

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

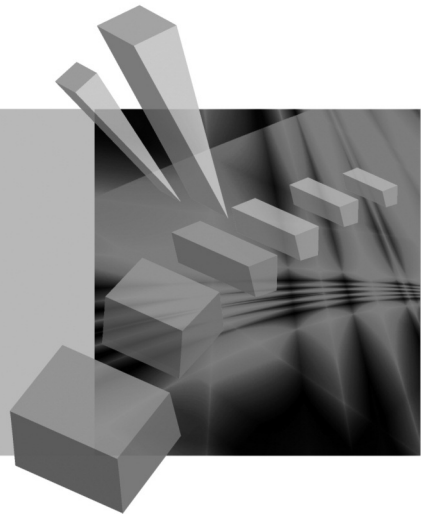
RESEARCH PAPERS

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,
Miroslaw 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 wspomaganie 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 Pawełczyk, 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

Hanna Dudek

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

SUBIEKTYWNE SKALE EKWIWALENTNOŚCI – ANALIZA NA PODSTAWIE DANYCH O SATYSFAKCJI Z OSIĄGANÝCH DOCHODÓW

Streszczenie: W artykule przedstawiono wyniki estymacji skal ekwiwalentności na podstawie danych o subiektywnej ocenie własnej sytuacji dochodowej przez gospodarstwa domowe. W tym celu zastosowano metodę częściowo uogólnionych uporządkowanych modeli logitowych. Analizę przeprowadzono na podstawie danych z badań budżetów gospodarstw domowych zrealizowanych przez GUS w 2009 r. Proponowana metodyka jest rozszerzeniem podejścia zaprezentowanego w publikacji [Schwarze 2003].

Słowa kluczowe: skala ekwiwalentności, dochody, gospodarstwa domowe.

1. Wstęp

Skale ekwiwalentności stanowią podstawowe narzędzie we wszelkich analizach dotyczących ubóstwa oraz nierówności dochodowych gospodarstw domowych o różnej wielkości. Odzwierciedlają one wpływ, jaki na koszty utrzymania gospodarstw ma ich skład demograficzny. Przy ich wyznaczaniu bierze się pod uwagę efekt skali wynikający ze wspólnego zamieszkiwania, gospodarowania i konsumpcji.

W literaturze przedmiotu prezentowane jest stanowisko, że nie ma jednej optymalnej, powszechnie akceptowanej metody szacowania skal ekwiwalentności¹. Do stosunkowo nowych propozycji należy metoda przedstawiona w publikacji [Schwarze 2003], w której wykorzystuje się dane dotyczące subiektywnego postrzegania swojej sytuacji dochodowej. Metodę tę, z pewnymi modyfikacjami, zastosowano w prezentowanym tu opracowaniu.

Podobnie jak w publikacji Schwarzego w celu analizy subiektywnej oceny dochodów gospodarstw domowych w pracy zastosowano skalę potęgową postaci:

$$S = n^e, \text{ gdzie } e \in [0; 1]. \quad (1)$$

¹ Przegląd różnych metod estymacji skal ekwiwalentności zawarto m.in. w pracach [Dudek 2011; Schröder 2009].

Wielu autorów uważa bowiem, że skale potęgowe można wykorzystać do przybliżania różnych typów skal ekwiwalentności [Buhmann i in. 1988; Kot 2000]. Parametr e nazywany jest elastycznością ekwiwalentności. Dzieląc dochód gospodarstwa przez wartość skali (1), uzyskuje się tzw. dochód ekwiwalentny służący do porównań zamożności gospodarstw domowych o różnej liczbie osób. Szczególnie w sytuacji, gdy $e = 0$, otrzymuje się dochód gospodarstwa domowego, a w wypadku gdy $e = 1$ – dochód w przeliczeniu na osobę. Podstawowym problemem napotkanym w zastosowaniach tych skal jest dobór parametru e . W niektórych opracowaniach przyjmuje się arbitralnie pewne wartości, np. w publikacji [Johnson, Smeeding, Torres 2005] wykorzystano skalę, w której $e = 0,5$, argumentując ten wybór kompromisem między skrajnymi wartościami 0 i 1, natomiast w monografii [Radziukiewicz 2006] decyzję o zastosowaniu we wzorze (1) wykładnika potęgi wynoszącego 0,72 uzasadniono autorytetem amerykańskich i europejskich ekspertów. W niniejszej pracy podjęto próbę oszacowania elastyczności ekwiwalentności przez estymację parametrów modelu, w którym zmienną objaśnianą stanowi zmienna wielomianowa odpowiadająca kategoriom uporządkowanym.

Celem pracy jest zbadanie możliwości zastosowania modeli kategorii uporządkowanych do oszacowania estymacji skal ekwiwalentności na podstawie danych o subiektywnych ocenach własnej sytuacji dochodowej przez polskie gospodarstwa domowe. Przedstawiono krytyczną analizę podejścia prezentowanego w pracy [Schwarze 2003] oraz zaproponowano pewne modyfikacje w tym zakresie.

2. Dane

W cytowanej pracy Schwarzego wykorzystano dane panelowe o gospodarstwach domowych. Takie podejście umożliwia uwzględnienie nieobserwowalnej heterogeniczności gospodarstw domowych. Ponadto informacje pozyskane w różnych okresach od tych samych jednostek są bardziej niezależne od chwilowych emocji respondentów niż dane pozyskane w jednym tylko okresie. Niestety, ze względu na brak tego typu danych w Polsce wykorzystano tu dane przekrojowe pochodzące z badań budżetów gospodarstw domowych prowadzonych przez GUS w 2009 r. Analizę przeprowadzono na podstawie informacji z 18 240 gospodarstw domowych pracowników.

W kwestionariuszach stosowanych w badaniach budżetów gospodarstw domowych uwzględniono m.in. pytanie następującej treści: „Czy z aktualnym dochodem Pana(i) gospodarstwo domowe wiąże koniec z końcem: 1) z wielką trudnością, 2) z trudnością, 3) z pewną trudnością, 4) raczej łatwo, 5) łatwo, 6) bardzo łatwo”. Osoba odniesienia (tzn. osoba, która ukończyła 16 lat i osiągnęła najwyższy dochód spośród wszystkich członków gospodarstwa domowego) dokonywała wyboru jednej z powyższych sześciu możliwości. W tabeli 1 przedstawiono podstawowe informacje dotyczące oceny własnych dochodów przez gospodarstwa domowe pracowników w 2009 r.

Tabela 1. Subiektywne oceny własnych dochodów

Z aktualnym dochodem gospodarstwo wiązało koniec z końcem	Kategoria	Odsetek gospodarstw domowych pracowników
Z wielką trudnością	$j = 1$	6,72
Z trudnością	$j = 2$	15,45
Z pewną trudnością	$j = 3$	42,30
Raczej łatwo	$j = 4$	26,89
Łatwo	$j = 5$	6,89
Bardzo łatwo	$j = 6$	1,76

Źródło: obliczenia własne.

Ze względu na to, że odsetek gospodarstw domowych pracowników oceniających, że z aktualnymi dochodami było im bardzo łatwo wiązać koniec z końcem, nie przekraczał 2%, gospodarstwa te połączono z grupą tych, którym łatwo przychodziło wiązanie końca z końcem. Ostatecznie zatem rozpatrywano nie 6, lecz 5 kategorii odpowiedzi.

3. Metoda

W celu oszacowania elastyczności skali ekwiwalentności w publikacji [Schwarze 2003] rozważono model:

$$u^* = \alpha_0 + \alpha_1 \ln y + \alpha_n \ln n + \sum_{k=1}^K \gamma_k S_k, \quad (2)$$

gdzie: u^* – satysfakcja z osiąganego dochodu,
 y – miesięczne dochody gospodarstwa domowego,
 n – liczba osób w gospodarstwie domowym,
 S_k – k -ta zmienna kontrolna², $k = 1, \dots, K$, odnosząca się np. do cech, takich jak: płeć, poziom wykształcenia, wiek, status cywilny, miejsce zamieszkania itp.

$\alpha_0, \alpha_1, \alpha_n, \gamma_1, \gamma_2, \dots, \gamma_K$ – parametry.

Ponieważ $\alpha_1 \ln y + \alpha_n \ln n = \ln \frac{y^{\alpha_1}}{n^{-\alpha_n}} = \alpha_1 \ln \frac{y}{n^{-\alpha_n/\alpha_1}}$, to elastyczność ekwiwa-

lentności e wynosi:

$$e = -\frac{\alpha_n}{\alpha_1}. \quad (3)$$

² W badaniach socjologicznych przez zmienne kontrolne rozumie się dodatkowe pośredniczące zmienne, które objaśniają kształtowanie się analizowanego zjawiska.

Do estymacji parametrów modelu w opracowaniu [Schwarze 2003] wykorzystano metody wielomianowych modeli kategorii uporządkowanych. Ogólna postać tych modeli to:

$$u_i^* = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i, \quad (4)$$

gdzie: u_i^* – zmienna nieobserwowalna odnosząca się i -tej obserwacji,

$\boldsymbol{\beta}$ – wektor kolumnowy parametrów,

\mathbf{x}_i – wektor wierszowy wartości zmiennych objaśniających dla i -tej obserwacji³,

ε_i – składnik losowy dla i -tej obserwacji, $i = 1, 2, \dots, N$, N – liczba obserwacji.

W powyższym podejściu ciągła nieobserwowalna zmienna u^* jest dyskretyzowana przez zbiór wartości progowych $\delta_0, \delta_1, \dots, \delta_m$, co prowadzi do zmiennej skategoryzowanej u określonej jako: $u_i = j$ ⁴, jeśli $\delta_j < u_i^* \leq \delta_{j+1}$, przy czym: $-\infty = \delta_0 < \delta_1 < \dots < \delta_m < \delta_{m+1} = \infty$. Prawdopodobieństwa $P(u_i = j | \mathbf{x}_i)$ w uporządkowanym modelu można określić jako: $P(u_i = j | \mathbf{x}_i) = F(\delta_j - \mathbf{x}_i \boldsymbol{\beta}) - F(\delta_{j-1} - \mathbf{x}_i \boldsymbol{\beta})$, dla $j = 1, 2, \dots, m$, gdzie F jest dystrybuantą składnika losowego ε . W zastosowaniach zwykle wykorzystuje się uporządkowany model logitowy i probitowy, w którym odpowiednio $F(z) = \Lambda(z) = \frac{1}{1 + \exp(-z)}$ oraz

$F(z) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$, gdzie: $z_i = \mathbf{x}_i \boldsymbol{\beta}$. Modele takie są dość restrykcyjne, zakładają bowiem, że:

$$P(u_i \leq j | \mathbf{x}_i) = F(\delta_j - \mathbf{x}_i \boldsymbol{\beta}), \quad (5)$$

czyli parametry przy zmiennych objaśniających nie zależą od kategorii j , $j = 1, 2, \dots, m$. Założenie to bywa nazywane założeniem regresji równoległych. W celu jego weryfikacji w niniejszej pracy zastosowano test zaproponowany w publikacji Branta [1990]. Idea tego testu polega na rozważeniu $m - 1$ modeli dwumianowych, w których zmienne objaśniane zdefiniowane są następująco⁵:

³ Wektor zmiennych objaśniających \mathbf{x} obejmuje zmienne y , n oraz s_1, s_2, \dots, s_K .

⁴ W niniejszej pracy $j = 1, 2, 3, 4, 5$, co odnosi się do odpowiedzi: 1 – z wielką trudnością, 2 – z trudnością, 3 – z pewną trudnością, 4 – raczej łatwo, 5 – łatwo i bardzo łatwo.

⁵ Opis testu Branta podano za publikacjami [Książek 2010; Long 1997].

$$u_j^{**} = \begin{cases} 1, & \text{gd}y\ u > j, \\ 0, & \text{w przeciwnych przypadkach,} \end{cases} \quad (6)$$

gdzie $j = 1, 2, \dots, m - 1$.

Po oszacowaniu łącznej macierzy wariancji i kowariancji parametrów we wszystkich $m - 1$ modelach dwumianowych przeprowadza się testy Walda – łączny oraz indywidualne testy dla poszczególnych zmiennych objaśniających. Pierwszy z testów wykorzystywany jest do weryfikacji hipotezy zerowej mówiącej o równości odpowiednich parametrów we wszystkich $m - 1$ regresjach dwumianowych łącznie dla całego zestawu zmiennych objaśniających⁶. Odrzucenie hipotezy zerowej oznacza, że co najmniej dla jednej zmiennej objaśniającej parametry istotnie różnią się w przynajmniej dwóch modelach dwumianowych, w następstwie czego założenie regresji równoległych nie jest spełnione. Przeprowadzenie indywidualnych testów Walda umożliwi wskazanie zmiennych „odpowiedzialnych” za naruszenie tego założenia. Hipoteza zerowa mówi wtedy o równości parametrów przy poszczególnej zmiennej objaśniającej we wszystkich modelach dwumianowych. W konsekwencji odrzucenia hipotezy zerowej model dla kategorii uporządkowanych nie powinien być stosowany i wtedy należy wykorzystać inne metody szacowania parametrów modelu (4). Do jednych z podejść możliwych do zastosowania w takiej sytuacji należy estymacja uogólnionego modelu uporządkowanego [Greene, Hensher 2010]:

$$P(u_i \leq j | \mathbf{x}_i) = F(\delta_j - \mathbf{x}_i \boldsymbol{\beta}_j), \text{ dla } j = 1, 2, \dots, m - 1. \quad (7)$$

W powyższym modelu dopuszcza się zatem, by parametry przy zmiennych objaśniających zależały od przynależności do kategorii zmiennej porządkowej u odpowiadającej nieobserwowalnej zmiennej objaśnianej. W przypadku, gdy indywidualne testy Walda nie nakazują odrzucenia hipotez o równości parametrów niektórych ze zmiennych objaśniających, możliwe jest zastosowanie częściowo uogólnionych modeli uporządkowanych⁷. W modelach tego typu we wzorze (7) parametry przy niektórych zmiennych objaśniających nie zależą od kategorii analizowanej zmiennej porządkowej. Takie podejście jest z jednej strony mniej restrykcyjne niż zastosowanie modelu uporządkowanego, z drugiej strony zaś – ze względu na uwzględnienie w modelu mniejszej liczby parametrów – bardziej „oszczędne” oraz umożliwiające łatwiejszą interpretację parametrów strukturalnych modelu. W pracy, w odróżnieniu od publikacji [Schwarze 2003], dokonano weryfikacji założenia regresji równoległych i zaproponowano pewne rozwiązania w tym zakresie.

⁶ Statystyka testowa w teście Branta ma asymptotyczny rozkład chi-kwadrat z $p(m - 2)$ stopniami swobody, gdzie p – liczba parametrów przy zmiennych objaśniających, m – liczba kategorii zmiennej porządkowej odpowiadającej nieobserwowalnej zmiennej objaśnianej.

⁷ W anglojęzycznej literaturze przedmiotu modele takie są nazywane mianem *partial proportional odds models for ordinal response variables* [Peterson, Harrell 1990; Williams 2006].

4. Wyniki

Na początku oszacowano model (2) zaproponowany w pracy [Schwarze 2003]. Po wykonaniu łącznego testu Branta stwierdzono, że założenie regresji równoległych powinno być odrzucone. Analiza wyników indywidualnego testu Branta wykazała, że „odpowiedzialność” za naruszenie tego założenia ponosi kilka zmiennych, w tym zmienna odnosząca się do dochodów gospodarstw domowych. W takiej sytuacji zastosowanie uogólnionych lub częściowo uogólnionych modeli uporządkowanych nie umożliwiłoby oszacowania elastyczności ekwiwalentności na podstawie wzoru (3). Dlatego też uwzględniono dodatkowe zmienne stanowiące interakcję między różnymi zmiennymi objaśniającymi. Ostatecznie otrzymano model⁸, którego wyniki weryfikacji zaprezentowano w tab. 2. W charakterze ilościowych zmiennych objaśniających uwzględniono logarytm dochodów, logarytm liczby osób w gospodarstwie, wiek osoby odniesienia oraz pewne interakcje wymienionych cech. Zmienne jakościowe (wymienione w czwartej kolumnie tej tabeli) są zmiennymi binarnymi przyjmującymi wartość 1 dla wariantu wskazanego nazwą oraz 0 w pozostałych przypadkach.

Tabela 2. Wyniki testu Branta dla uporządkowanego modelu logitowego

Zmienna	Wartość statystyki testowej	Wartość p	Zmienna	Wartość statystyki testowej	Wartość p
$\ln y$	5,90	0,117	wykształcenie wyższe	5,63	0,131
$\ln y \cdot \ln n$	4,63	0,201	wykształcenie średnie	5,41	0,144
$(\ln n)^2$	3,72	0,294	stanowisko robotnicze	6,56	0,087
$(\ln n)^3$	2,49	0,477	kobieta	9,62	0,022
wiek	0,38	0,945	związek formalny	12,84	0,005
wiek ²	0,20	0,978	miasto pow. 500 000 mieszkańców	17,04	0,001

Źródło: obliczenia własne w programie Stata v. 10.

Dla całego zestawu zmiennych prezentowanego w tab. 2 wartość statystyki testowej w łącznym teście Branta wyniosła 89,80, co świadczy o odrzuceniu założenia regresji równoległych (wartość krytyczna, odczytana z tablic rozkładu chi-kwadrat, dla 36 stopni swobody i poziomu istotności 0,05 jest równa 51,00). Z informacji przedstawionych w tab. 2 wynika, że dla zmiennych odnoszących się do logarytmu dochodów, wielkości gospodarstwa domowego oraz ich interakcji nie

⁸ Przy wyborze modelu kierowano się kryteriami merytorycznymi i statystycznymi. Wyniki prezentowane w tab. 2-5 uzyskano na podstawie danych dotyczących gospodarstw składających się z co najwyżej 7 osób, których logarytm dochodów kształtował się w granicach $[Q_1 - 2 \cdot IQR; Q_3 + 2 \cdot IQR]$, gdzie Q_1 i Q_3 – pierwszy i trzeci kwartyl, IQR – rozstęp ćwiartkowy. Pozostałe obserwacje, stanowiące ok. 1,5% próby, uznano za odstające i wyłączono z analizy.

ma podstaw do odrzucenia hipotezy, że parametry przy każdej z tych zmiennych nie zależą od kategorii. Dlatego też w celu oszacowania subiektywnych skal ekwiwalentności wykorzystano częściowo uogólniony model uporządkowany. Wyniki estymacji tych parametrów modelu, których wartości nie zależały⁹ od przynależności do kategorii zmiennej porządkowej, odnoszącej się do subiektywnej oceny dochodów, zaprezentowano w tab. 3.

Tabela 3. Wyniki estymacji parametrów częściowo uogólnionego uporządkowanego modelu logitowego

Zmienna	Ocena parametru	Błąd standardowy	Zmienna	Ocena parametru	Błąd standardowy
$\ln y$	2,84	0,05	wiek ²	0,0007	0,0001
$\ln y \cdot \ln n$	-0,33	0,03	wykształcenie wyższe	0,52	0,05
$(\ln n)^2$	1,43	0,28	wykształcenie średnie	0,23	0,04
$(\ln n)^3$	-0,45	0,10	stanowisko robotnicze	-0,19	0,04
wiek	-0,07	0,01	–	–	–

Źródło: obliczenia własne w programie Stata v. 10.

Oszacowania pozostałych parametrów częściowo uogólnionego uporządkowanego modelu logitowego zamieszczono w tab. 4.

Tabela 4. Wyniki estymacji parametrów częściowo uogólnionego uporządkowanego modelu logitowego, cd.

Zmienna	Dla $u > 1$	Dla $u > 2$	Dla $u > 3$	Dla $u > 4$
Kobieta	-0,41	-0,22	-0,26	-0,29
Związek formalny	0,39	0,46	0,26	0,44
Miasto pow. 500 000 mieszkańców	-0,60	-0,54	-0,37	-0,08*
Stała	-16,12	-17,90	-20,25	-22,76

* Oznacza, że parametr nie jest istotny na poziomie 0,05.

Źródło: obliczenia własne w programie Stata v. 10.

Na podstawie rezultatów przedstawionych w tab. 3 i 4 stwierdzono, że poza jednym parametrem (sytuację tę oznaczono symbolem * w tab. 4) pozostałe są statystycznie istotne na poziomie 0,05. Wyniki estymacji parametrów wykorzystano do oszacowania elastyczności ekwiwalentności. Ponieważ zamiast modelu (2)

⁹ W tabeli 3 zamieszczono oszacowania parametrów przy tych zmiennych, dla których wartość p w tab. 2 przekraczała 0,05.

proponowanego w pracy [Schwarze 2003] oszacowano model postaci:
 $u_i^* = \beta_0 + \alpha_y \ln y_i + \alpha_n \ln n_i + \alpha_{ny} \ln n_i \cdot \ln y_i + \alpha_{n2} (\ln n_i)^2 + \alpha_{n3} (\ln n_i)^3 + \mathbf{s}_i \boldsymbol{\gamma} + \varepsilon_i$,
 to elastyczność ekwiwalentności wyznaczono ze wzoru:

$$e = - \frac{\alpha_{ny} \cdot \ln y + \alpha_{n2} \cdot \ln n + \alpha_{n3} \cdot (\ln n)^2}{\alpha_y}. \text{ W następnym kroku wykorzystanie wzo-}$$

ru (1) umożliwiło estymację skal ekwiwalentności. Wartości oszacowań skal dla kilku przykładowych poziomów miesięcznych dochodów zaprezentowano w tab. 5.

Tabela 5. Wartości oszacowanych subiektywnych skal ekwiwalentności dla wybranych poziomów dochodów

Dochody (zł)	Wartość skali $S = n^e$ dla liczby osób:					
	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
2000	1,52	1,76	1,96	2,16	2,38	2,64
3000	1,57	1,86	2,09	2,33	2,59	2,89
4000	1,61	1,93	2,19	2,46	2,75	3,08
5000	1,64	1,98	2,27	2,56	2,88	3,24
6000	1,66	2,03	2,33	2,65	2,99	3,38
7000	1,68	2,07	2,39	2,73	3,09	3,50

Źródło: obliczenia własne.

Z informacji przedstawionych w tab. 5 wynika np., że aby zapewnić sobie taki sam poziom satysfakcji z osiągniętych dochodów jak gospodarstwa jednoosobowe wykazujące dochody wynoszące 5 tys. zł miesięcznie, gospodarstwa dwuosobowe powinny mieć dochód wyższy od nich o 64% (czyli $1,64 \times 5000 \text{ zł} = 8200 \text{ zł}$), gospodarstwa trzosobowe o 98%, czterosobowe o 127%, pięciosobowe o 156% itd.¹⁰

5. Wnioski

Na podstawie przeprowadzonej analizy stwierdzono, że „zwykły” uporządkowany model logitowy nie powinien być wykorzystywany do opisu subiektywnego postrzeżenia swojej sytuacji dochodowej przez gospodarstwa domowe pracowników w 2009 r. W konsekwencji tego, w odniesieniu do analizowanych danych, nie należy stosować metody wyznaczania subiektywnych skal ekwiwalentności w wersji proponowanej w pracy [Schwarze 2003]. W sytuacji, gdy nie ma podstaw do odrzucenia założenia równoległych regresji dla zmiennych dotyczących logarytmu dochodów, logarytmu liczby osób w gospodarstwie domowym oraz ich interakcji, do estymacji elastyczności ekwiwalentności można wykorzystać model częściowo uporządkowany.

¹⁰ Por. wartości w tab. 5 zaznaczone pogrubioną czcionką.

Elastyczność ekwiwalentności dla danych z gospodarstw domowych pracowników z 2009 r. jest funkcją ich dochodów oraz ich wielkości. Trzeba zatem stwierdzić, że formuły określające wartości skal ekwiwalentności są bardziej skomplikowane niż te, które zakłada się w niektórych pracach (por. [Johnson, Smeeding, Torres 2005; Radziukiewicz 2006]).

Należy także odnotować, że czynnikami determinującymi satysfakcję z osiągniętych dochodów w gospodarstwach domowych pracowników w 2009 r., oprócz dochodów ekwiwalentnych, były: poziom wykształcenia, płeć, wiek i fakt pozostawania w związku partnerskim osoby odniesienia oraz miejsce zamieszkania i przynależność do podgrupy społeczno-ekonomicznej.

Literatura

- Brant R., *Assessing proportionality in the proportional odds model for ordinal logistic regression*, „Biometrics” 1990, no 46 (4).
- Buhmann B., Rainwater L., Schmaus G., Smeeding T.M., *Equivalence scales well-Being, inequality, and poverty: sensitivity estimates across ten countries using the Luxembourg Income Study (LIS) database*, „Review of Income and Wealth” 1988, no 34.
- Dudek H., *Skale ekwiwalentności – estymacja na podstawie kompletnych modeli popytu*, Rozprawy i Monografie, nr 377, Wydawnictwo SGGW, 2011.
- Greene W.H., Hensher D.A., *Modeling ordered choices: a primer*, Cambridge, Cambridge University Press, 2010.
- Johnson D.S., Smeeding T.M., Torres B.B., *Economic inequality through the prisms of income and consumption*, „Monthly Labor Review” 2005, vol. 128, nr 4.
- Kot S.M., *Ekonometryczne modele dobrobytu*, PWN, Warszawa-Kraków 2000.
- Książek M., *Modele zmiennych wielomianowych uporządkowanych*, [w:] *Mikroekonometria*, M. Gruszczyński (red.), Wolters Kluwer Polska, Warszawa 2010.
- Long J.S., *Regression Models for Categorical and Limited Dependent Variables*, Sage Publications, Thousand Oaks, 1997.
- Peterson B., Harrell F.E. Jr., *Partial proportional odds models for ordinal response variables*, „Journal of the Royal Statistical Society. Series C” 1990, no 39 (2).
- Radziukiewicz M., *Zasięg ubóstwa w Polsce*, PWE, Warszawa 2006.
- Schröder C., *The Construction and Estimation of Equivalence Scales and Their Uses*, [w:] *Quantifying Consumer Preferences. Contributions to Economic Analysis*, D. Slottje (red.), Emerald, Bingley 2009.
- Schwarze J., *Using panel data on income satisfaction to estimate equivalence scale elasticity*, „Review of Income and Wealth” 2003, no 49.
- Williams R., *Generalized ordered logit/partial proportional odds models for ordinal dependent variables*, „The Stata Journal” 2006, no 6 (1).

SUBJECTIVE EQUIVALENCE SCALES – ANALYSIS BASED ON DATA ABOUT SATISFACTION WITH INCOMES

Summary: The article presents the results of the estimation of equivalence scales based on the subjective assessment of their own income situation of households. For this purpose, the method of partial generalized ordered logit models was used. The empirical investigation was based on data from household budget surveys carried out by GUS in 2009. The proposed methodology is an extension of the approach presented in [Schwarze 2003].

Keywords: equivalence scale, incomes, households.