

Aneta Ptak-Chmielewska

Szkoła Główna Handlowa

WYKORZYSTANIE MODELI PRZEŻYCIA I ANALIZY DYSKRYMINACYJNEJ DO OCENY RYZYKA UPADŁOŚCI PRZEDSIĘBIORSTW

Streszczenie: Artykuł przedstawia wykorzystanie regresji logistycznej, analizy dyskryminacyjnej oraz modelu Coxa do oceny zajścia zdarzenia upadłości przedsiębiorstwa. Obecnie standardem w praktyce jest wykorzystywanie analizy dyskryminacyjnej lub regresji logistycznej do modelowania ryzyka upadłości. Podejście takie ma charakter statyczny i nie uwzględnia czynnika czasu. Alternatywną propozycją jest wykorzystanie modeli hazardu. Modele te mają charakter dynamiczny, ponieważ uwzględniają wpływ czasu, po jakim nastąpiło zgłoszenie upadłości. Głównym przedmiotem zainteresowania oraz celem artykułu jest identyfikacja wpływu zmiennych współwystępujących na ryzyko upadłości oraz ewentualnych różnic w wynikach w zależności od przyjętego ujęcia: statycznego oraz dynamicznego. Przedstawione są wyniki badania empirycznego przeprowadzonego dla próby 1536 przedsiębiorstw (w tym 456 upadłości) na podstawie wybranych zmiennych finansowych.

Słowa kluczowe: analiza przeżycia, modele upadłości przedsiębiorstw, analiza dyskryminacyjna, regresja logistyczna.

1. Wstęp

W dostępnej literaturze istnieje wiele definicji systemów i modeli wczesnego ostrzegania przed upadłością przedsiębiorstw. Modele wczesnego ostrzegania przed upadłością są podstawowym narzędziem analitycznym dostarczającym jedynie informacji na temat istniejących zagrożeń, nie wskazują jednakże na sposoby rozwiązania zidentyfikowanych problemów. Systemy te powinny być wspierane przez bardziej zaawansowane metody monitorowania sytuacji finansowej przedsiębiorstwa. Modele wykorzystywane w analizie przedsiębiorstw dotyczą głównie skali mikro i bazują na wybranych wskaźnikach ekonomiczno-finansowych.

Pierwsze modele systemów wczesnego ostrzegania powstały na początku XX wieku. Za przełomowe uważa się prace E. Altmana z lat 60. XX wieku. Opracował on modele dostarczające informacji o zagrożeniach w funkcjonowaniu przedsiębiorstwa. Autor jest prekursorem przejścia od modeli jednowymiarowych do wielowymiarowych metod statystycznych analizy dyskryminacji. W latach 70. i 80. dynamicznie rozwijały się prace związane z wykorzystaniem w predykcji zagrożeń

finansowych modeli logitowych. W latach 90. w tym celu zaczęto używać metod sztucznej inteligencji oraz sieci neuronowych [Mączyńska 2009, s. 70-71]. W Polsce tematyką systemów wczesnego ostrzegania zaczęto zajmować się szerzej dopiero w latach 90. XX wieku, w następstwie transformacji gospodarczej. Mimo że od tego czasu poczyniono wiele w kierunku rozwoju systemów w Polsce, to nadal w praktyce zarządzania w przedsiębiorstwach nie są one popularnym narzędziem. Największy postęp w tej dziedzinie wykazują banki, którym tego typu systemy dostarczają niezwykle cennych informacji.

Zastosowanie modeli zagranicznych w warunkach polskich w świetle przeprowadzonych badań międzynarodowych nie jest możliwe bezpośrednio. Ich wykorzystanie jest możliwe tylko pod warunkiem, że zostaną one dostosowane do specyficznych warunków polskiej gospodarki. Oprócz tego zastosowanie modeli spośród modeli dotychczas skonstruowanych wymaga zweryfikowania poziomu trafności prognozowania tych modeli, zestawu wskaźników użytych w modelu w odniesieniu do konkretnej branży oraz prowadzonej działalności. Konieczne jest także dostosowanie w zakresie kryteriów uznania przedsiębiorstwa za zagrożone bankructwem bądź upadłe. Systemy wczesnego ostrzegania mają zarówno wady, jak i zalety, ale nie ma aktualnie szybszej metody dostarczającej informacji o standingu firmy oraz łatwiejszej w interpretacji. Obecnie istnieje wiele metod budowy systemów oraz wiele modeli.

2. Analiza historii zdarzeń **– semiparametryczny model regresji Coxa**

Analiza historii zdarzeń (analiza przeżycia) jest określona jako zbiór technik statystycznych służących opisowi i badaniu przebiegu życia jednostki, tj. częstości pewnych zdarzeń, ich sekwencji, rozkładu, czasu pobytu jednostki w różnych stanach. Przedmiotem analizy jest proces stochastyczny, w którym występują zdarzenia, określane również jako stany. Proces opisywany jest zazwyczaj w postaci czasów trwania, określających okres, jaki upływa od rozpoczęcia jakiegoś zjawiska, aż do momentu jego zakończenia, bądź do momentu dokonania pomiaru, jeśli poprzedza on moment zakończenia zjawiska. Przedmiotem badania jest czas upływający między zdarzeniami, określane mianem czasu trwania zjawiska bądź epizodu.

Ze względu na liczbę zdarzeń, które mogą zajść, wyróżnia się analizę pojedynczego epizodu oraz wielokrotnych epizodów, przy czym analiza pojedynczego epizodu jest najbardziej podstawowym modelem analizy historii zdarzeń. Porównując analizę historii zdarzeń z tradycyjnymi badaniami przekrojowymi bądź panelowymi, należy stwierdzić, że zaletą tej pierwszej jest dostarczenie informacji o dynamice badanego procesu (zjawiska).

Proces stochastyczny będący przedmiotem analizy rozpatrywany jest w trzech podstawowych płaszczyznach:

- czasów oczekiwania na wystąpienie wyróżnionych stanów (zdarzeń),
- intensywności przejść między wyróżnionymi stanami,
- liczby i kolejności zdarzeń.

Podstawową strukturę analityczną w analizie historii zdarzeń stanowią: przestrzeń stanów i oś czasu. Przestrzeń stanów jest dyskretna, pomiar czasu zaś może być ciągły bądź dyskretny. Sama oś czasu może zostać zdefiniowana na dwa sposoby: jako czas kalendarzowy bądź jako czas względny (wiek jednostki mierzony względem zdarzenia początkowego mającego miejsce w momencie t_0). Stan wejścia jest wspólny dla wszystkich jednostek badanej zbiorowości, która jest definiowana właśnie przez wspólne doświadczenie przez wszystkie jednostki w momencie t_0 pewnego zdarzenia (stanowi więc kohortę). Wystąpienie zdarzenia upadłości eliminuje przedsiębiorstwo z kohorty przedsiębiorstw działających. Stanowi więc ono zdarzenie kończące (stan wyjścia) dla modelu pojedynczego epizodu.

Analiza historii zdarzeń obejmuje wiele metod, których klasyfikacji można dokonać ze względu na różne kryteria. Ze względu na zastosowane metody estymacji można wyróżnić metody parametryczne i nieparametryczne. Rozróżnienie wynika z założeń przyjętych co do postaci funkcyjnej rozkładu czasu między zdarzeniami T . W przypadku braku takich założeń mamy do czynienia z analizą nieparametryczną, klasycznym przykładem są tu tablice trwania życia. Z kolei w podejściu parametrycznym zakłada się, że czas między zdarzeniami jest zmienną losową należącą do konkretnej rodziny rozkładów. W analizie parametrycznej wykorzystuje się metody regresji, co pozwala na oszacowanie wpływu zestawu dodatkowych zmiennych objaśniających na przejście między stanami, jak również uwzględnienie zagadnień związanych z tzw. heterogenicznością badanej populacji.

Połączeniem obu podejść są tzw. modele semiparametryczne. Najbardziej popularnym z tej grupy jest model proporcjonalnych hazardów Coxa, opisany w dalszej części artykułu. W modelach semiparametrycznych mamy do czynienia częściowo z podejściem parametrycznym, ponieważ wyspecyfikowana jest funkcja regresji, co daje możliwość badania interakcji między różnymi procesami. Jednocześnie w modelach tych nie zakłada się konkretnej postaci rozkładu czasu zdarzeń, co jest podejściem nieparametrycznym.

Podstawowym narzędziem do przeprowadzenia analizy historii zdarzeń jest model przeżycia, opisujący rozkład zmiennej losowej, jaką jest czas do wystąpienia określonego zdarzenia w pewnej zbiorowości. Przyjmijmy, że ciągła zmienna losowa T jest czasem, jaki upływa do momentu zajścia zdarzenia dla jednostki należącej do badanej kohorty w momencie t_0 . Zmienna T wyraża więc przyszłe trwanie życia jednostki mierzone w momencie t_0 . Rozkład zmiennej T można opisać na kilka sposobów, wykorzystując poza standardowo używanymi do opisu rozkładu zmiennych losowych funkcjami gęstości i dystrybuanty także funkcję przeżycia oraz hazardu.

- Funkcja przeżycia

$$S(t) = P(T > t),$$

gdzie $S(t)$ oznacza bezwarunkowe prawdopodobieństwo tego, że zdarzenie zajdzie po momencie t , a więc że kredytobiorca „dożyje” co najmniej do momentu t . Funkcja ta wyraża wzorzec przeżycia w badanej zbiorowości.

- Funkcja hazardu

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t},$$

gdzie $h(t)$ jest warunkową gęstością czasu do wystąpienia badanego zdarzenia (pod warunkiem, że zdarzenie nie wystąpiło do momentu t). Konsekwentnie iloczyn $h(t)\Delta t$ oznacza (przybliżone) prawdopodobieństwo tego, że zdarzenie wystąpi w krótkim przedziale czasu $(t, t + \Delta t)$, pod warunkiem że jednostka przeżyła co najmniej do czasu t .

Modelem najczęściej wykorzystywanym w praktyce jest model proporcjonalnych hazardów Coxa. Dla modelu Coxa funkcja hazardu wyraża się wzorem:

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

gdzie: $h_0(t)$ oznacza tzw. hazard bazowy; jest to niewyspecyfikowana parametrycznie funkcja czasu, natomiast X_1, X_2, \dots, X_k oznaczają zestaw zmiennych współtwarzających (mogą to być zmienne zależne od czasu).

Poza samą postacią modelu Cox zaproponował również specjalną metodę jego estymacji, tzw. metodę pseudowiarygodności. Metoda ta wykorzystuje możliwość rozdzielenia funkcji wiarygodności dla modelu proporcjonalnych hazardów na dwa składniki: pierwszy, zawierający informacje tylko o współczynnikach α_i , oraz drugi, zawierający informacje zarówno o współczynnikach α_i , jak i o funkcji hazardu. Zaletą modelu Coxa (jak również innych modeli semiparametrycznych) jest możliwość pomiaru wpływu wielu zmiennych, zarówno niezależnych, jak i zależnych od czasu, na badane zjawisko, bez konieczności specyfikacji bazowej funkcji hazardu $h_0(t)$. Pamiętać należy o poważnym ograniczeniu zastosowań takiej wersji modelu Coxa, jakim jest założenie proporcjonalnych hazardów. Oznacza ono, że dla każdej pary jednostek badanej populacji, w każdym momencie t stosunek wartości funkcji hazardu jest stały. W konsekwencji względny ranking jednostek populacji ze względu na ryzyko zajścia zdarzenia jest stały w czasie. Problem ten może być jednak rozwiązany przez włączenie do modelu dodatkowych zmiennych zależnych od czasu (np. będących iloczynem zmiennych współtwarzających oraz czasu x i t). Model w takim wariacie określany jest jako model nieproporcjonalnych hazardów Coxa. Model Coxa jest szczególnie atrakcyjny dla badacza w sytuacji [Blossfeld, Rohwer 2002]:

- braku znajomości kształtu zależności hazardu względem czasu,
- braku teoretycznych podstaw do specyfikacji modelu parametrycznego,
- mimo znajomości zależności hazardu względem czasu braku możliwości wskazania funkcji opisującej tę zależność,

- zainteresowania głównie wpływem zmiennych współwystępujących na czas trwania (a tym samym hazard).

Powyższe wskazania do praktycznego stosowania modelu Coxa pozwalają uznać, że model ten może być stosowany do oceny ryzyka upadłości przedsiębiorstw. Wynikiem estymacji modelu Coxa są współczynniki wskazujące wpływ zmiennych współtowarzyszących na prawdopodobieństwo doświadczenia zdarzenia oraz prawdopodobieństwo (hazard) bazowe, które jest takie samo dla wszystkich jednostek populacji i zależy jedynie od czasu.

3. Analiza dyskryminacyjna

Analiza dyskryminacyjna służy do określenia, które z analizowanych zmiennych mają zdolności do wyłonienia dwu lub więcej grup z analizowanego zbioru obserwacji. Pozwala ona na zidentyfikowanie tych zmiennych, które pozwalają zaklasyfikować obserwacje do różnych grup z większą trafnością niż przypadkowa. Zadaniem analizy dyskryminacyjnej jest prawidłowa klasyfikacja obserwacji do określanych jako grupy dwóch podprzestrzeni. Funkcją dyskryminującą jest maksymalizacja miary odległości pomiędzy populacjami [Staniec 2000, s. 90].

W metodach analizy dyskryminacyjnej klasyfikacja jednostek jako zagrożone bądź niezagrożone odbywa się na podstawie minimum dwóch zmiennych objaśniających. Następuje równoczesna analiza podmiotu pod kątem wszystkich wybranych wskaźników. Kluczowe jest tutaj odnalezienie zależności między zmiennymi umożliwiającą prawidłowe rozróżnienie podmiotów. W modelach analizy dyskryminacyjnej zmienna objaśniana jest zmienną jakościową. Klasyfikacja jednostek gospodarczych następuje na podstawie liniowej funkcji dyskryminacyjnej. Powstały w efekcie zastosowania modelu wskaźnik syntetyczny (wartość funkcji) umożliwia jednoznaczny klasyfikację podmiotu [Prusak 2009, s. 47-48]. Metody analizy dyskryminacyjnej mają pewne ograniczenia. Ich zastosowanie możliwe jest w przypadku, gdy analizowane wskaźniki mają rozkłady normalne. Konieczne jest także spełnienie założenia o ich niezależności i kompletności. Brak spełnienia założeń wpływa negatywnie na zdolności klasyfikacyjne modelu. Sprawdzenie spełnienia tych założeń przez dobrane wskaźniki może być zweryfikowane przez przeprowadzenie odpowiednich testów oraz statystycznych procedur.

Funkcje dyskryminacyjne, na podstawie których budowane są modele wielowymiarowe systemów ostrzegania przed upadłością, mogą przyjmować różną postać – mogą być funkcjami liniowymi, kwadratowymi itp. Liniowe funkcje dyskryminacyjne przyjmują zazwyczaj postać:

$$Z = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n$$

gdzie: Z – zmienna objaśniana (zależna),
 a_0 – stała,

$a_i, i = 1, 2, \dots, n$ – współczynniki (wagi) dyskryminacyjne,
 X_1, X_2, \dots, X_n – zmienne objaśniające (wskaźniki finansowe).

Przedstawiona postać funkcji dyskryminacyjnej nazywana jest funkcją dyskryminacyjną Fishera. Parametry a_i są zwane współczynnikami dyskryminacyjnymi (wagami). Po wyznaczeniu postaci funkcji dyskryminacyjnej następuje określenie wartości granicznej, pozwalającej na jednoznaczne zaklasyfikowanie jednostki jako zagrożonej bądź niezagrażonej finansowo. Najczęściej wyznaczana jest wartość średnia funkcji dyskryminacyjnej w poszczególnych grupach oraz wartość odcięcia w połowie między średnimi. Jeżeli wartość Z dla danego przedsiębiorstwa jest mniejsza niż $Z_{\text{odcięcia}}$, to wówczas jednostka ta klasyfikowana jest jako zagrożona upadłością, jeśli zaś większa, to jako jednostka zdrowa. Prawidłowość dokonanych klasyfikacji (efektywność modelu) oceniana jest za pomocą kategorii: błąd I rodzaju oraz błąd II rodzaju. Pierwsza z kategorii określa procent przedsiębiorstw zagrożonych zaklasyfikowanych jako przedsiębiorstwa zdrowe, natomiast druga kategoria przedstawia, ile procent zdrowych przedsiębiorstw zostało zaklasyfikowanych jako zagrożone.

Za prekursora modelu wielowymiarowego uważa się prof. Altmana, który przedstawił swój model w 1968 r. Model ten stanowi połączenie analizy wskaźnikowej oraz metody statystycznej – wielowymiarowej analizy dyskryminacyjnej. Autor przeprowadził analizę 22 wskaźników na grupie 66 przedsiębiorstw (33 upadłych i 33 nieupadłych). W kolejnych etapach analizy odrzucano wskaźniki o najniższych wartościach predykcyjnych, w efekcie czego do ostatecznego modelu weszło 5 wskaźników. W 1977 r. prof. Altman wraz z zespołem przeprowadził kolejne badania związane z predykcją upadłości przedsiębiorstw. Przeanalizowano 58 firm upadłych oraz 58 firm w dobrej kondycji finansowej. W efekcie powstał model składający się z 7 zmiennych, dla których tym razem nie określono wag, a co za tym idzie – nie wyznaczono postaci funkcji dyskryminacyjnej. Otrzymany model ZETA cechował się dużą zdolnością predykcyjną w ciągu 5 lat przed upadkiem jednostki – na rok przed upadkiem wynosił on 90%, a 5 lat przed upadkiem 70%. Kolejna wersja modelu Altmana opracowana została w 1983 r. E.I. Altman dokonał zmian wartości wag przypisanych do zmiennych z pierwszego modelu. Dla funkcji dyskryminacyjnej wartość błędu prognozy wynosiła 6%. Kolejne udoskonalenie modelu Altmana dotyczyło obniżenia wpływu koniunktury oraz specyfiki branży na wartość wskaźnika Z . Modele Altmana opracowane zostały dla przedsiębiorstw działających w realiach rynkowych amerykańskich. Jego zastosowanie dla przedsiębiorstw działających w innych uwarunkowaniach nie daje zadowalających efektów [Zaleska 2002, s. 28-33].

4. Analiza logitowa

Modele logitowe są bardzo popularną obecnie metodą wykorzystywaną do prognozowania bankructwa przedsiębiorstw. Mogą one występować jako modele dwumia-

nowe, gdzie jedna ze zmiennych jest dychotomiczna, lub jako modele wielomianowe uporządkowane, gdzie zmienna objaśniana może występować w więcej niż dwóch stanach. Do szacowania funkcji logitowych wykorzystywane są metoda ważona najmniejszych kwadratów lub metoda największej wiarygodności.

Funkcja logitowa w modelach dwumianowych przyjmuje postać:

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

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

β_0 – stała,

$\beta_i, i = 1, 2, \dots, k$ – wagi,

$x_i, i = 1, 2, \dots, k$ – zmienne niezależne – wskaźniki finansowe.

Wskaźnik $P(Y = 1)$ przyjmuje wartości z przedziału $\langle 0; 1 \rangle$, gdzie 0 oznacza przedsiębiorstwo zdrowe, 1 zaś zagrożone. Im wartość wskaźnika jest bliższa 0, tym mniejsze prawdopodobieństwo zaistnienia zagrożenia w stosunku do danego przedsiębiorstwa.

Istotną kwestią przy szacowaniu modelu dwumianowego jest właściwe wyznaczenie punktu granicznego. W przypadku modeli szacowanych na podstawie próby zbilansowanej wartość tego punktu równa jest zazwyczaj 0,5. Jednak nie zawsze się to sprawdza. Wpływ na wartość tego punktu ma struktura grupy (udział zagrożonych i zdrowych przedsiębiorstw). W interpretacji wyników otrzymanych za pomocą analizy logitowej dużą rolę odgrywa tzw. iloraz szans (*odds ratio*). Wskaźnik ten obliczany jest jako stosunek prawdopodobieństwa wystąpienia zdarzenia do prawdopodobieństwa jego niewystąpienia. Wielkość tego wskaźnika mówi, o jakie krotności ($\exp(c_i)$) zwiększają się średnio ilorazy szans, jeżeli zmienna zmienia się o jednostkę. Jeżeli $\exp(c_i) > 1$, wówczas można spodziewać się wzrostu ilorazu szans, jeżeli natomiast $\exp(c_i) < 1$, nastąpi sytuacja odwrotna. Należy pamiętać, że interpretacje obowiązują jedynie przy wybranym zestawie zmiennych niezależnych oraz założeniu *ceteris paribus* [Gruszczyński 2010, s. 67-68].

Model logitowy wymaga spełnienia wielu założeń, najistotniejszymi z nich są losowy charakter próby (o dużej liczebności), brak współliniowości zmiennych oraz niezależność obserwacji.

5. Wyniki empiryczne

Dokonano estymacji trzech opisanych modeli: regresji logistycznej, analizy dyskryminacyjnej i modelu Coxa. Jako datę upadłości przyjęto datę postanowienia sądu o upadłości układowej bądź likwidacyjnej. Złączono sprawozdania z różnych źródeł, w tym głównie z danych zakupionych od wywiadowni gospodarczej oraz da-

nych wewnętrznych o klientach jednego z banków w Polsce. Usunięto obserwacje z ujemnymi kapitałami własnymi oraz z przychodami poniżej 8 mln zł. Wybrano sprawozdania z przynajmniej 3 miesięcy przed datą upadłości. Usunięto sprawozdania dla deweloperów oraz sprawozdania z brakami danych. W rezultacie próba obejmowała 4297 sprawozdań za lata 2002-2011 dla 1536 podmiotów ze wszystkich branż działalności, w tym 456 zgłoszone upadłości. Obliczenia przeprowadzono w programie SAS 9.2 oraz R.

Zdefiniowano zestaw wskaźników zgodnie z publikacją M. Zaleskiej [2012] (por. tab. 1).

Tabela 1. Wskaźniki finansowe wykorzystane w modelach

Wskaźnik (etykieta)	Definicja
1	2
WSKAŹNIKI PŁYNNOŚCI	
PLYNNOŚĆ BIEŻĄCA [PL_PB]	aktywa obrotowe/zobowiązania krótkoterminowe
PLYNNOŚĆ SZYBKA [PL_PS]	aktywa obrotowe – zapasy – krótkoterminowe rozliczenia międzyokresowe/zobowiązania krótkoterminowe
PLYNNOŚĆ GOTÓWKOWA [PL_PG]	Inwestycje krótkoterminowe/zobowiązania krótkoterminowe
WSKAŹNIKI AKTYWNOŚCI	
CYKL ROZLICZENIOWY KAPITAŁU OBROTOWEGO [AK_CRKO]	kapitał obrotowy netto/przychody netto ze sprzedaży \times 360, kapitał obrotowy netto = majątek obrotowy – zobowiązania bieżące
CYKL ROZLICZENIOWY NALEŻNOŚCI [AK_CRM]	średni stan należności handlowych/przychody netto ze sprzedaży \times 365 dni
UDZIAŁ NALEŻNOŚCI W AKTYWACH OGÓLEM [AK_UNAO]	należności długoterminowe + należności krótkoterminowe/aktywa razem
CYKL ROZLICZENIOWY ZAPASÓW OGÓLEM [AK_CRZA]	średni stan zapasów/przychody netto ze sprzedaży \times 365 dni
CYKL ROZLICZENIOWY ZOBOWIĄZAŃ [AK_CRZK]	średni stan zobowiązań handlowych/przychody netto ze sprzedaży \times 365 dni
CYKL ROZLICZENIOWY ŚRODKÓW PIENIĘŻNYCH [AK_CRSP]	średni stan gotówki w kasie/przychody netto ze sprzedaży \times 365 dni
WSKAŹNIKI ZADŁUŻENIA	
WSKAŹNIK OGÓLNEGO ZADŁUŻENIA [SB_WOZ]	zobowiązania ogółem/pasywa razem
WSKAŹNIK POKRYCIA AKTYWÓW KAPITAŁEM/FUNDUSZEM WŁASNYM [SB_AKW]	kapitał własny/aktywa razem
WSKAŹNIK DŁGOTERMINOWEGO ZADŁUŻENIA [SB_WDZ]	zobowiązania długoterminowe /kapitał własny

Tabela 1, cd.

1	2
WSKAŹNIK OGÓLNEGO ZADŁUŻENIA KAPITAŁU/FUNDUSZU WŁASNEGO [SB_WOZKW]	zobowiązania ogółem/kapitał własny
WSKAŹNIK POKRYCIA OBSŁUGI DŁUGU Z WYNIKU FINANSOWEGO [OD_PODWF]	zysk netto/odsetki + suma spłat rat kapitałowych
WSKAŹNIK POKRYCIA OBSŁUGI DŁUGU Z EBIT [OD_PODE]	zysk netto + podatek dochodowy/odsetki + suma spłat rat kapitałowych
WSKAŹNIK POKRYCIA OBSŁUGI DŁUGU Z NADWYŻKI FINANSOWEJ NETTO [OD_PODNFN]	zysk netto + amortyzacja/odsetki + suma spłat rat kapitałowych
WSKAŹNIKI EFEKTYWNOŚCI	
WSKAŹNIK POZIOMU KOSZTÓW [EF_PK]	przychody netto ze sprzedaży – zysk netto – podatek dochodowy/przychody netto ze sprzedaży
WSKAŹNIK RENTOWNOŚCI SPRZEDAŻY [EF_ROS]	zysk netto/przychody netto ze sprzedaży
WSKAŹNIK RENTOWNOŚCI AKTYWÓW [EF_ROA]	zysk netto/aktywa razem
WSKAŹNIK RENTOWNOŚCI KAPITAŁÓW/FUNDUSZY WŁASNYCH [EF_ROE]	zysk netto/kapitał własny
Dodatkowo zdefiniowano zmienną opartą na przychodach w postaci: LOG_OBROTY	log(przychody ogółem)

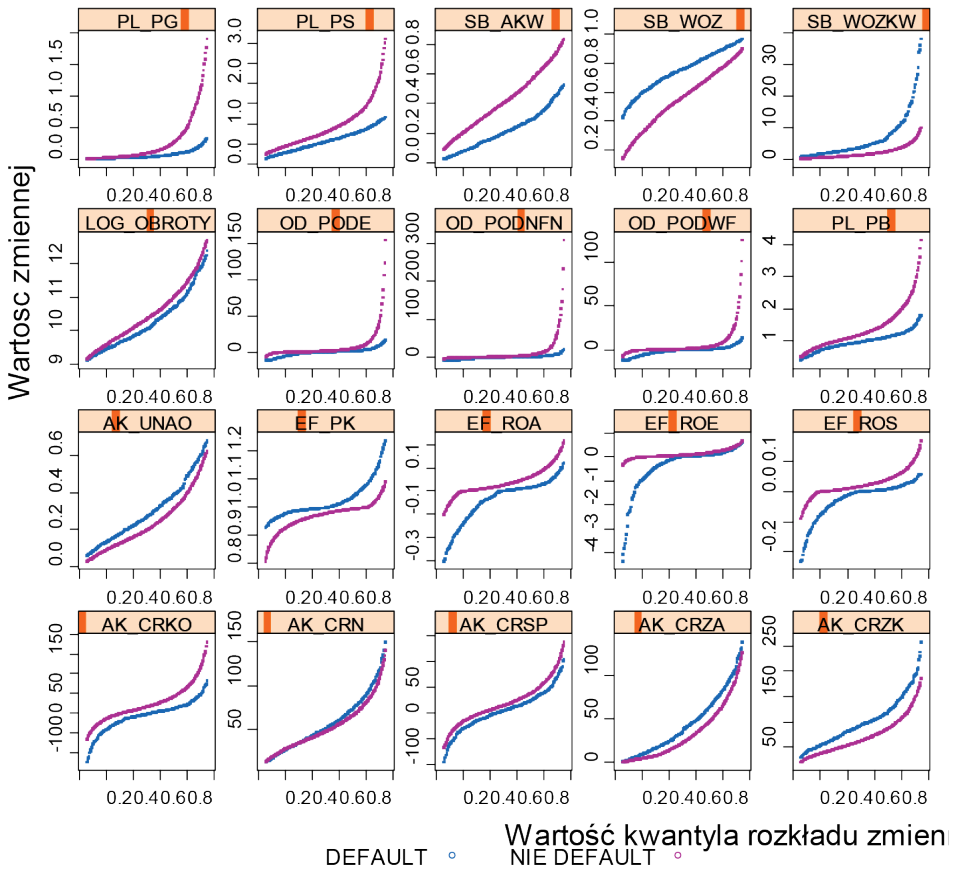
Źródło: opracowanie własne.

Dla wszystkich zmiennych, poza (*OD_PODWF*, *OD_PODE*, *OD_PODNFN*), zastosowano zamianę wartości skrajnych na wartości 1 i 99 percentyla w całej próbie. Dla zmiennych (*OD_PODWF*, *OD_PODE*, *OD_PODNFN*) zastosowano zamianę na wartości 5 i 95 percentyla w całej populacji ze względu na duży rozstęp wartości zmiennej. Dystrybuanty wartości zmiennych w próbie ze względu na wystąpienie zdarzenia upadłości w następnym roku przedstawia rys. 1.

W celu ustabilizowania wyników opartych na symulacjach zdefiniowano 100 próbek obserwacji określonych jako 50% całej próby obserwacji wylosowanych bez zwracania.

Zbadano korelacje pomiędzy wskaźnikami. Usunięto zmienne silnie skorelowane ze sobą, stosując kryterium $r > 0,7$ jako punkt odcięcia silnej korelacji. Mediana wartości AUC¹ dla wszystkich zmiennych w przypadku 100 symulacji wyniosła

¹ AUC (*Area Under the Curve*) – pole pod krzywą ROC – jest miarą syntetyczną mocy dyskryminacyjnej, czyli zdolności do rozróżniania podmiotów „złych” od „dobrych”. Wartość AUC = 1 oznacza idealny model, AUC=1/2 oznacza model losowy. W wymiarze prawdopodobieństw można AUC in-



Rys. 1. Dystrybuanty wartości zmiennych w próbie

Źródło: opracowanie własne.

od 0,24 do 0,74. Usunięto zmienne z niską wartością AUC. W celu oszacowania parametrów modeli wybrano obserwacje z lat 2001-2008, dla których zdarzenia upadłości występowały w latach 2002-2009. W celu oszacowania siły prognostycznej modelu (walidacja) wybrano obserwacje z lat 2009-2011, dla których zdarzenia upadłości występowały w latach 2010-2012.

Ostateczny model zawierał 9 wskaźników, dla których zbudowano trzy typy funkcji: logistyczną, dyskryminacyjną oraz hazardu (Coxa).

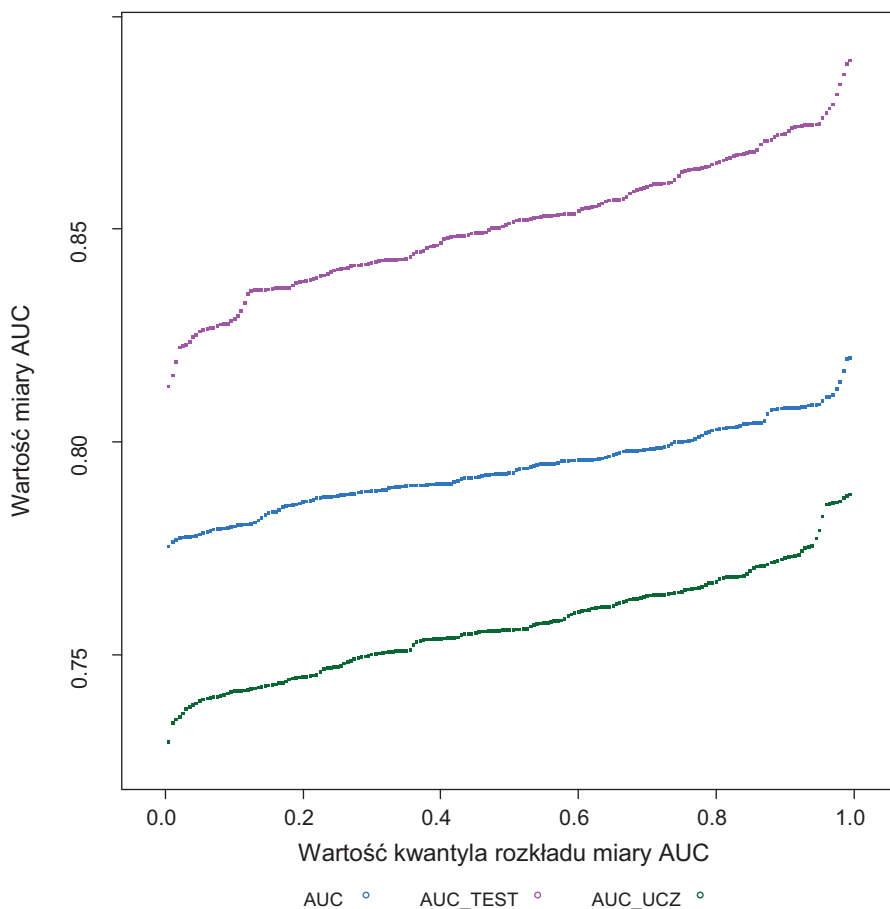
terpretować jako prawdopodobieństwo, że losowo wybrany „zły” podmiot będzie miał gorszą wartość wskaźnika (oceny punktowej w modelu) niż losowo wybrany „dobry” podmiot.

MODEL GLM(LOGIT):

$$\text{Logit}(Y) = -0,459 - 0,078 * \text{LOG_OBROTY} - 1,105 * \text{PL_PG} + 0,249 * \text{AK_UNAO} \\ + 0,001 * \text{AK_CRZA} + 0,0002 * \text{AK_CRZK} - 2,745 * \text{SB_AKW} + 0,020 * \\ * \text{SB_WOZKW} - 0,002 * \text{OD_PODWF} - 5,716 * \text{EF_ROS}.$$

Wszystkie znaki zmiennych są zgodne z interpretacją ekonomiczną. Zmienne: *LOG_OBROTY*, *AK_UNAO*, *AK_CRZA*, *AK_CRZK* oraz *OD_PODWF* okazały się statystycznie istotne na poziomie istotności 0,1.

Dystrybuantę rozkładu miary AUC dla symulacji próbek uczących, całych i testowych dla modelu oszacowanego na całej próbie uczącej przedstawia rys. 2.



Rys. 2. Dystrybuanta rozkładu miary AUC dla modelu regresji logistycznej

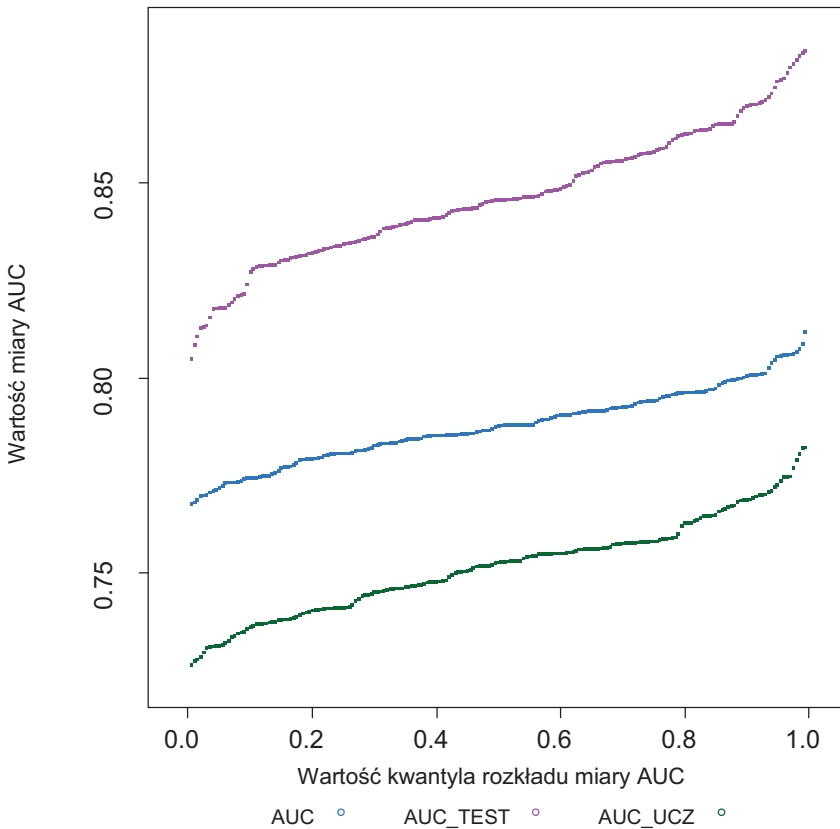
Źródło: opracowanie własne za pomocą SAS 9.2 i R.

Wzrost wartości AUC dla modelu testowego wynika z oparcia próby uczącej na upadłościach z okresu początku kryzysu i problemu opcji walutowych – w takich warunkach model uczący musiał wykazać się niższą sprawnością. Populacja testowa zawiera upadłości z kolejnych lat kryzysu, gdy przedsiębiorstwa upadają w sposób bardziej przewidywalny.

MODEL LDA (FUNKCJA DYSKRYMINACYJNA):

$$Y(1) = -0,091 * LOG_OBROTY + 0,021 * PL_PG + 0,142 * AK_UNAO + 0,001 * AK_CRZA + 0,002 * AK_CRZK - 1,582 * SB_AKW + 0,075 * SB_WOZKW + 0,00001 * OD_PODWF - 5,214 * EF_ROS.$$

Dystrybuantę rozkładu miary AUC dla symulacji próbek uczących, całych i testowych dla modelu oszacowanego dla całej próby uczącej przedstawia rys. 3.



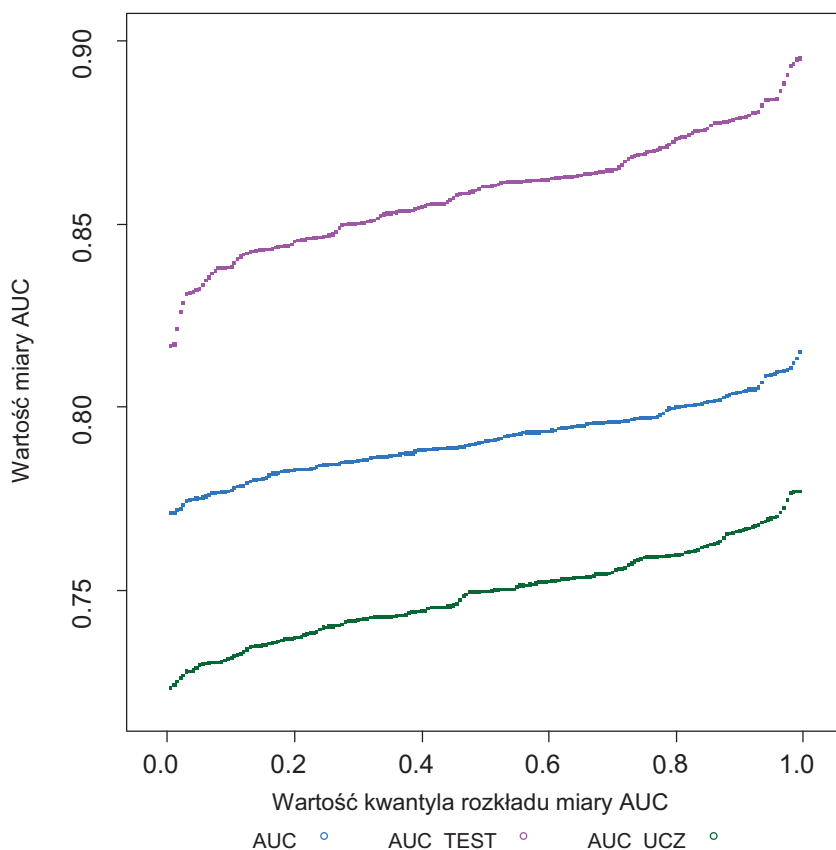
Rys. 3. Dystrybuanta rozkładu miary AUC dla modelu analizy dyskryminacyjnej

Źródło: opracowanie własne przy pomocy SAS 9.2 i R.

Model regresji Coxa:

Zastosowanie modelu Coxa wymagało dodania zmiennych zawierających wartości czasu ze względu na możliwość obserwacji danego podmiotu w kilku okresach. W celu uniknięcia możliwości ucięcia prognozy dla próby uczącej i testowej zastosowano odrębne daty startu tak, by pierwsza obserwacja dla danego klienta w każdej z grup zawsze zaczynała się od wartości zmiennej 'start' równej 0.

$$h(t) = h_0(t) \exp \left(\begin{array}{l} -0,202 * LOG_OBROTY - 0,827 * PL_PG + 0,473 * AK_UNAO \\ -0,0001 * AK_CRZA + 0,0005 * AK_CRZK - 2,716 * SB_AKW \\ + 0,0203 * SB_WOZKW + 0,00007 * OD_PODWF - 3,783 * EF_ROS \end{array} \right)$$



Rys. 4. Dystrybuanta rozkładu miary AUC dla modelu regresji Coxa

Źródło: opracowanie własne przy pomocy SAS 9.2 i R.

Wszystkie znaki zmiennych są zgodne z interpretacją ekonomiczną. Zmienne: AK_UNAO, AK_CRZA, AK_CRZK oraz OD_PODWF okazały się statystycznie

nieistotne na poziomie istotności 0,1. Dystrybuantę rozkładu miary AUC dla symulacji próbek uczących, całych i testowych dla modelu oszacowanego na całej próbie uczącej przedstawia rys. 4.

PORÓWNANIE AUC DLA MODELI LOGIT, LDA oraz COX:

Dla całej próby testowej uzyskano poniższe miary AUC dla modeli oszacowanych na całej próbie uczącej (por. tab. 2).

Tabela 2. Porównanie miary AUC dla trzech estymowanych modeli

	Próba ogółem	Próba ucząca	Próba testowa
LOGIT	0,7933552	0,7568146	0,8508912
LDA	0,790412	0,7484089	0,8454304
COXPH	0,780333	0,7370299	0,8507741

Źródło: opracowanie własne za pomocą SAS 9.2.

6. Podsumowanie i wnioski

Zastosowanie modeli klasyfikacyjnych typu LOGIT, LDA, COX nie prowadzi do zdecydowanego zróżnicowania siły prognostycznej. Występujące różnice są potencjalnie wynikiem odmiennych sposobów wyrażenia prognozy prawdopodobieństwa klasyfikacji w postaci przekształcenia logistycznego, prawdopodobieństwa posteriori lub ryzyka względnego. W porównaniu z modelem regresji logistycznej w modelu regresji Coxa dodatkowo zmienna wyrażająca wartość obrotów okazała się statystycznie istotna.

Z przedstawionego wcześniej opisu metod analizy historii zdarzeń wynika kilka powodów, dla których warto je zastosować jako alternatywę wobec tradycyjnie używanych modeli statycznych (np. modeli regresji logistycznej). Przede wszystkim należy podkreślić, że wykorzystanie metod analizy historii zdarzeń w modelowaniu ryzyka upadłości przedsiębiorstw pozwala rozszerzyć standardowe podejście statyczne o dynamikę zjawiska. Pozostałe zalety takiego podejścia są następujące:

- możliwość uwzględnienia danych cenzurowanych, jeśli przedsiębiorstwo z jakiegoś powodu zostanie wyeliminowane z obserwacji przed zarejestrowaniem zdarzenia upadłości (np. data badania),
- uniknięcie niestabilności spowodowanej koniecznością wyboru sztywno ustalonej długości okresu, w ciągu którego jest obserwowane zajście zdarzenia upadłości,
- estymacja długości okresu do zajścia zdarzenia umożliwia śledzenie w czasie przebiegu natężenia ryzyka wystąpienia zdarzenia,
- uzyskanie „dynamicznej” prognozy prawdopodobieństwa zajścia zdarzenia (wartość prognozy jest funkcją czasu), co jest bardzo użyteczne przy określaniu odpowiedniej strategii i polityki,

- możliwość uwzględnienia zmian w otoczeniu gospodarczym w ocenie ryzyka dzięki wykorzystaniu modeli umożliwiających włączenie zmiennych współtowarzyszących zależnych od czasu.

Podsumowując, należy powiedzieć, że zastosowanie metod analizy historii zdarzeń jako alternatywnego podejścia do modelowania ryzyka upadłości przedsiębiorstw daje szansę na szersze wykorzystanie wyników tych metod, niż jest to możliwe w przypadku zastosowania metod statycznych, a dzięki temu ulepszenie modeli prognozy upadłości przedsiębiorstw.

Literatura

- Allison P.D., *Regression for Longitudinal Event Data*, Sage University Paper Series on Quantitative Applications in the Social Sciences, nr 46, Publications, Sage 1984.
- Allison P.D., *Survival Analysis Using SAS*, A Practical Guide Second Edition, SAS Publishing 2010.
- Blossfeld H.P., Rohwer G., *Techniques of Event History Modeling. New Approaches to Causal Analysis*, Lawrence Elbaum Associates Publishers, London 2002
- Dec P., *Kompleksowy system wczesnego ostrzegania*, [w:] *Ryzyko w działalności przedsiębiorstw. Wybrane aspekty*, red. A. Fierla, Szkoła Główna Handlowa – Oficyna Wydawnicza, Warszawa 2009.
- Frączak E., *Analiza historii zdarzeń*, Oficyna Wydawnicza SGH, Warszawa 2005.
- Gruszczyński M., *Modele zmiennych jakościowych dwumianowych*, [w:] *Mikroekonometria. Modele metody analizy danych indywidualnych*, red. M. Gruszczyński, a Wolters Kluwer bussines, Warszawa 2010.
- Hadasik D., *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, [w:] *Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu, AE w Poznaniu*, Poznań 1998.
- Hamrol M., Chodakowski J., *Prognozowanie zagrożenia finansowego przedsiębiorstwa. Wartość predykcyjna polskich modeli analizy dyskryminacyjnej*, [w:] *Badania Operacyjne i Decyzje* 2008, nr 3.
- Hołda A., *Prognozowanie bankructwa jednostki w warunkach gospodarki polskiej z wykorzystaniem funkcji dyskryminacyjnej*, „Rachunkowość” 2001, nr 5.
- Mączyńska E., *Ocena ryzyka upadłości przedsiębiorstwa*, [w:] *Ryzyko w działalności przedsiębiorstw. Wybrane aspekty*, red. A. Fierli, Szkoła Główna Handlowa – Oficyna Wydawnicza, Warszawa 2009.
- Mączyńska E., *Wstęp*, [w:] *Meandry upadłości przedsiębiorstw*, red. E. Mączyńska, Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie, Warszawa 2010.
- Pieńkowska M., *Weryfikacja skuteczności funkcji dyskryminacyjnych opracowanych dla rynku polskiego*, [w:] *Zagrożenie upadłością*, red. K. Kuciński, E. Mączyńska, Szkoła Główna Handlowa, Warszawa 2005.
- Prusak B., *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Difin, Warszawa 2009.
- Ptak-Chmielewska A., Pęczkowski M., *Analiza dyskryminacji*, [w:] *Wielowymiarowa analiza statystyczna. Teoria – przykłady zastosowań z systemem SAS*, red. E. Frączak, Oficyna Wydawnicza SGH, Warszawa 2009.
- Staniec I., *Zastosowanie sztucznych sieci neuronowych i wybranych metod statystycznych do wspomaganie decyzji kredytowych*, praca doktorska, 2000.
- Stasiewski T., *Z-score – indeks przewidywanego upadku przedsiębiorstwa*, „Rachunkowość” 1996, nr 12.

- Zaleska M., *Identyfikacja ryzyka upadłości przedsiębiorstwa i banku. Systemy wczesnego ostrzegania*, Difin, Warszawa 2002.
- Zaleska M., *Ocena ekonomiczno-finansowa przedsiębiorstwa przez analityka bankowego*, Szkoła Główna Handlowa – Oficyna Wydawnicza, Warszawa 2012.
- Zarzecki D., *O metodach zagrożenia bankructwem i możliwościach ich wykorzystania w Polsce*, [w:] *Rynek kapitałowy*, red. W. Tarczyński, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin 2000.

APPLICATION OF SURVIVAL MODELS AND DISCRIMINANT ANALYSIS IN EVALUATION OF ENTERPRISES' BANKRUPTCY RISK

Summary: This paper presents the application of logistic regression, discriminant analysis and Cox regression model in the prediction of enterprises bankruptcy. A standard approach is the application of discriminant analysis and logistic regression in risk of bankruptcy modelling. This approach, however, is static and does not include the time factor. Alternative proposal is the application of hazard rate models. Those models are dynamic because they include the time to the event occurrence (bankruptcy). The main goal of this paper is an identification of the explanatory variables influence on the risk of bankruptcy and differences in modelling approaches: static and dynamic. An empirical example is presented based on financial results of the sample of 1536 enterprises (including 456 bankrupts).

Keywords: survival analysis, enterprises bankruptcy models, discriminant analysis, logistic regression.