

ANALIZA CZASU TRWANIA NA RYNKU NA PRZYKŁADZIE PRZEDSIĘBIORSTW Z WOJEWÓDZTWA MAZOWIECKIEGO

Dominika Urbańczyk  <https://orcid.org/0000-0002-5134-4693>

Wydział Zastosowań Informatyki i Matematyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: dominika_urbanczyk@sggw.pl

Streszczenie: Praca poświęcona jest analizie czasu trwania przedsiębiorstw na rynku. Próbę badawczą stanowiła obserwowana przez 10 lat kohorta firm z województwa mazowieckiego, które rozpoczęły działalność w 2007 r. Wykorzystano nieparametryczne oraz parametryczne metody analizy przeżycia, w tym modele hazardu. Zwrócono również uwagę na problem cenzurowania danych. Wyniki uzyskane w obu podejściach są zbieżne. Zauważono, że największe ryzyko zakończenia działalności występuje pomiędzy 2 a 5 rokiem istnienia firmy. Istotny wpływ na czas trwania przedsiębiorstw mają takie czynniki jak siedziba, forma własności, a także sekcja działalności według PKD.

Słowa kluczowe: analiza przeżycia, nieparametryczne i parametryczne modele hazardu, czas trwania przedsiębiorstw na rynku, dane cenzurowane, estymator Kaplana-Meiera

JEL classification: C4

WSTĘP

Kondycja firm jest kluczowym aspektem w ocenie sytuacji ekonomicznej i planowaniu polityki gospodarczej. Rozwój przedsiębiorczości jest korzystny nie tylko ze względu na poprawę wyników finansowych, ale także – przyczyniając się do większej aktywności zawodowej – wspomaga rozwiązywanie problemu bezrobocia. Nietrywialnym problemem jest również prawidłowa ocena sytuacji przedsiębiorstw w kontekście ich zdolności kredytowej lub też konstrukcji oferty ubezpieczeń. Tłumaczy to zainteresowanie zagadnieniem przetrwania firm na rynku [Boratyńska 2013; Dominiak, Mazurkiewicz 2011; Hadasik 1998; Kisielińska 2016;

<https://doi.org/10.22630/MIBE.2019.20.2.11>

Korol, Prusak 2015; Markowicz, Bieszk-Stolorz 2006; Markowicz 2016; Mikulec 2017; Ptak-Chmielewska 2012; Ptak-Chmielewska 2016].

Biorąc pod uwagę znaczenie sytuacji firm dla całej gospodarki, niepokojące wydają się publikowane przez Główny Urząd Statystyczny wyniki wskazujące na to, że duża część spośród małych i średnich przedsiębiorstw (MŚP) nie jest w stanie przetrwać na rynku pierwszych pięciu lat. W celu wprowadzenia skutecznych rozwiązań konieczne jest zdobycie rzetelnej wiedzy.

Zdolność prawidłowego prognozowania czasu trwania przedsiębiorstwa na rynku jest niewątpliwie istotnym problemem badawczym. Jednak konieczne jest również poznanie czynników warunkujących zdolność firm do kontynuowania działalności gospodarczej. Warty uwagi jest również problem identyfikacji okresu cyklu życia przedsiębiorstwa, w którym ryzyko zakończenia działalności jest największe. Kolejnym bardzo ważnym aspektem jest zapotrzebowanie na wiedzę o cechach populacji firm.

Podjmując temat przetrwania przedsiębiorstw na rynku należy mieć jednak świadomość występujących w tego rodzaju badaniach ograniczeń i trudności. Należy zwrócić uwagę na charakter danych dotyczących wszystkich czasów trwania. Przede wszystkim występuje tutaj problem niepełnej informacji oraz konieczność pracy z danymi uciętymi i cenzurowanymi. Brak wstępnej wiedzy na temat badanej populacji przedsiębiorstw uniemożliwia sformułowanie założeń *a priori* dotyczących rozkładu czasu trwania. Dodatkowo należy brać pod uwagę, że nietypowe indywidualne cechy firm (np. szczególne warunki otoczenia podmiotu) mogą mieć zaburzający wpływ na wynik i wnioski dotyczące całej populacji przedsiębiorstw.

Celem pracy jest zbadanie czasu trwania dla populacji firm z województwa mazowieckiego z wykorzystaniem podejścia opartego na analizie przeżycia – w tym modelach hazardu. Tego rodzaju metody statystyczne są zaprojektowane do badania nieujemnych zmiennych losowych, a w szczególności czasów trwania. Ich główną zaletą w stosunku do innych stosowanych w tego rodzaju badaniach metod jest możliwość uwzględnienia występowania danych cenzurowanych i uciętych, co pozwala na uzyskanie wyników obarczonych mniejszym błędem.

PROBLEM CENZUROWANIA DANYCH

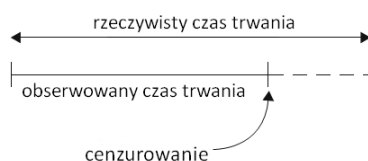
Niezależnie od aspektów będących przedmiotem zainteresowania, w pracach poświęconych badaniu czasów trwania (czyli czasu pomiędzy wystąpieniem określonych zdarzeń) pojawia się problem uwzględniania obserwacji, o których nie ma pełnej informacji. Mówimy wtedy o obserwacjach cenzurowanych i uciętych. Należy tutaj głównie zwrócić uwagę na przypadki, kiedy jedno ze zdarzeń ma miejsce przed rozpoczęciem obserwacji (lewostronne cenzurowanie), badanie zakończy się przed wystąpieniem zdarzenia końcowego, bądź też w trakcie prowadzenia badań stracimy możliwość obserwacji danego obiektu (cenzurowanie

prawostronne). W przypadku, kiedy występują oba zjawiska mówimy o cenzurowaniu obustronnym [Sokołowski 2010].

W takich sytuacjach nie możemy dokładnie określić czasu trwania dla danego epizodu, ale mamy częściową informację o tym, że w populacji wystąpiły takie zdarzenia i dla tych obserwacji możemy podjąć próbę określenia dolnej granicy czasu trwania. Zupełne pominięcie tego rodzaju obserwacji wydaje się traceniem cennych informacji o populacji. Z drugiej strony nie można uciętego czasu trwania traktować w sposób równoznaczny z dokładnie zmierzonymi czasami trwania dla innych obserwacji. Pozostaje więc pytanie jak radzić sobie z danymi cenzurowanymi i uciętymi, czy uwzględniać je w badaniach i jeśli tak to w jaki sposób.

W przypadku prowadzenia badań kohortowych, kiedy badanie danej populacji rozpoczyna się w momencie początku czasu trwania, eliminujemy problem cenzurowania lewostronnego. Natomiast nadal występują obserwacje cenzurowane prawostronnie (patrz rysunek 1). Wybrana metoda badawcza powinna więc umożliwiać uwzględnienie tego rodzaju zjawiska.

Rysunek 1. Czas trwania cenzurowany prawostronnie



Źródło: opracowanie własne

ANALIZA PRZEŻYCIA

Analiza przeżycia jest zbiorem metod statystycznych opracowanych do badania nieujemnych zmiennych losowych, w szczególności czasów trwania. Niech badany czas trwania będzie zmienną losową T , przyjmującą wartości z przedziału $(0, \infty)$. Wówczas funkcję S nazywamy funkcją przeżywalności i definiujemy następująco:

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

gdzie F jest dystrybuantą, a f funkcją gęstości prawdopodobieństwa zmiennej losowej T . Bezpośredni związek funkcji przeżywalności z dystrybuantą i funkcją gęstości prawdopodobieństwa wskazuje na to, że wyznaczenie funkcji S pozwala w pełni określić rozkład zmiennej losowej oraz stosować metody wnioskowania statystycznego w odniesieniu do tej funkcji [Klein i Moeschberger 2003].

Dla znanej funkcji przeżywalności wartość oczekiwaną zmiennej losowej T , czyli oczekiwany czas trwania definiuje się w następujący sposób:

$$E(T) = \int_0^{\infty} S(t) dt. \quad (2)$$

Ryzyko zajścia oczekiwanego zdarzenia końcowego wzdłuż rozkładu zmiennej T , może być opisywane poprzez analizę zmienności funkcji hazardu $h(t)$ (inaczej funkcji intensywności procesu) [Sokołowski 2010]. Jest ona wyrażona poprzez granicę prawdopodobieństwa warunkowego w następujący sposób [Klein i Moeschberger 2003]:

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt}. \quad (3)$$

Interpretacja wartości funkcji wykorzystywanych w analizie przeżycia w kontekście zagadnienia przetrwania przedsiębiorstw na rynku jest następująca: dla danego czasu t ryzyko wyjścia przedsiębiorstwa z rynku jest tym większe, im większą wartość przyjmuje funkcja hazardu $h(t)$. Natomiast prawdopodobieństwo, że firma będzie prowadziła działalność dłużej niż czas t jest tym większe, im większa jest wartość funkcji przeżywalności $S(t)$.

Tablice trwania życia

Jedną z najstarszych i najbardziej podstawowych metod analizy przeżycia jest metoda tablic trwania życia. W tej metodzie wartości czasu trwania dzieli się na przedziały jednakowej długości. Dla każdego z takich przedziałów można wyznaczyć liczbę obserwacji na początku przedziału (n_i) oraz liczbę obserwacji zakończonych (z_i) i uciętych (c_i) w danym przedziale. Na tej podstawie dla kolejnych przedziałów wyznacza się warunkowe prawdopodobieństwa przeżycia $\frac{n_{i+1}}{n_i}$, gdzie

dla każdego t : $n_{i+1} = n_i - z_i - c_i$.

Estymator Kaplana-Meiera funkcji przeżywalności

Estymator Kaplana-Meiera jest nieparametrycznym estymatorem największej wiarygodności dla funkcji przeżywalności. Metoda ta nie wymaga konstruowania podziału czasu obserwacji na klasy. Niech T będzie zmienną losową opisującą czas trwania. Wówczas dla analogicznych oznaczeń jak w przypadku tablic trwania życia estymator jest postaci:

$$\hat{S}(t) = \prod_{j=1}^i P(T > t_j | T \geq t_{j-1}) = \prod_{j=1}^i \hat{p}_j = \prod_{j=1}^i \left(1 - \frac{z_j}{n_j} \right). \quad (4)$$

Estymator pozwala na uzyskanie funkcji przeżywalności bez wcześniejszej znajomości rozkładu. Wyodrębnienie podgrup wewnątrz próby może być

zrealizowane poprzez estymację funkcji przeżywalności dla każdej z grup oddzielnie i porównanie ich postaci.

Modele regresyjne

Modele regresyjne wymagają zastosowania założeń *a priori* dotyczących rozkładu zmiennej losowej. Ich zaletą jest jednak możliwość włączenia do modelu zmiennych objaśniających, co pozwala na ilościową ocenę wpływu poszczególnych czynników na czas trwania. W ogólnej postaci model regresyjny dla funkcji przeżywalności jest postaci:

$$\hat{S}(t) = P(T > t | Z) = \hat{S}_0(t) \exp(\hat{\beta}Z), \quad (5)$$

gdzie $\hat{S}_0(t)$ jest bazową funkcją przeżycia, Z jest wektorem zmiennych objaśniających, a β wektorem parametrów.

WYNIKI BADAŃ

Wykorzystana baza danych

W pracy wykorzystano dane z rejestru REGON (prowadzonego przez Główny Urząd Statystyczny) dla przedsiębiorstw z województwa mazowieckiego, które rozpoczęły działalność w 2007 r. Obserwację zakończono na koniec 2017 roku. Badania zostały zrealizowane dla całej populacji, która zawierała 32 788 przedsiębiorstw. Spośród informacji zawartych w rejestrze REGON na potrzeby badania wybrano następujące charakterystyki dla każdej z firm:

Tabela 1. Zmienne wykorzystane w modelach

Nazwa zmiennej	Opis i wartości
data_roz poczecia	data dzienna rozpoczęcia działalności
data_zakonczenia	data dzienna zakończenia działalności
siedziba	adres siedziby firmy (na potrzeby badania przyjęto Warszawa dla siedziby w Warszawie, inne – dla pozostałych obszarów woj. mazowieckiego)
wlasnosc	forma własności (os. fizyczna, os. prawna, jedn. org. bez os. prawnej)
PKD2007	sektor działalności (wartości według głównych sekcji PKD 2007)

Źródło: opracowanie własne

Metody nieparametryczne

Analizę czasu trwania przedsiębiorstw na rynku rozpoczęto od zastosowania metod nieparametrycznych ze względu na brak konieczności stosowania założeń *a priori*, co było bardzo istotne, gdyż na początku badania Autor nie dysponował wiedzą na temat rozkładu czasu trwania.

W pierwszym etapie skonstruowano tablicę trwania dla firm w województwie mazowieckim (tabela 2).

Tabela 2. Tablica trwania firm w woj. mazowieckim (czas podano w miesiącach)

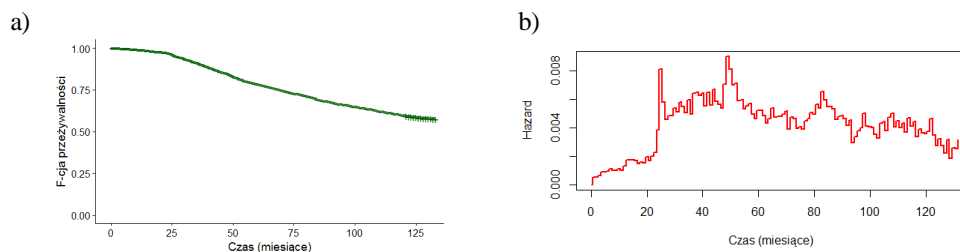
t	n _t	z _t	c _t	\hat{S}_t
0	32788	113	0	0,997
12	32367	42	0	0,986
24	31712	121	0	0,963
36	29729	146	0	0,902
48	27677	195	0	0,838
60	25648	133	0	0,778
72	24209	114	0	0,735
84	22850	135	0	0,693
96	21605	81	0	0,656
108	20574	102	0	0,624
120	19579	70	0	0,595
121	19509	7	32	0,593
122	19405	86	1345	0,590
126	12515	32	1460	0,583

Źródło: opracowanie własne na podstawie danych z REGON

Warto zauważyć, że ze względu na specyfikę próby badawczej (kohorta firm rozpoczynających działalność w 2007 roku) pierwsze obserwacje cenzurowane pojawiają się dla firm, które przetrwały minimum 10 lat, czyli dotrwały od 2007 roku do zakończenia obserwacji.

Następnie metodą Kaplana-Meiera oszacowano postać funkcji przeżywalności oraz hazardu dla całej próby przedsiębiorstw (rysunek 2).

Rysunek 2. Funkcja przeżywalności (a) i hazard (b) oszacowane metodą Kaplana-Meiera



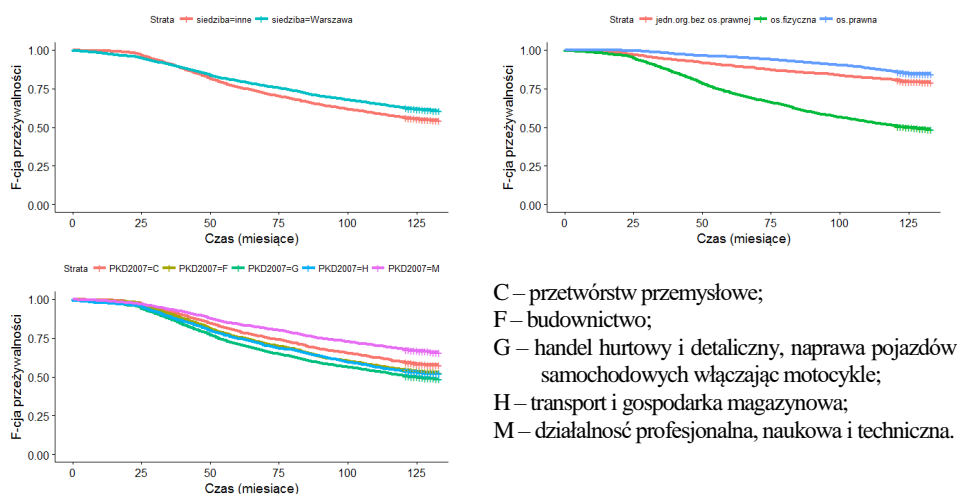
Źródło: opracowanie własne na podstawie danych z REGON

Widać, że w początkowym okresie trwania (do około 24 miesiąca) niewielka część przedsiębiorstw kończy działalność. Natomiast zarówno na podstawie wykresu funkcji przeżywalności jak i funkcji hazardu można zauważyć, że największe ryzyko zakończenia działalności występuje w okresie od 2 do 5 roku trwania firmy, po czym zaczyna maleć.

W kolejnym etapie badania w celu dokładniejszej analizy wpływu cech przedsiębiorstwa na ich zdolność przetrwania na rynku, przeprowadzono estymację funkcji przeżywalności dla podgrup. Jako czynniki podziału zastosowano siedzibę przedsiębiorstwa, formę własności oraz sektor działalności według PKD (rysunek 3). Prezentowane są wyniki dla wybranych pięciu sekcji działalności według PKD

2007 o dużej liczności i znacząco różnych krzywych przeżycia: C – przetwórstwo przemysłowe, F – budownictwo, G – handel hurtowy i detaliczny oraz naprawa pojazdów samochodowych (włączając motocykle), H – transport i gospodarka magazynowa, M – działalność profesjonalna, naukowa i techniczna. Wyniki nieparametrycznych testów na równość rozkładów [Sokołowski 2010] (zastosowano testy typu log-rank, test Wilcozona oraz test Fleminga-Harringtona) potwierdziły, że rozkłady dla podgrup wyodrębnionych ze względu na te zmienne są istotnie różne.

Rysunek 3. Funkcja przeżywalności dla podgrup



Źródło: opracowanie własne na podstawie danych z REGON

Metody parametryczne

Informacje na temat postaci funkcji hazardu uzyskane na podstawie wyników modeli nieparametrycznych, wykorzystano podczas estymacji funkcji przeżywalności i hazardu metodami parametrycznymi. W pierwszym etapie oszacowano modele w podstawowej postaci bez uwzględnienia zmiennych objaśniających. Ze względu na to, że oszacowana w modelach nieparametrycznych funkcja hazardu ma charakter niemonotoniczny zastosowano rozkład logarytmiczno-logistyczny, logarytmiczno-normalny oraz gamma. Uzyskano następujące wyniki:

Rozkład logarytmiczno-logistyczny	$S(t) = \frac{1}{1 + \lambda t^\alpha}$	$\alpha = 1,419$ $\lambda = 158,661$	$\ln L = -88358,24$ $AIC = 176720,5$
Rozkład logarytmiczno-normalny	$S(t) = 1 - \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$	$\mu = 5,241$ $\sigma = 1,494$	$\ln L = -89820,75$ $AIC = 179645,5$
Rozkład gamma	$S(t) = 1 - \int_0^t u^{\beta-1} \exp(-u) du / \Gamma(\beta), \quad u = \lambda t$	$\beta = 1,326$ $\lambda = 0,007$	$\ln L = -88464,15$ $AIC = 176932,3$

Następnie, biorąc pod uwagę kryteria dopasowania modelu (najniższa wartość AIC oraz najwyższa wartość $\ln L$), dla rozkładu logarytmiczno-logistycznego oszacowano model hazardu z uwzględnieniem zmiennych objaśniających. Dla tego rozkładu funkcja hazardu jest opisana wzorem: $h(t) = \frac{\lambda \alpha t^{\alpha-1}}{1 + \lambda t^\alpha}$. W modelu regresyjnym jest postaci $\hat{h}(t) = \hat{h}_0(t) \exp(\hat{\beta}X)$, gdzie h_0 jest hazardem bazowym, β – wektorem parametrów, a X – wektorem zmiennych objaśniających. W celu estymacji modelu dokonano analogicznego jak w podejściu nieparametrycznym wyboru sekcji działalności według PKD. Uzyskano następujące parametry oszacowanego modelu:

Tabela 3. Estymatory parametrów dla oszacowanego modelu hazardu

Zmienna	Parametr	$\exp(\beta)$
α	1,5511*	-
λ	324,9659*	-
siedziba = Warszawa	0,0848***	1,08850
własność = os.fizyczna	0,9007***	2,46133
własność = os.prawna	-0,3606***	0,69726
PKD2007= F	0,1728***	1,18863
PKD2007= G	0,2493***	1,28313
PKD2007= H	0,0600*	1,06184
PKD2007= M	-0,2479***	0,78044

Grupy referencyjne dla cech to siedziba = inne, własność = jedn. org. bez os. prawnej, PKD2007 = C). AIC = 116220, $\ln L = -58101$,
ozn. poziomu istotności: * to 0,1, ** to 0,01, *** to 0,001

Źródło: opracowanie własne na podstawie danych z REGON

Ze względu na spodziewany różny wpływ poszczególnych sekcji działalności w zależności od siedziby przedsiębiorstwa, oszacowano również osobne modele dla firm z Warszawy i pozostałej części województwa mazowieckiego.

Tabela 4. Estymatory parametrów dla modelu hazardu oszacowanego dla firm z Warszawy

Zmienna	Parametr	$\exp(\beta)$
α	1,44522*	-
λ	182,19801*	-
PKD2007= F	-0,04074*	0,96008
PKD2007= G	0,32635***	1,38590
PKD2007= H	0,40123***	1,49366
PKD2007= M	-0,10716**	0,89838

Grupa referencyjna to PKD2007 = C. AIC = 46169, $\ln L = -23078$,
oznaczenie poziomu istotności: * to 0,1, ** to 0,01, *** to 0,001

Źródło: opracowanie własne na podstawie danych z REGON

Tabela 5. Estymatory parametrów dla modelu hazardu oszacowanego dla firm poza Warszawą

Zmienna	Parametr	$\exp(\beta)$
α	1,6563*	-
λ	141,8764*	-
PKD2007= F	0,2046***	1,22703
PKD2007= G	0,2145***	1,23924
PKD2007= H	0,0110	1,01106
PKD2007= M	-0,2870***	0,75051

Grupa referencyjna to PKD2007 = C.

AIC = 70889, lnL = -35438, ozn. poziomu istotności: * to 0,1, ** to 0,01, *** to 0,001.

Źródło: opracowanie własne na podstawie danych z REGON

Zarówno w przypadku podejścia nieparametrycznego jak i oszacowanych modeli regresyjnych zaobserwowano różnice pomiędzy czasem trwania dla firm z siedzibą w Warszawie i poza granicami administracyjnymi miasta stołecznego. W pierwszych miesiącach prowadzenia działalności większą przeżywalnością charakteryzują się firmy spoza stolicy, natomiast w późniejszym okresie zauważalnie wyższą przeżywalność obserwowano dla firm z Warszawy (rysunek 3). Znacznie „bezpieczniejsze na rynku” są przedsiębiorstwa z osobowością prawną, co może być związane z tym, że są to większe firmy, często z sektora publicznego. Natomiast szczególnie narażone na ryzyko zakończenia działalności są raczej małe i średnie przedsiębiorstwa. W całym województwie mazowieckim wśród analizowanych sektorów największą zdolnością przetrwania charakteryzują się firmy zajmujące się działalnością profesjonalną, naukową i techniczną (tabela 3: dla tej zmiennej najniższa i mniejsza od 1 wartość $\exp(\beta)$, czyli ilorazu hazardu w przypadku rozważania wpływu prowadzenia działalności w tym sektorze przy założeniu *ceteris paribus*). Natomiast najbardziej narażoną na ryzyko sekcją działalności w Warszawie jest transport i gospodarka magazynowa (tabela 4), podczas gdy poza Warszawą nie zaobserwowano istotnego wpływu prowadzenia działalności tego typu na zdolność przetrwania na rynku (tabela 5). W pozostałej części województwa największe ryzyko zaobserwowano dla firm z sektora handlu, naprawy pojazdów samochodowych oraz budownictwa.

PODSUMOWANIE

Sytuacja firm staje się obecnie coraz istotniejszym zagadnieniem. Celem pracy była analiza czasu trwania przedsiębiorstw na rynku na przykładzie podmiotów gospodarczych z województwa mazowieckiego. Zrealizowano go przy wykorzystaniu metod analizy przeżycia, w szczególności modeli hazardu. Porównano wyniki podejścia nieparametrycznego oraz parametrycznego. Należy zauważyć, że istotną zaletą wybranych metod była możliwość uwzględnienia cenzurowania czasów trwania, co stanowi istotny problem w tego rodzaju badaniach.

Zastosowane metody nieparametryczne pozwoliły na oszacowanie postaci funkcji przeżywalności i hazardu. Na tej podstawie zauważno, że największe ryzyko zakończenia działalności występuje pomiędzy 2 a 5 rokiem istnienia przedsiębiorstwa, przy czym obserwuje się wyraźny wzrost tego ryzyka po zakończeniu 24 miesiąca. Biorąc pod uwagę duży udział w populacji podmiotów z sektora MŚP, można wnioskować, że wzrost ten jest powiązany z końcem preferencyjnego okresu dla nowych firm (np. obniżonej składki ZUS).

Zastosowane metody badawcze pozwoliły na zbadanie czasu trwania na rynku przedsiębiorstw z województwa mazowieckiego. Wyniki metod nieparametrycznych wykorzystano w dalszej części pracy, przy wyborze rozkładu podczas konstrukcji modeli parametrycznych. Wnioski sformułowane na podstawie metod nieparametrycznych dotyczące wpływu zmiennych na postać funkcji przeżywalności i hazardu zostały potwierdzone i uzupełnione przez ilościową analizę wykonaną za pomocą metod parametrycznych. Istotnymi czynnikami wpływającymi na czas trwania na rynku są siedziba przedsiębiorstwa, forma własności oraz sektor działalności według klasyfikacji PKD.

Dalsze prace badawcze poświęcone będą przeprowadzeniu analiz z uwzględnieniem różnych form zakończenia działalności firmy (modele ryzyk konkurencyjnych), co pozwoli na pełniejsze zrozumienie mechanizmów warunkujących zdolność przedsiębiorstwa do przetrwania na rynku.

BIBLIOGRAFIA

- Boratyńska K. (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, 27-32.
- Dominiak P., Mazurkiewicz M. (2011) Analysis of the Risk of Company's Bankruptcy in Polish Food and Beverage Production Sector using the Cox Regression. *Operations Research and Decisions*, 21(1), 19-31.
- Hadasik D. (1998) Upadłość przedsiębiorstw w Polsce i metody jej prognozowania. *Zeszyty Naukowe. Akademia Ekonomiczna w Poznaniu, Seria 2, Prace Habilitacyjne 1230-6673 z. 153*, 187-196.
- Kisielińska J. (2016) Skuteczność modeli upadłości przedsiębiorstw. *Studia Ekonomiczne i Regionalne*, 9(1), 5-17.
- Klein J. P., Moeschberger M. L. (2003) *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, 21-90.
- Korol T., Prusak B. (2015) Upadłość przedsiębiorstw a wykorzystanie sztucznej inteligencji. *CeDeWu*, Warszawa, 5-126.
- Markowicz I., Bieszk-Stolorz B. (2006) Wykorzystanie analizy historii zdarzeń do konstrukcji tablic żywotności firm. *Wiadomości Statystyczne*, 4, 1-7.
- Markowicz I. (2016) Analiza trwania firm w woj. zachodniopomorskim. *Wiadomości Statystyczne*, 1, 44-61.
- Mikulec A. (2017) Kohortowe tablice trwania przedsiębiorstw w województwie łódzkim – ujęcie kwartalne. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 468, 147-160.

- Ptak-Chmielewska A. (2012) Wykorzystanie modeli przeżycia i analizy dyskryminacyjnej do oceny ryzyka upadłości przedsiębiorstw. *Ekonometria*, 4, 157-172.
- Ptak-Chmielewska A. (2016) Determinanty przeżywalności mikro i małych przedsiębiorstw w Polsce. Oficyna Wydawnicza SGH. Warszawa, 10-200.
- Sokołowski A. (2010) Jak rozumieć i wykonywać analizę przeżycia. *StatSoftPolska*, www.statsoft.pl/czytelnia.html, 33-48.

SURVIVAL ANALYSIS OF ENTERPRISES ON THE EXAMPLE OF THE MAZOWIECKIE VOIVODESHIP

Abstract: This paper is devoted to investigate the enterprises duration on the market. The research sample consisted of a cohort of companies from the Mazowieckie voivodeship, which started their economic activity in 2007 (observed for 10 years). Non-parametric and parametric methods of survival analysis (including hazard models) were applied. Attention was also paid to the problem of censored data. The results obtained in both approaches are convergent. It was noted that the greatest risk of termination of business occurs between 2 and 5 years of the company existence. The factors such as the headquarters, ownership form and the business section have a significant impact on the duration of enterprises.

Keywords: survival analysis, non-parametric and parametric hazard models, enterprises duration on the market, censored data, Kaplan-Meier estimator

JEL classification: C4