

**Aneta Ptak-Chmielewska**

Szkoła Główna Handlowa w Warszawie

e-mail: aptak@sgh.waw.pl

---

## **MODELE PREDYKCJI UPADŁOŚCI MŚP W POLSCE – ANALIZA Z WYKORZYSTANIEM MODELU PRZEŻYCIA COXA I MODELU REGRESJI LOGISTYCZNEJ**

---

**Streszczenie:** Ryzyko kredytowe jest jednym z najważniejszych rodzajów ryzyka, na które wystawiony jest bank. Ryzyko upadłości przedsiębiorstw jest zazwyczaj modelowane z wykorzystaniem analizy dyskryminacyjnej lub regresji logistycznej. Jednakże poszukiwane są nowe sposoby i techniki efektywniejszej predykcji bankructwa. Jedną z takich metod jest analiza przeżycia. Analiza prezentowana w tym artykule ma na celu porównanie nowych metod predykcji bankructwa z tradycyjnymi modelami, takimi jak regresja logistyczna. Wskazane są zalety i wady tych metod oraz propozycja rozszerzenia analizy przeżycia o makrozmiennie, a modelu regresji – o zmienne nominalne. Włączenie makrozmiennych podnosi moc predykcyjną modeli. Oszacowano modele z wykorzystaniem próby polskich przedsiębiorstw MŚP zawierającej 1561 przedsiębiorstw, w tym 807 upadłości.

**Słowa kluczowe:** analiza przeżycia, makrozmiennie, model Coxa, ryzyko bankructwa.

DOI: 10.15611/ekt.2014.4.01

### **1. Wstęp**

Systemy i modele wczesnego ostrzegania przed upadłością cechuje mnogość definicji, ale niezależnie od definicji jest to podstawowe narzędzie analityczne dostarczające informacji o istniejących rodzajach ryzyka w działalności przedsiębiorstw. Modele wczesnego ostrzegania przed upadłością nie wskazują rozwiązania problemu, a jedynie kwantyfikują ryzyko wystąpienia problemu. Podstawowe modele bazują na wybranych wskaźnikach ekonomicznych i finansowych i wspierane są przez coraz bardziej zaawansowane metody i modele statystyczne i ekonometryczne.

Pierwsze modele wczesnego ostrzegania powstały na początku XX wieku. Model Altmana w latach 60. był prekursorem przejścia z modeli jednowymiarowych do wielowymiarowych modeli statystycznych wykorzystujących analizę dyskry-

minacyjną. W latach 70. i 80. obserwowano dynamiczny rozwój i wykorzystanie modeli logitowych. W latach 90. rozpoczęto natomiast stosowanie metod sztucznej inteligencji i sieci neuronowych. W Polsce systemy wczesnego ostrzegania zdobyły popularność w latach 90. w wyniku zmian gospodarczych (transformacja gospodarcza), jednak nie są w dalszym ciągu popularną metodą. Największy postęp w tym obszarze odnotowuje się raczej w działalności bankowej, gdzie ocena ryzyka upadłości związana jest z oceną i kwantyfikacją ryzyka kredytowego podmiotu jako kredytobiorcy.

Wykorzystanie zagranicznych modeli w polskich warunkach nie jest możliwe bezpośrednio ze względu na występowanie szczególnych warunków ekonomicznych gospodarki postkomunistycznej.

Celem artykułu jest porównanie nowych metod predykcji bankructwa (modeli przeżycia) z tradycyjnymi modelami, takimi jak regresja logistyczna, wskazanie zalet i wad tych metod oraz propozycja rozszerzenia analizy przeżycia o zmienne zależne od czasu (makrozmiennie), a modelu regresji o zmienne nominalne.

## 2. Przegląd literatury

W badaniu przeprowadzonym przez Laitinen [2002] wykorzystano analizę przeżywalności celem określenia modelu opisującego, jaki czas upływa do pierwszego braku spłaty przez przedsiębiorstwo (default). Zbiór danych zawiera informacje o przedsiębiorstwach działających na terenie Finlandii. Są to dane finansowe pokrywające pięcioletni okres historii przedsiębiorstwa oraz dodatkowo zbiór podzielony został na próbę uczącą i testową. Wyniki finansowe, wielkość, branża oraz czas istnienia przedsiębiorstwa stanowiły zmienne w modelu procesu przeżywalności poprzedzające pierwszy brak spłaty. Model hazardu został porównany z modelem regresji logistycznej dla predykcji bankructwa na danych w ciągu jednego roku przed zdarzeniem. Przedstawiano model proporcjonalnych hazardów celem ukazania prognozy braku spłaty na lata wcześniejsze, przed rozpoczęciem trudnej sytuacji finansowej w przedsiębiorstwie

W badaniach Jacobson i innych [2011] analizowany był związek pomiędzy wahaniami makroekonomicznymi (cyklami koniunkturalnymi) a defaultami przedsiębiorstw, biorąc również pod uwagę gałąź przemysłu oraz szeroki zakres czynników specyficznych dla przedsiębiorstwa. W badaniu wykorzystany został zbiór danych panelowych zawierający informację o niemalże wszystkich założonych w Szwecji w latach 1990-2009 firmach, czyli okresie zawierającym również kryzys w sektorze bankowym. Autorzy znaleźli mocne dowody na znaczny i stały wpływ zagregowanych wahań na defaulty przedsiębiorstw. Autorzy wykorzystali standardowy model regresji logistycznej oparty na wskaźnikach finansowych wzmocnionych czynnikami makroekonomicznymi. Efekty zmiennych makroekonomicznych różniły się w zależności od branży (zgodnie z oczekiwaniami). Modele były lepsze (na próbie *out-of-sample*) niż modele szeregów czasowych i modele bez zmiennych makro.

Whalen [1991] zajmował się modelami opartymi na modelach proporcjonalnych hazardów Coxa, które dostarczyły oszacowań prawdopodobieństw związanych z kryzysem bankowym w USA. Próba zawierała wszystkie banki, które upadły pomiędzy 1 stycznia 1987 a 31 października 1990 r., oraz próbę losowo wybranych 1500 banków w dobrej kondycji. Wykorzystano dostępny ogólnie zestaw danych – zmiennych objaśniających. Celem modelu była identyfikacja banków upadłych i w dobrej kondycji z jak najlepszą trafnością klasyfikacji. Siła dyskryminacyjna takiego modelu była bardzo wysoka.

Cole i Wu [2009] zbudowali dynamiczny model hazardu ze zmiennymi zależnymi od czasu do oceny wczesnego ostrzeżenia przed bankructwem banków, następnie przetestowali na próbie *out-of-sample* do oceny siły dyskryminacyjnej modelu, porównując z prostym modelem probitowym. Włączenie zmiennych zależnych od czasu pozwoliło na wykorzystanie zmiennych makroekonomicznych. Model okazał się znacznie lepszy od modelu probitowego zarówno ze zmiennymi, jak i bez zmiennych makroekonomicznych.

Próby budowy modeli ostrzeżenia przed upadłością były również podejmowane w Polsce. Implementacja zachodnich modeli predykcji zagrożenia upadłością do rynku przedsiębiorstw funkcjonujących w warunkach gospodarki w okresie transformacji takiej jak w Polsce nie powiodła się. Modele, oparte głównie na analizie dyskryminacyjnej, m.in. Altmanna, nie sprawdzają się w warunkach dynamicznych zmian politycznych i gospodarczych. Niezadowolające efekty wykorzystania w warunkach polskich modeli zagranicznych przyczyniły się do rozwoju badań nad modelami rodzimymi. Największą popularność, podobnie jak za granicą, zdobyły modele oparte na analizie dyskryminacyjnej. W latach 90. podjęto prace nad budową i wdrożeniem modeli dostosowanych do specyfiki gospodarki polskiej (były to między innymi prace takich autorów, jak D. Hadasik, E. Mączyńska, J. Gajdka i D. Stos, B. Prusak, M. Hamrol). Stosowano wielowymiarową analizę dyskryminacyjną, modele regresji logistycznej, modele sieci neuronowych. Wśród autorów modeli upadłości, którzy ogłosili swoje prace w latach późniejszych, można wymienić nazwiska, takie jak: T. Stępień i T. Strąk, A. Hołda, M. Gruszczyński, E. Mączyńska i M. Zawadzki, D. Appenzeller i wielu innych. W modelach tradycyjnych nie uwzględniane są jednak zmiany w czasie, które mogą być istotne. Zmiany w czasie uwzględniają modele przeżycia (tzw. analizy historii zdarzeń), których zastosowanie jest coraz częściej spotykane w pracach naukowych, m.in. autorstwa A. Ptak-Chmielewskiej i I. Schab. Wraz z rozwojem technik komputerowych i metod eksploracji danych (*data mining*) pojawiły się również prace wykorzystujące i konfrontujące tradycyjne modele analizy dyskryminacyjnej i regresji logistycznej z modelami opartymi na eksploracji danych typu: drzewa decyzyjne i sieci neuronowe. T. Korol pokazał przewagę sieci neuronowych nad analizą dyskryminacyjną na przykładzie 180 przedsiębiorstw produkcyjnych i sprawozdań z lat 1998-2001, natomiast T. Strąk wykorzystał drzewa klasyfikacyjne, wskazując potrzebę odejścia od trady-

cyjnych metod dyskryminacyjnych. W swoich pracach K. Dębowska porównała analizę dyskryminacyjną, regresję logistyczną i drzewa klasyfikacyjne na próbie 68 przedsiębiorstw ze sprawozdaniami z 2009 r. A. Ptak-Chmielewska pokazała przewagę analizy przeżycia i modelu Coxa w stosunku do regresji logistycznej i analizy dyskryminacyjnej. Często jednakże takie analizy prowadzone są bez refleksji nad metodami statystycznymi i w oderwaniu od podstawowych ograniczeń, jakie metody statystyczne ze sobą niosą.

Zgodnie z dokonaniem przeglądem literatury podstawowe pytania badawcze postawione w tej pracy są następujące:

- Czy model Coxa jest bardziej efektywny do pomiaru ryzyka upadłości przedsiębiorstw? Jakie są podstawowe zalety w wykorzystaniu modeli przeżycia?
- Czy uzupełnienie analizy o zmienne nominalne poprawia dyskryminację modelu? Czy rodzaj stosowanego modelu ma wpływ na siłę dyskryminacyjną predykcji?

### 3. Metody analizy

Dwa rodzaje modeli zostały wybrane do analizy: model przeżycia – regresja Coxa, oraz model regresji logistycznej. W porównaniu z tradycyjnym podejściem (regresja logistyczna) modele przeżycia dostarczają dynamicznej prognozy analizowanego ryzyka upadłości.

W modelu semiparametrycznym (model proporcjonalnych hazardów Coxa) tylko część dotycząca regresji jest parametrycznie wyspecyfikowana (interakcja pomiędzy procesami), natomiast rozkład czasu jest parametrycznie niewyspecyfikowany (podejście nieparametryczne). Zakłada się, iż zmienna ciągła  $T$  oznacza czas do wystąpienia zdarzenia. W modelu regresji Coxa czas jest estymowany jako funkcja hazardu:

$$h(t) = h_0(t)\exp(\alpha_1 x_1 + \dots + \alpha_k x_k),$$

gdzie:  $h_0(t)$  – hazard bazowy,  $X_1, X_2, \dots, X_k$  – zmienne objaśniające.

Cox [1972] zaproponował wykorzystanie metody częściowej największej wiarygodności do estymacji modeli semiparametrycznych. W tej metodzie funkcja wiarygodności jest podzielona na dwie części – pierwszą, zawierającą tylko parametry, i drugą, zawierającą parametry i funkcje hazardu. Podstawowym założeniem w modelu Coxa jest założenie proporcjonalności hazardów, czyli hazard jest proporcjonalny dla każdej pary jednostek i jest stały w czasie. Kiedy założenie nie jest spełnione, wówczas model staje się modelem nieproporcjonalnych hazardów przez włączenie interakcji pomiędzy zmienną  $X$  i czasem procesu  $t$  [Blossfeld, Rohwer 2002].

Drugim typem analizowanych modeli jest regresja logistyczna wykorzystująca do estymacji parametrów metodę największej wiarygodności.

Szacowane prawdopodobieństwo w modelu jest określone jako:

$$P(Y = 1) = 1/(1 + \exp\{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)\}),$$

gdzie:  $P(Y = 1)$  – zmienna zależna, prawdopodobieństwo upadłości,

$\beta_0$  – wyraz wolny,

$\beta_i$  for  $i = 1, 2, \dots, k$  – współczynniki,

$X_i$  for  $i = 1, 2, \dots, k$  – zmienne niezależne (wskaźniki finansowe).

$P(Y = 1)$  przyjmuje wartości z  $[0;1]$ .

*Odds ratio* informuje o względnym ilorazie szans zajścia zdarzenia (lub o zmianie szansy zajścia zdarzenia przy zmianie wartości zmiennej objaśniającej o jednostkę dla zmiennych ciągłych).

Model wymaga spełnienia założeń: losowość próby, liczne obserwacje, brak współliniowości zmiennych objaśniających, niezależność obserwacji.

### 3.1. Próba danych wykorzystana do badania

Próba poddana analizie zawiera dane o 1053 przedsiębiorstwach w dobrej kondycji i 494 w upadłości. Dane pochodzą z okresu od 2002 do 2010 (2004-2012 dla historii upadłości). Dla przedsiębiorstw dostępnych było 2910 sprawozdań finansowych. Ograniczono próbę tylko do przedsiębiorstw o obrotach rocznych w zakresie 8-140 milionów PLN.

Jako bazowy zastosowano model Z-Score(bis) Altmana oparty na formule:

$$Z = 0.717 \cdot X1 + 0.847 \cdot X2 + 3.107 \cdot X3 + 0.420 \cdot X4 + 0.998 \cdot X5,$$

gdzie:  $X1$  = kapitał pracujący/aktywa,

$X2$  = zatrzymane dochody/aktywa,

$X3$  = EBIT/aktywa,

$X4$  = wartość księgowa kapitału (*equity*)/wartość księgowa zadłużenia,

$X5$  = sprzedaż/aktywa,

W modelu zaproponowanym przez Altmana na próbie danych wykorzystanych przez autora model poprawnie zaklasyfikował 90,9% przedsiębiorstw w upadłości i 97,0% w dobrej kondycji. Próba zawierała tylko 66 firm (33 w upadłości oraz 33 w dobrej kondycji).

W pierwszym kroku zastosowano model Z-Score Altmana, otrzymując niezbyt wysoką siłę dyskryminacyjną modelu ( $AUC = 0,699$ ).

Zakłada się, iż wartość pola pod krzywą AUC powinna wynosić co najmniej 0,75-0,80, aby uznać siłę dyskryminacyjną modelu jako satysfakcjonującą.

Dla próby przedsiębiorstw polskich siła dyskryminacyjna tego modelu była więc na poziomie nisko średnim. Powodem mogła być heterogeniczność próby. Próba była ograniczona co do wielkości podmiotów, ale zawierała wszystkie typy działalności (branża).

### 3.2. Wyniki dla modelu regresji Coxa

Zastosowano model regresji Coxa dla oryginalnych wskaźników wybranych do funkcji Z-Score Altmana. Istotność wszystkich wskaźników potwierdziła się również w modelu przeżycia Coxa. Siła dyskryminacyjna tego modelu jest znacznie wyższa w porównaniu z oryginalnym modelem Z-Score Altmana i wynosi  $AUC = 0,746$ . Poziom siły dyskryminacyjnej tego modelu można uznać za średni. Sprawdzone również interakcje z czasem procesu w modelu Coxa – w celu sprawdzenia założenia o proporcjonalności hazardów.

Model Coxa staje się modelem nieproporcjonalnych hazardów, kiedy założenie o proporcjonalności nie jest spełnione (tak jak dla zmiennej X3, X4, X5). Siła dyskryminacyjna takiego modelu jest znacznie wyższa i wynosi  $AUC = 0,827$ .

Włączenie do modelu Coxa zmiennych makroekonomicznych podnosi siłę dyskryminacyjną modelu. Zmienne makroekonomiczne w modelu pokazują, iż wyższemu poziomowi PKB i bezrobocia towarzyszy niższe ryzyko upadłości. Kierunek wpływu dla zmiennej bezrobocie nie jest oczywisty i nie jest potwierdzony w wynikach innych badań. Siła dyskryminacyjna takiego modelu jest wyższa niż poprzedniego. Włączenie zmiennych makroekonomicznych (zależnych od cyklu koniunkturalnego) zwiększa efektywność modelu (por. tab. 1).

**Tabela 1.** Wynik dla modelu Coxa ze zmiennymi makroekonomicznymi

Parametr	SS	Oszacowanie parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi	Hazard Ratio
<b>X1</b>	1	-0,56683	0,08378	45,7800	<,0001	,567
<b>X2</b>	1	-1,39851	0,20301	47,4556	<,0001	,247
<b>X3</b>	1	-2,62719	0,22354	138,1217	<,0001	,072
<b>X4</b>	1	-0,45197	0,07127	40,2140	<,0001	,636
<b>X5</b>	1	-0,17275	0,03288	27,6053	<,0001	,841
<b>PKB</b>	1	-0,16940	0,02839	35,5974	<,0001	,844
<b>Bezrobocie</b>	1	-0,16784	0,02647	40,2175	<,0001	,845
<b>Inflacja</b>	1	0,08828	0,06511	1,8383	0,1752	,092

Źródło: opracowanie własne w SAS.

Siłę dyskryminacyjną takiego modelu można uznać za satysfakcjonującą. Interakcja z czasem jest istotna dla inflacji, pomimo iż sama zmienna nie jest istotna. Siła dyskryminacyjna modelu z interakcjami jest znacznie wyższa i wynosi  $AUC = 0,973$ .

### 3.3. Wyniki dla modelu regresji logistycznej

Również zastosowanie samych wskaźników z funkcji Z-Score w modelu regresji logistycznej z zastosowaniem innych wag (parametrów) podnosi siłę dyskryminacyjną modelu. W modelu zmienne X1 i X2 są statystycznie nieistotne.

**Tabela 2.** Model regresji logistycznej ze zmiennymi makroekonomicznymi

Parametr	SS	Oszacowanie parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi <sup>2</sup>
Wyraz wolny	1	2,1955	0,6535	11,2861	0,0008
X1	1	-0,5156	0,2049	6,3297	0,0119
X2	1	-1,2761	0,7157	3,1795	0,0746
X3	1	-4,4004	0,7092	38,4953	<,0001
X4	1	-0,6167	0,0983	39,3609	<,0001
X5	1	-0,1837	0,0444	17,1047	<,0001
PKB	1	-0,1832	0,0342	28,7606	<,0001
bezrobocie	1	-0,1965	0,0292	45,2945	<,0001
inflacja	1	0,0978	0,0736	1,7655	0,1839

Źródło: opracowanie własne w SAS.

Podobnie jak w modelu regresji Coxa dodanie zmiennych makroekonomicznych podnosi siłę dyskryminacyjną modelu do podobnego poziomu (por. tab. 2). Zmienna inflacja jest w modelu nieistotna. Włączenie zmiennych makroekonomicznych powoduje, iż zmienna X1 staje się istotna.

### 3.4. Wyniki dla modelu regresji logistycznej ze zmienną nominalną

Ostatnim krokiem w analizie było włączenie zmiennej nominalnej. Włączenie do modelu zmiennej nominalnej przyjmującej dużo różnych kategorii jest utrudnione ze względu na:

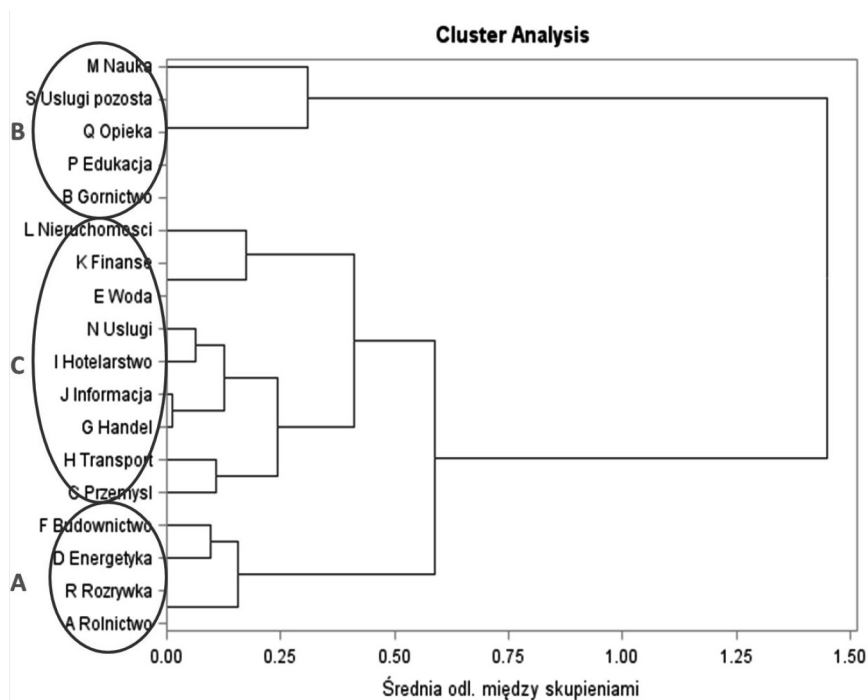
- zbyt duży wymiar analizy (zbyt dużo parametrów do oszacowania w modelu),
- występowanie kategorii bez upadłości,
- nieistotność licznych otrzymanych parametrów, trudności w interpretacji.

Do pogrupowania kategorii zmiennej PKD wykorzystano analizę skupień metodą hierarchiczną łączenia średnich odległości (połączeń). W metodzie tej grupowanie odbywa się na zasadzie najmniejszej średniej odległości między skupieniami. Zmienną użytą do grupowania jest udział przedsiębiorstw w upadłości. Obserwacje są łączone parami, a skupienia na zasadzie najmniejszej średniej odległości pomiędzy skupieniami (historię skupień przedstawiono w tab. 3 oraz na rys. 1).

W wyniku grupowania zaproponowano podział na 3 główne skupienia tak jak na rys. 1:

- Skupienie A: Budownictwo, Energetyka, Rozrywka.
- Skupienie B: Nauka, Usługi pozostałe, Opieka, Edukacja, Górnictwo.
- Skupienie C: Nieruchomości, Finanse, Woda, Usługi, Hotelarstwo, Informacja, Handel, Transport, Przemysł.





**Rys. 1.** Przebieg grupowania kategorii PKD metoda hierarchiczną – dendrogram

Źródło: opracowanie własne w SAS.

**Tabela 3.** PKD 2007 – grupowanie kategorii – metoda hierarchiczna średnich połączeń

Cluster History							
Number of Clusters	Clusters Joined		Freq	Pseudo F Statistic	Pseudo <i>t</i> -Squared	Normalized distance RMS	Tie
1	2		3	4	5	6	7
17	E Water	K Finance	2	.	.	0	T
16	B Mining	P Education	2	.	.	0	T
15	CL16	Q Care	3	.	.	0	T
14	A Agriculture	R Entertainment	2	.	.	0	T
13	CL15	S Other services	4	.	.	0	
12	G Trade	J Information	2	71E3	.	0,0115	
11	I Hotel Management	N Services	2	2877	.	0,0633	
10	D Energetics	F Construction	2	1147	.	0,095	
9	C industry	H Transportation	2	775	.	0,1071	



1	2		3	4	5	6	7
8	CL12	CL11	4	450	14,1	0,125	
7	CL14	CL10	4	315	9,9	0,1567	
6	CL17	L Real Estate	3	291	,	0,1748	
5	CL9	CL8	6	194	12,5	0,2442	
4	CL13	M Sciences	5	178	,	0,3108	
3	CL5	CL6	9	117	18,4	0,4128	
2	CL7	CL3	13	86,8	20,4	0,5869	
1	CL2	CL4	18	.	86,8	1,4487	

Źródło: opracowanie własne w SAS.

Kategorię C przyjęto jako kategorię odniesienia (najliczniejsza), charakteryzującą się średnim poziomem udziału upadłości. Kategoria A to kategoria o najwyższym poziomie udziału upadłości, natomiast kategoria B, najmniej liczna, charakteryzuje się najniższym poziomem udziału upadłości (por. tab. 4).

**Tabela 4.** Ostateczne grupowanie – wynik na podstawie analizy skupień

		Upadłość		Total
		nie	tak	
Grupa		369	99	468
A	frequency			
	percent	78,85	21,15	
B	frequency	90	2	92
	percent	97,83	2,17	
C	frequency	1957	393	2350
	percent	83,28	16,72	
Total	frequency	2416	494	2910
			16,97	

Źródło: opracowanie własne w SAS.

Ostatnim oszacowanym modelem był model regresji logistycznej ze zmienną nominalną. Siła dyskryminacyjna takiego modelu wzrosła tylko nieznacznie (por. tab. 5). Najwyższa wartość parametru wystąpiła dla zmiennej X3, wzrost o jednostkę dla tego wskaźnika powoduje spadek szansy zajęcia upadłości o prawie 99%. Ze względu jednak na konstrukcję wskaźnika (EBIT/AKTYWA) wzrost o 1 oznaczałby 100-procentowy wzrost dochodów przedsiębiorstwa.

Włączona zmienna PKD okazała się statystycznie istotna, podobnie różnice pomiędzy kategoriami A, B a kategorią odniesienia C również były statystycznie istotne. Ryzyko upadłości w grupie A jest 1,5 raza większe niż w grupie odniesienia.

nia C, natomiast w grupie B o 85% niższe niż w grupie odniesienia C. Najbardziej ryzykowne branże działalności to Budownictwo, Energetyka i Rozrywka. Najmniej ryzykowne natomiast okazały się branże Nauka, Usługi pozostałe, Opieka, Edukacja i Górnictwo.

**Tabela 5.** Wyniki oszacowania modelu regresji logistycznej ze zmiennymi makroekonomicznymi i zmienną nominalną PKD

Parametr	SS	Oszacowanie parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi <sup>2</sup>	Odds ratio
Wyraz wolny	1	1,7041	0,6980	5,9606	0,0146	
X1	1	-0,5506	0,2097	6,8918	0,0087	0,577
X2	1	-1,3357	0,9006	2,1997	0,1380	0,263
X3	1	-4,5680	0,8859	26,5904	<,0001	0,010
X4	1	-0,5658	0,0983	33,1540	<,0001	0,568
X5	1	-0,2113	0,0469	20,2819	<,0001	0,810
PKB	1	-0,1865	0,0345	29,3194	<,0001	0,830
bezrobocie	1	-0,1975	0,0294	45,1357	<,0001	0,821
inflacja	1	0,0924	0,0740	1,5562	0,2122	1,097
Grupa A od C	1	0,9543	0,2589	13,5821	0,0002	1,577
Grupa B od C	1	-1,4533	0,4887	8,8424	0,0029	0,142

Źródło: opracowanie własne w SAS.

Siła dyskryminacyjna takiego modelu jednak nie uległa poprawie. Wartość AUC wynosi 0,8081.

### 3.5. Porównanie siły dyskryminacyjnej modeli

Porównując siłę dyskryminacyjną modeli, wykorzystano miarę syntetyczną ROC, która jest wartością pola pod krzywą ROC, czyli AUC (*Area Under the Curve*).

**Tabela 6.** Porównanie siły dyskryminacyjnej modeli z wykorzystaniem pola pod krzywą ROC

Model	Regresja logistyczna	Model proporcjonalnych hazardów Coxa
Z-Score	0,6995	
X1-X5	0,7523	0,7460
X1-X5+Makro	<b>0,8025</b>	<b>0,7925</b>
X1-X5+Makro+Branża	0,8081	

Źródło: opracowanie własne w SAS.

#### 4. Podsumowanie i wnioski końcowe

Obecnie najczęściej wykorzystywane są modele analizy dyskryminacyjnej i modele regresji logistycznej do oceny ryzyka upadłości przedsiębiorstw z segmentu SME. Wydaje się jednak, iż włączenie modeli przeżycia do tego typu analiz jest uzasadnione ze względu na właściwości takich modeli.

W przeprowadzonych analizach i wynikach można zauważyć kilka powodów, dla których warto stosować alternatywne metody w porównaniu ze statyczną metodą regresji logistycznej.

Przed wszystkim modele przeżycia dostarczają dynamicznego oszacowania ryzyka upadłości, co jest znaczącym rozszerzeniem podejścia statycznego.

Pozostałe zalety stosowania modeli przeżycia są następujące:

- możliwość wykorzystania danych obciętych – zdarzenie zachodzi, kiedy przedsiębiorstwo jest już wyeliminowane z okna obserwacji,
- uniknięcie niestabilności oszacowań wynikających z przyjęcia stałego okna obserwacji,
- śledzenie oszacowania ryzyka w całym czasie trwania procesu,
- otrzymanie „dynamicznej” prognozy prawdopodobieństwa zajścia zdarzenia (funkcja czasu), użytecznej w określaniu strategii i polityki,
- włączenie zmian w otoczeniu makroekonomicznym – zmienne makroekonomiczne zależne od czasu.

Wnioski z zastosowania regresji logistycznej ze zmiennymi nominalnymi są następujące:

- Analiza skupień może być wykorzystana do grupowania kategorii zmiennej nominalnej w celu redukcji zróżnicowania zmiennej i wymiaru analizy.
- Grupowanie kategorii umożliwia włączenie zmiennej nominalnej bez problemu z *quasi*-kompletną separacją w regresji logistycznej.

Osiągnięto cel badania, porównując modele przeżycia Coxa z modelami regresji logistycznej i wskazując na wady i zalety tych dwóch podejść badawczych. Najwyższą siłą dyskryminacyjną otrzymano dla modelu regresji logistycznej i regresji Coxa włączającego zmienne makroekonomiczne. Siła dyskryminacyjna modelu regresji logistycznej ze zmienną nominalną wzrosła, ale wzrost był nieznaczny.

#### Literatura

- Allen L.N., Rose L.C., 2006, *Financial survival analysis of defaulted debtors*, “Journal of Research Society”, 57.
- Allison P., 2012, *Logistic Regression Using SAS: Theory and Application*, Second Edition.
- Altman E., Suggitt H.J., 2000, *Default rates in the syndicated bank loan market: A mortality analysis*, “Journal of Banking and Finance”, vol. 24, No. 1.
- Altman E.I., 1989, *Measuring corporate bond mortality and performance*, “Journal of Finance” 44(4).

- Appenzeller D. (red.), 2004, *Upadłość przedsiębiorstw w Polsce w latach 1990-2003. Teoria i praktyka*, Zeszyty Naukowe, nr 49/2004, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.
- Bauer J., Agarwal V., 2014, *Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test*, "Journal of Banking & Finance", vol. 40, pp. 432-442.
- Blossfeld H.P., G. Rohwer, 2002, *Techniques of Event History Modeling. New Approaches to Causal Analysis*, Lawrence Elbaum Associates Publishers, London.
- Cole R.A., Moshirian F., Wu Q., 2008, *Bank stock returns and economic growth*, "Journal of Banking and Finance" 32.
- Cole R.A., Wu Q., *Hazard versus Probit in Predicting U.S. Bank Failures: A Regulatory Perspective over Two Crises* (July 16, 2014), 22nd Australasian Finance and Banking Conference 2009, Available at SSRN: <http://ssrn.com/abstract=1460526> or <http://dx.doi.org/10.2139/ssrn.1460526>
- Cox D.R., 1972, *Regression models and life tables*, "Journal of the Royal Statistical Society (Series B)", no 34, pp.187-202.
- Dębrowska K., 2012, *Prognozowanie upadłości przedsiębiorstw za pomocą wybranych metod wielowymiarowej analizy statystycznej*, „Zarządzanie i Finanse” Vol. 10, nr 1.
- Figini S., Fantazzini D., *Random Survival Forests Models for SME Credit Risk Measurement (January 31, 2009). Methodology and Computing in Applied Probability, Forthcoming*. Available at SSRN: <http://ssrn.com/abstract=1335856>
- Gajdka J., Stos D., *Wykorzystanie analizy dyskryminacyjnej w ocenie kondycji finansowej przedsiębiorstw*, [w:] R. Borowiecki (red.) (1996), *Restrukturyzacja w procesie przekształceń i rozwoju przedsiębiorstw*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Gruszczyński M., 2005, *Zalety i słabości modeli bankructwa*, [w:] *Zagrożenie upadłościami*, Szkoła Główna Handlowa, Warszawa.
- Hadasik D., 1998, *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, [w:] Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu, AE w Poznaniu, Poznań.
- Hamrol M., Chodakowski J., 2008, *Prognozowanie zagrożenia finansowego przedsiębiorstwa. Wartość predykcyjna polskich modeli analizy dyskryminacyjnej*, „Badania Operacyjne i Decyzje” nr 3.
- Hołda A., 2001, *Prognozowanie bankructwa jednostki w warunkach gospodarki polskiej z wykorzystaniem funkcji dyskryminacyjnej*, „Rachunkowość” nr 5.
- Korol T., 2005, *Wykorzystanie sieci jednokierunkowej wielowarstwowej oraz sieci rekurencyjnej w prognozowaniu upadłości przedsiębiorstw*, [w:] *Zagrożenie upadłościami*, red. K. Kuciński, E. Mączyńska, SGH, Warszawa.
- Kuciński K., Mączyńska E. (red.), 2005, *Zagrożenie upadłościami*, Szkoła Główna Handlowa, Warszawa.
- Jacobson T., Linde J., Roszbach K., *Firm Default and Aggregate Fluctuations* (July 6, 2011), Sveriges Riksbank Working Paper Series No. 226. Available at SSRN: <http://ssrn.com/abstract=1471254> or <http://dx.doi.org/10.2139/ssrn.1471254>.
- Laitinen E.K., *A dynamic performance measurement system: evidence from small Finnish technology companies*, "Scandinavian Journal of Management", 18. 1. 2002, s. 65-99.
- Lando D., 1998, *On cox processes and credit risky securities*, "Review of Derivatives Research" 2.
- Mączyńska E. (red.), 2010, *Meandry upadłości przedsiębiorstw*, Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie, Warszawa.
- Mączyńska E., Zawadzki M., 2006, *Dyskryminacyjne modele predykcji bankructwa przedsiębiorstw*, „Ekonomista” 2.
- Malik M., Thomas L., 2009, *Modelling Credit Risk in Portfolios of Consumer Loans: Transition Matrix Model for Consumer Credit Ratings*, Southampton, UK, University of Southampton, 21pp. (Discussion Papers in Centre for Risk Research).
- McDonald R.A., Matuszyk A., Thomas L.C., 2010, *Application of survival analysis to cash flow modeling for mortgage products*, "OR Insight" 23.
- Narain, B., 1992, *Survival Analysis and the Credit Granting Decision*, [in:] L.C. Thomas, J. Crook N., D.B. Edelman (eds.), *Credit Scoring and Credit Control*, OUP, Oxford, U.K.

- Pierides Y.A., 1997, *The pricing of credit risk derivatives*, "Journal of Economic Dynamics and Control" 21(10).
- Prusak B., 2009, *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Difin, Warszawa.
- Ptak-Chmielewska A., Schab I., 2008, *Wykorzystanie modeli regresji logistycznej i hazardu do określenia determinant zaniechania zobowiązań*, [w:] J. Pocięcha (red.), *Współczesne problemy modelowania i prognozowania zjawisk społeczno-gospodarczych*, Studia i Prace Uniwersytetu Ekonomicznego w Krakowie nr 2, Kraków.
- Ptak-Chmielewska A., Pęczkowski M., 2009, *Analiza dyskryminacji*, [w:] *Wielowymiarowa analiza statystyczna. Teoria – przykłady zastosowań z systemem SAS*, E. Frątczak (red.), Oficyna Wydawnicza SGH, Warszawa.
- Ptak-Chmielewska A., 2012, *Wykorzystanie modeli przeżycia i analizy dyskryminacyjnej do oceny ryzyka upadłości przedsiębiorstw*, „*Ekonometria*” 4(38) 2012, Wrocław.
- Stepanova M., Thomas L.C., 2002, *Survival analysis methods for personal loan data*, "Journal of Operations Research" 50(2).
- Strąk T., 2005, *Wykorzystanie drzew klasyfikacyjnych do oceny zagrożenia bankructwem polskich przedsiębiorstw*, „*Monografie i Opracowania Naukowe*”, Szkoła Główna Handlowa w Warszawie. Kolegium Zarządzania i Finansów. Finanse przedsiębiorstwa.
- Whalen G., 1991, *A proportional hazards model of bank failure: An examination of its usefulness as an early warning model tool*, "Economic Review, Federal Reserve Bank of Cleveland".

## PREDICTION MODELS OF SME BANKRUPTCY IN POLAND – ANALYSIS USING COX SURVIVAL MODEL AND LOGISTIC REGRESSION MODEL

**Summary:** Credit risk is associated with the banking activity and is the most important type of the risk to which banks are exposed. Bankruptcy risk assessment is based on models using the discriminant analysis and logistic regression. However, the requirements for the models are changing and methods used in the banking sector are insufficient. Newer and more sophisticated techniques are developed. One of these methods is survival analysis the popularity of which has increased in the recent years. The aim of this study was to compare new technique (survival analysis) used in the credit models with the traditional ones, like logistic regression. The paper analyses the strengths and weaknesses of both methods and the possibility to use time-dependent variables in survival models (macro-variables) and nominal variables in logistic regression model. The inclusion of macro-variables increases the prediction power of such models. The number of models was built on the basis of the Polish SMEs (turnover above 2 million Euro) data set consisting of 1,561 enterprises including 807 defaults (FS 2002-2010).

**Keywords:** survival analysis, macrovariables, Cox model, bankruptcy risk.