

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

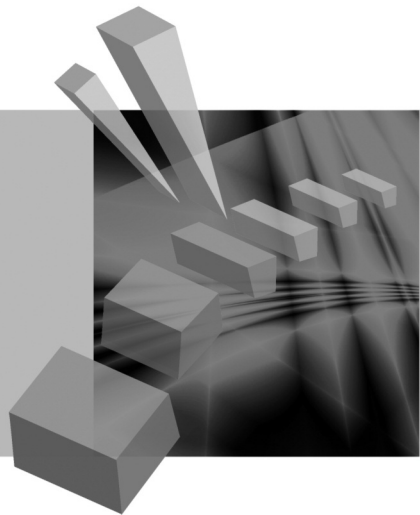
RESEARCH PAPERS

of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pocięcha,
Mirosław Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>
oraz w 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](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

Kopowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2012

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

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Wstęp	13
Stanisława Bartosiewicz , Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej	17
Andrzej Sokolowski , Q uniwersalna miara odległości	22
Eugeniusz Gatnar , Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP)	31
Marek Walesiak , Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania.....	39
Krzysztof Jajuga, Marek Walesiak , XXV lat konferencji taksonomicznych – fakty i refleksje	47
Józef Pocięcha, Barbara Pawelek , Model SEM w analizie zagrożenia bankructwem przedsiębiorstw w świetle koniunktury gospodarczej – problemy teoretyczne i praktyczne	50
Paweł Lula , Uczące się systemy pozyskiwania informacji z dokumentów tekstowych	58
Ewa Roszkowska , Zastosowanie metody TOPSIS do wspomagania procesu negocjacji.....	68
Andrzej Młodak , Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne	76
Andrzej Bąk , Modele kategorii nieuporządkowanych w badaniach preferencji	86
Jacek Kowalewski , Zintegrowany model optymalizacji badań statystycznych.....	96
Jan Paradysz, Karolina Paradysz , Obszary bezrobocia w Polsce – problem benchmarkowy.....	106
Tomasz Szubert , W co grać, aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich.....	116
Izabela Szamrej-Baran , Klasyfikacja krajów UE ze względu na ubóstwo energetyczne	126
Sylwia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , Analiza dojazdów do pracy za pomocą modelu grawitacji.....	135
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych	144
Hanna Dudek , Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów	153

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaročka , Segmentacja szkół wyższych metodą analizy skupień <i>versus</i> konkurencja technologiczna ustalona metodą DEA – studium komparatywne.....	163
Ewa Chodakowska , Wybrane metody klasyfikacji w konstrukcji ratingu szkół.....	173
Bartosz Soliński , Sektor energetyki odnawialnej w krajach Unii Europejskiej – klasyfikacja w świetle strategii zarządzania zmianą.....	182
Krzysztof Szwarz , Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną.....	192
Elżbieta Gołata, Grażyna Dehnel , Rejestry administracyjne w analizie przedsiębiorczości.....	202
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce.....	212
Katarzyna Dębowska , Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i regresji.....	222
Alina Bojan , Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji.....	231
Justyna Brzezińska , Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE.....	240
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Analiza klas ukrytych w badaniach satysfakcji studentów.....	247
Bartłomiej Jefmański , Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska R.....	256
Julita Stańczuk , Porównanie rezultatów wielostanowej klasyfikacji obiektów ekonomicznych z wykorzystaniem analizy dyskryminacyjnej oraz sieci neuronowych.....	265
Jerzy Krawczuk , Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500.....	275
Anna Czapkiewicz, Beata Basiura , Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych na podstawie modelu Copula-GARCH.....	283
Radosław Pietrzyk , Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych.....	291
Aleksandra Witkowska, Marek Witkowski , Zastosowanie metody Panzara-Rosse’a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych.....	306
Marcin Pelka , Podejście wielomodelowe z wykorzystaniem metody <i>boosting</i> w analizie danych symbolicznych.....	315
Justyna Wilk , Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych.....	323

Tomasz Bartłomowicz, Justyna Wilk , Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych.....	333
Kamila Migdał-Najman , Propozycja hybrydowej metody grupowania opartej na sieciach samouczących	342
Dorota Rozmus , Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	352
Krzysztof Najman , Grupowanie dynamiczne z wykorzystaniem samouczących się sieci GNG	361
Małgorzata Misztal , Wpływ wybranych metod uzupełniania brakujących danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności – ocena symulacyjna	370
Mariusz Kubus , Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych.....	380
Barbara Batóg, Jacek Batóg , Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym	387
Katarzyna Wójcik, Janusz Tuchowski , Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej	396
Iwona Staniec , Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach	406
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawelczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami	416
Iwona Foryś , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących atrakcyjność cenową mieszkań w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego.....	426
Ewa Genge , Analiza skupień oparta na mieszankach uciętych rozkładów normalnych.....	436
Jerzy Korzeniewski , Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień	444
Andrzej Dudek , SMS – propozycja nowego algorytmu analizy skupień	451
Artur Mikulec , Metody oceny wyniku grupowania w analizie skupień.....	460
Małgorzata Machowska-Szewczyk , Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych	469
Artur Zaborski , Analiza PROFIT i jej wykorzystanie w badaniu preferencji	479
Karolina Bartos , Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena	488

Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych .	496
Izabela Kurzawa , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto–wieś	505
Aleksandra Łuczak, Feliks Wysocki , Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych	513
Agnieszka Sompolska-Rzechuła , Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim	523
Joanna Banaś, Małgorzata Machowska-Szewczyk , Ocena intensywności wykorzystania skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego	532
Iwona Bąk , Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę	541
Aneta Becker , Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010	552
Katarzyna Dębowska , Klasyfikacja sektorów ze względu na ich kondycję finansową przy użyciu metod wielowymiarowej analizy statystycznej	562
Anna Domagała , Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru w przód).....	571
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej	580
Hanna Gruchociak , Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej.....	590
Tomasz Klimanek, Marcin Szymkowiak , Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy	601
Jarosław Lira , Prognozowanie opłacalności produkcji żywca wieprzowego w Polsce	610
Christian Lis , Wykorzystanie metody klasyfikacji w ocenie konkurencyjności portów południowego Bałtyku	619
Beata Bieszk-Stolorz, Iwona Markowicz , Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych .	628
Lucyna Przezbórska-Skobiej, Jarosław Lira , Przestrzeń agroturystyczna Polski i ocena jej atrakcyjności.....	637
Paweł Ulman , Model rozkładu wydatków a funkcje popytu.....	646
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Zastosowanie metod analizy statystycznej w badaniach mięczaków	655

Summaries

Stanisława Bartosiewicz , The effects of subjectivism in multivariate analysis revisited.....	21
Andrzej Sokółowski , Q universal distance measure	30
Eugeniusz Gatnar , Data quality in central banks' statistical systems (NBP example)	38
Marek Walesiak , Distance measures for ordinal data – strategies of proceedings.....	46
Krzysztof Jajuga, Marek Walesiak , XXV years of taxonomic conferences – some facts and remarks.....	49
Józef Pocięcha, Barbara Pawelek , General SEM model in researching corporate bankruptcy and business cycles – theoretical and practical problems.....	57
Paweł Lula , Learning-based systems of information extraction from textual resources	67
Ewa Roszkowska , The application of the TOPSIS method to support the negotiation process	75
Andrzej Młodak , Neighborhood of spatial areas in the physical and socio-economic context – a taxonomic approach.....	85
Andrzej Bąk , Models for unordered categories in preference analysis.....	95
Kowalewski Jacek , An integrated model of optimizing statistical surveys	105
Jan Paradysz, Karolina Paradysz , Areas of unemployment in Poland – benchmark problem	115
Tomasz Szubert , How to play to lose the least? Classification of systems in sports bets	125
Izabela Szamrej-Baran , Classification of EU member states in view of fuel poverty	134
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , An attempt to use the gravity model in the analysis of commuters.....	143
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Subsistence minimum versus factors influencing tendency to benefit from social care. Classification of households	152
Hanna Dudek , Subjective equivalence scales – analysis based on data about satisfaction with incomes.....	162
Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka , Segmentation of universities using cluster analysis versus technological competitors determined by the DEA method – a comparative study	172
Ewa Chodakowska , Selected methods of classification in schools' rating.....	181
Bartosz Soliński , Renewable energy sector in the European Union – classification in the light of change management strategy	191
Krzysztof Szwarc , Classification of Wielkopolska voivodeship due to the demographic situation	201

Elżbieta Gołata, Grażyna Dehnel , Administrative registers in business analysis.....	211
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Application of taxonomic methods in forecasting the profitability ratios of listed banks in Poland.....	221
Katarzyna Dębowska , Modeling bankruptcy of firms by using discrimination and regression methods.....	230
Alina Bojan , Identification of variables which influence attractiveness of given investments with the usage of multivariate analysis.....	239
Justyna Brzezińska , Log-linear analysis in the study of mortality in EU.....	246
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Latent class analysis in student satisfaction surveys.....	254
Bartłomiej Jefmański , The respondent's opinions measurement in the R program with an application of fuzzy sets theory.....	264
Julita Stańczuk , A comparison of the results of multistate classification of economic objects using discriminant analysis and artificial neural networks.....	274
Jerzy Krawczuk , Effectiveness of classification methods in S&P500 stock index direction changes forecasting.....	282
Anna Czapkiewicz, Beata Basiura , The simulation study of the utility of the Copula-GARCH models for clustering financial time series.....	290
Radosław Pietrzyk , Timing and selectivity in mutual funds performance measurement.....	305
Aleksandra Witkowska, Marek Witkowski , Use of the Panzar-Rosse method to assess of the competition level in the cooperative banks sector.....	314
Marcin Pelka , Ensemble learning with the application of <i>boosting</i> in symbolic data analysis.....	322
Justyna Wilk , Comparative study of symbolic data classification software.....	332
Tomasz Bartłomowicz, Justyna Wilk , Application of symbolic data analysis methods for domain database searching.....	341
Kamila Migdał-Najman , A proposal of hybrid clustering method based on self-learning networks.....	351
Dorota Rozmus , Comparison of accuracy of spectral clustering and cluster ensembles stability based on bagging idea.....	360
Krzysztof Najman , A dynamic grouping based on self-learning GNG networks.....	369
Małgorzata Misztal , Influence of data imputation methods on the results of object classification using classification trees in the case of small data sets – simulation assessment.....	379
Mariusz Kubus , The application of pre-conditioning of explanatory variable for feature selection.....	386
Barbara Batóg, Jacek Batóg , Application of discriminant analysis to the identification of factors determining the rate of return on the capital market.....	395

Katarzyna Wójcik, Janusz Tuchowski , Comparative analysis of text documents similarity measures based on frequency matrix and based on domain knowledge.....	405
Iwona Staniec , Factor analysis in the identification of areas that determine the improvement of management systems in Polish organizations.....	415
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Comparative analysis of selected data mining approaches to the classification of medical data with missing values (covariates).....	425
Iwona Foryś , The log-linear analysis using to select the factors determining the attractiveness of the price of flats on the secondary market on the example of local housing market.....	435
Ewa Genge , Trimming approach to the mixtures of normal distributions.....	443
Jerzy Korzeniewski , Efficiency assessment of Ichino method and mean value method of selecting variables in cluster analysis.....	450
Andrzej Dudek , SMS – proposal of new clustering algorithm.....	459
Artur Mikulec , Evaluation methods for the grouping result in cluster analysis.....	468
Małgorzata Machowska-Szewczyk , Fuzzy clustering algorithm for objects described by symbolic or fuzzy variables.....	478
Artur Zaborski , PROFIT analysis and its using in the research of preferences.....	487
Karolina Bartos , Cluster analysis of selected countries due to the structure of their citizens' consumer expenditures – the use of Kohonen networks.....	495
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Classification of households according to the impulses of concluding the insurance contract by means of qualitative variable models.....	504
Izabela Kurzawa , The application of LA/AIDS model to examine price elasticities of demand of households in the urban-rural relationship.....	512
Aleksandra Luczak, Feliks Wysocki , Linear ordering methods of objects described by a set of metric and ordinal characteristics.....	522
Agnieszka Sompolska-Rzechuła , The comparison of the classical and positional taxonomic analysis of the quality of life differentiation in Zachodniopomorskie voivodeship.....	531
Joanna Banaś, Małgorzata Machowska-Szewczyk , Evaluation of intensity of mailboxes using with the ordered probit model.....	540
Iwona Bąk , Segmentation of pensioners and annuitants households in terms of expenditures on recreation and culture.....	551
Aneta Becker , Application of ANP method to organize Polish voivodships in terms of dynamics of the use of ICT in 2008-2010.....	561
Katarzyna Dębowska , The classification of sectors' financial situation using the methods of multivariate statistical analysis.....	570

Anna Domagała , Proposal of a new method for variable selection in DEA models (combined forward stepwise selection method).....	579
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Statistical analysis in demand research of ICT services in mobile networks.....	589
Hanna Gruchociak , Construction of regression estimator for two-level data	600
Tomasz Klimanek, Marcin Szymkowiak , Application of spatial models in indirect estimation of some labor market characteristics	609
Jarosław Lira , Forecasting of hog livestock production profitability in Poland	618
Christian Lis , The utilization of taxonomic methods in the appraisal of competitiveness of south Baltic ports	627
Beata Bieszk-Stolorz, Iwona Markowicz , The application of the multinomial logit model in evaluating employment odds for the unemployed job seekers	636
Lucyna Przezbórska-Skobiej, Jarosław Lira , Agritourism space of Poland and its valuation.....	645
Paweł Ulman , Model of expenses distribution and demand functions.....	654
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Methods of statistical analysis in research of molluscs	663

Katarzyna Chudy, Marek Sobolewski, Kinga Stępień

Politechnika Rzeszowska

WYKORZYSTANIE METOD TAKSONOMICZNYCH W PROGNOZOWANIU WSKAŹNIKÓW RENTOWNOŚCI BANKÓW GIEŁDOWYCH W POLSCE*

Streszczenie: Badania dotyczące rentowności banków wymagają zgromadzenia odpowiednich danych finansowych. Informacje na temat działalności banków są prezentowane w różnych źródłach i różnych okresach sprawozdawczych. Główny problem, jaki występuje na tym etapie, to dostęp do ujednoliconych i aktualnych danych. Rozwiązaniem tego problemu może być zastosowanie metod prognostycznych bazujących na danych historycznych. Celem artykułu jest skonstruowanie prognozy wartości wskaźników rentowności ROA i ROE dla banków działających na GPW w Warszawie na lata 2012 i 2013. Analizą zostało objętych 14 banków giełdowych. Prognoza została ustalona dla **grup banków** (metoda Warda), które w 2007 i 2011 r. charakteryzowały się podobnymi wartościami wskaźników ROA i ROE.

Słowa kluczowe: rentowność, prognozowanie, taksonomia.

1. Wstęp

Celem artykułu jest podjęcie próby ustalenia prognozy wartości wskaźników rentowności ROA i ROE dla banków działających na GPW w Warszawie na lata 2012 i 2013. Analizą objęto 14 banków: PEKAO, PKO BP, Getin Noble Bank¹, BRE, Millennium, Kredyt Bank, BPH, ING Bank Śląski, Bank Handlowy w Warszawie, BZ WBK, Nordea, BOŚ i DZ Bank, BNP Paribas Bank².

Zazwyczaj prognozy ustalane są dla wskaźników dotyczących konkretnych przedsiębiorstw (w tym przypadku banków). W niniejszym artykule prognoza zostanie ustalona dla grup banków, które w 2007 i 2011 r. charakteryzowały się podobnymi wartościami wskaźników ROA i ROE. W pierwszym etapie przeprowa-

* Praca naukowa finansowana ze środków na naukę w latach 2007-2011 w ramach projektu badawczego „Kompleksowa analiza produktywności banków w Polsce – wyodrębnienie klastrów bankowych” (2602/B/H03/2007/33).

¹ W styczniu 2010 r. w wyniku połączenia Getin Banku i Noble Banku powstał Getin Noble Bank SA.

² W maju 2009 r. Fortis Bank został przyłączony do grupy BNP Paribas. W maju 2011 r. nazwę banku zmieniono na BNP Paribas Bank Polska.

dzone zostało grupowanie banków ze względu na przyjęte wskaźniki. Następnie po ustaleniu liczności i składu powstałych skupień został wyodrębniony średni trend. W etapie końcowym wykonano prognozę dla poszczególnych grup ze względu na wskaźniki ROA i ROE.

2. Metodologia

Do ustalenia liczności i składu skupień banków podobnych do siebie pod względem wskaźników rentowności ROA i ROE w latach 2007 i 2011 została użyta taksonomiczna metoda Warda. Do konstrukcji prognozy wykorzystano metody wygładzania wykładniczego – modelu Holta oraz trendu gasnącego.

Taksonomiczne metody hierarchiczne są jednymi z najczęściej wykorzystywanych w praktyce procedur grupowania. Prowadzą do wyodrębnienia pełnej hierarchii skupień z monotonicznie wzrastającym współczynnikiem ich podobieństwa. Uzyskiwane grupy wyższego rzędu zawierają w sobie rozłączne grupy niższych poziomów. W zależności od techniki grupowania wyróżnia się procedury aglomeracyjne i podziałowe. W metodach aglomeracyjnych wychodzi się z założenia, że każda jednostka stanowi początkowo odrębną grupę, a następnie w sposób sekwencyjny zmniejsza się sukcesywnie liczbę istniejących grup przez ich łączenie w grupy wyższego rzędu. Postępowanie kończy się w momencie otrzymania jednej grupy obejmującej wszystkie jednostki. Zaletą tej metody jest możliwość przedstawienia wyników klasyfikacji w zwartej formie graficznej za pomocą drzewa połączeń (dendrogramu). Podziału dendrogramu proponuje się dokonywać w miejscach o najmniejszej gęstości połączeń dendrogramu, tzn. odcinając najdłuższe gałęzie drzewa [Grabiński 1992]. Zazwyczaj dąży się do takiego podziału, w rezultacie którego z jednej strony liczba połączeń nie jest zbyt duża, a z drugiej strony skupienia nie zawierają zbyt dużej liczby obiektów. Zbyt duża liczba klas utrudnia wyłonienie charakterystycznych właściwości powstałych skupień. Gdy skupienie zawiera zbyt wiele obiektów, istnieje niebezpieczeństwo, że znajdą się w nim obiekty, które różnią się od siebie w sposób istotny [Ostasiewicz 1999]. Do przedstawienia wyników badań posłużono się metodą Warda. W metodzie tej odległość skupisk wyrażona jest różnicą pomiędzy sumami kwadratów odchyleń odległości poszczególnych jednostek od środka ciężkości grup, do których punkty te należą. Odległości między obiektami zostały wyznaczone za pomocą kwadratu odległości euklidesowej [Walesiak 1996].

Istota wygładzania wykładniczego polega na tym, że szereg czasowy zmiennej prognozowanej wygładza się za pomocą średniej ruchomej ważonej, a wagi ustalone są według funkcji wykładniczej. Prognoza oparta jest na średniej ważonej aktualnych i historycznych wartości szeregu [Aczel 2000]. Wygładzanie wykładnicze może być oparte na różnych modelach odpowiednich do rodzaju składowych analizowanego szeregu czasowego. Kluczowym etapem jest wybór wartości parametrów modelu. W literaturze przedmiotu podaje się kilka sposobów wyboru parametrów [Dittmann 1998; Cieślak 2008]. Jeśli poszczególne składowe zmieniają się szybko, to uważa

się, że wartości parametrów wygładzania należy ustalić na poziomie bliskim jedności, w przeciwnym wypadku na poziomie bliskim zera. Wartości parametrów można dobrać także metodą eksperymentu, minimalizując wybrane błędy.

W badaniach użyto dwóch modeli wygładzania: modelu Holta oraz trendu gasnącego. Model Holta znajduje zastosowanie w przypadku, gdy w szeregu czasowym występują: trend oraz wahania przypadkowe. Jest to model dwurównaniowy, w którym do opisu tendencji używa się wielomianu stopnia pierwszego. Model jest bardziej elastyczny ze względu na występowanie dwóch parametrów wygładzania: α oraz β .

$$F_{t-1} = \alpha y_{t-1} + (1 - \alpha)(F_{t-2} + S_{t-2}),$$

$$S_{t-1} = \beta (F_{t-1} - F_{t-2}) + (1 - \beta)S_{t-2},$$

gdzie: F_{t-1} – służy do wyznaczenia wygładzonych wartości szeregu czasowego,
 S_{t-1} – służy do wyznaczenia wygładzonych wartości przyrostu trendu,
 α, β – parametry modelu o wartościach z przedziału $[0; 1]$.

Model wygładzania wykładniczego z trendem gasnącym znajduje zastosowanie w przypadku, gdy w szeregu czasowym występuje tendencja o charakterze malejącym z wahaniami przypadkowymi. Oprócz standardowych parametrów wygładzania α i β w modelu występuje jeszcze parametr ϕ odpowiedzialny za wygładzanie wartości malejących (gasnących). W niniejszym opracowaniu większość wskaźników rentowności w analizowanych bankach ma właśnie charakter spadkowy (malejący, gasnący), stąd uzasadniony jest wybór tego właśnie modelu.

$$F_{t-1} = \alpha y_{t-2} + (1 - \alpha)(F_{t-2} \cdot S_{t-2})\phi,$$

$$S_{t-1} = \beta \left(\frac{F_{t-1}}{F_{t-2}} \right) + (1 - \beta)S_{t-2}\phi,$$

gdzie: F_{t-1} – wygładzona wartość zmiennej prognozowanej na okres lub moment $t-1$,
 S_{t-1} – wygładzona wartość przyrostu trendu na okres lub moment $t-1$,
 α, β, Φ – parametry wygładzania o wartościach z przedziału $[0, 1]$.

3. Wyniki badań

3.1. Ustalenie składu i liczności skupień banków giełdowych w 2011 r.

W celu ustalenia składu i liczności skupień banków w 2011 r. do analizy wzięto wskaźniki rentowności ROA i ROE zarówno z 2011 r., jak i wspomagająco z 2007 r. Rok 2007 charakteryzował się w większości banków maksymalnymi wartościami wskaźników w przyjętym do analizy zakresie danych szeregów czasowych obejmujących

lata 1997-2011, dla których w kolejnym etapie dokonana zostanie prognoza. Rok 2007 był ostatnim rokiem przed kryzysem i do tego czasu szeregi czasowe wskaźników ROA i ROE wykazywały tendencję wzrostową. W latach kryzysu wartości wskaźników znacznie się obniżyły do roku 2009, a w niektórych bankach spadek wartości wskaźników trwał do 2010 r. Jednak od 2009 r. ponownie wkroczyły na ścieżkę wzrostu, stąd przesłanka, aby do grupowania wprowadzić także wielkości wskaźników rentowności sprzed lat kryzysowych. Pozwoli to na bardziej dokładną i wrażliwą na zmiany analizę skupień, która uwzględni także różnice pomiędzy wskaźnikami z lat 2007 i 2011.

Analizując podstawowe charakterystyki opisowe wskaźników rentowności, zaobserwować można spadek ich wartości w roku 2011 w stosunku do roku 2007 (tab. 1).

Tabela 1. Podstawowe statystyki opisowe wskaźników ROA i ROE

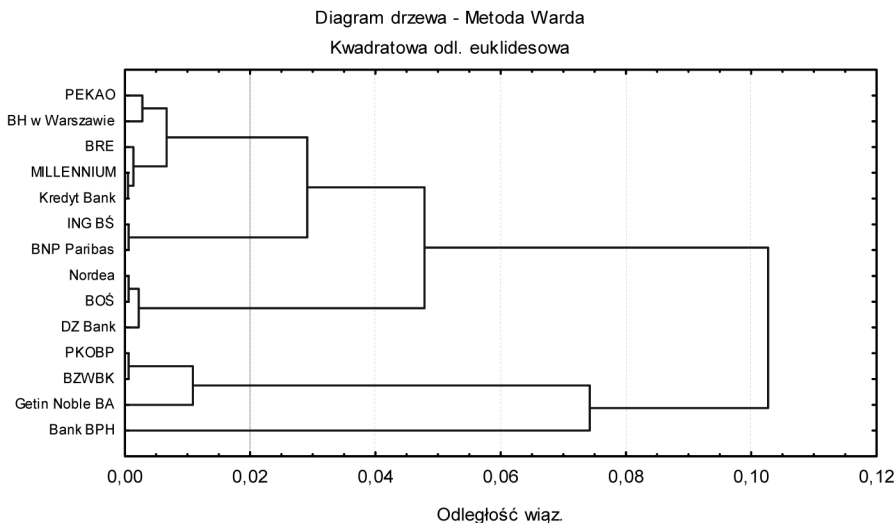
	Średnia	Me	Min	Maks	Odch. std.	Skośność	Vz
ROE 2011	10,6%	10%	1%	26,0%	6,8%	0,01	63,9%
ROE 2007	20,9%	19,9%	5,4%	36,9%	9,7%	0,19	46,7%
ROA 2011	2,1%	1,0%	0,0%	12%	3,0%	0,03	143,5%
ROA 2007	1,8%	1,6%	0,6%	4,0%	0,9%	0,82	50,9%

Źródło: opracowanie własne.

W przypadku wskaźnika ROE średnia wartość obniżyła się z poziomu ok. 21 do 11%, natomiast wskaźnika ROA o ok. 0,03%. Wartość minimalna obu omawianych wskaźników w 2007 r. była dodatnia, tylko w 2011 wskaźnik ROA w BOŚ Banku był równy 0. Współczynnik zmienności w 2011 r. przyjmuje duże wartości, co świadczy o znacznym zróżnicowaniu wskaźników rentowności i wskazuje na fakt, że banki w bardzo odmienny sposób reagowały na zmiany w gospodarce i wkroczenie w fazę kryzysu. Rozkłady wskaźników zarówno w 2007, jak i w 2011 r. wykazywały asymetrię prawostronną, co oznaczało, że większość badanych banków posiadała wskaźniki rentowności na poziomie niższym niż średnia.

Dla wybranych wskaźników rentowności ROA i ROE w latach 2007 i 2011 przeprowadzono grupowanie, którego celem było wyodrębnienie skupień banków o podobnych wielkościach wskaźników rentowności w 2007 i 2011 r. Za pomocą jednolitych skupień możliwe było skonstruowanie prognozy wskaźników rentowności dla każdej z grup. Wyniki grupowania przedstawia diagram drzewa połączeń (rys. 1).

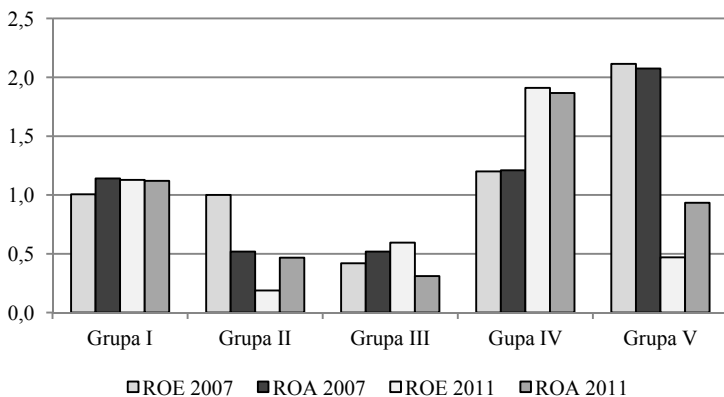
Wyodrębniono pięć skupień banków pod względem poziomu rentowności. Ich skład jest następujący: grupa I, najliczniejsza, obejmuje – PEKAO, BH w Warszawie, BRE Bank, Millennium Bank oraz Kredyt Bank. W skład grupy II weszły ING Bank Śląski oraz BNP Paribas. Najliczniejsze skupienie stanowi grupa III, do której zaklasyfikowały się: Nordea, BOŚ i DZ Bank. Grupa IV to PKO BP, BZ WBK oraz Getin Noble Bank. Ostatnie jednoelementowe skupienie stanowi Bank BPH.



Rys. 1. Wyniki grupowania banków giełdowych pod względem wskaźników rentowności

Źródło: opracowanie własne.

Na rysunku 2 przedstawiono wyniki średnich grupowych dla powstałych skupień banków. Najlepszą kondycją pod względem wskaźników rentowności charakteryzuje się grupa IV, dla której wskaźniki zarówno ROA, jak i ROE w latach 2007 i 2009 miały wartości największe.



Rys. 2. Wyniki średnich grupowych powstałych skupień banków giełdowych w 2011 r.

Źródło: opracowanie własne.

Wysokie wartości wskaźników rentowności wykazują także banki, które zostały sklasyfikowane do grupy I. Zarówno w 2007, jak i w 2011 r. są one na średnim

poziomie, można więc stwierdzić, że grupę tę charakteryzuje stabilność. Grupy II i III składają się z banków, których wskaźniki kształtowały się na poziomie niższym od średniego. W jednoelementowym skupieniu (grupa V) utworzonym przez Bank BPH w 2007 r. wskaźniki ROE i ROA były na najwyższym poziomie, natomiast w 2011 r. kształtowały się poniżej średniej. Charakter grupy będzie miał wpływ na sporządzanie prognozy dla poszczególnych skupień.

3.2. Prognoza wartości wskaźników rentowności ROE i ROA

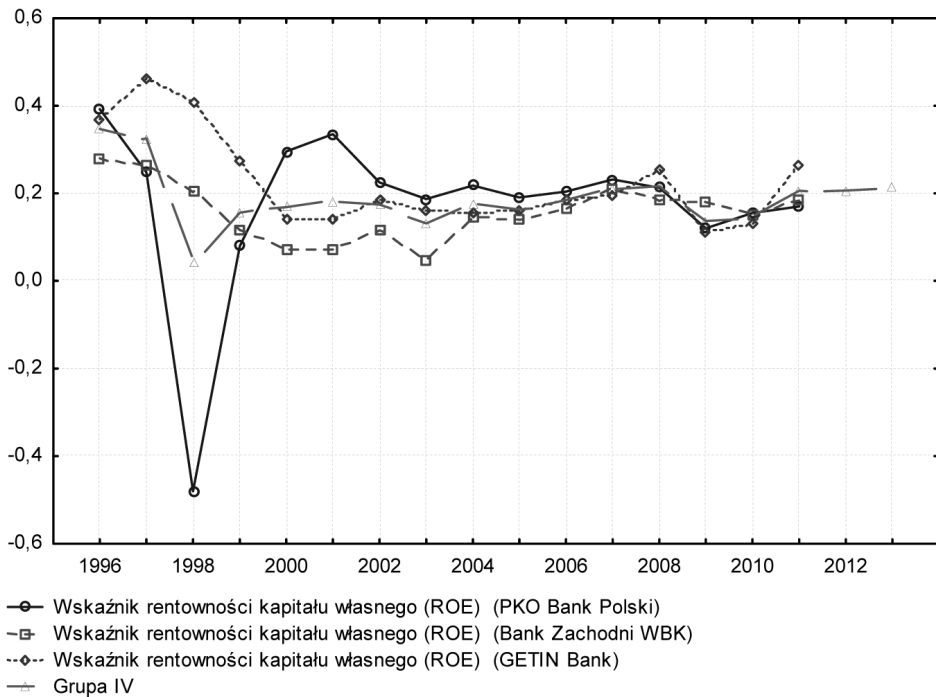
Po ustaleniu liczności i składu skupień możliwa była prognoza uśrednionych wskaźników rentowności dla każdego ze skupień. Banki wchodzące w skład wyodrębnionych skupień charakteryzują się podobną wielkością wskaźników rentowności, a szeregi czasowe dotyczące wskaźników mają podobną tendencję w badanym okresie od 1996 do 2011 r. Dla każdego skupienia dokonano prognozy uśrednionych wskaźników ROA i ROE. Wartości prognozy dla kolejnych lat 2012 i 2013 przedstawiono w tab. 2.

Tabela 2. Prognozowane wartości wskaźników rentowności ROA i ROE na lata 2012 i 2013 wraz z przyjętymi do prognozowania modelami i parametrami

	Wskaźnik	Model	Parametry	Średni błąd bezwzględny	Prognoza		Znak prognozy	Tendencja prognozy	
					2012	2013			
Grupa I	ROE	Wykładanie z trendem gasnącym	$\alpha=0,9; \beta=0,5; \phi=0,2$	0,0886	10,97%	11,07%	+	↑	
	ROA		$\alpha=0,9; \beta=0,1; \phi=0,2$	0,0091	1,087%	1,088%	+	⇔	
Grupa II	ROE		$\alpha=0,7; \beta=0,1; \phi=0,7$	0,0518	2,65%	2,48%	+	↓	
	ROA		$\alpha=0,9; \beta=0,9; \phi=0,6$	0,0055	0,829%	0,828%	+	⇔	
Grupa III	ROE		$\alpha=0,9; \beta=0,9; \phi=0,7$	0,0254	5,37%	4,63%	+	↓	
	ROA		$\alpha=0,7; \beta=0,9; \phi=0,7$	0,0032	0,6967%	0,6995%	+	↑	
Grupa IV	ROE		Model wykładniczy	$\alpha=0,9; \beta=0,1$	0,0442	20,66%	21,35%	+	↑
	ROA			$\alpha=0,9; \beta=0,2$	0,0035	2,19%	2,34%	+	↑
Grupa V	ROE		Wykładanie z trendem gasnącym	$\alpha=0,9; \beta=0,1; \phi=0,7$	0,0808	3,64%	3,48%	+	↓
	ROA			$\alpha=0,1; \beta=0,1; \phi=0,9$	0,0167	2,00%	1,98%	+	↓

Źródło: opracowanie własne.

Na rysunku 3 zobrazowano konstrukcję prognozy wskaźnika ROE dla grupy IV.



Rys. 3. Prognoza wskaźnika rentowności ROE dla grupy IV na rok 2012 i 2013

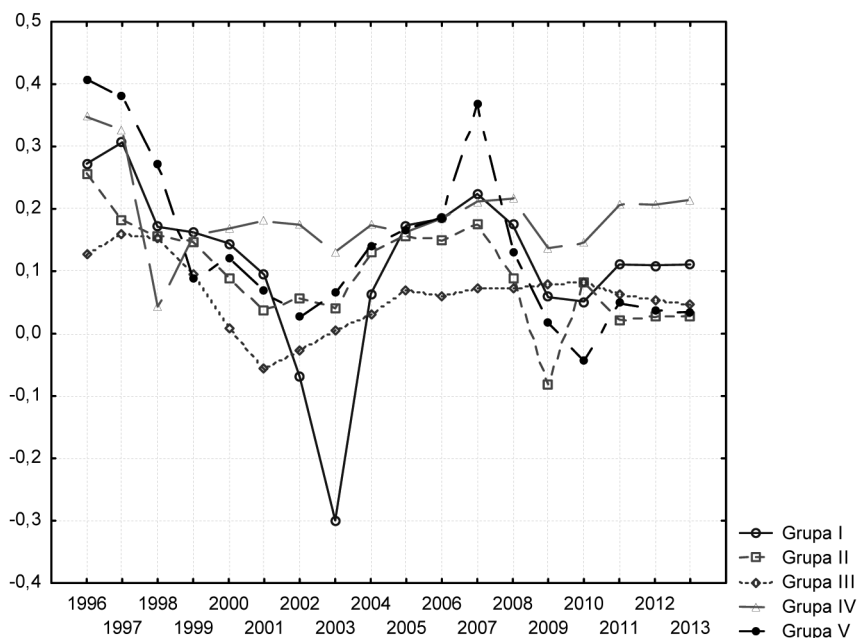
Źródło: opracowanie własne.

Grupa IV charakteryzowała się najlepszymi wynikami prognoz wskaźników rentowności. Prognoza wartości wskaźnika ROE dla tej grupy wykazuje niewielki wzrost w latach 2012 (20,66%) i 2013 (21,35%). Do prognozy użyto modelu wykładniczego z trendem wykładniczym $\alpha = 0,9$ i $\beta = 0,1$. Dla tych parametrów model wykazywał najlepsze dopasowanie do danych (PKO BP, BZ WBK oraz Getin Noble Bank).

Optymistyczne są także prognozy wskaźnika rentowności ROE dla grupy I, która stanowi najliczniejsze skupienie banków. Prognozy charakteryzują się dodatnimi wartościami oraz niewielką tendencją wzrostową. W pozostałych grupach prognozy wskaźnika ROE mają tendencję spadkową.

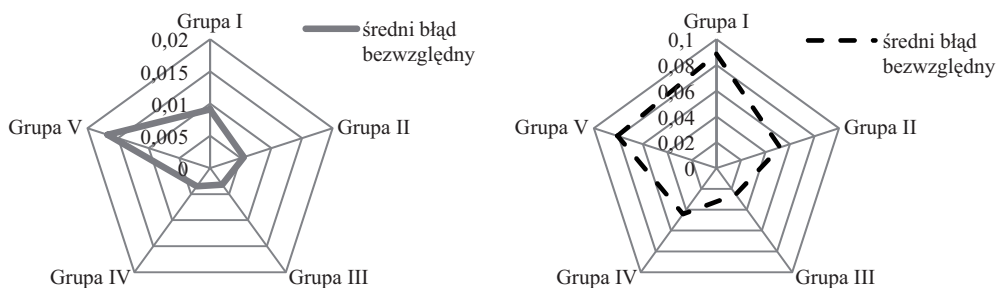
Najgorsze prognozy (największy spadek, ok. 0,74% w 2012 r. (5,37%) i w 2013 r. (4,63%)) odnotowano w grupie III, czyli dla banków Nordea, BOŚ i DZ Bank. W pozostałych grupach spadek wynosi ok. 0,16%.

Najmniejszym błędem charakteryzuje się model skonstruowany dla grupy III, natomiast największy błąd modelu wystąpił w przypadku grupy I (rys. 5).



Rys. 4. Prognoza wskaźników rentowności ROE dla powstałych grup banków dla lat 2012 i 2013

Źródło: opracowanie własne.



Rys. 5. Porównanie średnich błędów bezwzględnych dla wybranych modeli wskaźników ROA i ROE w poszczególnych grupach

Źródło: opracowanie własne.

W przypadku wskaźnika ROA we wszystkich grupach wartości prognoz są dodatnie. Największe wartości prognozy wskaźników prezentują grupy IV i V. Jednakże w grupie IV (PKO BP, BZ WBK oraz Getin Noble Bank) prognoza wskaźnika ROA ma charakter wzrostowy, a grupę V (Bank BPH) charakteryzuje spadkowa tendencja wskaźnika. W pierwszych trzech grupach prognozy wskaźnika ROA na

lata 2012 i 2013 nie wykazują większych zmian – utrzymują się na stałym poziomie (z bardzo niewielkimi zmianami).

Najmniejszym błędem charakteryzuje się model skonstruowany dla grupy III, natomiast największy błąd modelu występuje w przypadku grupy I (rys. 5).

4. Podsumowanie

W artykule przeprowadzono konstrukcję prognozy wskaźników ROA i ROE na lata 2012 i 2013 dla wybranych banków giełdowych. Zostały one wcześniej poddane procedurze grupowania ze względu na poziom wskaźników rentowności ROA i ROE z lat 2007 i 2011. Rok 2007 był ostatnim, w którym wskaźniki rentowności banków przyjmowały wartości maksymalne, w kolejnych latach poziom wskaźników zaczął się obniżać, co mogło być wynikiem pogarszania się sytuacji w sektorze bankowym na skutek kryzysu finansowego. Procedura grupowania pozwoliła na wyselekcjonowanie banków o podobnej tendencji i charakterze wskaźników rentowności. Dzięki temu można było zastosować pewne uproszczenie w konstrukcji prognozy. Dla każdej z grup uśredniono wskaźniki rentowności w badanym okresie, a prognoza została skonstruowana dla szeregu uśrednionego. W większości przypadków zastosowany został model wygładzania wykładniczego z tendencją gasnącą, zważywszy na charakter ostatnich lat kryzysowych, gdzie wskaźniki rentowności znacznie się obniżyły. W poszczególnych grupach spadek wskaźników był zróżnicowany. Wartości prognoz wskaźnika ROE i ROA wskazują na to, że w najlepszej kondycji znajdują się banki z grupy IV (PKO BP, BZ WBK oraz Getin Noble Bank). Prognozy mają znak dodatni, a tendencja wykazuje charakter wzrostowy. Zdecydowanie najgorszą sytuację w przypadku wskaźnika ROE prezentują banki z grupy III (Nordea, BOŚ i DZ Bank), gdzie spadek wskaźników był największy i jak wykazują prognozy, będzie się on pogłębiał. Natomiast najgorsze prognozy wskaźnika ROA uwidocznily się w grupie V (Bank BPH). Zastosowane metody opierają się na danych historycznych i nie biorą pod uwagę innych zjawisk gospodarczych wpływających na rozwój gospodarczy, tak więc do otrzymanych prognoz należy podchodzić z dużym dystansem. Proces prognozowania jest obciążony błędami natury zarówno obiektywnej, jak i subiektywnej. Otrzymane prognozy mogą jednak stanowić ważne źródło informacji dla podejmujących decyzje w zakresie rynku finansowego. Przeprowadzone badania mogą dostarczyć również informacji dotyczących słabych i mocnych stron badanego sektora.

Literatura

- Aczel A.D., *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa 2000.
Cieślak M., *Prognozowanie gospodarcze. Metody i zastosowania*, PWN, Warszawa 2008.
Dittmann P., *Metody prognozowania sprzedaży w przedsiębiorstwie*, Wydawnictwo AE, Wrocław 1998.

Grabiński T., *Metody taksonometrii*, Wydawnictwo AE, Kraków 1992.

Kopiński A., *Analiza finansowa banku*, PWE, Warszawa 2008.

Ostasiewicz W., *Statystyczne metody analizy danych*, Wydawnictwo AE, Wrocław 1999.

Walesiak M., *Metody analizy danych marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 1996.

APPLICATION OF TAXONOMIC METHODS IN FORECASTING THE PROFITABILITY RATIOS OF LISTED BANKS IN POLAND

Summary: Studies on the profitability of banks require gathering relevant financial data. Information on the activities of banks is presented in different sources and different reporting periods. The main problem that occurs at this stage is an access to standardized and up-to date data. The solution to this problem can be applied forecasting methods based on historical data. The article aims to construct a forecast of profitability ratios ROA and ROE for banks operating on the Warsaw Stock Exchange for the years 2012 and 2013. The analysis has been covered by the 14 listed banks. The forecast has been established for groups of banks (the Ward method) which in 2007 and 2011 were characterized by similar values of ROA and ROE.

Keywords: profitability, forecasting, taxonomy.