COMPETING RISKS MODELS FOR AN ENTERPRISES DURATION ON THE MARKET

Dominika M. Urbańczyk, MSc

Warsaw University of Life Sciences
Faculty of Economic, Department of Econometrics and Statistics
159 Nowoursynowska Street, 02-776 Warsaw, Poland
e-mail: dominika_urbanczyk@sggw.pl
ORCID: 0000-0002-5134-4693

Received 31 October 2019, Accepted 30 March 2020

Abstract

Research background: Enterprises are an important element of the economy, which explains that the analysis of their duration on the market is an important and willingly undertaken research topic. In the case of complex problems like this, considering only one type of event, which ends the duration, is often insufficient for full understanding.

Purpose: In this paper there is an analysis of the duration of enterprises on the market, taking into account various reasons for the termination of their business activity as well as their characteristics.

Research methodology: A survival analysis can be used to study duration on the market. However, the possibility of considering the waiting time for only one type of event is its important limitation. One solution is to use competing risks. Various competing risks models (naive Kaplan-Meier estimator, subdistribution model, subhazard and cause-specific hazard) are presented and compared with an indication of their advantages and weakness.

Results: The competing risks models are estimated to investigate the impact of the causes of an enterprises liquidation on duration distribution. The greatest risk concerns enterprises with a natural person as the owner (regardless of the reason of failure). For each of the competing risks, it is also indicated that there is a section of activity which adversely affects the ability of firms to survive on the market.

Novelty: A valuable result is considering the reasons for activity termination in the duration analysis for enterprises from the Mazowieckie Voivodeship.

Keywords: competing risks, duration analysis, enterprises survival on the market, Fine-Gray model, subdistribution

JEL classification: C40, C41, C46, C51, C55, M21

Introduction

Enterprises are an important element of the economy, and their condition greatly influences development both on a regional and global level. The importance of the condition of companies (especially those from the SME sector) is the reason for the broad interest in the problem of their ability to survive on the market (Boratyńska, Włoczewska, 2013; Kisielińska, 2016; Markowicz, 2018, 2019; Mikulec, 2017; Ptak-Chmielewska, 2013). In the study of this phenomenon, on the one hand, it is necessary to identify the factors that impact on the enterprises duration on the market. However, for a complete understanding of the issue of enterprises survival, various reasons for the termination of their business operations should also be considered.

A survival analysis, which is a set of statistical methods allowing determining the distribution of non-negative random variables, can be successfully used to study the duration of enterprises on the market. However, the basic survival analysis methods allow considering the waiting time for only one type of event to occur. In the case of complex phenomena, which include economic processes like the survival analysis of firms, this is a significant limitation of this method. In this situation, the solution is to extend survival analysis methods by applying competing risk models. In these kinds of models, the particular forms of business activity termination are considered as various risks.

The aim of this work is to explain the impact of the characteristics of the firms (as place of residence, ownership form and sector of activity) on the distribution of enterprises duration on the market as well as to analyze the duration distributions for different types of business activity termination (like legal bankruptcy or the suspension of business). The application of competing risks models allows this kind of study to be carried out. In addition, the results of various approaches to competing risks are compared with an indication of the main advantages and limitations of the methods.

This research is performed for a cohort of enterprises from the Mazowieckie Voivodeship, which started their businesses in 2007. The analysis is based on data from the National Official Business Register (REGON), additionally supplemented with the values of basic macroeconomic indicators.

1. Analysis methods

This section outlines the applied methodology. First, the basics of a survival analysis are presented. Then, there is explained the idea of the competing risks models that allow analyzing the differences in duration distributions for particular realizations of final event (various risks).

1.1. Basics of a survival analysis

A survival analysis is a set of statistical methods dedicated to study non-negative random variables, especially duration distributions. Let T be a random variable describing the waiting time for the occurrence of the expected event. Then the survival function S , describing the probability that the duration is longer than t , is defined as follows (Klein, Moeschberger, 2003):

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) = \int_t^{\infty} f(z) dz \quad (1)$$

where F is a cumulative distribution function, f is a density function of duration distribution.

In turn, the risk of occurrence of the expected event along the entire duration distribution can be a study by an analysis of the hazard function that is expressed in the following form (Klein, Moeschberger, 2003):

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt \mid T \geq t)}{dt} \quad (2)$$

The Kaplan-Meier estimator is a non-parametric maximum partial likelihood estimator for the survival function. Let t_i be time when at least event occurred, n_i be a number of observations at risk and d_i – a number of events that occurred at time t_i . Then the estimator is given by (Kaplan, Meier, 1958):

$$\hat{S}(t) = \prod_{k=1}^i P(T > t_k \mid T \geq t_{k-1}) = \prod_{k=1}^i \hat{p}_k = \prod_{k=1}^i \left(1 - \frac{d_k}{n_k} \right) \quad (3)$$

The above basic method of a survival analysis (like Kaplan-Meier estimator) can find its application in terms of enterprises duration on the market. However, it allows considering the waiting time for only one type of event to occur. In case of complex phenomena, like an enterprises survival, it is necessary to take into account several forms of the events occurrence. This requires extending basic survival analysis methods by applying competing risk models (Satagopan et al., 2004).

1.2. The extent of the Kaplan-Meier estimator (naive KM estimator)

The basic and intuitive approach to considering competing risks in a survival analysis is to construct the Kaplan-Meier estimator for various causes of the events occurrence. This kind

of estimator is called the naive Kaplan-Meier estimator and for each reason j belonged to the set of competing risks J is as follows (Austin, Lee, Fine, 2016):

$$\hat{S}_j(t) = \prod_{k=1}^i \left(1 - \frac{d_{kj}}{n_k} \right) \quad (4)$$

The estimator is determined separately for each event j excluding other competing risks. For this reason, the estimator often overestimates the probability of the occurrence of event j (Austin et al., 2016). Moreover, a significant limitation is also the necessity to make an assumption about the independence of competing risk in order to obtain a total distribution (including all of the competing risks):

$$\hat{S}(t) = \prod_{j=1}^J \hat{S}_j(t) \quad (5)$$

1.3. The idea of a subdistribution

Due to the limitations of the naive Kaplan-Meier estimator, finding other methods to consider competing risks is required. The idea of a subdistribution (Fine, Gray, 1999), also known as Fine-Gray model, could be one of them.

Let us assume that T is the waiting time for any of the competing events to occur j and ε defines the type of the event. Then for each event j the density function f_j may be defined as follows:

$$f_j(t) = \lim_{dt \rightarrow 0} \frac{P(T \in (t, t+dt) \wedge \varepsilon = j)}{dt} \quad (6)$$

The function f_j is called the subdensity function and it determines the subdistribution.

If a finite number of competing risks is considered, the density function for the distribution including all of the competing events can be obtained as follows:

$$\hat{f}(t) = \sum_{j=1}^J \hat{f}_j(t) \quad (7)$$

It is important that in this case it is not necessary to make an assumption about the independence of competing risks during defining the total distribution.

Similarly to the density function, we have for the cumulative distribution function F (CIF):

$$\hat{F}(t) = \sum_{j=1}^J \hat{F}_j(t) \quad (8)$$

where $\hat{F}_j(t)$ is the cumulative subdistribution function.

1.4. The non-parametric estimator of CIF

Considering subdistributions for competing risks, the non-parametric estimator of CIF is often used (Pintilie, 2006):

$$\hat{F}_j(t) = \sum_{k=1}^i \frac{d_{kj}}{n_k} \hat{S}(t_{k-1}) \quad (9)$$

where $\frac{d_{ij}}{n_i}$ is an estimator of the hazard for the risk j (d_{ij} is the number of entities for that the event j occurred at time t_i) and $\hat{S}$ is the Kaplan-Meier estimator.

The important advantage of the estimator is the fact that no assumption about the independence of competing risk is required.

Moreover, for every j there is:

$$\hat{F}_j(t) \leq 1 - \hat{S}_j(t) \quad (10)$$

where the equality occurs in the absence of competing risks or in general in the case of the independence of events. However, for a censored sample (like in this case) the independence of the events cannot be assumed. That is the reason why the identification of marginal distribution does not allow identifying the distribution of a multidimensional variable for censored samples (Peterson, 1976).

1.5. The two concepts of the hazard function

The hazard function can also be defined for the subdistribution. There are two concepts for determining hazard for competing risks. The way of considering the other competing risks than the one currently being investigated j , is the main difference between the two hazard models.

While determining a cause-specific hazard, events that occur due to j are considered, provided that the entity has survived until time t . If the duration ends for a reason other than j the duration time for this entity is considered as censored. This is partly inconsistent with the consideration of risks occurring for reasons other than j those is used for the cumulative subdistribution function. The hazard function in this case is as follows (Pintilie, 2006):

$$h_j(t) = \lim_{dt \rightarrow 0} \frac{P(T \in (t, t+dt) \wedge \varepsilon = j | T \geq t)}{dt} \quad (11)$$

In turn, in case of subdistribution hazard, we assume that if the event occurs, but for a reason other than that currently being investigated j , we do not consider the entity's duration

as censored in a normal way. This entity is still being considered. As a result of this, there is the same number of events to occur for reason j , but more entities at risk are considered. Hence the hazard values in this approach may be underestimated relative to the cause-specific hazard. It is significant that in this model there are the same approach and assumptions to explanatory variables and competing risks in hazard function and subdistribution CIF (Fine, Gray, 1999). The hazard function can then be given by (Gray, 1988):

$$\tilde{h}_j(t) = \lim_{dt \rightarrow 0} \frac{P(T \in (t, t+dt) \wedge \varepsilon = j | (T \geq t) \vee (T < t \wedge \varepsilon \neq j))}{dt} \quad (12)$$

1.6. The conditional probability (CP)

The conditional probability (CP) function determines the probability that event j occurs by time t , given that no other event has occurred by t (Pepe, Mori, 1993). In the case of two competing risks $\varepsilon = 1$ and $\varepsilon = 2$, it is defined as follows:

$$CP_1(t) = \frac{P(T \leq t, \varepsilon = 1)}{1 - P(T \leq t, \varepsilon = 2)} = \frac{F_1(t)}{1 - F_2(t)} = \frac{F_1(t)}{S(t) + F_1(t)} \quad (13)$$

2. Results of the empirical analysis

This section is devoted to presenting the results of the empirical analysis applying competing risks models. First, the database is described. Then, the cumulative distribution functions and conditional probabilities are estimated for each of the considered reasons of an enterprises business activity termination. Finally, the results of the regression subhazard and cause-specific hazard models are provided and compared.

2.1. Database

In this work data from the National Official Business Register (REGON) for enterprises from the Mazowieckie Voivodeship are applied (the database state at the end of 2017). The analysis relate to a cohort of firms that started their business in 2007. The database consists of a sample of 32,788 enterprises. The applied variables with their description are presented in Table 1.

Table 1. Description of the variables

Variable	Description
<i>date_of_starting</i>	date of starting the activity
<i>date_of_termination</i>	date of termination of the activity without a known reason (no due to bankruptcy or suspending)
<i>date_of_bankruptcy</i>	date of bankruptcy in legal terms
<i>date_of_suspension</i>	date of suspending the activity
<i>place_of_residence</i>	information of the firm's headquarters (in this study Warsaw or outside the administrative boundaries of the capital)
<i>ownership</i>	form of ownership (legal person, organizational entity without legal personality, natural person running economic activities)
<i>PKD2007</i>	section of activity by PKD2007 (Polish Classification of Activities) (A – agriculture; forestry, hunting and fishing; B – mining and quarrying; C – manufacturing; D – electricity, gas, steam, hot water and air conditioning manufacturing and supply; E – water supply, sewerage, waste management and remediation activities; F – construction; G – wholesale and retail trade, repair of motor vehicles including motorcycles; H – transportation and storage; I – accommodation and food service activities; J – information and communication; K – financial and insurance activities; L – real estate activities; M – professional, scientific and technical activities; N – administrative and support service activities; O – public administration and defence, compulsory social security; P – education; Q – human health and social work activities; R – arts, entertainment and recreation activities; S – other service activities; T – households as employers, goods- and services-producing activities of households for own use; U – extraterritorial organisations and bodies)

Source: own elaboration.

In addition, the data from REGON was supplemented with macroeconomic indicators like GDP growth rate, inflation and unemployment in individual years (data source: Statistics Poland). In order to include macroeconomic indicators for particular periods of time in the REGON data, the episode splitting method was used. In this method if a variable changes its value during an episode, the episode is split up at that point of time t_k ($k = 1, \dots, s$), creating two new sub episodes (with constant values of that variable). The episode can be split up several times (from 1 to s) if it is required. All sub episodes, except from the last one, are considered as right-censored. The last sub episode ends at the same state as an original episode (Blossfeld, 2014). As a result, we get a data set that can be used to build models for time-constant variables (such as parametric hazard models).

2.2. Comparison of CIFs, CPs and S functions

In the first stage of the research, the estimators of function S , CIF and CP are applied for each for considering competing risks (termination of the activity without a known reason, suspending the activity and legal bankruptcy). The naive Kaplan-Meier estimator (4) for

survival function and non-parametric estimator for CIF (9) are used. The results are shown in the (Figure 1).

Figure 1. The values of estimators of CP, $1 - S$ and CIF for competing risks: (a) termination of the activity without a known reason, (b) suspending the activity and (c) bankruptcy

Source: own elaboration.

Based on the results, we can conclude that the most frequently occurring cause of enterprise liquidation on the market, observed in the REGON database, is the termination of a business without a known reason. Suspension is less significant, whereas bankruptcy in legal terms is the least frequent.

The estimators of CP, CIF and survival function satisfy the following:

$$CP_j(t) \geq 1 - S_j(t) \geq F_j(t) \quad (14)$$

wherein it is known that equality occurs in the absence of competing risks or when the considered risks are independent (but it cannot be assumed in the case of the censored sample).

The obtained estimators for empirical data satisfy theoretical inequality (14). Moreover, it can be seen that the estimators values are not equal for any of the considered causes of termination of the activity (see Figure 1). Therefore, we cannot assume the mutual independence of competing risks in this case. Basic models like the naive Kaplan-Meier estimator are therefore insufficient to study the duration of enterprises on the market. It seems reasonable to look for more complicated competing risk models but that do not require the occurrence of event independence.

2.3. Empirical CIFs with the characteristics of the enterprises

CIFs for different causes of an enterprises failure on the market are estimated applying empirical data from REGON. In addition, the characteristics of the firms are taken into account. There are features considered such as form of ownership, information if the headquarters are in Warsaw or outside the administrative boundaries of the capital and section of activity by the Polish Classification of Activities (PKD 2007). The model with the activity sector as the explanatory variable has been estimated for all PKD2007 sections present in the database, but in order to increase the transparency of results, only selected sections are presented (the most numerous sections).

Firstly, CIFs are estimated without competing risks (time until termination is considered, depending only on the characteristics of the firm) (Figure 2). In this case, the CIF estimator is reduced to a cumulative distribution function of the duration distribution on the market for enterprises with the investigated features. In this way, we can observe the impact of particular enterprises characteristics on the form of distribution of their duration on the market.

Figure 2. CIFs for various features of enterprises: (a) ownership, (b) place of residence and (c) section of the activity

Source: own elaboration.

Then, in order to examine the duration distributions for different types of activity termination and the relationship between the kind of final event and the characteristics of enterprises, the competing risks model (the concept of subdistribution) is applied with the same explanatory variables. In this case each of the obtained CIF determines the subdistribution for enterprises that have the investigated characteristics and terminate for the result of one of the competing risks (Figures 3–5).

In the case of explanatory variables such as headquarters (Figure 4) and section of the activity (Figure 5), we can observe a clear separation of curves for particular competing risks. The occurrence of differences in distributions for particular cases was verified applying a statistical test for the equality of distributions dedicated to investigate the differences between CIF curves (Gray, 1988). For all cases, the *p-value* is much lower than the significance level $\alpha = 0.01$, therefore we accept the alternative hypothesis (the differences in duration distributions occur for the studied risks). The highest probability of an event occurring is for termination without a known reason, then for suspension and bankruptcy. In turn, for the form of ownership (Figure 3) still the most significant role is played by termination of activity without a known reason, but we cannot conclude that suspension is always more popular than bankruptcy in the legal terms. It is worth noticing that suspension is realized mainly by natural persons.

Figure 3. Results of the competing risks model (1 – termination of the activity without a known reason, 2 – suspension of the activity, 3 – bankruptcy in legal terms) with considered various forms of ownership

Source: own elaboration.

Figure 4. Results of the competing risks model (1 – termination of the activity without a known reason, 2 – suspension of the activity, 3 – bankruptcy in legal terms) with considered various headquarters

Source: own elaboration.

Figure 5. Results of the competing risks model (1 – termination of the activity without a known reason, 2 – suspension of the activity, 3 – bankruptcy in legal terms) with considered various sections of activity

Source: own elaboration.

Analyzing the results for particular competing risks and without taking them into account, it can be seen that the shape of the curves for the dominant risk (Figures 3–5) is similar to the total duration distribution form (Figure 2). Noteworthy, is that there is an intersection of the curves observed only for the termination of the activity without a known reason and an increase of the probability of this event occurring for enterprises localized outside the administrative borders of Warsaw between the third and fourth year of a firms existence. This phenomenon also occurs for total distribution, but for smaller duration values, which may be a result of the impact of other competing risks.

2.4. Cause-specific and subdistribution hazard regression models

Two different regression models of hazard are estimated to investigate the effect of explanatory variables. The cause-specific and subdistribution hazard models are applied. The way the competing risks are treated in both approaches is the difference between them (Zhang, 2017).

The cause-specific hazard regression model is fitted applying the Cox proportional regression model. The results of the regression model are presented in Table 2.

Table 2. The results of the cause-specific hazard regression model ($\alpha = 0.1$)

Variable	termination of the activity without a known reason		suspension of the activity		legal bankruptcy	
	$\exp(\beta)$	p -value	$\exp(\beta)$	p -value	$\exp(\beta)$	p -value
1	2	3	4	5	6	7
PKD 2007 section:						
B	4.60	$5.2 \cdot 10^{-6}$	—	$9.8 \cdot 10^{-1}$	11.10	$3.2 \cdot 10^{-6}$
C	3.67	$<2 \cdot 10^{-16}$	6.91	$3.6 \cdot 10^{-13}$	6.72	$5.0 \cdot 10^{-9}$
D	5.67	$9.0 \cdot 10^{-8}$	—	$9.8 \cdot 10^{-1}$	4.61	$2.1 \cdot 10^{-2}$
E	2.17	$3.3 \cdot 10^{-2}$	4.82	$2.3 \cdot 10^{-3}$	4.51	$1.1 \cdot 10^{-2}$
F	4.60	$<2 \cdot 10^{-16}$	9.56	$<2 \cdot 10^{-16}$	7.78	$1.6 \cdot 10^{-10}$
G	5.89	$<2 \cdot 10^{-16}$	7.26	$2.9 \cdot 10^{-14}$	7.60	$2.1 \cdot 10^{-10}$
H	4.02	$<2 \cdot 10^{-16}$	7.48	$2.8 \cdot 10^{-14}$	6.68	$5.5 \cdot 10^{-9}$
I	6.27	$<2 \cdot 10^{-16}$	8.39	$2.5 \cdot 10^{-14}$	6.50	$7.6 \cdot 10^{-8}$
J	2.58	$1.4 \cdot 10^{-9}$	6.65	$1.7 \cdot 10^{-12}$	6.48	$1.3 \cdot 10^{-8}$
K	6.06	$<2 \cdot 10^{-16}$	9.90	$<2 \cdot 10^{-16}$	8.11	$2.9 \cdot 10^{-10}$
L	3.07	$3.4 \cdot 10^{-12}$	6.39	$7.9 \cdot 10^{-11}$	4.95	$2.3 \cdot 10^{-6}$
M	2.52	$6.3 \cdot 10^{-10}$	5.39	$1.5 \cdot 10^{-10}$	4.68	$1.8 \cdot 10^{-6}$
N	4.61	$<2 \cdot 10^{-16}$	8.58	$1.6 \cdot 10^{-15}$	6.81	$9.6 \cdot 10^{-9}$
O	3.02	$1.9 \cdot 10^{-2}$	—	$9.8 \cdot 10^{-1}$	23.80	$3.7 \cdot 10^{-11}$
P	3.07	$4.1 \cdot 10^{-12}$	8.67	$1.7 \cdot 10^{-15}$	8.70	$7.9 \cdot 10^{-11}$
Q	1.33	$9.3 \cdot 10^{-2}$	3.52	$5.1 \cdot 10^{-6}$	2.55	$6.9 \cdot 10^{-3}$
R	4.46	$<2 \cdot 10^{-16}$	8.80	$6.9 \cdot 10^{-14}$	7.30	$2.4 \cdot 10^{-8}$
S	5.13	$<2 \cdot 10^{-16}$	6.92	$1.4 \cdot 10^{-12}$	5.04	$1.5 \cdot 10^{-6}$
U	—	$9.7 \cdot 10^{-1}$	—	$9.9 \cdot 10^{-1}$	—	$9.8 \cdot 10^{-1}$

1	2	3	4	5	6	7
Ownership:						
natural person	2.45	$<2 \cdot 10^{-16}$	43.40	$<2 \cdot 10^{-16}$	2.41	$<2 \cdot 10^{-16}$
legal person	0.56	$<2 \cdot 10^{-16}$	0.15	$7.1 \cdot 10^{-4}$	1.06	$5.6 \cdot 10^{-1}$
Place of residence:						
Warsaw	1.14	$4.8 \cdot 10^{-8}$	1.05	$1.5 \cdot 10^{-1}$	1.02	$6.4 \cdot 10^{-1}$
Macroeconomic indic.:						
GDP	0.98	$3.5 \cdot 10^{-2}$	0.85	$<2 \cdot 10^{-16}$	1.18	$1.0 \cdot 10^{-4}$
inflation	2.00	$<2 \cdot 10^{-16}$	2.12	$<2 \cdot 10^{-16}$	1.85	$<2 \cdot 10^{-16}$
unemployment	1.09	$<2 \cdot 10^{-16}$	0.98	$5.6 \cdot 10^{-2}$	3.89	$<2 \cdot 10^{-16}$

Source: own elaboration.

The subdistribution hazard function is also related to the Cox proportional regression model. However, the entities at risk are considered in a different way than in the case of a cause-specific hazard (Fine, Gray, 1999). The results of the regression model for the subdistribution hazard are shown in Table 3.

Table 3. The results of the subdistribution hazard regression model ($\alpha = 0.1$)

Variable	termination of the activity without a known reason		suspension of the activity		legal bankruptcy	
	$\exp(\beta)$	p -value	$\exp(\beta)$	p -value	$\exp(\beta)$	p -value
1	2	3	4	5	6	7
PKD 2007 section:						
B	4.47	$5.2 \cdot 10^{-6}$	0.01	$<2 \cdot 10^{-16}$	9.17	$1.8 \cdot 10^{-5}$
C	3.48	$2.2 \cdot 10^{-16}$	6.54	$1.5 \cdot 10^{-12}$	5.78	$6.8 \cdot 10^{-8}$
D	5.38	$1.8 \cdot 10^{-7}$	0.03	$<2 \cdot 10^{-16}$	3.99	$3.6 \cdot 10^{-2}$
E	2.09	$4.2 \cdot 10^{-2}$	4.67	$2.8 \cdot 10^{-3}$	4.14	$1.5 \cdot 10^{-2}$
F	4.29	$<2 \cdot 10^{-16}$	8.87	$<2 \cdot 10^{-16}$	6.42	$6.3 \cdot 10^{-9}$
G	5.55	$<2 \cdot 10^{-16}$	6.60	$4.4 \cdot 10^{-13}$	6.16	$1.2 \cdot 10^{-8}$
H	3.81	$<2 \cdot 10^{-16}$	7.02	$1.7 \cdot 10^{-13}$	5.72	$7.9 \cdot 10^{-8}$
I	5.89	$<2 \cdot 10^{-16}$	7.58	$3.9 \cdot 10^{-13}$	5.24	$1.8 \cdot 10^{-6}$
J	2.47	$7.4 \cdot 10^{-9}$	6.48	$3.1 \cdot 10^{-12}$	5.70	$1.1 \cdot 10^{-78}$
K	5.65	$<2 \cdot 10^{-16}$	9.00	$2.2 \cdot 10^{-16}$	6.66	$1.0 \cdot 10^{-8}$
L	2.93	$2.7 \cdot 10^{-11}$	6.14	$2.1 \cdot 10^{-10}$	4.24	$1.8 \cdot 10^{-5}$
M	2.44	$2.3 \cdot 10^{-9}$	5.28	$2.7 \cdot 10^{-10}$	4.27	$6.7 \cdot 10^{-6}$
N	4.33	$<2 \cdot 10^{-16}$	8.01	$1.2 \cdot 10^{-14}$	5.64	$2.1 \cdot 10^{-79}$
O	2.81	$2.9 \cdot 10^{-2}$	0.02	$<2 \cdot 10^{-16}$	20.95	$2.5 \cdot 10^{-10}$
P	2.90	$4.3 \cdot 10^{-11}$	8.32	$5.8 \cdot 10^{-15}$	7.45	$1.4 \cdot 10^{-9}$
Q	1.31	$1.1 \cdot 10^{-1}$	3.52	$5.3 \cdot 10^{-6}$	2.49	$8.5 \cdot 10^{-3}$
R	4.16	$<2 \cdot 10^{-16}$	8.17	$4.5 \cdot 10^{-13}$	6.04	$4.1 \cdot 10^{-7}$
S	4.87	$<2 \cdot 10^{-16}$	6.36	$1.3 \cdot 10^{-11}$	4.18	$2.0 \cdot 10^{-5}$
U	0.11	$5.1 \cdot 10^{-11}$	0.05	$2.1 \cdot 10^{-9}$	0.06	$1.4 \cdot 10^{-8}$
Ownership:						
natural person	2.31	$<2 \cdot 10^{-16}$	41.02	$<2 \cdot 10^{-16}$	2.05	$2.2 \cdot 10^{-16}$
legal person	0.56	$<2 \cdot 10^{-16}$	0.15	$8.3 \cdot 10^{-4}$	1.08	$4.2 \cdot 10^{-1}$
Place of residence:						
Warsaw	1.14	$7.2 \cdot 10^{-8}$	1.03	$3.8 \cdot 10^{-1}$	1.02	$5.9 \cdot 10^{-1}$

1	2	3	4	5	6	7
Macroeconomic indic.:						
GDP	0.98	$1.2 \cdot 10^{-1}$	0.86	$< 2 \cdot 10^{-16}$	1.15	$4.7 \cdot 10^{-11}$
inflation	1.86	$< 2 \cdot 10^{-16}$	1.92	$< 2 \cdot 10^{-16}$	0.96	$1.2 \cdot 10^{-3}$
unemployment	1.06	$5.1 \cdot 10^{-14}$	0.97	$1.8 \cdot 10^{-2}$	2.21	$< 2 \cdot 10^{-16}$

Source: own elaboration.

In Tables 2 and 3 – for particular types of termination – the exponent of the estimated values of coefficients β (with *p-values*) are presented. For each explanatory variable the value of $\exp(\beta)$ means how many times the hazard increases when the value of the variable increases by one unit while maintaining the assumption of *ceteris paribus*. The variables that characterize the enterprises (PKD section, ownership and place of residence) take the value 1 when the company has this feature and 0 otherwise, so the value of $\exp(\beta)$ determines how the hazard changes when the firm has the given feature.

As technical comments to the results, we can see that the following values have been treated as a reference: for section PKD2007 – section A, for the form of ownership – organizational entity without legal personality, and for the place of residence – outside the administrative boundaries of Warsaw. It should also be noted that in the case of the cause-specific hazard model for some sections, the coefficients with the expected level of significance could not be estimated. This is due to the low number of enterprises representing these sections in the database. In addition, the *p-values* for the coefficients of these variables are much higher than the level of significance, so no conclusions can be made on their basis. Considering the results for macroeconomic indicators, it is also worth emphasizing that the time dependence of these variables was taken into account applying the episode splitting method (a brief description of the method in subsection 3.1), which allowed to meet the assumption of the proportionality of hazard.

For both estimated hazard models, model verification was carried out applying the convergence criterion of Newton-Raphson iterative algorithm. In addition, Schoenfeld residuals were analyzed. The residuals oscillate around zero. The systematic time dependence of residuals is not observed for any of the variables. This proves that the assumption of the proportionality of hazard is met.

Analysing the results, we can see that the estimated values of coefficients are a little smaller in the case of the subdistribution hazard model (Table 3). This is for the reason of different approaches to competing risks considering (other risks as censored observations in the cause-specific hazard model) (Haller, Schmidt, Ulm, 2013). However, it is worth emphasizing that the results of both of the estimated models are consistent, i.e. the same firm characteristics for

particular causes of termination of the activity are treated as increasing hazard to a greater extent (i.e. dangerous to the enterprise).

We can conclude that for activity termination without a known reason the most dangerous PKD2007 section is I (accommodation and food service activities), for suspending section K (financial and insurance activities) and for legal bankruptcy section O (public administration and defence). However, one must be aware of the differences in the specifics of doing business in these sectors. The surprisingly high coefficient for section O (for legal bankruptcy) does not necessarily mean that these enterprises are particularly at risk in general (for other risks the coefficients are smaller), but for example that in this sector full bankruptcy proceedings must be carried out in the event that the enterprises terminate their operations. However, this does not change the fact that for companies from other sectors legal declaration of bankruptcy is less common. Moreover, we can see that for each of the considered causes of termination a natural person is the form of ownership that most clearly causes the increase in hazard values. It may be related to the fact that for small enterprises (which most often are characterized by this form of ownership) the dynamics of market activities is generally greater. Based on the results of regression, the place of residence of the enterprise headquarters only slightly affects hazard in the case of termination of activity without a known reason (Warsaw slightly increases the risk). For both other risks this variable is not statistically significant.

The results regarding macroeconomic indicators are similar for termination of activity without a known reason and suspension of activity (both for the subdistribution and cause-specific hazard models). Higher GDP growth reduces the risk of termination of operations, and rising inflation has an adverse effect. The impact of unemployment is insignificant (but this variable is statistically significant). On the other hand, in the case of legal bankruptcy, a growth of GDP increases hazard and increasing inflation lowers it (although the impact is insignificant in both of these cases). We observed a significant impact of unemployment when this growth increases hazard. It should be noted here that legal bankruptcy is represented by the least numerous group in the sample and, on the basis of the other results, it can be seen that it behaves in a different way than the other two considered competing risks.

Conclusions

This paper is focused on the presentation of competing risks models and applying a selected number of them to investigate an enterprises survival on the market. The most common competing risks models are discussed with an indication of their advantages and weaknesses. It is

noteworthy that in the case of empirical data, when we do not have enough knowledge about the investigated phenomenon to make *a priori* assumptions, the essential advantage of the models is a lack of the need to apply assumptions about the mutual independence of competing risks. This problem occurs in the case of the naive Kaplan-Meier estimator, while the solution is the subdistribution concept proposed by Fine and Gray.

This approach is used to estimate CIFs. The obtained results show a large diversity in the distribution of the enterprises duration on the market, both due to the reason for termination of activity and the characteristics of the firms. It is worth noticing, that distribution without competing risks is similar to the distribution of the dominant risk, which is in agreement with intuition. The use of competing risks also allows examining the distribution for other causes of liquidation. In each of the considered cases, the form of ownership that is associated with the shortest duration of the enterprises is a natural person. The dominant risk is the termination of activity without a known reason, although the suspension of business operations by a natural person as an owner is characterized by higher values of CIF than other forms of ownership in the case of termination of activity without a known reason.

In addition, the application of the regression models allows indicating the differences between them, as well as a more detailed analysis of particular causes of termination of business activity by enterprises. The results are consistent with those obtained earlier in the study. For each of the considered risks, the features of enterprises that have a negative impact on their duration on the market are indicated. It was emphasised, that in the analysis of the results, a broader view and awareness of the specifics of business activities for various business entities are needed. The termination of business activity without a known reason and suspension of activity behave differently from legal bankruptcy, which emphasizes the completely different nature of these forms of enterprise and thus their failure on the market.

In summary, it is important to emphasize that competing risks models allow for more accurate results and a more comprehensive understanding of the phenomenon. Moreover, if models that do not require assumptions about the mutual independence of competitive risks are chosen, then it is possible to estimate the distribution of enterprises duration on the market (both for reasons and the total distribution) without making *a priori* assumptions. This is extremely valuable in the case of complex problems that we have to deal with in microeconomic research concerning for example the situation of enterprises on the market.

The obtained results are valuable because of a lack of analysis of duration on the market for firms from the Mazowieckie Voivodeship with considering various reasons for their termination of activity. It is worth considering performing an analogous study for cohorts of firms that

started their business activity in other years than 2007 and then a comparison of their results could be analysed.

References

- Austin, P.C., Lee, D.S., Fine, J. P. (2016). Introduction to the Analysis of Survival Data In the Presence of Competing Risks. *Circulation*, 133 (6), 601–609. DOI: 10.1161/CIRCULATIONAHA.115.017719.
- Blossfeld, H.P., Hamerle, A., Mayer, K.U. (2014). *Event History Analysis. Statistical Theory and Application in the Social Sciences*. New York, Psychology Press.
- Boratyńska, K., Włoczevska, E. (2013). Selected methods of corporate bankruptcy risk assessment – the cases of the Polish chosen meat sector companies. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, XV (6), 27–32.
- Fine, J., Gray, R. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risks. *Journal of the American Statistical Association*, 94 (446), 496–509. DOI: 10.1080/01621459.1999.10474144.
- Gray, R. (1988). A class of k-sample tests for comparing the cumulative incidence function in the presence of competing risks. *Annals of Statistics*, 16, 1141–1154.
- Haller, B., Schmidt, G., Ulm, K. (2013). Applying competing risks regression models: an overview. *Lifetime Data Anal*, 19, 33–58. DOI: 10.1007/s10985-012-9230-8.
- Kaplan, E.L., Meier, P. (1958). Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, 53 (282), 457–481. DOI: 10.1080/01621459.1958.10501452.
- Klein, J.P., Moeschberger, M.L. (2003). *Survival analysis: techniques for censored and truncated data*. New York: Springer.
- Kisielińska, J. (2016). Skuteczność modeli upadłości przedsiębiorstw. *Studia Ekonomiczne i Regionalne*, 9 (1), 5–17.
- Markowicz, I. (2018). Modeling the Survival Time of Trading Companies in the Zachodniopomorskie Voivodship. *Folia Oeconomica, Acta Universitatis Lodzensis*, 4 (337), 85–97. DOI: 10.18778/0208-6018.337.06.
- Markowicz, I. (2019). Analysis of the Risk of Liquidation Depending on the Age of the Company: A Study of Entities Established In Szczecin In Period 1990–2010. *ECONOMETRIS, Advances in Applied Data Analysis*, 23 (2), 49–62. DOI: 10.15611/eada.2019.2.04.

- Mikulec, A. (2017). Cohort Tables of Enterprise Duration in Lodzkie Voivodeship – Quarterly Basis. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Taksonomia*, 468 (28), 147–160. DOI: 10.15611/pn.2017.468.15.
- Pepe, M.S., Mori, M. (1993). Kaplan-Meier, Marginal or Conditional Probability Curves in Summarizing Competing Risks Failure Time Data? *Statistics in Medicine*, 12 (8), 737–751. DOI: 10.1002/sim.4780120803.
- Peterson, A.V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. *PNAS*, 73 (1), 11–13. DOI: 10.1073/pnas.73.1.11.
- Pintilie, M. (2006). *Competing risks: a practical perspective*. Chichester: Wiley.
- Satagopan, J.M., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D., Auerbach, A.D. (2004). A note on competing risks in survival data analysis. *British Journal of Cancer*, 91, 1229–1235. DOI: 10.1038/sj.bjc.6602102.
- Ptak-Chmielewska, A. (2013). Semiparametric Cox Regression Model in Estimation of Small and Micro Enterprises' Survival in the Malopolska Voivodeship. *Quantitative Methods In Economics, XIV* (2), 169–180.
- Zhang, Z. (2017). Survival analysis in the presence of competing risks. *Ann Transl Med.*, 5 (3), 47–55. DOI: 10.21037/atm.2016.08.62.