HAZARD FUNCTION AS A TOOL TO DIAGNOSE BUSINESS LIQUIDATION

Iwona Markowicz, Assoc. Prof.

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

Received 30 October 2013, Accepted 17 January 2014

Abstract

Research of firms cohorts, such as human cohorts, are carried out by means of the methods of survival analysis (duration). The paper presents the possibility of using expectancy tables and intensity function (hazard) of firms liquidation. This function may take a typical shape such as shape of a “bathtub” – the intensity of deaths, or the shape of an inverted U – the intensity of firms liquidations.

Keywords: firms, expectancy tables, hazard (intensity).

JEL classification: M13, C14.
Introduction

The research into the population of economic entities provide important information about both emerging and closing down firms which reflect not only the economic conditions but also the entrepreneurs’ moods. Empirical studies are the basis for creating and developing theories which explain the probability of company survival. Out of the many trends in company survival theory K. Poznańska chose the trend deriving from the theory about company organization, a trend deriving from the economic of industry and a trend deriving from research about labour market. From the viewpoint of this article topic, the most accurate is the first of the above mentioned trends where the determinants of survival and development of companies are structural features of an organisation, such as their age. The most important concepts developed under this trend are:

a) liability of smallness – small firms have less chances to survive than the large ones (the benefits of scale in production, scientific research, demand and distribution, greater chances to attract skilled workers by guaranteeing job security and development opportunities);

b) liability of newness – the most important survival factor is a length of a business life on the market (new firms have a tendency to leave the market fast);

c) ecological economy – changes in industrial branches are similar to changes in biological populations. The structure and dynamics of the economic entities population are influenced by: the probability of a new company to be set up, the probability of a new company to survive and the growth rate of the companies which have survived.

As it was mentioned above, the concepts of company survival appear and develop basing on the research which contributes to developing knowledge in this topic.

The purpose of this article is to present the possibilities of using the survival analysis methods in the analysis of the company lifespan and to find out whether the hazard function for companies in Szczecin has the inverted-U shape. The study covers the firms established in Szczecin in 2001–2006.

1. Analysis of the company lifespan

In Poland the issues of company lifespan (the vitality of company cohorts including their status, structure, dynamics, intensity and lifespan) were studied by for example:

a) G. Dehnel (2010) – who deals with micro-firms; beside the small areas statistic method, used to evaluate those companies development, she used such indicators as:
the dynamics of the number of companies, the entry rate, the exit rate, the turnover and survival rates; she focused on Polish micro-firms (employing less than 10 workers) operating in 2001–2005;
b) K. Poznańska (2007) – with J. Jackson, J. Klich and K. Poznańska did research based on the GUS database concerning firms established in Poland in 1990–1997 and employing more than five workers; that database allowed to observe the company cohorts (starting their business in a given year) in terms of their survival over the next years including their size of employment, value of sales, earnings, province, section;
c) P. Dominiak (2005) – presented the definitions of concepts related to “demography” of firms, a model of factors shaping the size and structure of the population of companies and the results of analyses mainly for European countries, but also for Japan and the USA;
d) F. Bławat (2004) – with F. Bławat, N. Daszkiewicz, T. Korol, J.Cz. Ossowski, B. Prusak and J. Wasilczuk did the research of 185 small and medium firms in the Gdańskie voivodship in 1992 and 1996; the aim of their research was to define the firms’ and entrepreneurs’ features having influence on their ability to survive and develop;
e) I. Markowicz (2000) – except apart from the small companies’ profitability, she was also involved in their vitality; her studies covered small entities in the Szczecińskie voivodship in 1990–1997 and she analysed the number of established and closed down firms, their lifespan and indicators of the establishment and liquidation;

The analysis of corporate vitality is a mix of statistic methods of the ratio analysis and the analysis of duration applied to conduct comprehensive research of the status, structure, dynamics, intensity and modeling the lifespan of the population of companies\(^2\). The article presented the possibility to use the company lifespan tables belonging to the methods of survival analysis. Other authors emphasize that the separation of surviving analysis from the standard procedures of data analysis is dictated by the fact that the data concerning the duration of an investigated phenomenon have a non-symmetric distribution and are usually incomplete.

The company vitality analysis includes the research into the lifespan of economic entities on the national market. The company lifespan is a random variable, which is denoted as \(T\) and which expresses the time from the day of entry in the REGON registry to the day of its sign off. Length of the companies’ lifespan and the intensity of their liquidation is not constant over space and time. It is impossible to answer the question: how long will exist a firm which has just been
established? Therefore the company lifespan is regarded as a random variable. The attempts have been made to adapt the methods used in the analysis of human lifespan (in demography, insurance or medicine) to the analysis of a variety of phenomena duration, including the analysis of the company lifespan. Further in the article the author will discuss the basic measures used in the analysis of the company lifespan. The cumulative distribution of the random variable \( T \) can be written as:

\[
F(t) = P(T \leq t), \quad t \geq 0
\]

This is a probability that the random variable \( T \) will take the value smaller than or equal to \( t \), i.e. the probability that since the moment of its set up the firm will survive for a maximum of the time \( t \). The function \( F(t) \) is called the survival function. Because the event complementary to the event “survival by the time \( t \)” is an event “survival beyond the time \( t \)”, this is why the function complementary to the survival function is an outliving function, which is written:

\[
S(t) = 1 - F(t) = P(T > t), \quad t \geq 0
\]

This is a probability of survival (staying in cohort) beyond the time \( t \), i.e. the likelihood that an event called failure will appear after the time \( t \), i.e. the firm will be liquidated after the time \( t \).

At the moment \( t = 0 \), that is at the moment of setting up the firm, \( S(t) = 1 \), and with the passage of time this function is decreasing. The rate at which the survival function is decreasing depends on the value \( t \), i.e. on the lifespan of a given firm and is defined as the hazard intensity function (hazard function):

\[
h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}
\]

where \( f(t) \) – the function of density probability expressed by the formula:

\[
f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}
\]

In demography the function \( h(t) \) is called the intensity function of mortality and expresses the probability of death by persons in some age for short increments of time \((\Delta t)^t \). In the analysis of firms survival it is an intensity function of firms liquidation, which means the probability of their exit in a short period of time provide that it does not take place before the moment \( t \).

Survival models can be classified by the type of intensity function\( ^5 \). This function can be constant, decreasing or increasing. Many surveys conducted in different fields, have confirmed that units are usually not subject to the same law of intensity throughout the whole period
of their existence. A typical example is a diagram of the course of the mortality intensity function which takes a so called bathtub shape. It follows three different stages: fast decreasing mortality intensity after leaving infancy, constant mortality intensity of young persons and increasing mortality intensity among the elderly.

The function $S(t)$ characterizes the probability distribution of the random variable $T$ and defines the survival model and the cumulative distribution function $F(t)$ – the survival model. If the distribution of both the random variable $T$ and the form of analytic function $S(t)$ are known, then the surviving model is parametrical. But when this distribution is not known, then the model is non-parametrical or semi-parametrical (the function is partly specified).

The distributions of lifespans can be different depending on the type of research process, concrete population, subpopulation, individual reactions of the research subjects or the conditions of the process. As A. Balicki says, these circumstances result in the fact that it is impossible to determine one theoretical distribution of the random variable which would be an appropriate model of many empirical distributions of lifespans. In different fields of science we evaluate the probabilistic functions of duration (survival) models in a way that is non-parametric for given cohort or census data. Of course, there have been some attempts to formulate the laws that would at least approximately describe distributions observed in reality. For example, the following distributions were used as parametric models of survival: steady, exponential, Weibull’s (generalization of exponential distributions), normal, gamma, Erlang’s (special example of gamma distribution), logarithmically – normal, logistic, loglogistic, Gompertz’s, Gompertz-Makeham’s. These distributions and formulated laws concerned the research into demography (mortality), medicine (morbidity) and technology (materials durability). The research typically dealt with the mortality of human population, but as M. Matloka put it, “too many causes of death made it impossible to formulate a theoretical distribution which could be used to approximate the distributions of live duration.” This is why there are so many analytical formulas of the mortality intensity or the survival function.

2. Non-parametric model

Modelling of the company lifespan, that is estimating the distribution of the survival probability, makes it possible to identify the risk factors that affect the survival time distribution. In analysis of phenomenon history, we can distinguish such models as: parametric, semi-parametric and non-parametric. As it has been mentioned before, building the parametric models requires the adoption of a theoretical distribution of the observed variable, which in case of
firms is not possible. In article the author uses a non-parametric model, that is the survival tables (a continuous time model).

Main purpose of the survival analysis is to assess the survival model in order to find the regularity of the survival process. The survival tables create a non-parametric model, because they do not include the assumptions about the analytical form of the distribution of the time of expecting for the event to happen\textsuperscript{11}. In these tables there are given such quantities as the number of units experiencing a given event or the number of units surviving for definite intervals of time. This model, which is called tabular, is mainly used in demography as tables of survival (mortality). Then, usually the yearly range of age is adopted, and these tables are called the full tables. If the data are taken in the range longer than a year, mostly this is a five year range, then we can talk about the abbreviated tables\textsuperscript{12}. Adopting the “correct” length of survival range in the research into the company vitality is difficult and it is impossible to find a explicit instructions in the literature. The author has decided on a 3-month range for cohort tables and a yearly range for cross-section tables.

Apart from the classification into the full and abbreviated tables, we can also use the categories of unclassified and model tables\textsuperscript{13}. In unclassified tables the value of table functions are calculated on the basis of crude statistical material, while in the model ones (or levelled, hypothetical or perspective\textsuperscript{14}) the function values result from theoretical assumptions adopted for to research needs.

Because it is not possible to specify the correct length of the survival time, the author has decided not to use the division of tables into the full and abbreviated ones. Instead, they are specified as 3 monthly or yearly (in accordance with the adopted division into the cohort and cross-section tables). All the proposed tables are unclassified (empirical). In the studies on company survival, it is important to divide the tables into the cohort (longitudinal) and cross-section (transverse, current, census) ones\textsuperscript{15}. A cohort is a group of people or objects, singled out on the basis of the jointly experienced event (such as a demographic or social even), in a specified place and time. If this shared event is a birthday (set-up), then this cohort is called a generation. The changes having place in a generation could be considered according to a common or an individual calendar (at the moment of birth the time has the value of 0). It is worth highlighting that each of these approaches is unique and they can be combined to make the analysis of the phenomenon range wider. The cohort tables are the result of the long-term observation of a given generation from its start until the moment when all the objects leave the cohort or when the observation comes to an end, which means that the considerations have to
include censored units as well. The cross-section tables are an effect of the observation of the units from different generations over a short period of time.

3. Cohort tables and intensity function of hazard

The cohort tables containing the intensity function of firms liquidation were calculated for entities established in Szczecin in 2001–2006 and observed until 2007. In subsequent years 5575, 4483, 4120, 4178, 4586 and 5067 companies were established respectively.

The cohort tables presenting the process of changes occurring over time in the cohort’s composition. A cohort is made of companies set up in a given year. What is recorded is the fact of the company close down, i.e. its exit from the cohort, as well as the time from the moment the observation was launched to the moment of a company leaving the cohort. The cohort tables belong to models with continuous time, but presented here estimates of functions are discrete, which is why the value of the time variable, grouped in equal intervals, was given as the beginning of the interval \( t \). The number of firms that survived \( (n_t) \) was given at the beginning of the interval and calculated as \( n_t = n_{t-1} - (z_t + c_t) \), wherein \( n_t \) for \( t = 0 \) \( (n_0) \) denotes the initial number of companies in the cohort; \( z_t \) marks the number of firms liquidated in the interval \( \langle t, t + 1 \rangle \), that is the number of units which experienced the analysed event in a given interval of time; \( c_t \) denotes the number of firms which did not experience that event by the end of the cohort observation time. That is why for firms from the 2005 cohort (Table 1), i.e. set up in 2005 and observed by the end of 2007, the last interval of time in the Table is the interval of \( \langle 33, 36 \rangle \) months. The firms established in the period from January to March 2005 and not liquidated until 31st December 2007 were acknowledged as censored in the interval starting at the time \( t = 33 \). Similarly, the firms established from April to June 2005 and not liquidated are censored in the interval starting at the time \( t = 30 \), firms established from July to September 2005 – are censored in the interval starting at the time \( t = 27 \), and those established from October to December 2005 – in interval starting at the time \( t = 24 \). For individual cohorts, the time of observation is different, so the number of time intervals is different too\(^{16}\).

The next two values in the cohort table of survival were estimated according to the discrete approach because they can be calculated only for the \( f \) time interval. First of them means the probability of a firm to be liquidated in the time interval \( f_t \) (the conditional probability that the unit is experiencing the event, its probability of exit), defined as the conditional probability of a firm to be liquidated in the interval of survival time \( \langle t, t + 1 \rangle \), providing that the firm has not been liquidated by the time \( t \). The distribution of the company survival cannot be assigned to
any known type of probability distribution. This is why the functions describing the process of the company survival are not known, and the tables of survival contain them their estimates calculated on the basis on empirical data. The probability estimator of firm liquidation in the time interval $\hat{f}_t$ is a ratio of the number of liquidated firms in a given interval of time $z_t$ to the number of firms which survived by the beginning of the interval $n_t$:

$$\hat{f}_t = \frac{z_t}{n_t}$$

(5)

Opposite to the probability of firm liquidation in the time interval is the probability of firm survival in the time interval $p_t$, (the conditional probability that the unit will not experience the event, the probability of survival), defined as the conditional probability for the firm to survive through the interval $\langle t, t + 1 \rangle$, provided that the firm had not been liquidated by the time $t$. The probability estimator of a firm survival in the interval $\hat{p}_t$ is a ratio of the number of firms surviving by the beginnings of the interval $t + 1$ ($n_{t+1}$) to the number of firms which survived by the beginning of the interval $t$ ($n_t$):

$$\hat{p}_t = 1 - \hat{f}_t = \frac{n_{t+1}}{n_t}$$

(6)

The probability of survival $F_t$, of outliving $S_t$ and the hazard intensity $h_t$ are functions continuous in their nature, but in the tables they are presented in a discrete approach. The probability of outliving $F_t$ beyond the interval $\langle t, t + 1 \rangle$ means that since its set up a firm will survive no longer than by the time $t + 1$. Its estimator is a ratio of the number of firms liquidated in the time $\langle 0, t + 1 \rangle$ to the initial number of companies in the cohort:

$$\hat{F}_t = \frac{n_0 - n_{t+1}}{n_0}$$

(7)

A function complementary to the probability of survival is the probability of outliving $S_t$ which, when calculated for interval $\langle t, t + 1 \rangle$, is the probability that the firm will be liquidated after the time $t + 1$. The ratio of the number of companies which survived by the time $t + 1$ to the initial number of companies in the cohort, is the estimator of the outliving probability function $\hat{S}_t$:

$$\hat{S}_t = 1 - \hat{F}_t = \frac{n_{t+1}}{n_0}$$

(8)
In the moment $t = 0$, i.e. in moment of setting up a firm, $S_t = 1$ and this function is decreasing over time. The rate at which the outliving function is decreasing depends on the value of $t$, i.e. on the length of a given firm’s lifespan, and is defined as a function of hazard intensity $\hat{h}_t$. The estimator $\hat{h}_t$ of this function is calculated as the ratio of the probability estimator of a company liquidation in the interval $\langle t, t + 1 \rangle$ to the half of the sum of the estimators of the company probability to outlive beyond the intervals $\langle t, t + 1 \rangle$ and $\langle t - 1, t \rangle$:

$$\hat{h}_t = \frac{\hat{j}_t}{(\hat{S}_t + \hat{S}_{t-1})/2}$$  \hspace{1cm} (9)

A sample cohort table of survival for firms set up in Szczecin in 2005 is shown in Table 1. The graph of the estimated intensity of hazard (firm liquidation) $\hat{h}_t$ is presented in Figure 1.

**Table 1. The cohort table of survival of firms established in Szczecin in 2005 – observation by the end of 2007**

<table>
<thead>
<tr>
<th>Lifespan [months]</th>
<th>Number of firms, which survived</th>
<th>Number of liquidated firms in the interval $\langle t, t + 1 \rangle$</th>
<th>Number of censored firms in the interval $\langle t, t + 1 \rangle$</th>
<th>Estimation of the probability of liquidation of a firm in the interval $\langle t, t + 1 \rangle$</th>
<th>Estimation of the probability of surviving in the interval $\langle t, t + 1 \rangle$</th>
<th>Estimation of the hazard intensity $\hat{h}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4586</td>
<td>89</td>
<td>0</td>
<td>0.0194</td>
<td>0.9806</td>
<td>0.0196</td>
</tr>
<tr>
<td>3</td>
<td>4497</td>
<td>135</td>
<td>0</td>
<td>0.0300</td>
<td>0.9512</td>
<td>0.0311</td>
</tr>
<tr>
<td>6</td>
<td>4362</td>
<td>181</td>
<td>0</td>
<td>0.0415</td>
<td>0.9117</td>
<td>0.0446</td>
</tr>
<tr>
<td>9</td>
<td>4181</td>
<td>176</td>
<td>0</td>
<td>0.0421</td>
<td>0.8733</td>
<td>0.0472</td>
</tr>
<tr>
<td>12</td>
<td>4005</td>
<td>183</td>
<td>0</td>
<td>0.0457</td>
<td>0.8334</td>
<td>0.0535</td>
</tr>
<tr>
<td>15</td>
<td>3822</td>
<td>164</td>
<td>0</td>
<td>0.0429</td>
<td>0.7977</td>
<td>0.0526</td>
</tr>
<tr>
<td>18</td>
<td>3658</td>
<td>157</td>
<td>0</td>
<td>0.0429</td>
<td>0.7634</td>
<td>0.0550</td>
</tr>
<tr>
<td>21</td>
<td>3501</td>
<td>131</td>
<td>0</td>
<td>0.0374</td>
<td>0.7349</td>
<td>0.0500</td>
</tr>
<tr>
<td>24</td>
<td>3370</td>
<td>122</td>
<td>711</td>
<td>0.0362</td>
<td>0.7082</td>
<td>0.0502</td>
</tr>
<tr>
<td>27</td>
<td>2537</td>
<td>61</td>
<td>695</td>
<td>0.0240</td>
<td>0.6949</td>
<td>0.0343</td>
</tr>
<tr>
<td>30</td>
<td>1781</td>
<td>23</td>
<td>812</td>
<td>0.0129</td>
<td>0.6899</td>
<td>0.0187</td>
</tr>
<tr>
<td>33</td>
<td>946</td>
<td>7</td>
<td>939</td>
<td>0.0074</td>
<td>0.6884</td>
<td>0.0107</td>
</tr>
</tbody>
</table>

Source: own study (STATISTICA).

The rate at which the outliving function is decreasing depends on the value of $t$, i.e. on lifespan of a given firm, and it is defined as the function of hazard intensity $h_t$. This is the conditional hazard, because it concerns the interval of time $\langle t, t + 1 \rangle$ provided that the firm has not been liquidated by the time $t$. 
Hazard intensity functions estimated for subsequent cohorts of firms are more and more similar to the inverted-U shape. The hazard function is increasing to a certain maximum and then it was decreasing. Such a shape of the smoothed hazard function has been confirmed by other authors’ study results:

a) D.B. Audretsch, E. Santarelli, M. Vivarelli (1999) – Italian manufacturing firms with its maximum of two years at the point of establishment;

b) A. Bhattacharjee (2005) – listed companies in UK with its maximum at the point of three years post-listing;

c) J. Wagner (1994) – small German manufacturing firms with its maximum around three years after entry;

d) E. Bartelsman, S. Scarpetta, F. Schivardi (2005) – find also a clear inverted-U shaped hazard function for the UK, Italy and the US;

e) C.M. Van Praag (2003) – young, small firms in the US with its maximum at the point of 27 months;

f) P. López-Garcia, S. Puente (2006) – Spanish firms (born between 1995 and 2002 across all sectors in the business economy) with its maximum around the fourth year at the point of establishment;

g) A. Nunes, E. de Morais Sarmento (2009) – Portuguese firms; the hazard function exhibits an inverted-U shape, with a maximum around the sixth year of activity.

In the English-language literature the authors indicate that the hazard function with the inverted-U shape is consistent with the theoretical models of active and passive learning if one assumes that the new firms need some time to learn about their efficiency\textsuperscript{19}. In the initial
period, increased costs of activity are predictable and covered by collected resources. If after some time a firm is profitable, it can survive. If it is not, its survival in the market won't be possible. Experience gained by an entrepreneur in the initial period of enterprise presence in the market are very important and are called the liability of the adolescence, which could be translated as “mature responsibility.” In literature we can also find theoretical models based on the assumption that a hazard function reaches its maximum at the beginning of the firm’s operation to decrease over time.

The cohort studies on firms in Szczecin show that their hazard function maximum is falling. The function of the entities’ liquidation intensity for the 2001 cohorts has not taken a typical inverted-U shape yet, although the highest intensity is observed in the intervals of 57–60 and 60–63 months. The functions for other cohorts are getting more and more similar to the shape discussed in the literature, which means that initially the intensity is increasing and after crossing a certain maximum it starts decreasing. This maximum is getting smaller. On the basis on these results we can come to the conclusion that the firms from other cohorts, i.e. set up later, were adapting to the environment at a faster rate. The intensity function maximum, i.e. a certain critical point, was crossed sooner by subsequent cohorts – firms established in 2001 reached that point after around five years of activity, while the firms set up in 2006 – after a year.

Conclusions

Survival models concern mainly people cohorts, but nowadays they are more and more often applied in many other fields of science. They are used in company lifespan analyses. These models are classified by the type of intensity function which, integrally or in stages, can be constant, decreasing or increasing. A typical example is a course of the mortality function which takes a ‘bathtub’ shape. The results of this study and the research presented in the literature confirm that the typical graphic representation of the intensity function is an inverted-U shaped curve.

Notes

References


