Beata Bieszk-Stolorz

University of Szczecin, Institute of Economics and Finance, beata.bieszk-stolorz@usz.edu.pl

Models of Multiple Events in the Analysis of Subsequent Registrations in the Labour Office

Abstract: In many fields of science, it is necessary to analyse recurrent events. In medical science, the problem is to assess the risk of chronic disease recurrence. In economic and social sciences, it is possible to analyse the time of entering and leaving the sphere of poverty, the time of subsequent guarantee or insurance claims, as well as the time of subsequent periods of unemployment. In these studies, there are different ways of defining risk intervals, i.e. the time frame over which an event is at risk (or likely to occur) for an entity.

Research on registered unemployment in Poland shows a high percentage of people returning to the labour office and registering again. The aim of the article is assessment of the risk of subsequent registrations in the labour office depending on selected characteristics of the unemployed: gender, age, education, and seniority.

In the study, methods of survival analysis were used. The results obtained for four models being an extension of the Cox proportional hazard model were compared. The Anderson-Gil model does not distinguish between first and next events. The number of events that occurred is important. Two Prentince-Williams-Peterson conditional models and the Wei, Lin and Weissfeld models are based on the Cox stratified model. The strata are consecutive events. They differ in the way risk intervals are determined.

In the analysed period, only age and education influenced the risk of multiple registrations at the Poviat Labour Office in Szczecin. Gender and seniority did not have a significant impact on this risk. The analysis performed for subsequent registrations confirmed the impact of the same features on the first subsequent registration. In general, it can be stated that the analysed characteristics of the unemployed did not have a significant impact on the second and subsequent returns to the labour office.

Keywords: Cox regression model, recurrent events models, risk of registration in the labour office

JEL: C41, J64

1. Introduction

Since 2013 the economic situation in the Polish labour market has been improving. This is reflected in the decreasing unemployment rate, the number of persons registered as unemployed and the increased outflow of persons registered in labour offices. From the point of view of social policy, it is important to activate unemployed persons to take up employment. The best solution would be to take up a permanent job or start a long-term business activity. Unfortunately, statistical data show a high percentage of people returning to the labour office and registering again. In 2004, 2090.4 thousand people were registered, of whom 1417.7 thousand were re-registered. It constituted 68% of registered persons and it was a minimum value in the period 2004–2018. Since then, this percentage has increased and remains relatively stable. In 2016, it was 83% (the maximum value in the presented period), and in 2017–2018 – 82% (Figure 1).
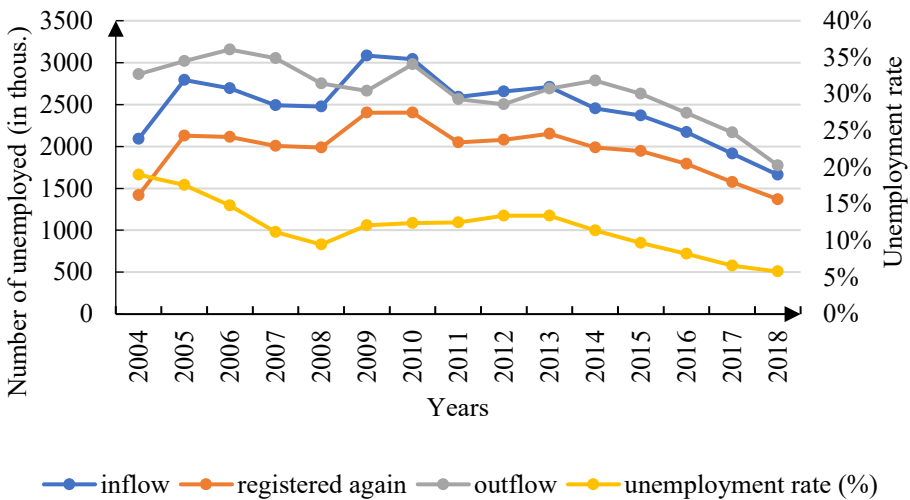

Figure 1. Inflow and outflow of the registered unemployed and the registered unemployment rate in Poland in the years 2004–2018

Source: own calculations

In the presented study, methods derived from the survival analysis were used to analyse subsequent registrations in the labour office. The duration of the unit in a specific state is analysed until the moment of occurrence of the specific event ending the observation. They are derived from demography and technical sciences. In the past, they were used to study the duration of human life and failure-free operation of devices. They are now also used in the analysis of the duration of economic phenomena. The investigated phenomenon may result in the occurrence of a specific event (death, illness, failure, company failure, entry into unemploy-

ment, entry into or exit from the poverty sphere). The event may not occur before the end of the observation period. Such observations in the analysis of duration shall be taken as censored. In addition to observations not yet completed before the end of the observation period, censorship includes those where the observed entity disappears from sight or there is an observation ending event that excludes the occurrence of a relevant event (Pepe, 1991), i.e. a competing event (Bieszk-Stolorz, 2018c).

A very important issue is the analysis of processes during which the unit can be in the state defined by the study several times. These processes can be analysed using selected methods from the area of survival analysis. These methods in the technical sciences are used for testing downtime on assembly lines or for analysing software fault detection and troubleshooting processes. In medical science, the time until the relapse of disease symptoms is examined (Sagara et al., 2014). In economic and social sciences, it is possible to analyse the time of entering and leaving the sphere of poverty (Sączewska-Piotrowska, 2015), the time of subsequent guarantee or insurance claims or subsequent periods of unemployment (Gałecka-Burdziak, 2016). The analysed random variable T in such studies is the time to occurrence of the event.

The aim of the article is assessment of the risk of subsequent registrations in the labour office depending on selected characteristics of the unemployed: gender, age, education, and seniority.

2. Risk intervals in the analysis of recurrent events

Multiple events (recurrent events) are defined as processes that repeatedly generate specific events (Klein, Goel, 1992; Hosmer, Lemeshow, 1999; Therneau, Grambsch, 2000; Machin, Cheung, Parmar, 2006; Cook, Lawless, 2007; Aalen, Borgan, Gjessing, 2008). Figure 2 shows examples of objects with recurrent events. Objects 1, 2 and 3 have experienced one, two and three full events respectively. Subject 4 has suffered two complete incidents and subject 3 is being censored. This censorship is related to the end of the observation period. Object 5 has suffered an event that initiated the process, but has not suffered a final event. In this case, the observation is also censored because of the loss of the unit from the observation field. Intervals between successive events are often referred to as episodes in the literature.

In the case of recurrent events, the time is indexed on two scales: calendar and between consecutive events (Sączewska-Piotrowska, 2015). In this case, risk ranges (episodes) shall be defined, i.e. the ranges within which the unit is exposed to an event along a given time scale. There are three types of such intervals: the time gap, the total time and the counting process (Prentice, Williams, Peterson, 1981). Intercurrence time (gap time) is time between events. In order to determine

it, the clock is restarted after each subsequent event (Figure 3). In the case of total time, the clock is not restarted, i.e. the time is counted from the selected point on the time scale. Most often it is the beginning of the observation of the unit (Figure 4). The counting process is a combination of calendar time and time gaps. In this case, the time is calculated on the same scale as the total time, but the fact that the beginning of each subsequent period coincides with the end of the previous period is taken into account (Figure 5). The common feature of all presented methods of determining the time is the same risk range for the first event.
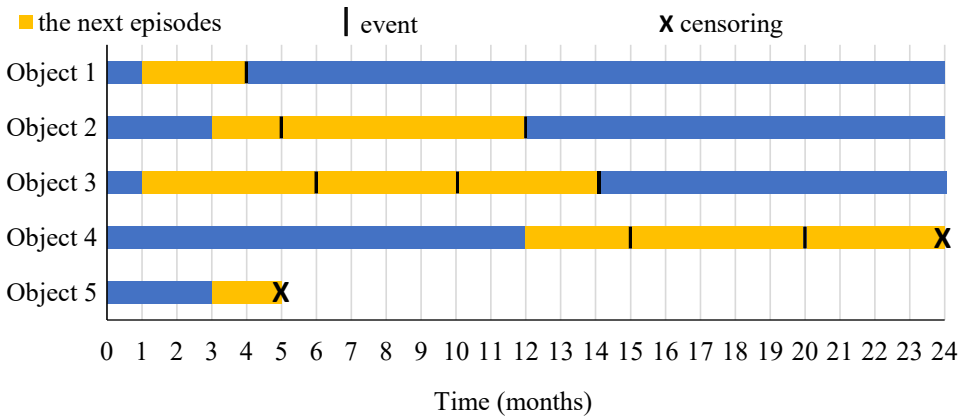

Figure 2. Examples of objects with multiple events
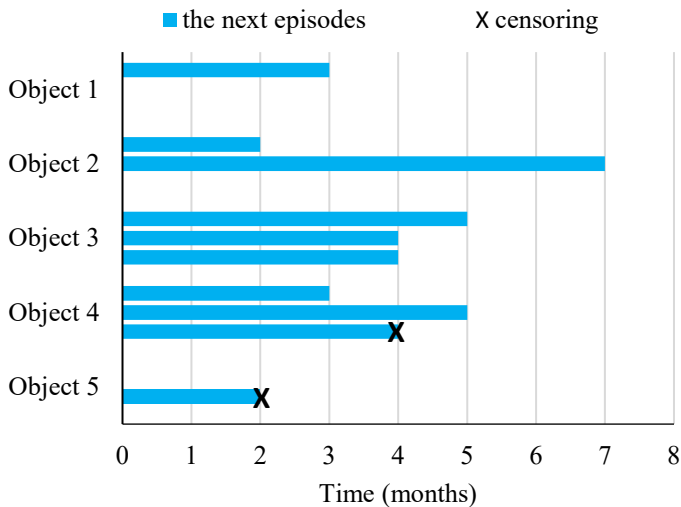Source: own calculations

Figure 3. Risk intervals for the gap time case
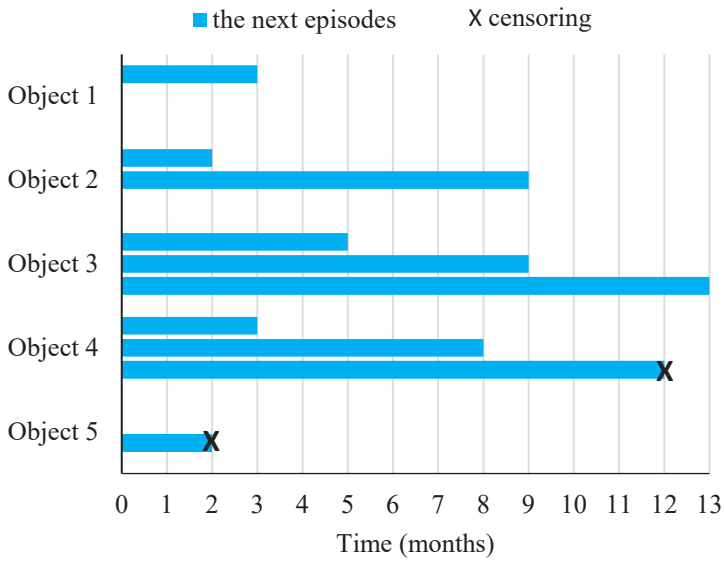Source: own calculations

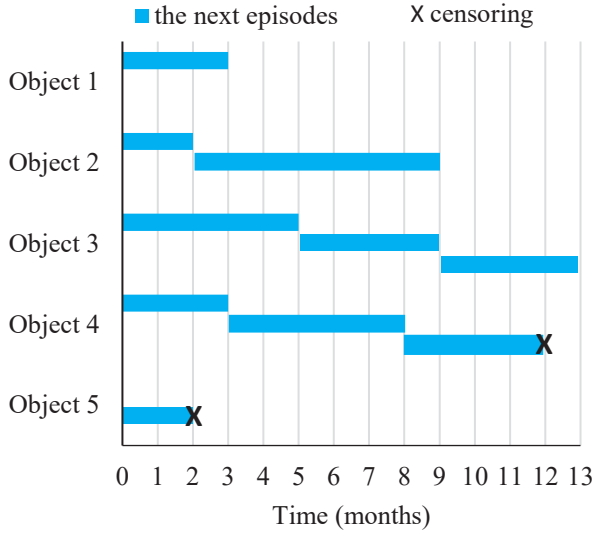Figure 4. Risk intervals for total time
Source: own calculations

Figure 5. Risk intervals for the counting process
Source: own calculations

3. Research methodology

The basic concept of the duration analysis is the duration function (survival function) which determines the probability that an event will not occur at least up to time t. It is defined as follows:

$$S(t) = P(t < T) = 1 - F(t), \tag{1}$$

where:
T – duration of the phenomenon,
$F(T)$ – cumulative density distribution of the random variable T. The most commonly used estimator of the survival function is the Kaplan-Meier estimator (Kaplan, Meier, 1958):

$$\hat{S}(t) = \prod\nolimits_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right), \tag{2}$$

where:
d_j – the number of events at the moment t_j,
n_j – the number of threatened entities until the moment of t_j.
   The second function used in the survival analysis is the hazard function. It describes the intensity of the event, i.e. the probability of the event occurring at moment t provided that it survives till time t, and is defined as follows (Kleinbaum, Klein, 2012):

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{3}$$

   In this case, the semi-parametric model of Cox proportional hazard, defined by the formula (Cox, 1972; 1975), is popular:

$$h(t, X) = h_0(t) \exp\left(\sum\nolimits_{i=1}^{m} \beta_i X_i\right), \tag{4}$$

where:
t – time,
$X = [X_1, X_2, …, X_m]$ – vector of explanatory variables,
$h_0(t)$ – basic hazard.
   If we do not have a parametrically specified function for baseline hazard, the traditional method of maximum likelihood cannot be used. The basic method for estimating the parameters of the Cox model is the partial likelihood method (Cox, 1972; 1975). Let $Y_t(t)$ represent the set of objects at risk at time t. For the model (4), this set is defined as follows:

$$Y_i(t) = I(t_i \geq t). \tag{5}$$

The partial likelihood function proposed by Cox is defined as follows (Ozga, Kieser, Rauch, 2018):

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(\beta X_i)}{\sum_{l \in Y_i(t)} \exp(\beta X_l)} \right)^{\delta_i}, \tag{6}$$

where δ_i takes the value 1 if the observation is full and 0 if the observation is censored. The partial likelihood function (or any of its approximation) is then maximised iteratively.

In the case of recurrent events in the study, models that are an extension of the Cox regression model may be used. Replacement of recurrent events models has been applied: Prentice, Williams and Peterson (1981) (counting process model – PWP-CP, time gap model – PWP-GT), Andersen and Gill (1982, AG), Wei, Lin and Weissfeld (1989, WLW). Each unit (i) present in the study is assigned successive events that form strata. The model (4) therefore takes one of two forms (Sousa-Ferreira, Abreu, 2019):

$$h(t, X_{ik}) = h_0(t) \exp\left(\sum_{j=1}^{m} \beta_j X_{jik} \right), \tag{7}$$

$$h(t, X_{ik}) = h_{0k}(t) \exp\left(\sum_{j=1}^{m} \beta_j X_{jik} \right), \tag{8}$$

where:
t – time,
$\beta = [\beta_1, \beta_2, \ldots, \beta_m]$ – vector of model's parameters,
$X = [X_1, X_2, \ldots, X_m]$ – vector of explanatory variables,
$h_0(t)$ – basic hazard for all events,
$h_{0k}(t)$ – basic hazard for all events from the episode (stratum) k,
i – the number of the object,
k – the number of the subsequent episode (stratum).

Let s be the number of strata. Then the partial likelihood function can be written as the product of the partial likelihood function specific to subsequent layers (Ozga, Kieser, Rauch, 2018):

$$L(\beta) = \prod_{k=1}^{s} L_k(\beta), \tag{9}$$

where:

$$L_k\left(\beta\right)=\prod_{i=1}^{n}\left(\frac{\exp\left(\beta X_i\right)}{\sum_{l\in Y_{ik}(t)}\exp\left(\beta X_l\right)}\right)^{\delta_{ik}}, \tag{10}$$

where δ_{ik} takes the value 1 if the observation is full and 0 if the observation is censored. The risk sets $Y_{ik}(t)$ are defined separately for each stratum.

To assess the relative intensity of the event, the hazard ratio (*HR*) determined by the following formula is used:

$$HR=\exp\beta_j. \tag{11}$$

In addition to basic hazard, these models differ in the time formula used, which is linked to different risk ranges.

In the Prentice, Williams and Peterson (PWP) models, the hazard function has the form (8). It is assumed that for each episode (stratum) k the basic hazard is different. Thus, the intensity of the event is considered in strata, i.e. the events that occur in an orderly manner are analysed. The risk of another event occurring is affected by the previous event. The PWP model allows us to save risk intervals on two possible time scales: the counting process or the time gap. In the model, the larger the sequence of events, the smaller the size of the subsequent strata, which may lead to unreliable estimates (Cai, Schaubel, 2003). Therefore, in order to avoid such a situation, the number of strata should be carefully selected.

In the Prentice, Williams and Peterson models, for the Counting Process (PWP-CP model), intervals are counted. The counting process is the determination of a time from the start of a study where the initial time of each risk interval coincides with the end of the previous event. In this case, the risk set indicator is defined as:

$$Y_{ik}\left(t\right)=I(t_{i(k-1)}<t<t_{ik}). \tag{12}$$

The Prentice, Williams and Peterson gap time models (Gap Time, PWP-GT) are based on the time interval between two events. The clock restarts when each event occurs. The indicator of the risk set is defined as:

$$Y_{ik}(t)=I(g_{ik}\geq t), \tag{13}$$

where $g_{ik}=t_{ik}-t_{i(k-1)}$ represents the observed time of interruption between two consecutive events. Instead of basic hazard $h_{0k}(t)$, basic hazard $h_{0k}(t_{ik}-t_{i(k-1)})$ is therefore being considered.

In the next model – Andersen and Gill (AG model) – the hazard function has a form (7). In this case, strata are not considered and the assumption is that the basic hazard is the same. The unit is still assigned several events, but they are not ordered. Events have the same risk of occurrence. Similarly to the PWP-CP model, the periods are counted. The risk set indicator is defined as:

$$Y_{ik}(t) = I\left(t_{i(k-1)} < t < t_{ik}\right). \tag{14}$$

This model was developed for situations in which events do not depend on the observed time from the last event or on the number of events that occurred before. Although counting the formulation of the process requires a conditional structure of dependencies, among the events it is assumed that the times between them are independent. Some authors believe that the AG model is the simplest model, but the one with the strongest assumptions (Therneau, Grambsch, 2000; Cai, Schaubel, 2003).

The last of the presented models – the Wei, Lin and Weissfeld model (WLW model) has the hazard function calculated by means of equation (8). It is therefore a stratified model with a differentiated function of basic hazard. The risk intervals are based on the total time, i.e. the time from the start of the observation. The indicator of the risk set is defined as:

$$Y_{ik}(t) = I(t_{ik} \geq t). \tag{15}$$

There are similarities and differences between the presented models. The PWP and WLW models are the stratified ones and have the same hazard function, but differ in the definition of risk intervals. The AG and PWP-CP models have the same defined risk interval, but differ in their basic hazard function. The main limitation of PWP models is that they can give unbelievable results for higher-order events. As the sequence of events increases, the number of objects in the risk range decreases. The construction of the data set in the WLW model allows us to avoid this problem.

The analyses of the labour market can examine subsequent periods of unemployment or employment. The studies on the duration of unemployment in the labour market in Szczecin show that for the first episodes gender was a strong determinant of the intensity of de-registrations for any reason, taking up employment and removal. In the case of the second episodes, gender determined the intensity of de-registration for any reason and removal from the labour market. For the third episodes, gender was the only determinant of de-registration. It did not determine the intensity of the fourth and subsequent de-registrations and de-registrations for other reasons for each episode. In these studies, a stratified model of proportional Cox hazard in baseline and alternative versions were used to assess the intensity of the registrations (Bieszk-Stolorz, 2018a; 2018b).

4. Data used in the study

In the study, individual data on the unemployed persons registered in the Poviat Labour Office in Szczecin, generated from the Syriusz system, were used. The cohort consisted of persons registered for the first time in 2016 and were observed until the end of 2017. The entire history of their registrations was analysed. Their subsequent registration in the office was accepted as an event. As a starting point for the observation of each person ($t = 0$), his/her first registration was accepted. Each history consists of events, i.e. successive registrations in the office. These are recurrent events. Among the observed persons, the highest number of them consisted of persons that did not register again in the analysed period (2808 persons) and there were 836 returning persons. A total of 3644 histories were analysed. After a preliminary analysis of the number of events in the history of registrations, it was decided to divide them into four groups: without subsequent return, with one, two and three or more events. The separation of the latter group resulted from the small number of people registered with at least three events. The number of separated groups is presented in Table 1. Due to such defined events, censored observations appeared. These are observations in which a given unit suffered the k-th event subsequent registration at the office for $k = 1, 2, 3$), and did not experience, by the end of 2017, an event with the number $k + 1$. In the case of the study, all observations are of the same nature as object 4 in Figure 2. The observation period ended at the end of 2017, but the observed persons may register again with the labour office in the future. The calculations were performed in the R environment by means of the `survival` package.

Table 1. History of registrations by the number of events

Number of events (subsequent registrations)	Number of histories
0	2,808
1	648
2	144
3 and more	44
Total	3,644

Source: own calculations

The study took into account four characteristics of the unemployed: gender – a dichotomous variable: gender: women (1), men (0, reference group); education – five levels: at most lower secondary (S_1, reference group), basic vocational (S_2), general secondary (S_3), secondary vocational (S_4) and higher (S_5); age – six age groups: 18–24 (W_1, reference group), 25–34 (W_2), 35–44 (W_3), 45–54 (W_4), 55–59 (W_5) and 60+ (W_6); seniority (D) – two groups: people with no professional experience (0, reference group) and people with professional experience (1).

5. Analysis of subsequent registrations in the labour office

The study was carried out in two stages. The first one consisted in a joint analysis of events, and the second one was an analysis for subsequent events. Both stages were preceded by an initial analysis of the duration of individual episodes in order to justify the choice of the methods used. First, the median and the maximum duration of subsequent event-end episodes were determined (without taking into account censored observations). Both the median and the maximum time in the analysed period were shortened. It follows that if a person re-registered in the office, they did so more and more quickly (Table 2). In the next step, the Kaplan-Meier estimators were calculated (Figure 6).

Table 2. Median and maximum time to occurrence of subsequent events

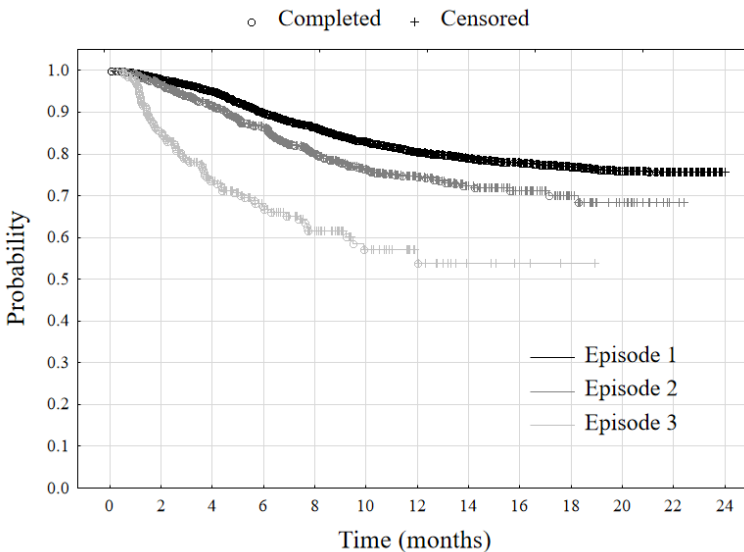| Number of subsequent deregistrations	Median (months)	Maximum (months)
1	6.7	21.3
2	5.1	18.3
3 and more	2.2	12.0

Source: own calculations

Figure 6. Kaplan-Meier estimators
Source: own calculations

The analysis of Kaplan-Meier's estimators indicates that the probability of survival decreases with the observation of recurrent events. Subsequent events carry an increasing risk of occurrence. Differences between the survival curves (confirmed by a log-rank test for each pair of duration curves) therefore justify the use of methods for analysing multiple events.

In the first stage of the study, a total analysis of the events was carried out using all four models: PWP-CP, PWP-GT, AG and WLW. It was examined whether there were significant differences between the successive events and whether the characteristics of the unemployed influenced the risk of subsequent registrations in the office. The results are presented in Table 3.

Table 3. Hazard ratios for all episodes combined based
on the PWP-CP, PWP-GT, AG and WLW models

Variable	PWP-CP	PWP-GT	AG	WLW
P	0.9752	0.9702	0.9740	0.9799
S_2	0.7037**	0.7005**	0.6615**	0.6345**
S_3	0.6718***	0.6810***	0.6422***	0.6182***
S_4	0.6059***	0.6128***	0.5661***	0.5397***
S_5	0.8225*	0.8330	0.7935*	0.7744*
W_2	0.9013	0.9067	0.9457	0.9485
W_3	0.6979**	0.6948**	0.6839**	0.6649**
W_4	0.7884	0.7901	0.7891	0.7671
W_5	0.6225**	0.6231**	0.6100**	0.5947*
W_6	0.8419	0.8309	0.8567	0.8380
D	0.9605	0.9655	0.9415	0.9318

*** $p = 0.001$, ** $p = 0.01$, * $p = 0.05$.

Source: own calculations

All the models gave similar results. Gender and seniority did not determine the risk of subsequent registration in the office. People with at most lower secondary education or up to 24 years of age were most at risk of being subsequently registered in the office. This risk decreased as education levels increased. However, for people with higher education, it increased again. In the case of age groups, the lowest risk was observed in the 55–59 age group. In both models, the risk was 38% lower than for people aged 18–24. For people aged 60+, the parameters of the model were statistically insignificant.

In the second stage of the study, the relative intensity of subsequent events was analysed. It was examined whether the characteristics of unemployed people influenced the risk of k-th registration in the office. In this case, models based

on Cox's layered regression model, i.e. models, were used: PWP-CP, PWP-GT and WLW. Table 4 shows the hazard quotients for subsequent episodes determined using these models.

Table 4. Hazard ratios for subsequent episodes based on PWP-CP, PWP-GT and WLW models

Variable	Model	Episode 1	Episode 2	Episode 3
P	PWP-CP	1.0131	1.0042	0.7311
	PWP-GT	1.0131	1.0106	0.6478
	WLW	1.0131	1.0395	0.6416
S_2	PWP-CP	0.6220***	1.0532	0.7836
	PWP-GT	0.6220***	1.0242	0.6246
	WLW	0.6220***	0.7195	0.5013
S_3	PWP-CP	0.5829***	1.0947	0.8557
	PWP-GT	0.5829***	1.1150	0.9083
	WLW	0.5829***	0.7373	0.8006
S_4	PWP-CP	0.5387***	0.9447	0.6452*
	PWP-GT	0.5387***	0.9460	0.6084
	WLW	0.5387***	0.5902*	0.3161
S_5	PWP-CP	0.7051***	1.0081	1.6640
	PWP-GT	0.7051***	1.0553	1.5302
	WLW	0.7051***	0.7865	0.7865
W_2	PWP-CP	0.8270*	1.0879	1.7620
	PWP-GT	0.8270*	1.0879	1.9246
	WLW	0.8270*	1.0835	3.2777*
W_3	PWP-CP	0.6727**	0.8615	0.5989
	PWP-GT	0.6727**	0.8586	0.5746
	WLW	0.6727**	0.6859	0.4383
W_4	PWP-CP	0.7140*	0.8704	2.3243
	PWP-GT	0.7140*	0.8330	2.6201
	WLW	0.7140*	0.6507	3.0756
W_5	PWP-CP	0.5556**	0.8786	0.9628
	PWP-GT	0.5556**	0.9088	0.9712
	WLW	0.5556**	0.7490	0.7295
W_6	PWP-CP	0.6381	1.5373	2.3446
	PWP-GT	0.6381	1.5251	2.1455
	WLW	0.6381	1.1591	4.9112*
D	PWP-CP	0.9680	0.9291	1.2199
	PWP-GT	0.9680	0.9435	1.2767
	WLW	0.9680	0.8161	0.9230

*** $p = 0.001$, ** $p = 0.01$, * $p = 0.05$.

Source: own calculations

According to the assumptions, the estimators for the first event are the same. Women did not differ significantly from men in terms of the risk of subsequent

registration for each episode. Education and age (except for the oldest persons) differentiated between the unemployed only in the case of the first event. Professional experience was not a determinant of the intensity of unemployment exit. The analysed features were not determinants (with small exceptions) of the intensity of second, third and subsequent registrations in the labour office.

6. Conclusions

The paper presents a review of four methods useful in the analysis of data of subsequent events and applies them to model the event consisting in subsequent registrations in the labour office. The declining median and the maximum value of time until the next registration indicate that people who repeatedly register at the office have problems with finding a permanent job or are not interested in it. In the analysed period, only age and education influenced the risk of multiple registrations at the Poviat Labour Office in Szczecin. Gender and seniority did not have a significant impact. The analysis performed in each stratum, i.e. for subsequent registrations, confirmed the impact of the same features in the first stratum, i.e. on the first subsequent registration. In general, it can be stated that the analysed characteristics of the unemployed did not have a significant impact on the second and subsequent returns to the labour office. The risk of subsequent registrations was the highest in the case of people with low education (and then with higher education) and aged up to 24 (and then at the age of 25–34). The result could be influenced by the number of strata. The number of observations decreased with the next event. This may have affected the reliability of estimates in the last strata, especially in the PWP models. However, similar results were also obtained by means of the WLW model, where the construction of the data set eliminated the problem of decreasing number of events in particular strata.

Some researchers assume that, if there are significant differences in the course of the duration curves for subsequent episodes, it is appropriate to use the PWP or WLW model. If, in addition, it is known that there is a relationship between subsequent episodes, the PWP (CP or GT) model (Sousa-Ferreira, Abreu, 2019) gives better results. In the presented study, the first assumption is satisfied. Nothing is known about the satisfaction of the second one. The conducted study has not provided a clear answer to the question which model is better. This is probably due to the fact that the analysed features did not significantly affect the subsequent episodes. Perhaps the differences in the course of the duration curves are caused by other features of the examined units.

Acknowledgements

References

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

Andersen P., Gill R. (1982), *Cox's Regression Model for Counting Processes: A Large Sample Study*, "The Annals of Statistics", vol. 10, no. 4, pp. 1100–1120, http://dx.doi.org/10.1214/aos/1176345976

Bieszk-Stolorz B. (2018a), *Badanie czasu trwania w bezrobociu z wykorzystaniem modeli regresji dla zdarzeń powtarzających się*, "Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu", no. 507, pp. 21–29, http://dx.doi.org/10.15611/pn.2018.507.02

Bieszk-Stolorz B. (2018b), *Stratified Cox Model with Interactions in Analysis of Recurrent Events*, "Acta Universitatis Lodziensis. Folia Oeconomica", vol. 3, no. 335, pp. 207–218, http://dx.doi.org/10.18778/0208-6018.335.14

Bieszk-Stolorz B. (2018c), *Ocena wpływu płci na formę wyjścia z bezrobocia*, "Wiadomości Statystyczne", vol. 6, no. 685, pp. 23–38.

Cai J., Schaubel D. (2003), *Analysis of Recurrent Event Data*, [in:] N. Balakrishnan, C. Rao (eds.), *Handbook of Statistics: Advances in Survival Analysis*, vol. 23, Elsevier, North Holland, pp. 603–623.

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

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

Cox D. (1975), *Partial likelihood*, "Biometrika", vol. 62, no. 2, pp. 269–276, http://dx.doi.org/10.1093/biomet/62.2.269

Gałecka-Burdziak E. (2016), *Multiple unemployment spells duration in Poland*, Collegium of Economic Analysis SGH – Working Papers, no. 19/10.

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

Kaplan E.L., Meier P. (1958), *Non-parametric estimation from incomplete observations*, "Journal of American Statistical Association", vol. 53, no. 282, pp. 457–481, http://dx.doi.org/10.2307/2281868

Klein J., Goel P. (eds.) (1992), *Survival Analysis: State of the Art*, Springer Netherlands, Dordrecht.

Kleinbaum D., Klein M. (2012), *Survival Analysis. A Self-Learning Text, Third Edition*, Springer-Verlag, New York.

Machin D., Cheung Y.B., Parmar M.K.B. (2006), *Survival Analysis. A Practical Approach. Second Edition*, John Wiley & Sons, Chichester.

Ozga A., Kieser M., Rauch G. (2018), *A systematic comparison of recurrent event models for application to composite endpoints*, "BMC Medical Research Methodology", vol. 18, no. 2, https://doi.org/10.1186/s12874-017-0462-x

Pepe M.S. (1991), *Inference for Events With Dependent Risks in Multiple Endpoint Studies*, "Journal of the American Statistical Association", vol. 86, no. 415, pp. 770–778, http://dx.doi.org/10.1080/01621459.1991.10475108

Prentice R. L., Williams B. J., Peterson A. V. (1981), *On the regression analysis of multivariate failure time data*, "Biometrika", vol. 68, no. 2, pp. 373–379, https://dx.doi.org/10.1093/biomet/68.2.373

Sączewska-Piotrowska A. (2015), *Badanie ubóstwa z zastosowaniem nieparametrycznej estymacji funkcji przeżycia dla zdarzeń powtarzających się*, "Przegląd Statystyczny", R. LXII, z. 1, pp. 29–51.

Sagara I., Giorgi R., Doumbo O. K., Piarroux R., Gaudart J. (2014), *Modelling recurrent events: comparison of statistical models with continuous and discontinuous risk intervals on recurrent malaria episodes data*, "Malaria Journal", vol. 13, no. 293, https://doi.org/10.1186/1475-2875-13-293

Sousa-Ferreira I., Abreu A. M. (2019), *A review of Cox's model extensions for multiple events*, "IJRDO – Journal of Applied Science", vol. 5, no. 2, pp. 47–62.

Therneau T., Grambsch P. (2000), *Modeling, Survival Data: Extending the Cox Model*, Springer Science & Business Media, New York.

Wei L., Lin D., Weissfeld L. (1989), *Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions*, "Journal of the American Statistical Association", vol. 84, no. 408, pp. 1065–1073, http://dx.doi.org/10.2307/2290084

Modele zdarzeń wielokrotnych w analizie powtórnych rejestracji w urzędzie pracy

Streszczenie: W wielu dziedzinach nauki zachodzi konieczność analizy powtarzających się zdarzeń. W naukach medycznych problemem jest ocena ryzyka nawrotu przewlekłej choroby. W naukach ekonomiczno-społecznych analizować można czas kolejnych wejść i wyjść w sferę ubóstwa, czas kolejnych roszczeń gwarancyjnych lub ubezpieczeniowych, a także czas kolejnych okresów bezrobocia. W badaniach tych w różny sposób można definiować przedziały ryzyka, czyli przedział czasu, w którym dla danej jednostki istnieje ryzyko (lub szansa) wystąpienia zdarzenia.

Badania bezrobocia rejestrowanego w Polsce wskazują na duży odsetek osób powracających do urzędu pracy i rejestrujących się ponownie. Celem artykułu jest analiza ryzyka kolejnych zarejestrowań w urzędzie pracy w zależności od wybranych cech osób bezrobotnych: płci, wieku, wykształcenia oraz stażu pracy.

W badaniu wykorzystano metody analizy trwania. Porównano wyniki otrzymane dla czterech modeli będących rozszerzeniem modelu proporcjonalnego hazardu Coxa. W modelu Andersona-Gila nie rozróżnia się, które zdarzenie wystąpiło jako pierwsze, a które jako kolejne. Istotna jest liczba zdarzeń, które wystąpiły. Dwa modele warunkowe Prentince'a, Williamsa i Petersona oraz Wei, Lin i Weissfelda bazują na warstwowym modelu Coxa. Warstwami są kolejne zdarzenia. Modele te różnią się sposobem wyznaczania przedziałów ryzyka.

W analizowanym okresie tylko wiek i wykształcenie wpływały na ryzyko wielokrotnych zarejestrowań w Powiatowym Urzędzie Pracy w Szczecinie. Płeć i staż pracy nie miały na to istotnego wpływu. Analiza wykonana dla kolejnych zarejestrowań potwierdziła wpływ tych samych cech na pierwsze z kolejnych zarejestrowań. Ogólnie można stwierdzić, że analizowane cechy osób bezrobotnych nie miały istotnego wpływu na drugi i kolejne powroty do urzędu pracy.

Słowa kluczowe: model regresji Coxa, modele zdarzeń wielokrotnych, ryzyko zarejestrowania w urzędzie pracy

JEL: C41, J64

COPE
Member since 2018
JM13703

This journal adheres to the COPE's Core Practices
https://publicationethics.org/core-practices