ECONOMETRICS. EKONOMETRIA
Advances in Applied Data Analysis
Year 2018, Vol. 22, No. 3
ISSN 1507-3866; e-ISSN 2449-9994

DOES THE TYPE OF BUSINESS ACTIVITY AND THE ENTERPRISE LOCATION AFFECT A FIRM'S SURVIVAL? RESULTS OF AN ANALYSIS FOR NATURAL PERSONS CONDUCTING ECONOMIC ACTIVITY IN THE ŁÓDZKIE VOIVODSHIP

Artur Mikulec
University of Lodz, Lodz, Poland
e-mail: amikulec@uni.lodz.pl
ORCID: 0000-0001-8249-2296

Małgorzata Misztal
University of Lodz, Lodz, Poland
e-mail: mmisztal@uni.lodz.pl
ORCID: 0000-0002-8719-2097

DOI: 10.15611/eada.2018.3.02
JEL Classification: C10, C14, C41

Abstract: The article presents the results of the duration analysis for 21,163 enterprises (natural persons conducting economic activity) established in the Łódzkie Voivodship in 2010 and observed until December 31, 2015. The Kaplan-Meier estimation of the survival function, the Cox proportional-hazards model and the recursive partitioning method (the CTree algorithm) are applied to achieve the goal of the conducted research i.e. to answer the following question: does the type of business activity and location of the enterprise affect its duration? Prediction error curves based on the bootstrap cross-validation estimates of the prediction error are used to assess and compare predictions obtained from all three models. On the basis of the analysis results it can be assumed that the type of business activity makes firms more varied due to their duration compared to their location.

Keywords: enterprises, duration analysis, Kaplan-Meier survival curve, Cox proportional-hazards model, survival trees.

1. Introduction

This paper investigates some determinants of the firms' duration with the use of selected survival analysis methods. Survival analysis is a collection of various techniques and statistical methods originally developed for analyzing lifetime data in medicine and allied areas of research. Currently, all these methods are widely used in many scientific disciplines including engineering, social sciences and economics. Survival analysis deals with survival time, i.e. the time to the occurrence of an event of interest. In medical research the event of interest can be, for example, the death of a patient after the diagnosis, in engineering – the failure of a machine part, and in economics – exiting unemployment.

In our study, the event of interest is the liquidation of an enterprise. The key issue in the analysis of the firms' duration is the search for variables significantly affecting their duration. The occurrence of censored observations precludes the use of traditional statistical methods such as regression analysis. The most common methods used in the survival analysis are the Kaplan-Meier estimation of the survival function and the Cox proportional-hazards regression model. As an alternative to the Kaplan-Meier estimation and the Cox model, a regression tree approach (based on a recursive partitioning method) could be proposed. All these methods were used to achieve the goal of the conducted research i.e. to answer the following question: does the type of business activity and location of the enterprise affect its duration?

The problem of the duration of enterprises and the factors determining their survival have been investigated by researchers from various countries. The results of survival analysis of enterprises, in which the type of business activity or business sector were considered, are presented, among others, by: Bartelsman et al. [2005] for enterprises from 10 OECD countries; Fritsch et al. [2006] and Fertala [2008] for German enterprises; López-Garcia and Puente [2006] for firms in Spain; Velucchi and Viviani [2007] for firms in Italy; Lin and Huang [2008] for firms in Taiwan; Huiban [2011] for firms in France; Jackson et al. [1999, 2000], Śmiech [2011], Markowicz [2012], Gurgul and Zając [2016] and Ptak-Chmielewska [2016] for Polish enterprises. One of the latest studies in this area is the analysis of enterprises from the so called Visegrad Group (the Czech Republic, Hungary, Poland and Slovakia) – Baumöhl et al. [2017] and the analysis of enterprises in Chile – López et al. [2017]. From amongst the studies that analyze geographical location as a factor influencing the survival of enterprises the following can be mentioned: Fotopoulos and Louri [2000] – firms in Greece; Stuart and Sorenson [2003] and Folta et al. [2006] – firms in USA; Fertala [2008] – firms in Germany and Huiban [2011] – firms in France. Also Jackson et al. [1999, 2000] and Szymański [2011] in their research took into account the aspect of enterprise location (voivodship, 'powiat', i.e. county) as a factor having impact on its survival. An attempt of the analysis of enterprises according to regional industry is presented in Dehnel [2010], where selected methods of small area statistics are implemented to analyze the development of microenterprises. Hence our article (according to the authors' knowledge) can be seen as the first in Polish literature on this subject, where the combined impact of both factors – type of business activity and location – on the survival of enterprises was analyzed[1]. The results of the conducted analysis were compared with the other results for companies in Poland, and additional national and international literature review about duration analysis are presented in Markowicz [2015].

2. Data characteristics

The individual data on 21,163 of enterprises – natural persons conducting economic activity[2] established in the Łódzkie Voivodship in 2010 were used in the work. Out of this number, 9,373 went into liquidation (44.3%) by the end of 2015, and 11,790 (55.7%) continued their economic activity. These firms are treated as censored data. The structures of enterprises by Polish Classification of Activities and location are presented in Tables 1 and 2.

Table 1. The number of natural persons conducting economic activity established in the Łódzkie Voivodship in 2010 by Polish Classification of Activities (PKD 2007)

PKD 2007 section code	Enterprises established		PKD 2007 section code	Enterprises established	
	N	%		*N*	%
A – Agriculture, forestry and fishing	54	0.3	L – Real estate activities	171	0.8
B+C+D+E – Industry	2,303	10.9	M – Professional, scientific and technical activities	1,886	8.9
F – Construction	2,805	13.3	N – Administrative and support service activities	723	3.4
G – Trade; repair of motor vehicles	7,350	34.7	P – Education	541	2.6
H – Transportation and storage	1,078	5.1	Q – Human health and social work activities	855	4.0
I – Accommodation and food service activities	683	3.2	R – Arts, entertainment and recreation	262	1.2
J – Information and communication	679	3.2	S – Other service activities	1,007	4.8
K – Financial and insurance activities	766	3.6	TOTAL	21,163	100.0

Source: own elaboration.

[1] Other results of the analysis of a firm's duration analysis for the Łódzkie Voivodship with the implementation of the recursive partitioning method are presented in Mikulec, Misztal [2018].

[2] The enterprises were selected according to the methodology of survey *SP-3 Report on economic activity of enterprises* (annual survey of enterprises with up to nine persons employed), conducted by the Central Statistical Office and the Statistical Office in Łódź. The authors omitted the criterion of the size of the enterprise (number of employed persons), because the analysis was to concern all enterprises conducting manufacturing, trade or service activity on a free market basis – for profit and for self-employment (regardless of their size).

Among the analyzed enterprises established in 2010 the largest number of entities conducted their economic activities in section G (*Trade; repair of motor vehicles*) 34.7%, in section F (*Construction*) 13.3% and in sections B+C+D+E (*Industry*) 10.9% of enterprises in total. The smallest number of natural persons conducting economic activity was in section A (*Agriculture, forestry and fishing*) 0.3%, in section L (*Real estate activities*) 0.8% and in section R (*Arts, entertainment and recreation*) 1.2%.

Table 2. The number of natural persons conducting economic activity established in the Łódzkie Voivodship in 2010 by location

'Powiat' (county) code	Enterprises established		'Powiat' (county) code	Enterprises established	
	N	%		*N*	%
01 – bełchatowski	754	3.6	14 – sieradzki	830	3.9
02 – kutnowski	679	3.2	15 – skierniewicki	186	0.9
03 – łaski	409	1.9	16 – tomaszowski	870	4.1
04 – łęczycki	245	1.2	17 – wieluński	599	2.8
05 – łowicki	496	2.3	18 – wieruszowski	289	1.4
06 – łódzki wschodni	666	3.1	19 – zduńskowolski	641	3.0
07 – opoczyński	459	2.2	20 – zgierski	1,354	6.4
08 – pabianicki	1,143	5.4	21 – brzeziński	262	1.2
09 – pajęczański	294	1.4	61 – Łódź city	7,658	36.3
10 – piotrkowski	501	2.4	62 – Piotrków Trybunalski city	746	3.5
11 – poddębicki	323	1.5	63 – Skierniewice city	458	2.2
12 – radomszczański	957	4.5	TOTAL	21,163	100.0
13 – rawski	344	1.6			

Source: own elaboration.

However, taking into account the location ('powiat') of the enterprises established in the Łódzkie Voivodship in 2010 it can be noticed that the largest number of natural persons started their business activity in the city of Łódź – 36.3% and in two neighbouring counties ('powiat' – zgierski and pabianicki, 6.4% and 5.4%, respectively. The smallest number of natural persons started their economic activity in the brzeziński (1.2%), łęczycki (1.2%) and skierniewicki (0.9%) counties – located in the north of the Łódzkie Voivodship.

3. Methods

The analysed dependent variable is the enterprise duration time in days, i.e. the time up to the liquidation of the enterprise. The data are right-censored: enterprises enter the study in different months of 2010, some of them experience the event (liquidation) before the end of the study and some are observed until December 31, 2015 (the end of the study) without the occurrence of the event.

A key concept in survival analysis is the survivor function, denoted by $S(t)$, which represents the probability that an object survives longer than some specified time point t. In other words, $S(t)$ gives the probability that the random variable T exceeds the specified time t [Kleinbaum, Klein 2005]. A simple and commonly used non-parametric procedure for the estimation of the survivor function of a homogenous right-censored data is the Kaplan-Meier (KM) estimation [Kaplan, Meier 1958]. The KM estimator of the survivor function at time t can be obtained with the use of the following product limit formula:

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \frac{n_i - d_i}{n_i} \quad (1)$$

with the convention that $\hat{S}(t) = 1$ if $t < t_{(1)}$ and where n_i denotes the number at risk of dying at $t_{(i)}$ and d_i – the observed number of deaths.

A graphical representation of the KM survival probability against time is the Kaplan-Meier survival curve that can be used to estimate measures such as median survival time. To compare KM survival curves for different groups of objects the log-rank test can be applied (see e.g. [Kleinbaum, Klein 2005]).

However, the KM estimation does not provide the possibility to model survival times as a function of a set of covariates. This can be done with the use of the Cox proportional-hazards (PH) model [Cox 1972]:

$$h(t, \mathbf{x}) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p), \quad (2)$$

where: $h(t, \mathbf{x})$ denotes the outcome hazard at a given time t, $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ is the vector of p covariates (explanatory variables) and $h_0(t)$ is the baseline hazard function. The Cox model is reduced to the baseline hazard when all the explanatory variables are equal to zero, or when there are no explanatory variables in the model.

As an alternative, to model the relationship between a survival time and a set of covariates, a regression tree approach can be used. The model building process is based on a recursive partitioning of the set of objects into homogenous subsets and the Kaplan-Meier estimate of the survivor function or the Cox regression model is reported in each terminal node. The graphical representation of the model takes the form of survival trees. A detailed review of the survival trees algorithms is presented in [Bou-Hamad et al. 2011]. The advantage of the survival trees over the Cox model lies in their great flexibility, no assumptions on distributions of the survival times and the possibility to automatically detect interactions between covariates without the need to specify them beforehand. As was also emphasized by Bou-Hamad et al. [2011, p. 45]: "a single tree can naturally group subjects according to their survival behaviour based on their covariates". In this paper we propose to use the CTree (*Conditional Inference Tree*) algorithm [Hothorn et al. 2006] to build the survival tree. CTree is a non-parametric class of regression trees embedding recursive binary partitioning into a well-defined theory of conditional inference procedures with stopping rule based on multiple test procedures (for details see: [Hothorn et al. 2006]).

For an assessment and comparison of predictions in survival analysis one can use (among others) prediction error curves [Mogensen et al. 2012]. Prediction error curves are time dependent estimates of the population average Brier score. At a given time point *t*, the Brier score for the single object is defined as [Mogensen et al. 2012, p. 10]:

$$BS(t,\hat{S}) = E\left(Y_i(t) - \hat{S}(t/\boldsymbol{X}_i)\right)^2, \tag{3}$$

where the expectation is taken with respect to the data of an object *i* which does not belong to the training set, $Y_i(t) = I(T_i \geq t)$ is the true status of the object *i* and $\hat{S}(t/\boldsymbol{X}_i)$ is the predicted survival probability at time *t* for the object *i* with covariates $\boldsymbol{X}_i$. For the estimation of the expected Brier score, the true status is replaced by the observed status, defined as $\tilde{Y}_i(t) = I(\tilde{T}_i \geq t)$, and the squared residuals are weighted with the use of inverse probability of censoring weights (IPCW, see: [Gerds, Schumacher 2007]) given by:

$$\widehat{W}_i(t) = \frac{(1-\tilde{Y}_i(t))\Delta_i}{\hat{G}(\tilde{T}_i - |\boldsymbol{X}_i)} + \frac{\tilde{Y}_i(t)}{\hat{G}(t/\boldsymbol{X}_i)}, \tag{4}$$

where $\hat{G}(t/x) \approx P(C_i > t|\boldsymbol{X}_i = \boldsymbol{x})$ is an estimate of the conditional survival function of the censoring times. For new observations or an independent test data set $\widetilde{D}_M$, the expected Brier score is estimated by:

$$\widehat{BS}(t,\hat{S}) = \frac{1}{M}\sum_{i\in\widetilde{D}_M} \widehat{W}_i(t)\{\tilde{Y}_i(t) - \hat{S}(t/\boldsymbol{X}_i)\}^2, \tag{5}$$

where *M* is the number of objects in $\widetilde{D}_M$ and $\hat{S}$ is based on a training data.

Several data splitting algorithms based on cross-validation and bootstrap have been proposed to estimate the prediction accuracy of a model in a typical situation where a single data set has to be used to build the prediction models and again to estimate the prediction performance (see e.g.: [Efron, Tibshirani 1997; Gerds, Schumacher 2007; Adler, Lausen 2009]). In our research we use the bootstrap cross-validation approach where the data set D_N is split into many bootstrap training samples D_b and corresponding test samples D_N/D_b ($b = 1, ..., B$). Bootstrap samples can be drawn with or without replacement from the original data. Then, models $\hat{S}_b$ are trained with the bootstrap training data D_b, the corresponding test samples are predicted and residuals are computed. Finally, the bootstrap cross-validation estimate of the prediction error is calculated by averaging over the test data sets using the formula:

$$BootCvErr(t,\hat{S}) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{M_b}\sum_{i\in D_N\setminus D_b}\widehat{W}_i(t)\{\tilde{Y}_i(t)-\hat{S}_b(t/\boldsymbol{X}_i)\}^2, \quad (6)$$

where M_b is a fixed user defined number smaller than N and the same for each b, and is the size of the bootstrap samples for resampling without replacement, or – for bootstrap with replacement – M_b is the number of objects not drawn in the bootstrap sample D_b. Some other estimators of the prediction accuracy are described in [Mogensen et al. 2012]. The graphical representation of these estimators takes the form of prediction error curves. The prediction error curves can be summarized with the Integrated Brier Score defined as:

$$IBS(predErr,\tau) = \frac{1}{\tau}\int_0^{\tau} predErr(u,\hat{S})du, \quad (7)$$

where *predErr* refers to any method of estimation of the predictive performance and $\tau > 0$ can be set to any value smaller than the minimum of the maximum times for which estimated prediction errors can be evaluated in each bootstrap sample [Mogensen et al. 2012, p. 13].

All the calculations were done with the use of the following packages: STATISTICA ver. 12.5, SPSS ver. 24 and R-project ver. 3.4.1 (packages: rms, party, pec).

4. Results

Figure 1 shows the Kaplan-Meier survival curve for all the analyzed enterprises in the Łódzkie Voivodship. The probability of survival of 2190 days is 0.5315. A sharper decrease of the KM survival curve for all firms can be observed after the first two years (730 days).

The Kaplan-Meier survival curves for enterprises by type of their economic activity (Polish Classification of Activities) are presented in Figure 2. For the sake of simplicity, only selected KM curves are described by the respective PKD codes.

When comparing the survival curves for the PKD classification, it can be noticed that the KM curves are rather close to each other during the first two years of observation; they start to vary at time $t = 800$ days. The KM curve for the Q section (*Human health and social work activities*) is higher than the others starting from approximately 900 days. The probability of surviving 2190 days for the Q section is 0.734.

It is easier to compare all the KM curves by analyzing the probability of survival of 2190 days for enterprises belonging to each PKD section. These results, ranked according to the probability of surviving 2190 days, are presented in Table 3. It is also possible to estimate the median survival time for firms belonging to some of the PKD sections.

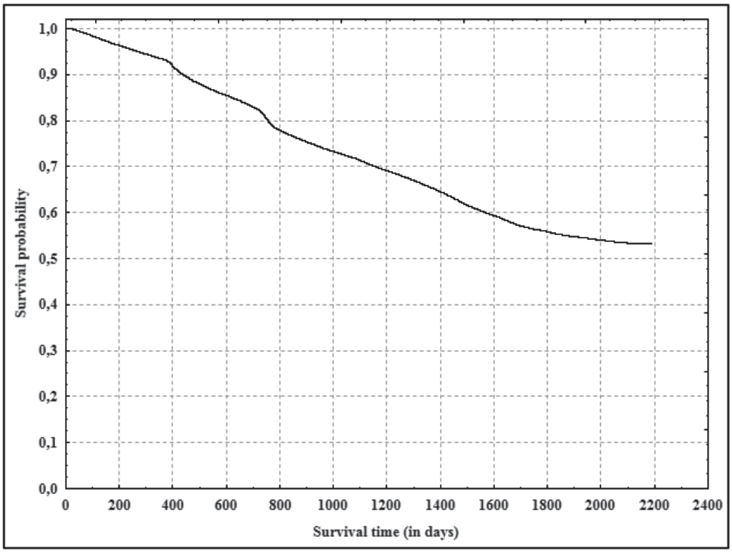
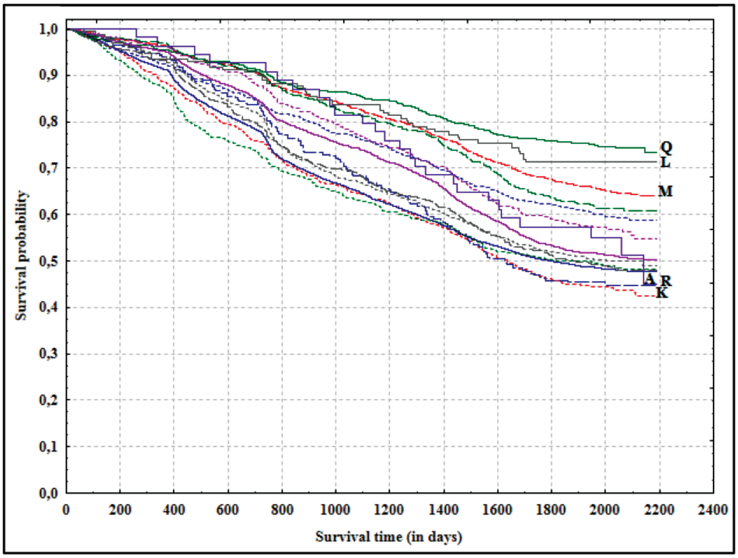


Fig. 1. The Kaplan-Meier survival curve for all the enterprises observed between 2010-2015

Source: own elaboration.



Key: Q, L, M, A, R, K – described in Table 1.

Fig. 2. The Kaplan-Meier survival curves for enterprises by type of economic activity (PKD 2007)

Source: own elaboration.

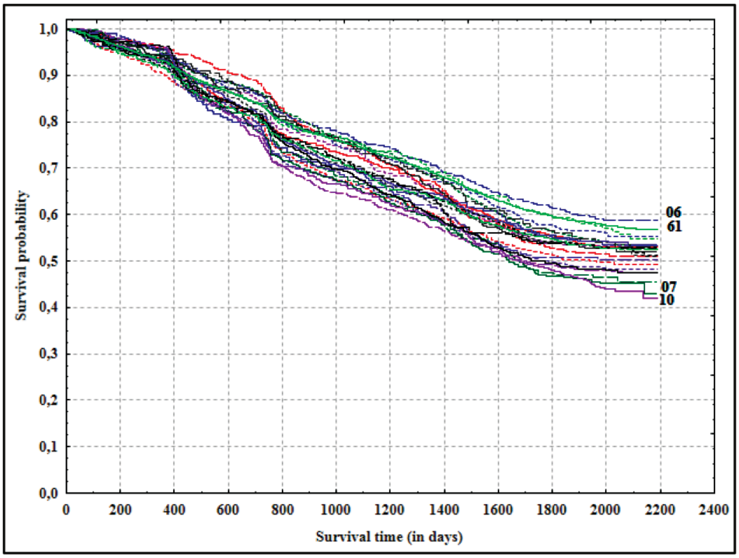
Table 3. Probability of surviving 2190 days for enterprises conducting a particular economic activity

PKD 2007 section code*	Probability of survival of 2190 days	Median survival time	PKD 2007 section code*	Probability of survival of 2190 days	Median survival time
Q	0.734	x	S	0.491	2002
L	0.713	x	I	0.482	1893
M	0.641	x	N	0.481	1891
J	0.609	x	G	0.477	1801
B+C+D+E	0.603	x	A	0.448	2142
H	0.587	x	R	0.446	1624
P	0.547	x	K	0.425	1633
F	0.503	x			

* PKD 2007 codes described in Table 1.

Source: own elaboration (SPSS package).

The Kaplan-Meier survival curves for enterprises by location are presented in Figure 3. Only selected KM curves are described by the respective county's code. It can be noticed that the KM curves are close to each other.



Key: 'powiat' (county) codes: 06, 61, 07, 10 described in Table 2.

Fig. 3. The Kaplan-Meier survival curves for enterprises by location

Source: own elaboration.

The probability of surviving 2190 days for enterprises by location is given in Table 4. Estimates of the median survival time for firms located in seven counties (‘powiat’: 02 – kutnowski, 07 – opoczyński, 09 – pajęczański, 10 – piotrkowski, 12 – radomszczański, 17 – wieluński, 19 – zduńskowolski) are also presented.

Table 4. Probability of surviving 2190 days for enterprises by location

‘Powiat’ code*	Probability of survival of 2190 days	Median survival time	‘Powiat’ code*	Probability of survival of 2190 days	Median survival time
06	0.589	x	13	0.521	x
61	0.568	x	14	0.514	x
05	0.552	x	20	0.510	x
63	0.549	x	03	0.510	x
04	0.536	x	18	0.502	x
08	0.533	x	02	0.494	1937
16	0.533	x	17	0.482	1780
11	0.532	x	19	0.475	1738
01	0.531	x	12	0.475	1703
15	0.531	x	09	0.455	1678
21	0.528	x	07	0.430	1640
62	0.526	x	10	0.420	1673

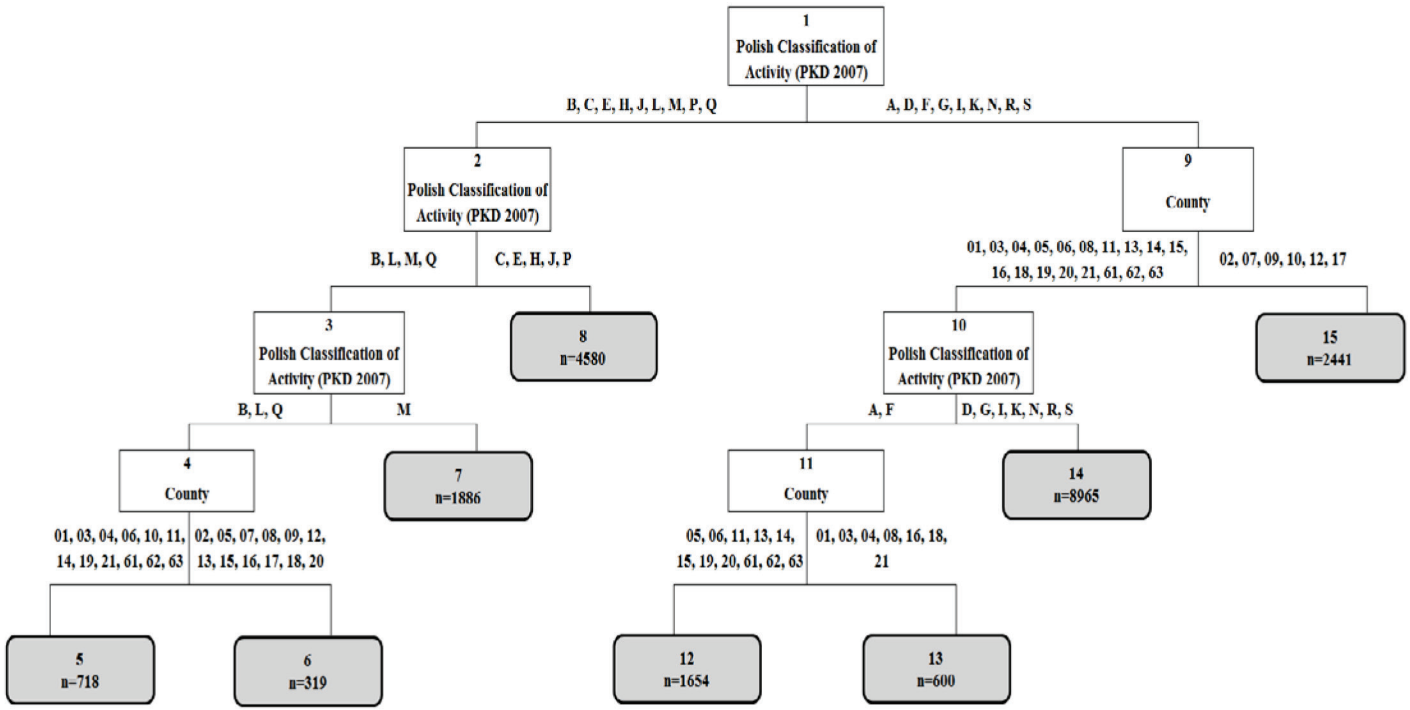
* ‘Powiat’ (county) codes described in Table 2.

Source: own elaboration (SPSS package).

Analyzing the KM curves for enterprises by their location and by economic activity, it can be assumed that the type of business activity makes firms more varied due to their duration.

Conducting business activity in certain sections of the Polish Classification of Activities and location of the firm in some counties turn out to be the factors significantly influencing the risk of enterprise going into liquidation, also in the case of using the Cox proportional-hazards model. Due to the limited length of the paper the presentation of detailed results is omitted, only statistically significant results are briefly outlined.

The risk of enterprise going into liquidation significantly increases for: firms from section K (*Financial and insurance activities*) according to PKD 2007 (Hazard Rate: HR=1.511; 95%CI: 1.016-2.248; $p = 0.0416$) and location of the company in counties: opoczyński (HR = 1.250; 95%CI: 1.063-1.470; $p = 0.0070$), piotrkowski (HR = 1.287; 95%CI: 1.100-1.506; $p = 0.0016$) and radomszczański (HR = 1.180; 95%CI: 1.029-1.352; $p = 0.0179$). The factors that considerably reduce the risk of enterprise going into liquidation include: conducting business activity in the PKD



Key: PKD 2007 codes and 'powiat' (county) codes described in Table 1 and Table 2.

Fig. 4. Survival tree (CTree algorithm)

Source: own elaboration.

section Q (*Human health and social work activities*) (HR=0.552; 95%CI: 0.367-0.950; p = 0.0042) and location of the enterprise in the 'powiat łódzki wschodni' (HR = 0.811; 95%CI: 0.692-0.950; p = 0.0096). The whole model is statistically significant (LR χ^2 = 687.58; df = 37; p < 0.0001).

Figure 4 shows the survival tree obtained using the CTree algorithm. Both the type of economic activity and location are used for splitting the nodes. There are eight terminal nodes (leaves). Sample size (n) is presented for each leaf.

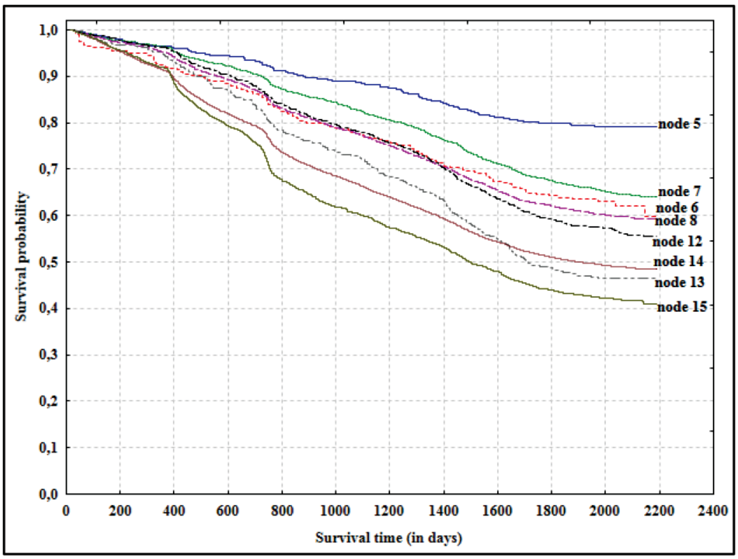


Fig. 5. The Kaplan-Meier survival curves for eight groups of enterprises separated in the leaves of the CTree survival tree

Source: own elaboration.

Figure 5 presents the Kaplan-Meier survival curves for eight groups of enterprises separated in the leaves of the CTree survival tree. When comparing the survival curves for the CTree survival tree terminal nodes, it can be observed that the KM curve for node no 5 is consistently higher than the KM curves for all the other nodes. Only 713 natural persons conducting economic activity in section L (*Real estate activities*) or section Q (*Human health and social work activities*) and located in one of the following counties: 01 – bełchatowski,03 – łaski, 04 – łęczycki, 06 – łódzki wschodni, 10 – piotrkowski, 11 – poddębicki, 14 – sieradzki, 19 – zduńskowolski, 21 – brzeziński, 61 – Łódź city, 62 – Piotrków Trybunalski city, 63 – Skierniewice city, are observed in that node. The probability of survival of 2190 days is 0.791.

The lowest-lying KM curve refers to node no 15. The probability of surviving 2190 days is 0.410 and the median survival time is equal to 1498 days. There are

2440 enterprises in this node that are characterized by belonging to one of the following PKD sections: A (*Agriculture, forestry and fishing*); F (*Construction*); G (*Trade; repair of motor vehicles*); I (*Accommodation and food service activities*); K (*Financial and insurance activities*); N (A*dministrative and support service activities*); R (*Arts, entertainment and recreation*); S (*Other service activities*) and the firm's location in one of the counties: 02 – kutnowski, 07 – opoczyński, 09 – pajęczański, 10 – piotrkowski, 12 – radomszczański, 17 – wieluński.

A sharper decrease of the KM survival curve for the firms in node 15 compared to KM survival curves in all the other nodes can be observed in the first two years (730 days).

The probability of surviving 2190 days and median survival time for enterprises separated in the leaves of the survival tree is presented in Table 5.

Table 5. Probability of surviving 2190 days for enterprises grouped in the CTree leaves

Node	Probability of survival of 2190 days	Median survival time	Node	Probability of survival of 2190 days	Median survival time
5	0.791	x	12	0.554	x
6	0.598	x	13	0.465	1708
7	0.641	x	14	0.486	1898
8	0.594	x	15	0.410	1498

Source: own elaboration (SPSS package).

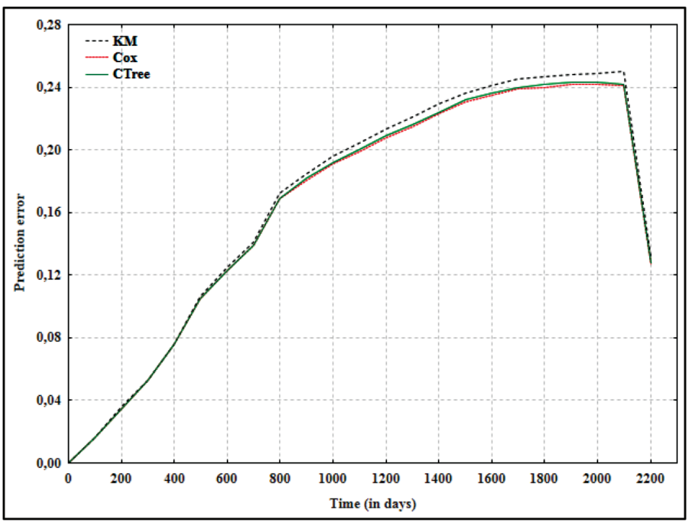


Fig. 6. Comparison of prediction error curves (the bootstrap cross-validation estimates)

Source: own elaboration.

To assess and compare predictions in the proposed survival analysis methods, prediction error curves were determined based on the bootstrap cross-validation estimates of the prediction error (see Figure 6). All three curves (KM ignoring covariates, the Cox model and the CTree) are practically the same during the first year of observation and then one can see the advantage of the Cox model and the CTree survival tree over the KM estimation of the survival function.

The Integrated Brier Scores between 0 and 2190 days for the bootstrap cross-validation estimates of the prediction error are the same for the Cox model and the CTree survival tree and are equal to 0.164, so both the models perform slightly better than Kaplan-Meier (IBS = 0.168).

5. Conclusions

The results of the conducted analysis for enterprises established in the Łódzkie Voivodship in 2010 and observed until the end of 2015 allow to formulate the following conclusions.

The Kaplan-Meier survival curve (Figure 1) proves that economic activity of natural persons is often liquidated just after the first two years. This is a period in which the entrepreneur, fulfilling statutorily defined conditions, benefits from preferential rates of social security payments. After this period the level of contributions increases 4.5 times, and can be the reason for the loss of financial liquidity and as a consequence the business closure. Another cause for ceasing economic activity by natural persons after the first two years can be the loss of subsidies granted to new entrepreneurs by the Employment Office. It is also noteworthy that the liquidation of business activity conducted by a natural person, in contrast to a legal person, does not require a winding-up procedure. A similar situation, i.e. the sharp decrease of survival curve after the second year of its activity, can be seen for enterprises in the Łódzkie Voivodship according to PKD sections, or location (see Figures 2 and 3).

The Kaplan-Meier survival curves for the three PKD sections presented in Figure 2, i.e. Q, L, M are much higher than the KM curve for all the entities in the voivodship, in which the largest part is taken by companies conducting trade activity (section G). The above-mentioned sections cover companies performing economic activities in the following sectors: *Human health and social work activities*, *Real estate activities* and *Professional, scientific and technical activities*. The authors claim that these sections are of the high minimum efficient scale of production[3]. They will advise the companies to give a lot of thought and prepare well before they decide to start a business, as well as to gather and commit the capital of significant value at the start-up (high market entry barriers – entities from these sections represented 13.8% of newly established companies). In the subject literature it is believed that

[3] The minimum efficient scale of production – quantity of output that when exceeded becomes the opposite to economy of scale, thus the increase of average long term total cost.

the higher the minimum efficient scale of output, the higher the probability of a firm's survival [Szymański 2011]. It can be claimed that the highest survival on the market were characterized by natural persons: (in section Q) conducting general and comprehensive health care services, (in section L) conducting real estate activity; and advertising agencies (in section M). The population of the Łódzkie Voivodship is a rapidly ageing society, hence there is a huge demand for medical and health care services. Moreover, the results obtained for section Q are consistent with the results of analysis in Markowicz [2012], for enterprises in Szczecin.

The situation is different in the case of enterprises in section A (*Agriculture, forestry and fishing*), R (*Arts, entertainment and recreation*), K (*Financial and insurance activities*) for which the KM survival curves were the lowest (see Figure 2). Entities in these sections represented 5.1% of newly established companies. According to the authors in these sections occur the low costs of enterprise's establishment, so that there are more companies entering (some of them randomly), thus more enterprises are liquidated. What is more, these sections are characterized by the low level of company innovation and competition, compared to the other sections. Thus, it can be stated that the lowest survival on the market is characterized by natural persons conducting: (in section A) forestry services, (in section R) entertainment and recreation, as well as insurance agents and brokers (in section K). The results obtained in section K are similar to the results presented in Gurgul and Zając [2016], for enterprises located in the Małopolskie Voivodship.

In the literature of the subject the effect of economic activities location on the probability of company going into liquidation is not explicit [Ptak-Chmielewska 2016]. The results of analysis for Poland, including the Łódzkie Voivodship [Szymański 2011; Mikulec, Misztal 2018] indicate that enterprise location in a large urban area or nearby, in big cities (despite greater competition) has a positive impact on enterprise duration. The higher level of the development of a given 'powiat'[4] and the higher level of urbanization translate into the higher probability of surviving, longer duration. Also the population density, and consequently the market size, has an impact on a firm's success [Jackson et al. 1999, 2000]. The above conclusions are confirmed by the statistical data of enterprises from the Łódzkie Voivodship. Hence, the highest Kaplan-Meier survival curves were observed for enterprises located in counties ('powiat') – łódzki wschodni (06), Łódź city (61), while entities from opoczyński (10), piotrkowski (07) had the lowest probability of surviving. Favorable conditions for business over the long term in Łódź city and adjacent county of 'powiat łódzki wschodni' are e.g. the high concentration of entities (including newly

[4] The level of county's development can be measured by: the number of national economic entities per 10 thousand residents (+); the number of new companies registered (natural persons conducting economic activity) per 1000 persons of working age (+); the average gross monthly salaries (+); the length of hard-surfaced roads in km per 100 km^2 (+) and the share of registered unemployed people per 1000 persons of working age (–). The given signs show the direction of the indicators' influence on the level of the county's development.

established companies), market size, wages and salaries, road density, and the labour market situation (low unemployment). The piotrkowski 'powiat' is an example of the lowest level of urbanization in the voivodship (7.0% in 2010).

The conclusions drawn on the basis of the Kaplan-Meier curves were confirmed and supplemented by the results obtained by the Cox proportional-hazards regression model. The Cox regression, apart from the confirmation of the impact on enterprises' duration such factors as conducting activities within sections Q and L, as well as the location of the firms in the counties: łódzki wschodni, opoczyński and piotrkowski, additionally indicates that the location of entities in the radomszczański 'powiat' is a factor causing a significant increase of the risk of enterprises going into liquidation.

The cumulative impact of business activity and location on the survival of enterprises established in the Łódzkie Voivodship in 2010 is presented on the CTree survival tree. Of particular interest is the analysis of results of node no 5 with the highest survival curve and node no 15 with the lowest survival curve (see Figures 4 and 5). The highest probability of surviving (0.791) period of six years (2190 days) is characterized by natural persons conducting economic activity in section L (*Real estate activities*) and Q (*Human health and social work activities*); located in 12 counties situated in the eastern central part of the voivodship, including the cities with the rights of a county – Łódź, Piotrków Trybunalski and Skierniewice. By contrast, the lowest probability of surviving (0.410) period of six years (2190 days) is characterized by natural persons conducting economic activity in sections: A (*Agriculture, forestry and fishing*); F (*Construction*); G (*Trade; repair of motor vehicles*); I (*Accommodation and food service activities*); K (*Financial and insurance activities*); N (A*dministrative and support service activities*); R (*Arts, entertainment and recreation*); S (*Other service activities*); located in kutnowski 'powiat' and 5 others situated in the southern part of the voivodship. The share of enterprises which went into liquidation in these was high and similar (from 54.5% in kutnowski to 61.5% in the piotrkowski 'powiat') and these were natural persons conducting economic activity mainly from sections G and F (77.4% of the enterprises in total). The possible factors that could affect the relatively short duration of these entities are: low demand for their products and services due to the high unemployment rate, low wages and salaries (except for the counties: kutnowski and pajęczański) and a large number of super and hypermarkets (except for the opoczyński, pajęczański and piotrkowski counties).

Comparing the models used by the prediction error curves (see Figure 6) and the size of the Integrated Brier Score shows no significant difference between the prediction accuracy of enterprise duration using the Cox model and the CTree algorithm. The advantage of the recursive partitioning method over the Cox model consists in a clear, graphical representation of the obtained results in the form of a survival tree. On the basis of simple classification rules it is possible to identify groups of companies similar from the point of view of duration and then, for each subgroup, make an additional detailed analysis based, in the current case, on the Kaplan-Meier estimation survival curves method.

Bibliography

Adler W., Lausen B., 2009, *Bootstrap Estimated True and False Positive Rates and ROC Curve*, Computational Statistics & Data Analysis, 53(3), pp. 718-729.

Bartelsman E., Scarpetta S., Schivardi F., 2005, *Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries*, Industrial and Corporate Change, 14(3), pp. 365-391.

Baumöhl E., Iwasaki I., Kočenda E., 2017, *Firms' survival in the new EU member state*, Centre of Economic Institutions Working Paper Series, 2017-5, http://hermes-ir.lib.hit-u.ac.jp/rs/bitstream/10086/28883/1/wp2017-5.pdf, (30.07.2018).

Bou-Hamad I., Larocque D., Ben-Ameur H., 2011, *A review of survival trees*, Statistics Surveys, 5, pp. 44-71.

Cox D.R., 1972, *Regression models and life tables (with discussion)*, Journal of the Royal Statistical Society, Series B, 34, pp. 187-220.

Dehnel G., 2010, *Rozwój mikroprzedsiębiorczości w Polsce w świetle estymacji dla małych domen*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań.

Efron B., Tibshirani R., 1997, *Improvements on Cross-Validation: The .632+ Bootstrap Method*, Journal of the American Statistical Association, 92(438), pp. 548-560.

Fertala N., 2008, *The shadow of death: do regional differences matter for firm survival across native and immigrant entrepreneurs?*, Empirica, 35(1), pp. 59-80.

Folta T.B., Cooper A.C., Baik Y.-S., 2006, *Geographic cluster size and firm performance*, Journal of Business Venturing, 21(2), pp. 217-242.

Fotopoulos G., Louri H., 2000, *Location and survival of a new entry*, Small Business Economics, 14(4), pp. 311-321.

Fritsch M., Brixy U., Falck O., 2006, *The effect of industry, region, and time on new business survival – a multi-dimensional analysis*, Review of Industrial Organization, 28(3), pp. 285-306.

Gerds T.A., Schumacher M., 2007, *Efron-type measures of prediction error for survival analysis*, Biometrics, 63(4), pp. 1283-1287.

Gurgul H., Zając P., 2016, *Modelowanie nieparametryczne czasu przeżycia przedsiębiorstw w Małopolsce*, Wiadomości Statystyczne, 12(667), pp. 7-29.

Hothorn T., Hornik K., Zeileis A., 2006, *Unbiased Recursive Partitioning: A Conditional Inference Framework*, Journal of Computational and Graphical Statistics, 15(3), pp. 651-674.

Huiban J.-P., 2011, *The spatial demography of new plants: urban creation and rural survival*, Small Business Economics, 37(1), pp. 73-86.

Jackson J.E., Klich J., Poznańska K., 1999, *Firm creation and economic transitions*, Journal of Business Venturing, 14(5-6), pp. 427-450.

Jackson J.E., Klich J., Poznańska K., 2000, *Nowe przedsiębiorstwa w transformacji gospodarki polskiej*, Gospodarka Narodowa, 5-6, pp. 42-77.

Kaplan E.L., Meier P., 1958, *Nonparametric estimation from incomplete observations*, Journal of American Statistical Association, 53, pp. 457-481.

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

Lin P.C., Huang D.S., 2008, *Technological regimes and firm survival: evidence across sectors and over time*, Small Business Economics, 30(2), pp. 175-186.

López-Garcia P., Puente S., 2006, *Business demography in Spain: determinants of firm survival*, Documentos de Trabajo No. 0608, Banco de España, http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/06/Fic/dt0608e.pdf, (30.07.2018).

López D., O´Farrill D., Pérez J., Velasquez B., 2017, *Determinants on firm survival in Chile: Evidence from cohort 2010 for the period 2011-2015*, Central Bank of Chile, IFC-Bank Indonesia Satellite

Seminar on "Big Data" at the ISI Regional Statistics Conference 2017, Bali, Indonesia, 21 March 2017, https://www.bis.org/ifc/publ/ifcb44h.pdf, (30.07.2018).

Mikulec A., Misztal M., 2018, *Zastosowanie metody rekurencyjnego podziału w analizie trwania przedsiębiorstw województwa łódzkiego*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 507, pp. 179-190.

Mogensen U.B., Ishwaran H., Gerds T.A., 2012, *Evaluating random forest for survival analysis using prediction error curves*, Journal of Statistical Software, 50(11), pp. 1-23.

Markowicz I., 2012, *Statystyczna analiza żywotności firm*, Rozprawy i Studia, t. (CMIX) 835, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.

Markowicz I., 2015, *Badanie kohort firm w analizie trwania*, Ekonometria, 4(50), pp. 105-125.

Ptak-Chmielewska, A., 2016, *Determinanty przeżywalności mikro- i małych przedsiębiorstw w Polsce*, Oficyna Wydawnicza SGH, Warszawa.

Stuart T., Sorenson O., 2003, *The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms*, Research Policy, 32(2), pp. 229-253.

Szymański D., 2011, *Badanie żywotności nowo powstałych mikroprzedsiębiorstw w Polsce w latach 2002-2007*, rozprawa doktorska, http://depotuw.ceon.pl/handle/item/177 (11.04.2018).

Śmiech S., 2011, *Analiza przeżycia podmiotów gospodarczych w województwie małopolskim w latach 2002-2008*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie, 876, pp. 121-132.

Velucchi M., Viviani A., 2007, *Firm's survival and competitiveness: a case study in Tuscany*, Statistica Applicata, 19(4), pp. 325-342.

CZY RODZAJ PROWADZONEJ DZIAŁALNOŚCI I LOKALIZACJA PRZEDSIĘBIORSTWA WPŁYWAJĄ NA CZAS JEGO TRWANIA? WYNIKI ANALIZY DLA OSÓB FIZYCZNYCH PROWADZĄCYCH DZIAŁALNOŚĆ GOSPODARCZĄ W WOJEWÓDZTWIE ŁÓDZKIM

Streszczenie: W artykule przedstawiono wyniki analizy czasu trwania dla 21,163 przedsiębiorstw (osób fizycznych prowadzących działalność gospodarczą) powstałych w województwie łódzkim w roku 2010 i obserwowanych do 31 grudnia 2015 roku. W celu realizacji celu badania, czyli uzyskania odpowiedzi na pytanie: czy rodzaj prowadzonej działalności i lokalizacja przedsiębiorstwa wpływa na czas jego trwania, wykorzystano krzywe przeżycia Kaplana-Meiera, model proporcjonalnego hazardu Coxa oraz metodę rekurencyjnego podziału (algorytm CTree). Do oceny i porównania tych trzech modeli ze względu na ich zdolność prognostyczną zastosowano krzywe błędu predykcji oparte na bootstrapowym sprawdzianie krzyżowym. Na podstawie uzyskanych wyników można przypuszczać, że rodzaj prowadzonej działalności silniej różnicuje przedsiębiorstwa z punktu widzenia ich czasu trwania w porównaniu z wpływem lokalizacji siedziby firmy.

Słowa kluczowe: przedsiębiorstwa, analiza trwania, krzywa przeżycia Kaplana-Meiera, model proporcjonalnego hazardu Coxa, drzewa przeżycia.