Event Data, John Wiley & Sons, New York.

Jiang S.T, Landers T.L. Rhoads T.R. (2006), Proportional Intensity Models Robustness with Overhaul Intervals, “Quality and Reliability Engineering International”, vol. 22, issue 3, pp. 251–263.

Kaplan E.L., Meier P. (1958), Non-parametric estimation from incomplete observations, “Journal of American Statistical Association”, vol. 53, pp. 457–481.

Kleinbaum D., Klein M. (2005), Survival Analysis. A Self-Learning Text, Springer, New York.

Machin D., Cheung Y.B., Parmar M.K.B. (2006), Survival Analysis. A Practical Approach. Second Edition, John Wiley & Sons, Chichester.
Warstwowy model Coxa z interakcjami w analizie zdarzeń powtarzających się

Streszczenie: Celem artykułu jest ocena względnej intensywności wychodzenia z bezrobocia rejestrowanego za pomocą analizy zdarzeń powtarzających się i porównanie wyników z wynikami otrzymanymi dla pojedynczych epizodów. W analizie wykorzystano warstwowy model Coxa z interakcjami. Dane statystyczne urzędów pracy wskazują, że duży odsetek osób bezrobotnych wielokrotnie korzysta z pośrednictwa urzędów w poszukiwaniu pracy. Wiele z tych osób rezygnuje z tego pośrednictwa i zostaje wykreślonych z rejestru. W artykule porównano intensywność wyrejestrowania z różnych powodów dla kobiet i mężczyzn. W badaniu wykorzystano dane indywidualne o osobach zarejestrowanych w Powiatowym Urzędzie Pracy w Szczecinie w 2013 roku. Obserwowano historię ich rejestracji do końca 2014 roku. Płeć osób bezrobotnych wpływała na intensywność wyrejestrowania z dowolnego powodu, z powodu podjęcia pracy i wykreślzenia w pierwszych, częściowo drugich i częściowo trzecich epizodach. Natomiast nie wpływała na intensywność wyrejestrowania w czwartych i dalszych epizodach. W przypadku wyrejestrowania z pozostałych powodów różnice w intensywnościach również nie były istotne statystycznie. Zaproponowana analiza jest ważna z punktu widzenia polityki rynku pracy. Interesująca jest identyfikacja nie tylko osób podejmujących zatrudnienie, ale również tych, które rezygnują z pośrednictwa urzędu pracy.

Słowa kluczowe: analiza trwania, zdarzenia powtarzające się, warstwowy model Coxa, niestyczne przedziały ryzyka, bezrobocie

JEL: C41, J64