

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 327

Taksonomia 22

**Klasyfikacja i analiza danych –
teoria i zastosowania**

Redaktorzy naukowci

Krzysztof Jajuga, Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

w Dolnośląskiej Bibliotece Cyfrowej www.dbc.wroc.pl,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2014

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	9
Eugeniusz Gatnar , Balance of payments statistics and external competitiveness of Poland.....	15
Andrzej Sokolowski, Magdalena Czaja , Efektywność metody k -średnich w zależności od separowalności grup.....	23
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Wielosektorowa analiza ukrytych przejść w modelowaniu zagrożenia upadłością polskich przedsiębiorstw	30
Elżbieta Gołata , Zróżnicowanie procesu starzenia i struktur demograficznych w Poznaniu i aglomeracji poznańskiej na tle wybranych dużych miast Polski w latach 2002-2011.....	39
Aleksandra Łuczak, Feliks Wysocki , Ustalanie systemu wag dla cech w zagadnieniach porządkowania liniowego obiektów	49
Marek Walesiak , Wzmacnianie skali pomiaru dla danych porządkowych w statystycznej analizie wielowymiarowej	60
Paweł Lula , Identyfikacja słów i fraz kluczowych w tekstach polskojęzycznych za pomocą algorytmu <i>RAKE</i>	69
Mariusz Kubus , Propozycja modyfikacji metody złagodzonego LASSO.....	77
Andrzej Bąk, Tomasz Bartłomowicz , Wielomianowe modele logitowe wyborów dyskretnych i ich implementacja w pakiecie <i>DiscreteChoice</i> programu R.....	85
Justyna Brzezińska , Wykorzystanie modeli logarytmiczno-liniowych do analizy bezrobocia w Polsce w latach 2004-2012.....	95
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Zastosowanie pakietu <i>dcMNM</i> programu R w badaniach preferencji konsumentów wódki	104
Barbara Batóg, Jacek Batóg , Analiza stabilności klasyfikacji polskich województw według sektorowej wydajności pracy w latach 2002-2010	113
Małgorzata Markowska, Danuta Strahl , Klasyfikacja europejskiej przestrzeni regionalnej ze względu na filary inteligentnego rozwoju z wykorzystaniem referencyjnego systemu granicznego.....	121
Kamila Migdał-Najman, Krzysztof Najman , Formalna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	131
Kamila Migdał-Najman, Krzysztof Najman , Graficzna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	139
Beata Basiura, Anna Czapkiewicz , Badanie jakości klasyfikacji szeregów czasowych	148
Michał Trzęsiok , Wybrane metody identyfikacji obserwacji oddalonych.....	157

Grażyna Dehnel, Tomasz Klimanek , Taksonomiczne aspekty estymacji pośredniej uwzględniającej autokorelację przestrzenną w statystyce gospodarczej.....	167
Michał Bernard Pietrzak, Justyna Wilk , Odległość ekonomiczna w modelowaniu zjawisk przestrzennych z wykorzystaniem modelu grawitacji.....	177
Maciej Beręsewicz , Próba zastosowania różnych miar odległości w uogólnionym estymatorze Petersena.....	186
Marcin Szymkowiak, Tomasz Józefowski , Konstrukcja i praktyczne wykorzystanie estymatorów typu SPREE na przykładzie dwuwymiarowych tabel kontyngencji.....	195
Marcin Pelka , Klasyfikacja pojęciowa danych symbolicznych w podejściu wielomodelowym.....	202
Małgorzata Machowska-Szewczyk , Ocena klas w rozmytej klasyfikacji obiektów symbolicznych.....	210
Justyna Wilk , Problem wyboru liczby klas w taksonomicznej analizie danych symbolicznych.....	220
Andrzej Dudek , Metody analizy skupień w klasyfikacji markerów map Google.....	229
Ewa Roszkowska , Ocena ofert negocjacyjnych w słabo ustrukturyzowanych problemach negocjacyjnych z wykorzystaniem rozmytej procedury SAW.....	237
Marcin Szymkowiak, Marek Witkowski , Zastosowanie analizy korespondencji do badania kondycji finansowej banków spółdzielczych.....	248
Bartłomiej Jefmański , Budowa rozmytych indeksów satysfakcji klientów z zastosowaniem programu R.....	257
Karolina Bartos , Odkrywanie wzorców zachowań konsumentów za pomocą analizy koszykowej danych transakcyjnych.....	266
Joanna Trzęsiok , Taksonomiczna analiza krajów pod względem dzietności kobiet oraz innych czynników demograficznych.....	275
Beata Bal-Domańska , Próba identyfikacji większych skupisk regionalnych oraz ich konwergencja.....	285
Beata Bieszk-Stolorz, Iwona Markowicz , Wpływ zasiłku na proces poszukiwania pracy.....	294
Marta Dziechciarz-Duda, Klaudia Przybysz , Wykształcenie a potrzeby rynku pracy. Klasyfikacja absolwentów wyższych uczelni.....	303
Tomasz Klimanek , Problem pomiaru procesu dezagrarnizacji wsi polskiej w świetle wielowymiarowych metod statystycznych.....	313
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Wybrane metody analizy danych wzdluznych.....	321
Artur Zaborski , Zastosowanie miar odległości dla danych porządkowych do agregacji preferencji indywidualnych.....	330
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Zastosowanie analizy korespondencji do badania sytuacji mieszkańców strefy podmiejskiej Warszawy na rynku pracy.....	338

Katarzyna Wawrzyniak , Klasyfikacja województw według stopnia realizacji priorytetów Strategii Rozwoju Kraju 2007-2015 z wykorzystaniem wartości centrum wierszowego	346
---	-----

Summaries

Eugeniusz Gatnar , Statystyka bilansu płatniczego a konkurencyjność gospodarki Polski	22
Andrzej Sokółowski, Magdalena Czaja , Cluster separability and the effectiveness of k -means method	29
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Multisectoral analysis of latent transitions in bankruptcy prediction models.....	38
Elżbieta Golata , Differences in the process of aging and demographic structures in Poznań and the agglomeration compared to selected Polish cities in the years 2002-2011	48
Aleksandra Łuczak, Feliks Wysocki , Determination of weights for features in problems of linear ordering of objects	59
Marek Walesiak , Reinforcing measurement scale for ordinal data in multivariate statistical analysis	68
Paweł Lula , Automatic identification of keywords and keyphrases in documents written in Polish.....	76
Mariusz Kubus , The proposition of modification of the relaxed LASSO method.....	84
Andrzej Bąk, Tomasz Bartłomowicz , Microeconomic multinomial logit models and their implementation in the <code>DiscreteChoice</code> R package .	94
Justyna Brzezińska , The analysis of unemployment data in Poland in 2004-2012 with application of log-linear models	103
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Application of the MMLM package of R software for vodka consumers preference analysis.....	112
Barbara Batóg, Jacek Batóg , Analysis of the stability of classification of Polish voivodeships in 2002-2010 according to the sectoral labour productivity	120
Małgorzata Markowska, Danuta Strahl , Classification of the European regional space in terms of smart growth pillars using the reference limit system.....	130
Kamila Migdał Najman, Krzysztof Najman , Formal quality assessment of group structure mapping on the Kohonen's map	138
Kamila Migdał Najman, Krzysztof Najman , Graphical quality assessment of group structure mapping on the Kohonen's map	147
Beata Basiura, Anna Czapkiewicz , Validation of time series clustering	156
Michał Trzęsiok , Selected methods for outlier detection.....	166

Grażyna Dehnel, Tomasz Klimanek , Taxonomic aspects of indirect estimation accounting for spatial correlation in enterprise statistics	176
Michał Bernard Pietrzak, Justyna Wilk , Economic distance in modeling spatial phenomena with the application of gravity model.....	185
Maciej Beręsewicz , An attempt to use different distance measures in the Generalized Petersen estimator	194
Marcin Szymkowiak, Tomasz Józefowski , Construction and practical using of SPREE estimators for two-dimensional contingency tables.....	201
Marcin Pelka , The ensemble conceptual clustering for symbolic data.....	209
Małgorzata Machowska-Szewczyk , Evaluation of clusters obtained by fuzzy classification methods for symbolic objects.....	219
Justyna Wilk , Problem of determining the number of clusters in taxonomic analysis of symbolic data	228
Andrzej Dudek , Clustering techniques for Google maps markers.....	236
Ewa Roszkowska , The evaluation of negotiation offers in ill structure negotiation problems with the application of fuzzy SAW procedure	247
Marcin Szymkowiak, Marek Witkowski , The use of correspondence analysis in analysing the financial situation of cooperative banks.....	256
Bartłomiej Jefmański , The construction of fuzzy customer satisfaction indexes using R program.....	265
Karolina Bartos , Discovering patterns of consumer behaviour by market basket analysis of the transactional data.....	274
Joanna Trzęsiok , Cluster analysis of countries with respect to fertility rate and other demographic factors	284
Beata Bal-Domańska , An attempt to identify major regional clusters and their convergence	293
Beata Bieszk-Stolorz, Iwona Markowicz , The influence of benefit on the job finding process	302
Marta Dziechciarz-Duda, Klaudia Przybysz , Education and labor market needs. Classification of university graduates	312
Tomasz Klimanek , The problem of measuring deagrarianisation process in rural areas in Poland using multivariate statistical methods.....	320
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Selected methods for an analysis of longitudinal data.....	329
Artur Zaborski , The application of distance measures for ordinal data for aggregation individual preferences	337
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Application of correspondence analysis to examine the situation of the inhabitants of Warsaw suburban area in the labour market	345
Katarzyna Wawrzyniak , Classification of voivodeships according to the level of the realization of priorities of <i>the National Development Strategy 2007-2015</i> with using the values of centroid of the rows	355

Barbara Pawelek, Józef Pocięcha, Adam Sagan

Uniwersytet Ekonomiczny w Krakowie

WIELOSEKTOROWA ANALIZA UKRYTYCH PRZEJŚĆ W MODELOWANIU ZAGROŻENIA UPADŁOŚCIĄ POLSKICH PRZEDSIĘBIORSTW¹

Streszczenie: W pracy został podjęty aktualny i ważny dla dalszego rozwoju polskiej gospodarki problem zagrożenia upadłością polskich przedsiębiorstw z sekcji przetwórstwo przemysłowe. Celem pracy jest ukazanie przydatności wielosektorowej analizy ukrytych przejść do oceny zagrożenia upadłością polskich przedsiębiorstw z tej sekcji. W modelowaniu ukrytych przejść wykorzystano wskaźniki finansowe z lat 2007 i 2009, szczególnie istotnych z punktu widzenia kształtowania się sytuacji kryzysowej i prawdopodobieństwa bankructwa.

Słowa kluczowe: analiza ukrytych przejść, analiza wielosektorowa, zagrożenie upadłością.

1. Wstęp

Upadłość przedsiębiorstw jako zjawisko społeczno-gospodarcze jest jednym z elementów gospodarki rynkowej. Wielość i różnorodność negatywnych konsekwencji tego zjawiska, które dotyczą zarówno społeczeństwa, jak i funkcjonowania gospodarki kraju, przyczyniają się do rozwoju metod umożliwiających modelowanie i prognozowanie zagrożenia upadłością przedsiębiorstw [Altman 1968; Bellovary, Giacomino, Akers 2007]. W prognozowaniu bankructwa przedsiębiorstw wykorzystywane są różne modele statystyczne. Najczęściej są to modele wielowymiarowej analizy dyskryminacyjnej, modele logitowe, modele sztucznych sieci neuronowych [Aziz, Dar 2006; McKee 2000]. W modelach tych uwzględnia się wskaźniki finansowe, które w literaturze z zakresu rachunkowości są wskazywane jako syntetyczne miary odzwierciedlające sytuację przedsiębiorstwa [Wędzki 2009, s. 31].

W dotychczas stosowanych modelach bankructwa firm zakłada się, że zjawiska występujące w teorii ekonomicznej są bezpośrednio obserwowane. W badaniu zagrożenia upadłością przedsiębiorstw takie zjawiska, jak płynność, zadłużenie,

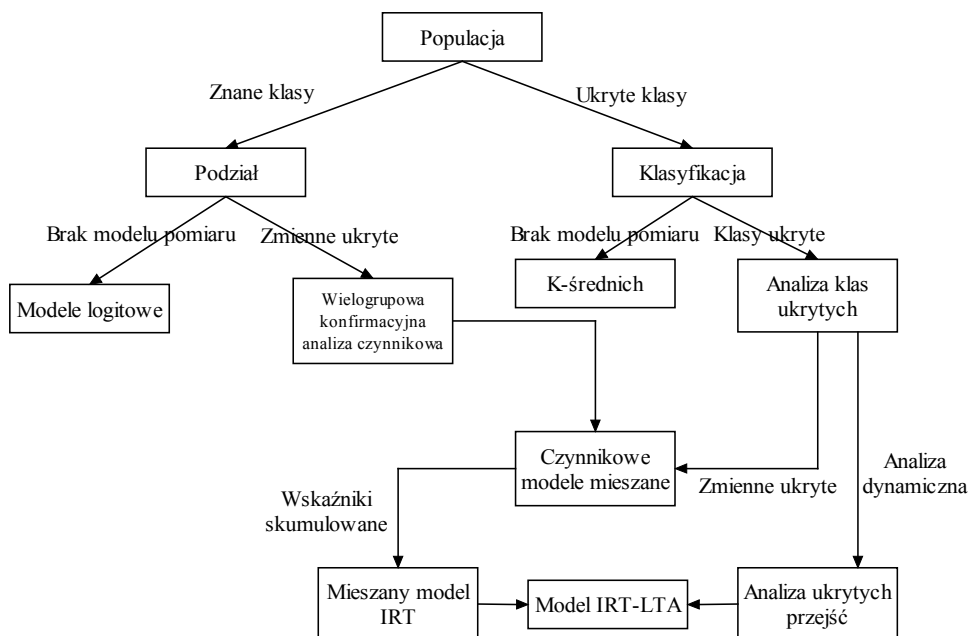
¹ Praca powstała w ramach realizacji grantu badawczego nr 5401/B/H03/2011/40 pt. Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej.

rentowność i sprawność działania, można zmierzyć jedynie przez ich obserwowalne wskaźniki [Pawełek, Pocięcha 2012].

Celem pracy jest przedstawienie propozycji wykorzystania analizy ukrytych przejść (*Latent Transition Analysis* – *LTA*) do modelowania zagrożenia upadłością polskich przedsiębiorstw przemysłu przetwórczego w podziale na sektory działalności gospodarczej².

2. Model pomiaru kondycji finansowej przedsiębiorstw – podejście IRT

Analiza sektorowej ukrytej trajektorii bankructwa wynika z połączenia modelu czynnikowego pomiarowego IRT (*Item Response Theory*) i modelu ukrytych przejść, będącego dynamiczną wersją analizy klas ukrytych (*Latent Class Analysis* – *LCA*).



Rys. 1. Typologia metod klasyfikacyjnych

Źródło: opracowanie własne.

² Termin *latent transition* (ukryte przejścia) wynika z wielopoziomowego charakteru modelu, w którym kolejne pomiary są zagnieżdżone w analizowanych obiektach i ich zmienność w czasie wynika z nieobserwowanej (ukrytej) heterogeniczności.

Generalnie analizy segmentacyjne wiążą się z pytaniem o charakter identyfikowanych klas oraz z istnieniem w ich ramach modelu pomiarowego wyjaśniającego wzajemne kowariancje między wskaźnikami dla klas. W przypadku nieznanych *a priori* klas i braku modeli pomiarowych do identyfikacji segmentów wykorzystywane są metody analizy skupień lub klas ukrytych (rys. 1). W sytuacji, gdy w ramach klas ukrytych występuje model pomiarowy wyjaśniający pewną część wariancji wskaźników, wówczas analiza klasyfikacyjna ma postać czynnikowych modeli mieszanych (*factor mixtures model*). Specyficzna wersja tych modeli, uwzględniających skumulowany charakter wskaźników, stanowi mieszane modele IRT. Modele te wykorzystywane w badaniach z powtarzanymi pomiarami są modelami czynnikowymi ukrytych przejść. Mają one zastosowanie w sytuacji, gdy struktura i liczba klas ukrytych nie jest znana i cechują się one wewnętrzną dynamiką zmian w czasie.

3. Wybór wskaźników finansowych i liczby klas ukrytych

Prezentowane badanie składało się z dwóch etapów. Celem badań przeprowadzonych w pierwszym etapie było między innymi wytypowanie spośród 30 rozważanych wskaźników finansowych tych, które są przydatne do przewidywania bankructwa firm. Baza danych obejmowała 941 przedsiębiorstw przemysłu przetwórczego w Polsce (w tym 28 bankrutów). Dane pochodziły z lat 2007 i 2008. Źródłem danych były Monitor B i baza EMIS.

W badaniu rozważono 30 wskaźników finansowych w podziale na wskaźniki: płynności (4), zadłużenia (10), rentowności (7) i sprawności działania (9) oraz zmienną zero-jedynkową przyjmującą wartości: 1 – gdy firma zbankrutowała w 2009 r. i 0 – w przeciwnym przypadku.

W celu zastosowania modelu IRT ciągle wskaźniki finansowe poddano binaryzacji, przyjmując za punkt progowy wartość mediany obliczanej dla $Me_j(B)$ (mediana dla bankrutów) i $Me_j(NB)$ (mediana dla nie-bankrutów). Wartość 1 oznaczała, że poziom wskaźnika finansowego danego przedsiębiorstwa jest z zakresu wartości bardziej odpowiadających bankrotom niż nie-bankrotom (wysokie ryzyko upadłości), natomiast wartość 0 – z zakresu wartości bardziej odpowiadających nie-bankrotom niż bankrotom (niskie ryzyko upadłości).

Na podstawie nieparametrycznego modelu Mokkena i dwuparametrycznego modelu IRT Birnbauma dla lat 2007 i 2008 wybrano następujące wskaźniki finansowe: R04 = kapitał pracujący / aktywa ogółem (wskaźnik płynności), R11 = (wynik netto + amortyzacja) / zobowiązania (wskaźnik zadłużenia), R12 = kapitał własny / zobowiązania (wskaźnik zadłużenia), R16 = EBIT / aktywa ogółem (wskaźnik rentowności), R20 = wynik netto / aktywa ogółem (wskaźnik rentowności). Cztery spośród pięciu wytypowanych wskaźników (oprócz R11) można uznać za odpowiedniki zmiennych występujących w modelu Altmana [Altman 1968].

Następnie przeprowadzono analizę klas ukrytych, co pozwoliło na wskazanie trzech klas ukrytych. Przy wyborze liczby klas ukrytych wykorzystano miary oparte na statystyce chi-kwadrat oraz kryteria informacyjne AIC i BIC³.

4. Struktura danych

W drugim etapie badań, których wyniki zostały zaprezentowane w niniejszej pracy, powrócono do danych ciągłych. Podjęto próbę przeprowadzenia analizy ukrytych przejść bez utraty informacji związanej z przejściem z danych ciągłych na dyskretne.

Tabela 1. Liczebności działów i sektorów w bazie danych

Liczba firm	Sektor	Dział	Nazwa działu	Liczba firm
283	spożywczy	15	produkcja artykułów spożywczych	283
27	włókienniczy	17	włókiennictwo	27
89	drzewny	20	produkcja drewna i wyrobów z drewna oraz korka (oprócz mebli), artykułów ze słomy i materiałów używanych do wyplatania	49
		21	produkcja masy włóknistej, papieru oraz wyrobów z papieru	9
		22	działalność wydawnicza; poligrafia i reprodukcja zapisanych nośników informacji	31
154	chemiczny	24	produkcja wyrobów chemicznych	13
		25	produkcja wyrobów gumowych i z tworzyw sztucznych pozostałych	141
72	niemetaliczny	26	produkcja wyrobów z surowców niemetalicznych pozostałych	72
210	metalowy	27	produkcja metali	48
		28	produkcja metalowych wyrobów gotowych z wyjątkiem maszyn i urządzeń	162
124	maszynowy	29	produkcja maszyn i urządzeń, gdzie indziej niesklasyfikowana	113
		30	produkcja maszyn biurowych i komputerów	11
48	elektryczny	31	produkcja maszyn i aparatury elektrycznej, gdzie indziej niesklasyfikowana	26
		32	produkcja sprzętu i urządzeń radiowych, telewizyjnych i telekomunikacyjnych	7
		33	produkcja instrumentów medycznych, precyzyjnych i optycznych, zegarów i zegarków	15
89	transportowy	34	produkcja pojazdów mechanicznych, przyczep i naczep	80
		35	produkcja pozostałego sprzętu transportowego	9
70	meblarski	36	produkcja mebli; działalność produkcyjna, gdzie indziej niesklasyfikowana	70

Źródło: opracowanie własne.

³ Wyniki analiz z pierwszego etapu badań zaprezentowano w referacie [Pawełek, Pocięcha, Sagan 2013].

Na podstawie wyników wcześniejszych badań, przyjęto pięć wymienionych powyżej wskaźników finansowych i założono, że są trzy klasy ukryte.

Zbiór obiektów zawierał $N = 1166$ przedsiębiorstw przemysłu przetwórczego w Polsce, które nie zbankrutowały do 2009 roku włącznie. Dane dotyczyły lat 2007 i 2009. Źródłem danych były ponownie Monitor B i baza EMIS. W próbie znajdowały się przedsiębiorstwa z 18 działów według PKD 2004, opisujących działalność z zakresu przemysłu przetwórczego. Z uwagi na małą reprezentację w bazie niektórych działów PKD, podjęto decyzję o utworzeniu sektorów, obejmujących działy reprezentujące podobną działalność gospodarczą (tab. 1).

W celu uwzględnienia ilościowego charakteru wskaźników finansowych w budowie finalnego modelu pomiarowego zastosowano model ciągłych reakcji (*Continuous Response Model* – CRM) Samejimy [1973]. Pozwala on na uwzględnienie metrycznego charakteru wskaźników finansowych w budowie modelu pomiaru. Model ten jest przedstawiony równaniem [Wang, Zeng 1998]:

$$P(X_{ij} \geq x | \theta_i, a_j, b_j, \alpha_j) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^v e^{-\frac{t^2}{2}} dt. \quad (1)$$

Na podstawie parametrów pozycji wskaźników obliczone są wartości skalowe obiektów (przedsiębiorstw) zajmujących dane położenie na continuum cechy ukrytej (θ_i), które są funkcją parametrów trudności (a_j), dyskryminacji (b_j) i skalowania (α_j) oraz znormalizowanych wartości wskaźników x_{ij} ($z_{ij} \in (0, 100)$) [Samejima 1973]:

$$MLE(\theta_i) = \frac{\sum_{j=1}^m \left[a_j^2 \left(b_j + \frac{z_{ij}}{\alpha_j} \right) \right]}{\sum_{j=1}^m \alpha_j^2}. \quad (2)$$

Biorąc pod uwagę wyniki modeli pomiarowych CRM (tab. 2), należy podkreślić bardzo podobne wartości parametrów trudności i dyskryminacji. Istotnie różne parametry dyskryminacji dotyczą tylko wskaźnika R11, co oznacza brak spełnienia przez ten wskaźnik założenia inwariancji metrycznej w przekroju lat [Cho i in. 2010].

Uwzględniając wartości parametrów oraz relacje między wskaźnikami finansowymi a zmienną ukrytą *theta* (kondycją finansową), można zbudować ranking wskaźników finansowych z punktu widzenia trudności ich osiągnięcia przez firmę. Najłatwiejszym wskaźnikiem (najniższy parametr trudności) jest R16, dla którego firmy o słabej kondycji finansowej (niska wartość parametru *theta*) osiągają wysokie wartości tego wskaźnika. Z drugiej strony wskaźnik R4 jest wskaźnikiem najtrudniejszym, co powoduje, że wysokie wartości tego wskaźnika są osiągalne jedy-

Tabela 2. Parametry modelu Samejimy

Wskaźnik	Parametr dyskryminacji	Parametr trudności	Parametr skalowania
R4_07	0,358	0,097	0,210
R11_07	64,052	-0,031	0,457
R12_07	1,096	-0,053	0,347
R16_07	0,573	-0,084	0,321
R20_07	0,801	-0,007	0,387
R4_09	0,315	0,134	0,203
R11_09	36,705	-0,027	0,447
R12_09	0,962	-0,044	0,307
R16_09	0,430	-0,058	0,356
R20_09	0,784	-0,002	0,349

Źródło: obliczenia własne z wykorzystaniem programu R.

nie przez firmy o silnej kondycji finansowej (wysoka wartość *theta*). Z punktu widzenia siły dyskryminacyjnej wskaźników R11 cechuje się bardzo silnym stopniem dyskryminacji (niewielka zmiana w kondycji finansowej prowadzi do dużej zmiany w wartościach tego wskaźnika), natomiast najsłabszą siłą dyskryminacyjną cechuje się wskaźnik R4 (duże zmiany w wartościach kondycji finansowej wywołują niewielkie zmiany w wartościach wskaźnika).

5. Wielosektorowa analiza LTA

Na podstawie modelu pomiarowego IRT i oszacowanych wartości zmiennych ukrytych (*theta*) dla poszczególnych lat został zbudowany model ukrytych przejść [Lubke, Muthen 2005; Cho i in. 2010]. Model składa się z dwóch kategoryalnych zmiennych ukrytych C1 i C2, reprezentujących poszczególne fale badań (badania w roku 2007 i 2009), składających się z trzech klas ukrytych występujących w każdym z badanych lat⁴. Przed zbudowaniem modelu zostało sprawdzone założenie wielowymiarowej normalności rozkładu zmiennych ukrytych. Wartości statystyk wielowymiarowej skośności i wielowymiarowej kurtozy wskazują, że założenie normalności rozkładu należy odrzucić (rozkład jest leptokurtyczny). Na podstawie analizy rozrzutu punktów stwierdzono jednak, że rozkłady zmiennych ukrytych można uznać za symetryczne. Biorąc także pod uwagę rozmiar badanej próby, podjęto decyzję o budowie modelu ukrytych przejść.

W celu uwzględnienia struktury przepływów między statusami ukrytymi wprowadzona została do modelu zmienna kowariancyjna opisująca rodzaj sektora,

⁴ W modelach LTA kategoryjalne zmienne ukryte (C1, C2) reprezentują nieobserwowalną zmienność wskaźników w dwóch powtarzanych pomiarach. W ramach każdego pomiaru zostały wyodrębnione trzy klasy ukryte („bankruci”, „przejsiowi” i „zdrowi”), które w odróżnieniu od klasycznych modeli klas ukrytych nie zakładają lokalnej niezależności wskaźników, lecz w ramach danej klasy wskaźniki finansowe są składowymi czynnikowego modelu pomiarowego IRT.

do którego należy przedsiębiorstwo⁵. Uwzględnienie tej zmiennej pozwoliło na oszacowanie prawdopodobieństw warunkowych przynależności do klas ukrytych oraz warunkowych prawdopodobieństw ukrytych przejść między statusami w zależności od typu sektora. Na poziomie warunków początkowych (tab. 3) największe prawdopodobieństwa przynależności do klasy „bankrutów” występują w przypadku przedsiębiorstw z sektora spożywczego, elektrycznego i meblarskiego, a do klasy „zdrowych” – niemetalicznego i chemicznego.

Tabela 3. Prawdopodobieństwa warunkowe przynależności do klas ukrytych w 2007 roku

Sektor	Klasa „bankrutów”	Klasa „przejściowych”	Klasa „zdrowych”
CG1 – spożywczy	0,63	0,26	0,11
CG2 – włókienniczy	0,54	0,37	0,09
CG3 – drzewny	0,52	0,30	0,18
CG4 – chemiczny	0,51	0,27	0,22
CG5 – niemetaliczny	0,46	0,29	0,25
CG6 – metalowy	0,52	0,33	0,15
CG7 – maszynowy	0,54	0,32	0,13
CG8 – elektryczny	0,61	0,30	0,09
CG9 – transportowy	0,56	0,30	0,14
CG10 – meblarski	0,63	0,18	0,19

Źródło: obliczenia własne z wykorzystaniem programu Mplus.

Tabela 4. Prawdopodobieństwa ukrytych przejść między statusami

Prawdopodobieństwa przejść	C21	C22	C23
C11	0,93	0,05	0,02
C12	0,08	0,81	0,10
C13	0,05	0,05	0,90

Uwaga: C11 – pierwsza klasa ukryta w 2007 roku, C12 – druga klasa ukryta w 2007 roku, C13 – trzecia klasa ukryta w 2007 roku, C21 – pierwsza klasa ukryta w 2009 roku, C22 – druga klasa ukryta w 2009 roku, C23 – trzecia klasa ukryta w 2009 roku.

Źródło: obliczenia własne z wykorzystaniem programu Mplus.

Z punktu widzenia efektów dynamicznych obliczone zostały prawdopodobieństwa ukrytych przejść (tab. 4), które wskazują na wysokie prawdopodobieństwa pozostawania w klasach ukrytych (szczególnie dla klasy „bankrutów” $p = 0,93$). Względnie wyższe prawdopodobieństwa przejść występują dla zmian położenia z klasy „przejściowej” do „zdrowej” ($p = 0,10$) i z „przejściowej” do „bankrutów” ($p = 0,08$).

⁵ Pojęcie ukrytego statusu (*latent statuses*) jest charakterystyczne dla analizy LTA. Odzwierciedla ono teoretyczną liczebność analizowanych obiektów (estymowaną na podstawie modelu), które zmieniają swoją przynależność z danej klasy ukrytej w czasie t_1 do innej (lub tej samej) klasy ukrytej w czasie t_2 .

6. Podsumowanie

W czasach postępującej globalizacji procesów społeczno-gospodarczych zmienia się liczba, intensywność i rozmiar występowania różnego rodzaju zjawisk utrudniających lub uniemożliwiających prowadzenie działalności gospodarczej. Jednym z tych zjawisk jest zagrożenie upadłością przedsiębiorstw. Negatywne konsekwencje bankructwa powodują, że wiedza o zagrożeniu upadłością przedsiębiorstwa jest pilnie poszukiwana przez praktykę gospodarczą. Możliwość przewidywania pogorszenia się kondycji finansowej przedsiębiorstwa lub nawet jego bankructwa pozwala na ewentualne podjęcie przez zarząd firmy działań naprawczych, które nie dopuszczają do zrealizowania się negatywnego dla przedsiębiorstwa scenariusza.

Prezentowane badanie wpisuje się w ciągle aktualny nurt poszukiwań narzędzi statystycznych służących do modelowania i prognozowania zagrożenia upadłością przedsiębiorstw.

W pracy zaproponowano wykorzystanie podejścia probabilistycznego do modelowania tytułowego problemu. Zastosowanie analizy ukrytych przejść pozwoliło oszacować prawdopodobieństwa warunkowe przynależności przedsiębiorstw do trzech klas ukrytych, odzwierciedlających stopień zagrożenia upadłością przedsiębiorstw w 2007 roku oraz prawdopodobieństwa ukrytych przejść między stanami w latach 2007 i 2009. Uwzględnienie w analizie typu prowadzonej działalności gospodarczej umożliwiło wskazanie sektorów bardziej lub mniej zagrożonych upadłością w badanych latach.

Literatura

- Altman E.I. (1968), *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*, „The Journal of Finance”, September, Vol. XXIII, No. 4, s. 589-609.
- Aziz M.A., Dar H.A. (2006), *Predicting corporate bankruptcy: where we stand?*, „Corporate Governance”, Vol. 6, No. 1, s. 18-33.
- Bellovary J., Giacomin D., Akers M. (2007), *A review of bankruptcy prediction studies: 1930 to present*, „Journal of Financial Education”, Winter, Vol. 33.
- Cho S.J., Cohen A.S., Kim S.H., Bottge B. (2010), *Latent transition analysis with a mixture item response theory measurement model*, „Applied Psychological Measurement”, 2010, 34(7), s. 483-504.
- Lubke G. H., Muthen B. (2005), *Investigating Population Heterogeneity With Factor Mixture Models*, „Psychological Methods”, 10(1), s. 21-39.
- McKee T.E. (2000), *Developing a bankruptcy prediction model via rough sets theory*, „International Journal of Intelligent Systems in Accounting, Finance and Management”, No. 9, s. 159-173.
- Pawełek B., Pocięcha J. (2012), *General SEM Model in Researching Corporate Bankruptcy and Business Cycles*, [w:] *Data Analysis Methods and Its Applications*, red. J. Pocięcha, R. Decker, C.H. Beck, Warsaw, s. 215-231.
- Pawełek B., Pocięcha J., Sagan A. (2013), *Latent Transitions with Mixture Rasch Model of Bankruptcy Risk in the Classification of Polish Firms*, [w:] *Program and Book of Abstracts*, Proceedings of Conference of the International Federation of Classification Societies IFCS, 14-17 July, Tilburg, the Netherlands.

- Samejima F. (1973), *Homogeneous case of the continuous response model*, „Psychometrika”, 38(2), s. 203-219.
- Wang T., Zeng L. (1998), *Item parameter estimation for a continuous response model using an EM algorithm*, „Applied Psychological Measurement”, 22(4), s. 333-343.
- Wędzki D. (2009), *Analiza wskaźnikowa sprawozdania finansowego*, t. 2, wyd. II poprawione i rozszerzone, Oficyna a Wolters Kluwer business, Kraków.
- Zopluoglu C. (2012), *EstCRM: An R package for Samejima's continuous IRT model*, „Applied Psychological Measurement”, 36, s. 149-150.

MULTISECTORAL ANALYSIS OF LATENT TRANSITIONS IN BANKRUPTCY PREDICTION MODELS

Summary: Bankruptcy prediction models are useful tools for the evaluation of the economic situation of the business environment in a country. Business theory and practice formulate various types of bankruptcy prediction. The aim of the paper is the presentation of applicability of multisectoral analysis of latent transitions in bankruptcy prediction models for Polish enterprises which belong to processing industry sector of the economy. Latent transitions modelling is based on the data included in financial ratios for years 2007 (before economic crisis) and 2009 (in time of crisis). It allows to compare bankruptcy risk in different economic environment.

Keywords: latent transitions analysis, multisectoral analysis, bankruptcy risk.