PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 384

Taksonomia 24
Klasyfikacja i analiza danych – teoria i zastosowania

Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronie internetowej Wydawnictwa
www.pracenaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 384 • 2015
Taksonomia 24 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania e-ISSN 2392-0041

Józef Pociecha, Mateusz Baryła, Barbara Pawełek
Uniwersytet Ekonomiczny w Krakowie
e-mails: {jozef.pociecha, mateusz.baryla, barbara.pawelek}@uek.krakow.pl

PORÓWNANIE SKUTECZNOŚCI KLASYFIKACYJNEJ WYBRANYCH METOD PROGNOZOWANIA BANKRUCTWA PRZEDSIĘBIORSTW PRZY LOSOWYM I NIELOSOWYM DOBORZE PRÓB[1]

Streszczenie: Jednym ze źródeł błędów występujących w prognozowaniu bankructwa przedsiębiorstw jest metoda doboru prób. Przy konstruowaniu prób o zbilansowanej strukturze najczęściej wykorzystywana jest technika o charakterze nielosowym, polegająca na dobieraniu parami przedsiębiorstw bankrutów i niebankrutów. Alternatywą dla metody parowania może być losowanie niezależne. W artykule zaprezentowano analizę porównawczą wybranych modeli prognozowania bankructwa, uwzględniając losowy i nielosowy dobór prób. Rozważono dwa podziały zbioru danych na zbiór uczący i testowy w stosunku 7:3 i 6:4. Zaprezentowano modyfikację modelu logitowego, polegającą na wprowadzeniu do modelu czynnika czasu. Rozważono użyteczność tej modyfikacji w kontekście możliwości poprawy skuteczności klasyfikacyjnej modeli z uwzględnieniem losowego i nielosowego doboru prób.

Słowa kluczowe: bankructwo, prognozowanie, dobór próby, model logitowy, koniunktura gospodarcza.

DOI: 10.15611/pn.2015.384.25

1. Wstęp

Zjawisko upadłości przedsiębiorstw jest elementem gospodarki rynkowej. Ze względu na konsekwencje społeczno-ekonomiczne, jakie towarzyszą występowaniu tego zjawiska, jest ono przedmiotem zainteresowania zarówno praktyków gospodarczych, jak i badaczy. Jednym z przejawów tego zainteresowania jest rozwój metod prognozowania zagrożenia upadłością przedsiębiorstw.

Metody służące do przewidywania bankructwa przedsiębiorstw powinny charakteryzować się wysoką jakością prognostyczną. Wśród możliwych źródeł błędów

[1] Udział w konferencji SKAD 2014 był możliwy dzięki środkom finansowym z projektów badawczych 028/WZ-KS/01/2014/S/4224 i 031/WZ-KS/04/2014/S/4227.

w prognozowaniu bankructwa przedsiębiorstw wymienia się [Pawełek, Pociecha 2012]: metodę doboru prób i niestabilny charakter badanych populacji.

Jedną z najpopularniejszych technik doboru prób zbilansowanych z populacji przedsiębiorstw upadłych i nieupadłych jest metoda kojarzenia parami podmiotów, nazywana także metodą parowania. Zgodnie z tą techniką za próbę bankrutów przyjmuje się wszystkie przedsiębiorstwa upadłe w danym okresie. Następnie do tak uzyskanej próby bankrutów dobiera się przedsiębiorstwa, które kontynuują działalność gospodarczą, zgodnie z przyjętym kryterium. Wobec pewnych wątpliwości, jakie rodzi metoda parowania, alternatywą dla tej nielosowej techniki może być losowanie niezależne (zob. np. [Pociecha, Pawełek 2011]).

Głównym celem artykułu jest porównanie skuteczności klasyfikacyjnej czterech najczęściej stosowanych typów modeli w badaniach nad upadłością przedsiębiorstw, przy zastosowaniu dwóch wspomnianych metod pobierania prób bankrutów i niebankrutów. Ponadto dla obu metod doboru obiektów do próby sprawdzono, czy wprowadzenie do modelu logitowego czynnika czasu, odzwierciedlającego zmiany zachodzące w otoczeniu gospodarczym przedsiębiorstw, może przyczynić się do poprawy skuteczności klasyfikacyjnej modelu logitowego służącego do prognozowania bankructwa przedsiębiorstw przetwórstwa przemysłowego w Polsce.

2. Charakterystyka danych i opis procedury badawczej

Podstawę prowadzenia badań stanowiła baza danych składająca się z 7329 rekordów, zawierających informacje na temat 1852 przedsiębiorstw przetwórstwa przemysłowego w Polsce. Dane finansowe zostały zaczerpnięte z Monitora Polski B oraz serwisu EMIS i obejmowały one lata 2005-2009. Każdy rekord bazy był opisany przez 35 zmiennych będących wskaźnikami finansowymi (w tym: wskaźnikami płynności, zadłużenia, rentowności, sprawności działania) oraz zmienną binarną, która przyjmowała wartość 1, gdy przedsiębiorstwo zbankrutowało w latach 2007-2010, oraz wartość 0, gdy przedsiębiorstwo nie zbankrutowało w latach 2005-2010. Przyjęto, że dane finansowe na temat tego samego przedsiębiorstwa w różnych latach będą traktowane jako informacje o różnych przedsiębiorstwach, co pozwoliło utożsamić rekordy bazy danych z przedsiębiorstwami. Ostatecznie dysponowano danymi o 7329 przedsiębiorstwach, w tym 182 bankrutach (2,5%) i 7147 przedsiębiorstwach dobrze prosperujących (97,5%).

W prezentowanej analizie rozważono następujące warianty badania:

- wariant W_1: obejmował dane za okres 2006-2009 i dotyczył prognozowania bankructwa przedsiębiorstw z rocznym wyprzedzeniem,
- wariant W_2: obejmował dane za okres 2005-2008 i dotyczył prognozowania bankructwa przedsiębiorstw z dwuletnim wyprzedzeniem.

W przeprowadzonym badaniu wszystkie przedsiębiorstwa upadłe oraz nieupadłe w danym okresie, który związany był z rozważanym wariantem badania, potraktowano jako populacje bankrutów i niebankrutów.

Konstruowanie prób odbywało się przy wykorzystaniu dwóch technik, tj. metody parowania oraz losowania niezależnego. Stosując pierwszą z wymienionych metod, jako próbę bankrutów przyjęto wszystkie przedsiębiorstwa, które upadły w rozważanym okresie. Następnie podmioty upadłe łączono w pary z przedsiębiorstwami znajdującymi się w dobrej kondycji finansowej, kierując się przy tym podobną wielkością przedsiębiorstwa oraz tym samym działem PKD. Kojarzenie przedsiębiorstw parami dotyczyło tego samego roku. Tak oto dla wariantów W_1 i W_2 uzyskano próby zbilansowane o liczności równej odpowiednio: 118 i 246 przedsiębiorstw. Zastosowanie drugiej techniki sprowadzało się do wylosowania z populacji bankrutów i niebankrutów określonej liczby przedsiębiorstw, z zastosowaniem przy tym losowania niezależnego. Liczba wylosowanych podmiotów z danej grupy pozostawała w zgodzie z liczbą przedsiębiorstw, które utworzyły próby przy zastosowaniu metody parowania.

W celu przeprowadzenia analizy poczyniono dodatkowe założenia, które dotyczyły: typu zastosowanych modeli, struktury podziału próby na część uczącą i testową oraz przyjętej techniki wyboru zmiennych do modelu. W badaniu zastosowano najczęściej wykorzystywane cztery typy modeli predykcji bankructwa przedsiębiorstw, a mianowicie: liniową funkcję dyskryminacyjną, model logitowy, sieć neuronową posiadającą architekturę perceptronu wielowarstwowego oraz drzewo klasyfikacyjne uzyskane za pomocą algorytmu CART. W przypadku sieci neuronowych rozpatrywano jedynie perceptrony trójwarstwowe, składające się z warstwy wejściowej, jednej warstwy ukrytej oraz warstwy wyjściowej. Pod uwagę brano następujące funkcje aktywacji: funkcję liniową, tangens hiperboliczny, funkcję wykładniczą, funkcję logistyczną, funkcję softmax. Podczas konstruowania drzew klasyfikacyjnych do oceny jakości podziału obiektów w węzłach drzewa wykorzystano wskaźnik Giniego.

Rozważono dwa podziały próby na zbiór uczący i testowy, w stosunku 6:4 (zbiór uczący zawierał w przypadku wariantu W_1 – 70 obiektów, W_2 – 148 obiektów) i 7:3 (zbiór uczący obejmował, odpowiednio: 82 i 172 obiekty). Dzieląc losowo próbę na oba podzbiory, zachowywano równoliczność bankrutów i niebankrutów w obrębie tych dwóch grup.

Wyboru zmiennych w przypadku modeli statystycznych (modelu dyskryminacyjnego i logitowego) dokonano z wykorzystaniem metody krokowej „w przód" oraz „w tył" w ramach analizy dyskryminacyjnej i logitowej. Dla drzew klasyfikacyjnych zastosowanie algorytmu CART automatycznie prowadziło do redukcji liczby zmiennych w modelu. Zmienne, jakie wyselekcjonowano wspomnianymi metodami, zostały również wykorzystane do konstrukcji sieci neuronowych.

W prowadzonej analizie zastosowano następującą krokową procedurę. Uzyskane próby (w ramach danego wariantu badania i przy zastosowaniu określonej metody pobierania prób bankrutów i niebankrutów) dzielono wielokrotnie w sposób losowy w odpowiedniej proporcji na dwie podpróby (uczącą i testową) aż do

momentu uzyskania 10 modeli danego typu, które spełniały określone warunki. Były to modele:

- posiadające nie więcej niż 6 zmiennych niezależnych (wskaźników finansowych);
- dla których wartość mierników sprawności I rodzaju (definiowanej jako procent bankrutów poprawnie zaklasyfikowanych przez model do zbioru bankrutów) oraz II rodzaju (definiowanej jako procent niebankrutów poprawnie zaklasyfikowanych przez model do zbioru niebankrutów), na zbiorze zarówno uczącym, jak i testowym, była wyższa niż 50%;
- z parametrami statystycznie istotnymi na poziomie istotności $\alpha = 0{,}05$ (dotyczy tylko modeli statystycznych).

W kolejnym kroku z grona tak otrzymanych 10 modeli danego typu wybierano model o najwyższych zdolnościach prognostycznych (na podstawie sprawności klasyfikacyjnej na zbiorze testowym). W pierwszej kolejności, przy wyborze najlepszego modelu, kierowano się maksymalizacją wartości miary SP I (tj. sprawności I rodzaju) na zbiorze testowym (zob. np. [Bellovary i in. 2007]). W sytuacji, gdy kilka modeli miało tę samą wartość miary SP I, jako najlepszy model wybierano ten, który charakteryzował się najwyższą wartością miary SP II (tj. sprawności II rodzaju) na zbiorze testowym.

3. Wyniki empiryczne dla modeli tradycyjnych

Zastosowanie opisanej w punkcie 2 procedury doprowadziło do uzyskania 32 modeli o najwyższych zdolnościach prognostycznych, które zestawiono w postaci rankingów i zaprezentowano w tab. 1 i 2. Trzy ostatnie kolumny tych tabel przedstawiają wartości miar SP I, SP II oraz SP (sprawność ogólna, tj. procent wszystkich przedsiębiorstw poprawnie zaklasyfikowanych przez model) obliczonych dla danych, które utworzyły zbiór testowy. Podczas sporządzania zestawień przydzielanie poszczególnych miejsc rankingowych odbywało się według tej samej zasady, która znalazła swoje zastosowanie przy wyborze najlepszych modeli.

W celu sprawdzenia, jak zastosowana technika doboru próby wpływa na zdolności prognostyczne uzyskanych modeli, zdecydowano się na porównywanie modeli parami. Podczas dokonywania takich zestawień brano pod uwagę modele tego samego typu, powstałe dla tego samego wariantu badania, przy zastosowaniu tej samej struktury podziału próby, lecz przy wykorzystaniu innej techniki pobierania próby bankrutów i niebankrutów. Z takich porównań par modeli zwycięsko wychodziły te spośród nich, które odznaczały się wyższą wartością miary SP I na zbiorze testowym, a w przypadku takiej samej wartości tej miary kierowano się maksymalizacją wartości miernika SP II (na zbiorze testowym).

Tabela 1. Rankingi najlepszych modeli przy nielosowej technice doboru przedsiębiorstw do próby

Wariant badania	Typ podziału próby	Miejsce w rankingu	Rodzaj modelu*	Zbiór testowy		
				SP I	SP II	SP
W_1	6:4	1	SN_1 (3)	95,83	83,33	89,58
		2	D_1 (3)	95,83	75,00	85,42
		3	DK_1 (1)	95,83	70,83	83,33
		4	L_1 (2)	91,67	75,00	83,33
	7:3	1	SN_2 (3)	94,44	72,22	83,33
		2	L_2 (2)	88,89	83,33	86,11
		3	D_2 (3)	88,89	77,78	83,33
		4	DK_2 (1)	88,89	66,67	77,78
W_2	6:4	1	SN_3 (2)	83,67	69,39	76,53
		2	DK_3 (1)	83,67	63,27	73,47
		3	L_3 (2)	77,55	69,39	73,47
		4	D_3 (2)	67,35	71,43	69,39
	7:3	1	SN_4 (2)	81,08	81,08	81,08
		2	DK_4 (1)	81,08	75,68	78,38
		3	L_4 (2)	72,97	72,97	72,97
		4	D_4 (3)	70,27	54,05	62,16

* *SN* oznacza sieć neuronową, *DK* − drzewo klasyfikacyjne, *L* − model logitowy, *D* − model dyskryminacyjny. W nawiasie podano liczbę wskaźników finansowych występujących w modelu.

Źródło: obliczenia własne.

Tabela 2. Rankingi najlepszych modeli przy losowej technice doboru przedsiębiorstw do próby

Wariant badania	Typ podziału próby	Miejsce w rankingu	Rodzaj modelu*	Zbiór testowy		
				SP I	SP II	SP
W_1	6:4	1	SN_5 (2)	100,00	91,67	95,83
		2	DK_5 (1)	100,00	79,17	89,58
		3	D_5 (4)	91,67	100,00	95,83
		4	L_5 (2)	87,50	66,67	77,08
	7:3	1	SN_6 (4)	100,00	83,33	91,67
		2	D_6 (2)	100,00	72,22	86,11
		3	DK_6 (1)	94,44	88,89	91,67
		4	L_6 (2)	94,44	83,33	88,89
W_2	6:4	1	DK_7 (1)	89,80	61,22	75,51
		2	SN_7 (4)	87,76	67,35	77,55
		3	L_7 (4)	81,63	61,22	71,43
		4	D_7 (4)	73,47	75,51	74,49
	7:3	1	SN_8 (2)	89,19	64,86	77,03
		2	DK_8 (1)	83,78	67,57	75,68
		3	D_8 (2)	81,08	62,16	71,62
		4	L_8 (2)	75,68	72,97	74,32

* jak pod tab. 1.

Źródło: obliczenia własne.

Tabela 3. Wyniki porównań par modeli ze względu na zastosowaną technikę doboru przedsiębiorstw do próby

Wariant badania	Typ podziału próby	Liczba zwycięskich porównań	Wynik porównania
W_1	6:4	2	na korzyść losowania niezależnego
		2	na korzyść metody parowania
	7:3	4	na korzyść losowania niezależnego
		0	na korzyść metody parowania
W_2	6:4	4	na korzyść losowania niezależnego
		0	na korzyść metody parowania
	7:3	4	na korzyść losowania niezależnego
		0	na korzyść metody parowania

Źródło: opracowanie własne.

Analizując otrzymane rezultaty (tab. 3), można zauważyć, że na 16 dokonanych porównań par modeli aż w 14 przypadkach zwyciężyły modele uzyskane na gruncie prób losowych. W grupie czterech rozważanych typów modeli losowanie niezależne przyczyniło się do otrzymania lepszych prognoz w przypadku wariantu badania W_1 przy podziale próby na część uczącą i testową w stosunku 7:3 oraz wariantu W_2. Jednoznacznego rozstrzygnięcia nie uzyskano w przypadku podejścia badawczego W_1 przy podziale danych w stosunku 6:4. Warto zwrócić uwagę na to, że tam, gdzie zwyciężały modele budowane na podstawie prób uzyskanych metodą parowania, miało to miejsce jedynie w przypadku dwóch modeli statystycznych (D_1, L_1).

4. Wyniki empiryczne dla zmodyfikowanych modeli logitowych

Wśród potencjalnych źródeł błędów popełnianych w prognozowaniu bankructwa przedsiębiorstw wymienia się, oprócz metody doboru prób, także niestabilny charakter badanych populacji (w tym: brak uwzględniania stanu koniunktury gospodarczej).

W prognozowaniu bankructwa przedsiębiorstw wykorzystuje się dane pobrane ze sprawozdań finansowych przedsiębiorstw bankrutów i niebankrutów. Dane te pochodzą bardzo często z kilku lat. Spowodowane jest to zwykle brakiem możliwości zebrania dostatecznie dużego zbioru danych dla jednego roku. W literaturze przedmiotu można znaleźć rozważania dotyczące problemów pojawiających się przy budowaniu modeli dla binarnej zmiennej zależnej na podstawie danych pochodzących z różnych okresów (np. [Beck, Katz, Tucker 1998]). W pracach dotyczących tego zagadnienia proponuje się zastępowanie tradycyjnych modeli statycznych modelami uwzględniającymi zmiany w czasie obserwowanych wartości (np. [Chava, Jarrow 2004; Shumway 2001]).

Celem badań związanych z prognozowaniem bankructwa przedsiębiorstw jest m.in. zbudowanie modelu charakteryzującego się wysoką zdolnością prognostyczną. Bazując na danych, które odzwierciedlają sytuację finansową przedsiębiorstw w różnych latach, niekiedy przy różnym stanie koniunktury gospodarczej w danym kraju, należy zadać pytanie: czy model z ocenami parametrów uzyskanymi bez uwzględniania zmian w otoczeniu gospodarczym przedsiębiorstw może być podstawą wiarygodnego przewidywania bankructwa?

Modyfikacji modelu, służącego do prognozowania bankructwa przedsiębiorstw, poprzez wprowadzenie czynnika czasu, który reprezentuje zmiany w otoczeniu gospodarczym przedsiębiorstw, dokonano dla modelu logitowego. Podobne próby dynamizacji modeli przewidywania zagrożenia upadłością przedsiębiorstw można znaleźć w literaturze przedmiotu (np. [De Leonardis, Rocci 2014]).

Modyfikacja modelu logitowego polegała na rozszerzeniu zbioru zmiennych objaśniających o zmienne sztuczne [Maddala 2008, s. 349-359] w postaci:

- zmiennych zero-jedynkowych:

$$Y^t = \begin{cases} 1 & \text{gdy} \quad \text{rok} = t \\ 0 & \text{gdy} \quad \text{rok} \neq t \end{cases} \quad \left(t = \begin{cases} 2007, 2008, 2009 \text{ dla } W_1 \\ 2006, 2007, 2008 \text{ dla } W_2 \end{cases} \right), \tag{1}$$

identyfikujących rok, z którego pochodzi sprawozdanie finansowe,

- zmiennych jakościowo-ilościowych:

$$R_i^t = \begin{cases} R_i & \text{gdy} \quad \text{rok} = t \\ 0 & \text{gdy} \quad \text{rok} \neq t \end{cases} \quad \left(i = 01,...,33; t = \begin{cases} 2007, 2008, 2009 \text{ dla } W_1 \\ 2006, 2007, 2008 \text{ dla } W_2 \end{cases} \right), \tag{2}$$

które odzwierciedlają zmieniające się w czasie znaczenie wskaźników finansowych R_i dla prognozowania bankructwa przedsiębiorstw.

Głównym celem dokonania modyfikacji modelu logitowego jest przezwyciężenie trudności wynikających z niestabilnego charakteru badanych populacji. Dodatkowym celem, oprócz wspomnianej próby dynamizacji modelu, jest sprawdzenie użyteczności rozważanego rozwiązania w kontekście problemu doboru prób.

Wyniki badań zaprezentowane w tab. 4 i 5 wskazują na użyteczność dokonanej modyfikacji tradycyjnego modelu logitowego służącego prognozowaniu bankructwa przedsiębiorstw przetwórstwa przemysłowego w Polsce. Porównując wyniki otrzymane dla prób dobieranych w nielosowy i losowy sposób, można zauważyć, że poprawę sprawności klasyfikacyjnej uzyskano częściej w przypadku doboru losowego próby niż doboru nielosowego próby.

Przeprowadzone badanie, na przykładzie modelu logitowego, wskazuje na większą użyteczność dokonanej modyfikacji modeli w przypadku prób uzyskanych metodą losowania niezależnego (tab. 5) niż metodą dobierania parami (tab. 4).

Tabela 4. Porównanie modeli logitowych przy zastosowaniu metody parowania

Typ modelu logitowego (liczba zmiennych objaśniających)	Rodzaj próby	Zbiór testowy		
		SP I	SP II	SP
Tradycyjny (2)	W_1 – 6:4	91,67	75,00	83,33
Zmodyfikowany (5)	W_1 – 6:4	87,50	**83,33**	**85,42**
Tradycyjny (2)	W_1 – 7:3	88,89	83,33	86,11
Zmodyfikowany (–)	W_1 – 7:3	–	–	–
Tradycyjny (2)	W_2 – 6:4	77,55	69,39	73,47
Zmodyfikowany (–)	W_2 – 6:4	–	–	–
Tradycyjny (2)	W_2 – 7:3	72,97	72,97	72,97
Zmodyfikowany (3)	W_2 – 7:3	**75,68**	**72,97**	**74,32**

Uwaga: Symbol „–" oznacza, że zmodyfikowany model logitowy charakteryzował się niższymi wartościami wszystkich mierników sprawności niż tradycyjny model logitowy. Wartości mierników sprawności obliczonych dla zmodyfikowanych modeli logitowych, które są co najmniej równe wartościom obliczonym dla odpowiadających im modeli tradycyjnych, zapisano czcionką pogrubioną.

Źródło: obliczenia własne.

Tabela 5. Porównanie modeli logitowych przy zastosowaniu losowania niezależnego

Typ modelu logitowego (liczba zmiennych objaśniających)	Rodzaj próby	Zbiór testowy		
		SP I	SP II	SP
Tradycyjny (2)	W_1 – 6:4	87,50	66,67	77,08
Zmodyfikowany (3)	W_1 – 6:4	**91,67**	**70,83**	**81,25**
Tradycyjny (2)	W_1 – 7:3	94,44	83,33	88,89
Zmodyfikowany (–)	W_1 – 7:3	–	–	–
Tradycyjny (4)	W_2 – 6:4	81,63	61,22	71,43
Zmodyfikowany (3)	W_2 – 6:4	**89,80**	**69,39**	**79,59**
Tradycyjny (2)	W_2 – 7:3	75,68	72,97	74,32
Zmodyfikowany (4)	W_2 – 7:3	**81,08**	67,57	**74,32**

Uwaga: jak pod tab. 4.

Źródło: obliczenia własne.

Porównując pary tradycyjnych modeli logitowych (tab. 1 i 2), oszacowanych dla prób uzyskanych w wyniku zastosowania różnych technik doboru obiektów, otrzymujemy, że w 3 na 4 przypadki wyższą zdolnością prognostyczną charakteryzowały się modele oparte na próbach dobieranych metodą losowania niezależnego. W tym kontekście rośnie znaczenie wyników otrzymanych dla zmodyfikowanych modeli logitowych. Uwzględnienie zmian w otoczeniu gospodarczym przedsiębiorstw w modelu prognozowania bankructwa przedsiębiorstw, poprzez wprowadzenie zmiennych sztucznych, może poprawić skuteczność klasyfikacyjną tych modeli.

5. Zakończenie

Podsumowując zaprezentowane wyniki badań, można stwierdzić, że:

- Losowanie niezależne sprzyjało uzyskiwaniu lepszych prognoz w przypadku drzew klasyfikacyjnych i sztucznych sieci neuronowych niż przy nielosowym doborze prób.
- Większość oszacowanych modeli statystycznych odznaczała się lepszymi zdolnościami prognostycznymi przy zastosowaniu losowania niezależnego jako metody pobierania prób.
- Wprowadzenie czynnika czasu do modelu logitowego wpłynęło na poprawę skuteczności klasyfikacyjnej niektórych z rozważanych modeli. Wzrost zdolności prognostycznej był obserwowany przede wszystkim w przypadku prób uzyskanych w wyniku dokonania losowania niezależnego.

Literatura

Beck N., Katz J.N., Tucker R., 1998, *Taking time seriously: time-series–cross-section analysis with a binary dependent variable*, American Journal of Political Science, vol. 42, no. 4, s. 1260-1288.

Bellovary J., Giacomino D., Akers M., 2007, *A review of bankruptcy prediction studies: 1930 to present*, Journal of Financial Education, vol. 33, s. 1-42.

Chava S., Jarrow R.A., 2004, *Bankruptcy prediction with industry effects*, http://dx.doi.org/10.2139/ssrn.287474.

De Leonardis D., Rocci R., 2014, *Default risk analysis via a discrete-time cure rate model*, Applied Stochastic Models in Business and Industry, vol. 30, no. 5, s. 529-543.

Maddala G.S., 2008, *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa.

Pawełek B., Pociecha J., 2012, *General SEM Model in Researching Corporate Bankruptcy and Business Cycles*, [w:] Pociecha J., Decker R. (red.), *Data Analysis Methods and its Applications*, C.H. Beck, Warszawa, s. 215-231.

Pociecha J., Pawełek B., 2011, *Prognozowanie bankructwa a koniunktura gospodarcza*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie – Metody analizy danych, nr 873, s. 5-27.

Shumway T., 2001, *Forecasting bankruptcy more accurately: a simple hazard model*, The Journal of Business, vol. 74, no. 1, s. 101-124.

COMPARISON OF CLASSIFICATION ACCURACY OF SELECTED BANKRUPTCY PREDICTION METHODS IN THE CASE OF RANDOM AND NON-RANDOM SAMPLING TECHNIQUE

Summary: One of the sources of errors being committed in the process of bankruptcy prediction is a method for selecting samples. During the construction of a sample of balanced structure, the most popular non-random approach is based on pairing up

bankrupt companies with non-bankrupt ones. The alternative to pair-matched sampling is simple random sampling with replacement. The article presents a comparative study of selected failure prediction models, taking into account the random and non-random technique of samples selection. Data was divided into a training group and a testing group in a ratio of both 7:3 and 6:4. A modification of Logit model consisting in introducing a time factor into a model is also presented. The usefulness of this modification in the context of its classification accuracy improvement for two aforementioned techniques of sampling was verified.

Keywords: bankruptcy, prediction, sample selection, logit model, economic situation.