

Beata Bieszk-Stolorz

Geometric method of determining hazard for the continuous survival function

Folia Oeconomica Stetinensia 15(23)/1, 22-33

2015

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.



DE GRUYTER
OPEN

Folia Oeconomica Stetinensia

DOI: 10.1515/fofi-2015-0031



GEOMETRIC METHOD OF DETERMINING HAZARD FOR THE CONTINUOUS SURVIVAL FUNCTION

Beata Bieszk-Stolorz, Ph.D. Associate Professor

*University of Szczecin
Department of Econometrics and Statistics
Faculty of Economics and Management
Mickiewicza 64, 71-101 Szczecin, Poland
e-mail: beatus@wneiz.pl*

Received 23 January 2015, Accepted 29 April 2015

Abstract

A basic assumption in proportional intensity models is the proportionality, that each covariate has a multiplicative effect on the intensity. The proportionality assumption is a strong assumption which is not always necessarily reasonable and thus needs to be checked. The survival analysis often employs graphic methods to study hazard proportionality. In this paper a geometrical method for determining the value of the hazard function on the basis of the continuous survival function was proposed. This method can be used to compare the intensity of the event for objects belonging to two subgroups of the analysed population. If we have graphs of survival function, then an analysis of the tangents at a specific time and their roots enables us to find the intensity and to study the relationship between them for different subgroups. This method can also be useful when studying the proportionality of hazard. It is a condition for the use of the Cox proportional hazards model. The above method was used to evaluate the effect of unemployment benefit and gender on unemployment and on the intensity of finding a job.

Keywords: non-proportional hazard, continuous survival function, geometric method, unemployment.

JEL classification: C41, J64, J65.

Introduction

Methods of event history analysis are increasingly popular in the modelling of socio-economic phenomena. The survival analysis, depending on the area of application also called the duration or reliability analysis is one of the most important ones. The survival analysis is concerned with studying the time between entry to a study and a subsequent event, such as lifetime (the time from birth to death), the job seeking time or the time of a machine's operation until its breakdown. Subjects are observed from the moment when a given process starts to its termination. The advantage of these methods is that we can include in the observation those subjects that, for various reasons, have dropped out from the study. They form a group of censored observations. These methods can also be used when we do not know the distribution of the random variable T – either discontinuous or continuous – of the time to event.

In this article I propose a geometric method to determine the value of the intensity function on the basis of the behaviour of a survival function when we do not know the distribution of the phenomenon duration.¹ The method can be used to compare the event intensity for subjects belonging to two subgroups of a population of interest. In this article I intend to apply the proposed method to evaluating the probability of failed job search and of employment intensity by the unemployed individuals registered by the Poviát Labour Office in Szczecin over 2012 and 2013.

1. The survival function and the hazard function

A primary function in a survival analysis is the survival function. It is defined as:

$$S(t) = P(T > t) = 1 - F(t) \quad \text{for } t \geq 0 \quad (1)$$

where $F(t)$ is a distribution function of the random variable T which denotes the time to event. Depending on the definition of the distribution function, the survival function is also described as $S(t) = P(T \geq t)$ (Collett, 2003; Selvin, 2008). Thus, the survival function denotes the probability of an event that the subject will survive longer than t , i.e. that the event will not happen sooner than t . The survival function is non-increasing and meets the conditions: $S(0) = 1$ and $\lim_{t \rightarrow +\infty} S(t) = 0$. Because $F(t) = \int_0^t f(u) du$, where $f(t)$ is the probability density function, we have (Klein, Moeschberger, 2003):⁰

¹ See more on the analytic form of the survival function and the intensity for the known survival time distributions in: Huzurbazar (2005).

$$S(t) = \int_t^{+\infty} f(u) du \quad \text{for } t \geq 0 \quad (2)$$

$$f(t) = -\frac{dS(t)}{dt} \quad \text{for } t \geq 0 \quad (3)$$

The second very important function in the survival analysis is the hazard function, sometimes called the intensity function. Here it is assumed that in a given interval of time $(t, t + \Delta t)$ the subject experienced a given event, under the condition that it did not happen until the time t . The probability of such an event to happen equals $P(t < T \leq t + \Delta t \mid T > t)$. In case when $\Delta t \rightarrow 0$ the hazard function $h(t)$ is defined as (Aalen et al., 2008: 6):

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t \mid T > t)}{\Delta t} \quad \text{for } t \geq 0 \quad (4)$$

Formula (4) can be also written:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{S(t) - S(t + \Delta t)}{S(t)} \quad \text{for } t \geq 0 \quad (5)$$

therefore:

$$h(t) = -\frac{\frac{dS(t)}{dt}}{S(t)} \quad \text{for } t \geq 0 \quad (6)$$

Since the first function derivative is interpreted as velocity, the formula (5) expresses relative velocity taking the reversed sign. The hazard function is a conditional density function of the event to happen at the time t given that the event did not take place until this time. Because the survival function is non-increasing, its derivative is non-positive. Therefore the hazard $h(t)$ is a non-negative function meeting the conditions (Kalbfleisch, Prentice, 2002: 7):

$$\int_0^t h(u) du < +\infty \quad \text{for } t \geq 0 \quad (7)$$

$$\int_0^{+\infty} h(u) du = +\infty \quad (8)$$

Contrary to the survival function, the hazard function does not have to be monotonic (Kleinbaum, Klein, 2005: 12–13).

2. Geometric method of determining the intensity of events

A statistical test for checking the proportionality assumption has been considered by a number of authors (Cox, 1972; Schoenfeld, 1982; Wei, 1984; O'Quigley, Pessione, 1989; Lin et al., 1993; Grambsch, Thernau, 1994; Scheike, Martinussen, 2004; Kvaløy, Neef, 2004; Kraus, 2007). The survival analysis often employs graphic methods to study hazard proportionality (Andersen et al., 1993). It is one of the elementary conditions when using the Cox proportional hazard model for modelling the intensity of events. The most common method consists in testing if the transformed survival curves (determined by means of, for instance, the Kaplan-Meier estimator) for individual categories of the variable in question are parallel (Kleinbaum, Klein, 2005: 137–145). To examine the 'parallelism' we transform the survival function values $S(t, X)$ into $-\ln(-\ln S(t, X))$ (or $\ln(-\ln S(t, X))$). The transformation was first proposed by Kalbfleisch and Prentice (2002).

In case of a continuous survival function we can assess the hazard value looking at the survival function diagram. Figure 1 shows the curve (a continuous line) of an optional function $y = S(t)$. A broken line represents the tangent to this curve at the point t . We know that the slope of this tangent is equal to the derivative of the function at the tangency point, thus:

$$\tan\alpha = \frac{dS(t)}{dt} \quad (9)$$

Let the angle β complement the angle α to straight angle. Then, according to reduction formulas, we have a relationship:

$$\tan\beta = \tan(180^\circ - \alpha) = -\tan\alpha \quad (10)$$

therefore:

$$\tan\beta = -\frac{dS(t)}{dt} \quad (11)$$

Let $x = t_1 - t$, where t_1 is a point where the tangent intersects with the x -axis. According to the triangle properties, we have the equality:

$$\frac{S(t)}{x} = \tan\alpha \quad (12)$$

or, comparing formulas (11) and (12):

$$x = -\frac{S(t)}{\frac{dS(t)}{dt}} = \frac{1}{h(t)} \quad (13)$$

Thus, the gap between the analysed moment in time and the tangent zero point equals the reversed hazard calculated at this moment. Therefore, the larger the value x , the smaller the value of hazard $h(t)$. The faster the rate of the survival function decrease, the larger the value of the event intensity.

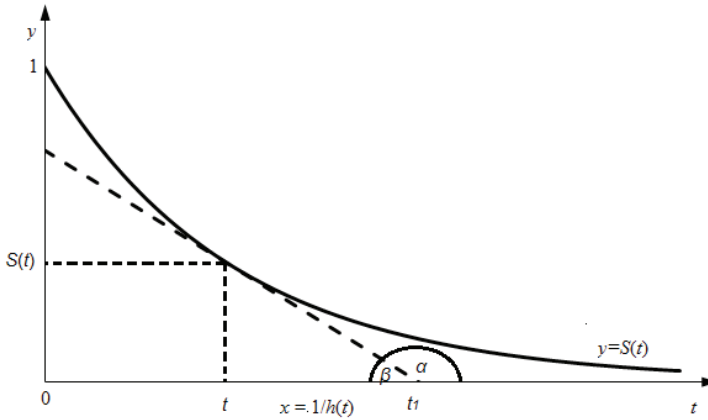


Fig. 1. The relation between the survival curve and the hazard

Source: own study.

3. Geometric method for a comparison of the intensity of events

Formula (13) can be used for comparing the hazard in two groups of the observed population at the time t when we have just the diagrams of the survival function. Figure 2 presents the diagrams of two continuous survival functions $y = S_1(t)$ and $y = S_2(t)$. The tangents to both curves at the time t intersect the x -axis at the points t_{01} and t_{02} , respectively. When analysing the values $t_{01} - t$ and $t_{02} - t$ which, according to Formula (13), corresponds to the values $\frac{1}{h_1(t)}$ and $\frac{1}{h_2(t)}$, we can conclude that $h_1(t) < h_2(t)$.

The values of the hazard function change over the time of observation. In a specific case it may happen that $t_{01} = t_{02}$, or $t_{01} - t = t_{02} - t$. Thus $h_1(t) = h_2(t)$. It means that the event intensities in both subgroups at the time t become equal. Such a situation is presented in Figure 3.

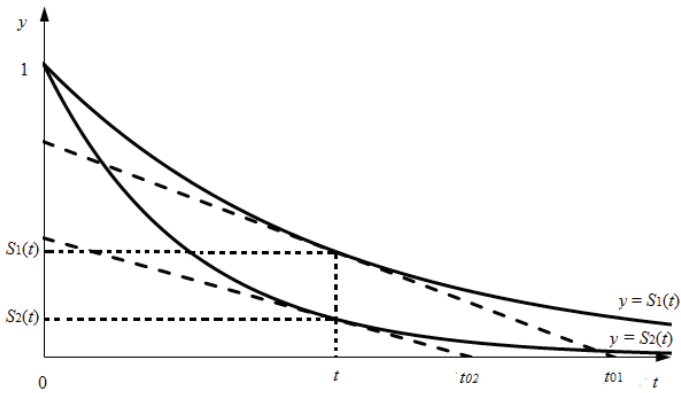


Fig. 2. The relation between the survival curves and the hazards in the two subgroups of the observed population

Source: own study.

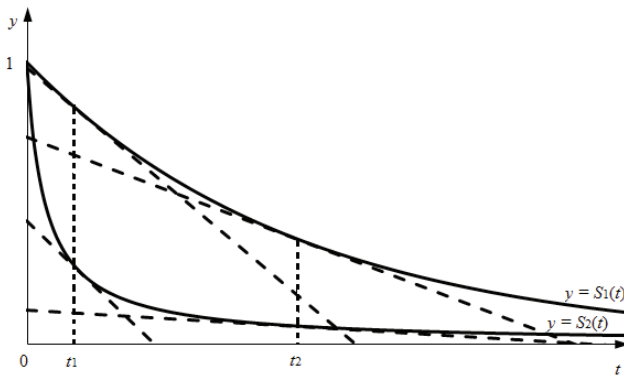


Fig. 3. The analysis of tangents to the survival curves when the hazards in the two subgroups are equal

Source: own study.

This method can be useful when studying hazard proportionality. It is a necessary preliminary element when we want to use the Cox proportional hazard model (Hosmer, Lemeshow, 1999; Machin et al., 2006; O’Quigley, 2008). Figure 4 presents the situation where at the time t_1 we have $h_1(t) > h_2(t)$, while at the time t_2 we have $h_1(t) < h_2(t)$. In this case the survival curves are

intersecting. Due the fact that there is a point where these curves intersect we know that the relation between the event intensities in both subgroups has changed. Remember, however, that the hazards have not become equal at the very point of intersection because we have $S_1(t) = S_2(t)$, but the tangents to the survival curves at the point of intersection are inclined with respect to the

x -axis at different angles, so $\frac{dS_1(t)}{dt} \neq \frac{dS_2(t)}{dt}$, therefore $h_1(t) \neq h_2(t)$. Generally, the hazards

will become equal at these points of time where $-\frac{\frac{dS_1(t)}{dt}}{S_1(t)} = -\frac{\frac{dS_2(t)}{dt}}{S_2(t)}$. It will happen when the

survival curves coincide in a certain interval. Then the function values and their derivatives will be equal. However, it is possible that the equality will be satisfied for other points of the survival curves as well – see Figure 4. Both survival functions are continuous, so the slopes of tangents drawn at subsequent points of these curves change in a continuous manner as well. It can be concluded that the changes in the hazard function value are also continuous. If $h_1(t) > h_2(t)$ changes into $h_1(t) < h_2(t)$, there is a point t_0 at which $h_1(t_0) = h_2(t_0)$ and it is not the point where the given survival curves intersect.

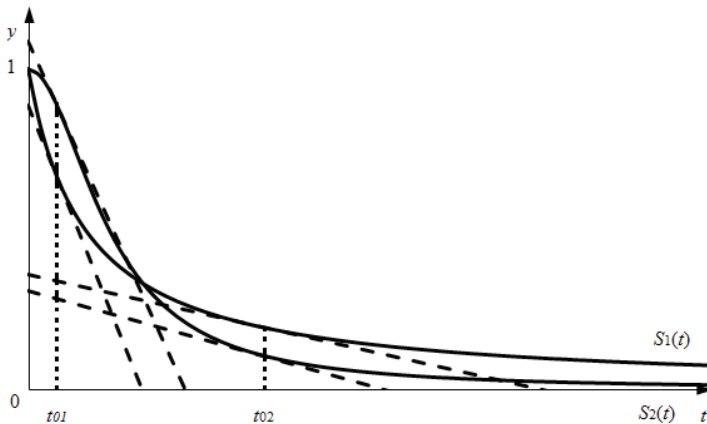


Fig. 4. The analysis of tangents to the survival curves when the relation between the hazards in the two subgroups has changed

Source: own study.

4. Unemployment benefit and gender as determinants of unemployment leaving intensity

The research was based on the individual data of 21, 398 jobseekers registered by the Poviát Labour Office in Szczecin in 2012 and observed by the end of 2013. The data pertain to the seventh most populated city in Poland (approximately 400 thousand in 2013). Due to varying characteristics of the regions, the findings of this study cannot be discussed in a broader context. From the social policy point of view, however, this kind of research is highly recommendable. The author determined the dates of the unemployed person's registration. The observation was considered to be complete if it ended with the termination of unemployment. If the de-registration had not taken place before the end of 2013 or if it had taken place before the end of 2013 due to a reason other than finding a job, the data were regarded as censored. Additionally, the author looked at the impact of unemployment benefit on the spell of unemployment duration and employment intensity. Figure 5 shows the curves of survival in unemployment, while Figure 6 – the graphs of hazards in two groups of the observed cohorts of the unemployed: those with and without unemployment benefits. The survival curves are intersecting, which is analogous to the situation presented in Figure 4. Before the 11th month the probability of not finding a job by the jobseekers who were granted unemployment benefit was higher than in the case of the ones without the benefit. In the 11th month their employment odds levelled off, which was then followed by a reversal of the trend. The change in the employment intensity had taken place earlier. The jobseekers not entitled to employment benefit were finding jobs with higher intensity by the second month, when the trend was reversed. Therefore, we can conclude that in this particular situation the proportionality of hazards does not take place.

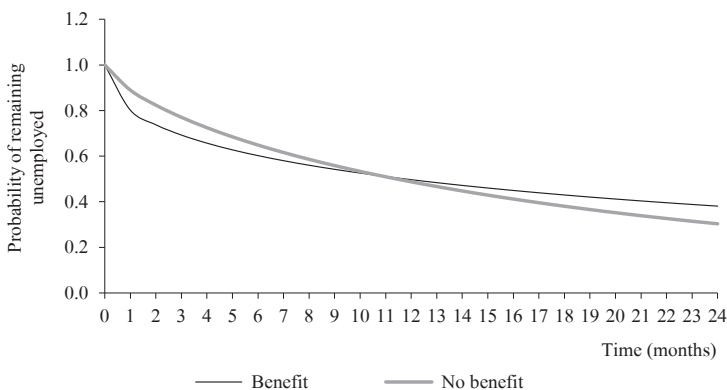


Fig. 5. The curves of survival in unemployment depending on unemployment benefit

Source: own study based on the data from the Poviát Labour Office Szczecin.

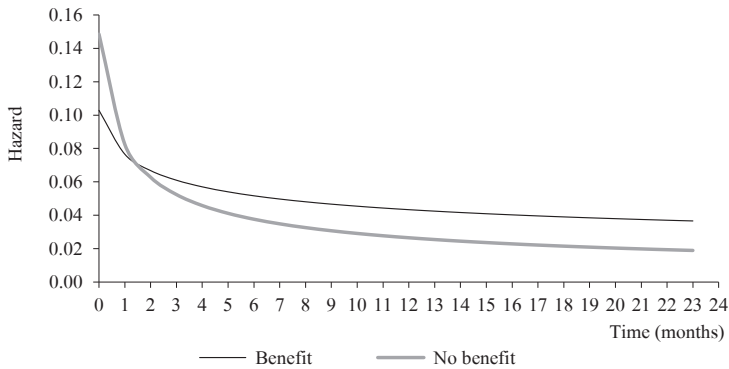


Fig. 6. Intensity of finding employment by unemployment benefit

Source: own study based on the data from the Poviát Labour of Szczecin.

The author also examined the effect of gender on the survival in unemployment and job finding intensity. In the period of observation women were more successful in job seeking than men. The absence of points where the survival curves intersected means that the above relation continued throughout the observation time (Figure 7). The hazard functions determined for both gender groups were declining and parallel. This proves that the hazard is proportional (Figure 8). Women's job seeking efforts were more intense than men's and the hazard coefficient (women/men) over the period of 24 months was approximately 1.12.

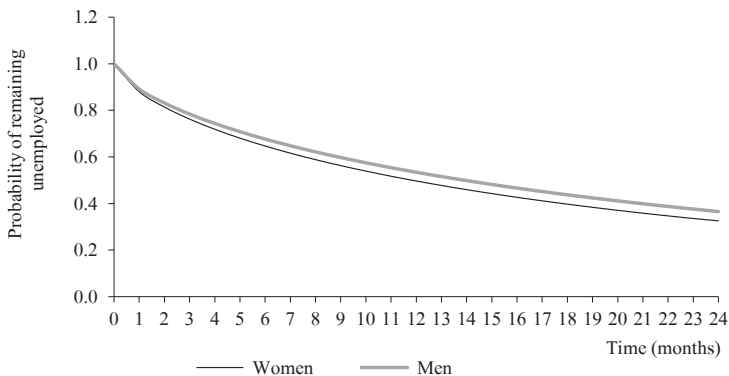


Fig. 7. The curves of survival in unemployment depending on gender

Source: own study based on data from the Poviát Labour of Szczecin.

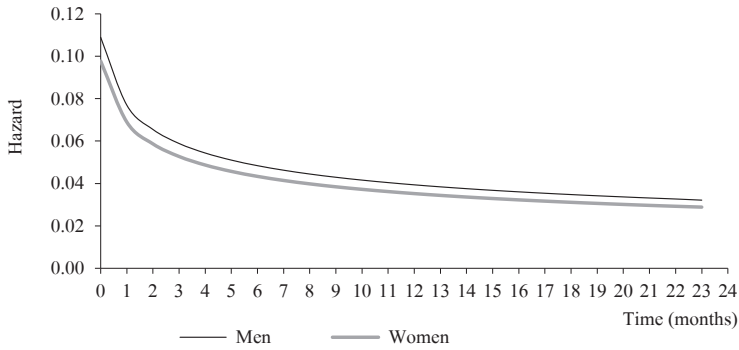


Fig. 8. Intensity of finding employment by gender

Source: own study based on data from the Poviatt Labour of Szczecin.

Conclusions

In the survival analysis we can study the hazard proportionality by means of a graphic method where the curves of the observed subgroups' survival are transformed into $-\ln(-\ln S(t))$. If the curves that have been transformed in this manner are parallel, we can use the Cox proportional hazard model to model the intensity of events. In this case the parallelism of the curves means that the distances calculated along the y -axis are the same. In practice it is sufficient for these curves to be close to each other. Then, we assume that the hazard quotients of both subgroups are constant throughout the time of the event. The graphic method mentioned above allows to assess the hazard rate and to examine the relation between hazards in the two subgroups of the observed population. Also, it is possible to assess the proportionality while not having to take a double logarithm of the survival function. In the case when there is no proportionality, we can preliminarily estimate these moments of time when the intensities have got equal, which may help us choose a proper form of the Cox non-proportional hazard model. Additionally, the analysis of the length of job seeking spells shows that in the observed period the jobseekers not entitled to unemployment benefit found jobs sooner (by the 11th month of registration) and with higher intensity (by the 2nd month of registration). As far as the gender perspective is concerned, women were leaving unemployment faster and with higher intensity (by 12%) than men.

References

- Aalen, O.O., Borgen, O. & Gjessing H.K. (2008). *Survival and Event History Analysis. A Process Point of View*. New York: Springer.
- Andersen, P.K., Borgan, Ø, Gill, R.D. & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.
- Bieszk-Stolorz, B. (2013). *Analiza historii zdarzeń w badaniu bezrobocia*. Szczecin: Volumina.pl.
- Collett, D. (2003). *Modelling Survival Data in Medical Research*. Florida: Chapman and Hall.
- Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B (Methodological)*, 34: 187–220.
- Grambsch, P.M. & Therneau, T.M. (1994). Proportional hazards test and diagnostics based on weighted residuals. *Biometrika*, 81: 515–526.
- Huzurbazar, A.V. (2005). *Flowgraph Models for Multistate Time-to-Event-Data*. New Jersey: Wiley.
- Hosmer, D.W. & Lemeshow, S. (1999). *Applied Survival Analysis. Regression Modeling of Time to Event Data*. New York: Wiley.
- Kalbfleisch, J.D. & Prentice, R.L. (2002). *The statistical analysis of failure time data. 2nd Edition*. New York: Wiley.
- Klein, J.P. & Moeschberger, M.L. (2003). *Survival Analysis. Techniques for Censored and Truncated Data. 2nd Edition*. New York: Springer.
- Kleinbaum, D.G. & Klein, M. (2005). *Survival Analysis. 2nd Edition*, New York: Springer.
- Kraus, D. (2007). Data-driven smooth tests of the proportional hazards assumption. *Lifetime Data Analysis*, 13: 1–16.
- Kvaløy, J.T. & Neef, L.R. (2004). Tests for the Proportional Intensity Assumption Based on the Score Process. *Lifetime Data Analysis*, 10: 139–157.
- Lin, D.Y., Wei, L.J. & Ying, Z. (1993). Checking the Cox model with cumulative sums of the martingale-based residuals. *Biometrika*, 80: 557–572.
- Machin, D., Cheung, Y.B. & Parmar, M.K. (2006). *Survival Analysis. A practical Approach. 2nd Edition*. Chichester: Wiley.
- O’Quigley, J. & Pessione, F. (1989). Score tests for homogeneity effects in the proportional hazards model. *Biometrics*, 45: 135–144.
- O’Quigley, J. (2008). *Proportional Hazard Regression*, New York: Springer.
- Scheike, T. & Martinussen, T. (2004). On estimation and tests of time-varying effects in the proportional hazards model. *Scandinavian Journal of Statistics*, 31: 51–62.

- Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69: 239–241.
- Selvin, S. (2008). *Survival Analysis for Epidemiologic and Medical Research. A Practical Guide*. Cambridge: Cambridge University Press.
- Wei, L.J. (1984). Testing goodness of fit for the proportional hazards model with censored observations. *Journal of the American Statistical Association*, 79: 649–652.