

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

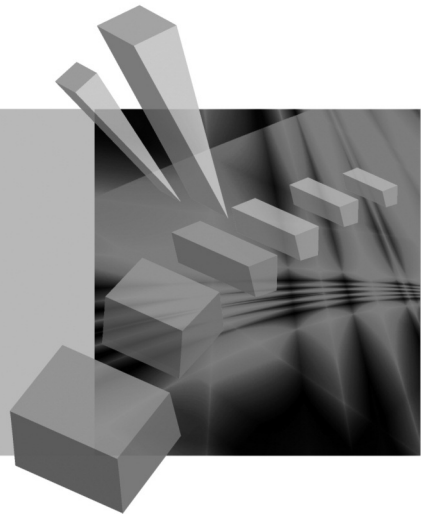
**RESEARCH PAPERS**

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](http://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](http://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](http://www.wydawnictwo.ue.wroc.pl)

Kopiowanie 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

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

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

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

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

## Summaries

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

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



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

---

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

**Anna Domagała**

Uniwersytet Ekonomiczny w Poznaniu

---

## **PROPOZYCJA METODY DOBORU ZMIENNYCH DO MODELI DEA (PROCEDURA KOMBINOWANEGO DOBORU W PRZÓD)**

---

**Streszczenie:** Artykuł przedstawia propozycję metody doboru zmiennych (nakładów oraz wyników) do modeli DEA (*Data Envelopment Analysis*). Metoda polega na stopniowym dodawaniu zmiennych do modelu DEA, zgodnie z opracowanym algorytmem. Decyzję o wprowadzeniu danej zmiennej podejmuje się na podstawie dwóch kryteriów – procentowej zmiany średniej efektywności, wywołanej dodaniem takiej zmiennej do układu, oraz współczynnika korelacji pomiędzy rezultatami badania bez tej zmiennej a wynikami badania z jej uwzględnieniem.

**Słowa kluczowe:** DEA, efektywność, krokowy dobór zmiennych.

### **1. Wstęp**

W trakcie prowadzenia analiz wielowymiarowych zazwyczaj pojawia się pytanie, czy możliwe jest zredukowanie liczby wykorzystywanych zmiennych. Celem jest maksymalne uproszczenie modelu, gdyż – jak wiadomo – każda dodatkowa zmienna zwiększa koszty przeprowadzanego badania. Znalezienie zmiennej, która nie ma wyraźnego wpływu na wyniki analizy, oznacza, iż nie wnosi ona dodatkowych informacji i można usunąć ją z badania. Takie procedury są szeroko stosowane np. w analizie regresji, czego przykładem jest procedura regresji krokowej w przód lub wstecz<sup>1</sup> czy metoda pojemności informacyjnej Z. Hellwiga<sup>2</sup>.

W badaniach efektywności z wykorzystaniem metody DEA rzadko stosuje się metody doboru zmiennych – mimo iż w literaturze znaleźć można propozycje takich procedur dla analiz typu DEA. Przykładem są prace Normana i Stokera [1991], Ser-

---

<sup>1</sup> Regresja krokowa w przód (lub wstecz) to najbardziej popularna metoda doboru zmiennych do modelu ekonometrycznego. Oparta jest ona na testach istotności (np.  $F$  lub  $t$ -Studenta).

<sup>2</sup> Oparta na współczynnikach korelacji metoda Z. Hellwiga polega na poszukiwaniu podzbioru zmiennych o największej tzw. pojemności informacyjnej [Hellwig 1969].

rano-Cinci, Mar-Molinera i Chaparro-Garcii [2002] czy Wagner i Shimshaka [2007], a w polskiej literaturze Guzika [2009].

Przeprowadzone przez autorkę badania<sup>3</sup>, wykorzystujące metodę DEA do empirycznego ustalania efektywności obiektów, skłoniły do opracowania nowego sposobu doboru zmiennych do wykorzystywanego tam modelu SBM-I<sup>4</sup>.

Celem opracowania jest propozycja nowej metody doboru zmiennych oraz próba weryfikacji postawionej hipotezy badawczej, według której proponowana metoda może znaleźć praktyczne zastosowanie w badaniach wykorzystujących różne modele typu DEA.

Zaproponowaną metodę doboru zmiennych nazwano *procedurą kombinowanego doboru w przód*. Zainspirowana została pracami Normana i Stokera [1991], Serrano-Cinci, Mar-Molinera i Chaparro-Garcii [2002] oraz Wagner i Shimshaka [2007]. Należy podkreślić, iż w każdej z metod prezentowanych przez powyższych autorów istnieje jeden wspólny „koszyk” (nakładów i wyników), z którego wybiera się dodawane do modelu zmienne. Oznacza to, iż metody te przebiegają jednotorowo.

## 2. Propozycja procedury ustalania listy nakładów i wyników w modelu DEA

W prezentowanej tu procedurze kombinowanego doboru w przód selekcja zmiennych również polega na stopniowym dodawaniu zmiennych do prostego modelu początkowego. Jednakże w odróżnieniu od metod prezentowanych przez wspomnianych wyżej autorów proponowana procedura przebiega dwutorowo.

Osobno dobiera się nakłady i osobno wyniki, dzięki czemu powstają dwa zbiory zmiennych istotnych w modelu: zbiór nakładów i zbiór wyników. Przy ustalaniu zbioru nakładów lista wyników jest pełna i nie zmienia się, a procedura polega na stopniowym dodawaniu kolejnych nakładów. To samo dzieje się po stronie wyników – lista nakładów jest pełna i nie zmienia się, a metoda polega na stopniowym dodawaniu kolejnych wyników.

Warunki, na podstawie których podejmuje się decyzję o dołączeniu danej zmiennej (a więc warunki uznania zmiennej za „istotną”), zostały zainspirowane kryteriami zaproponowanymi w pracach Normana i Stokera [1991], Serrano-Cinci, Mar-Molinera i Chaparro-Garcii [2002] oraz Wagner i Shimshaka [2007].

Decyzję o uznaniu zmiennej za „istotną” w modelu podejmuje się na podstawie zmiany poziomu średniej efektywności (wyznaczonej jako moduł<sup>5</sup> procentowej zmiany średniej efektywności uzyskanej z modelu uwzględniającego dodawaną w danym kroku zmienną w stosunku do średniej efektywności z modelu bez tej

<sup>3</sup> W ramach niepublikowanej rozprawy doktorskiej autorki.

<sup>4</sup> Zorientowany na nakłady nieradialny model SBM (*Slack-Based Measure*), należący do grupy modeli typu DEA.

<sup>5</sup> Interesująca jest bowiem „siła” zmiany, a nie jej kierunek.

zmiennej<sup>6</sup>) oraz współczynnika korelacji (między rezultatami analizy po wprowadzeniu danej zmiennej do modelu a wynikami analizy przed jej dodaniem<sup>7</sup>).

Zmienna, która spełnia powyższe warunki, dalej nazywana jest zmienną istotną w modelu. Uznano, iż warto połączyć oba powyższe kryteria (a więc wprowadzić warunek, iż muszą być one spełnione równocześnie), gdyż każde z nich pozwala zaobserwować inny aspekt zmiany rezultatów badania efektywności wywołanej dodaniem nowej zmiennej do modelu.

Algorytm proponowanej procedury przedstawiono na schemacie zamieszczonym na rys. 1. Występujące tam ramki prostokątne (lub kwadratowe) symbolizują kolejne konstruowane modele. Ramki elipsoidalne to pytania, na które należy odpowiedzieć, opierając się na uzyskanych rezultatach z danego modelu. Z kolei ramki w kształcie równoległoboku zawierają odpowiedź lub decyzję, którą należy podjąć, a ramki „gwiazdziste” pełnią funkcję informacyjną. Etapy procedury kombinowanego doboru omówiono poniżej, na przykładzie toru „nakładów” (N)<sup>8</sup>.

**Etap I:** Konstruuje się tyle prostych modeli (nazwanych modelami typu N1), ile jest nakładów. Każdy taki model zawiera wybrany jeden nakład i pełną wstępną listę wyników. Jeżeli liczba nakładów wynosi  $m$ , to modeli typu N1 będzie oczywiście  $m$ . Na podstawie tych modeli przeprowadza się analizę efektywności z wykorzystaniem przyjętego modelu DEA. Do następnego etapu przechodzi model, dla którego średnia efektywność (po wszystkich badanych obiektach) jest najwyższa. Oznaczmy go jako N1\*.

**Etap II:** Do ustalonego w kroku poprzednim modelu N1\* dodaje się kolejno jeden z pozostałych nakładów. Powstaje w ten sposób  $m - 1$  modeli typu N2, o dwóch nakładach i wszystkich wynikach. Bada się procentową zmianę średniej efektywności wywołaną dodaniem danego nakładu oraz poziom współczynnika korelacji pomiędzy rezultatami bez tego nakładu a rezultatami z uwzględnieniem tego nakładu. Ustala się w ten sposób, czy takie rozszerzenie jest „istotne”.

Rozszerzenie modelu jest istotne, gdy wartość bezwzględna procentowej zmiany średniej efektywności jest wyraźna, a więc przekracza pewien próg krytyczny  $\Delta_{kryt}$ , a współczynnik korelacji między wynikami dwóch analiz jest niższy od wcześniej ustalonego progu  $r_{kryt}$ <sup>9</sup>. Dany nakład można dodać do modelu tylko wtedy, gdy oba kryteria spełnione są równocześnie.

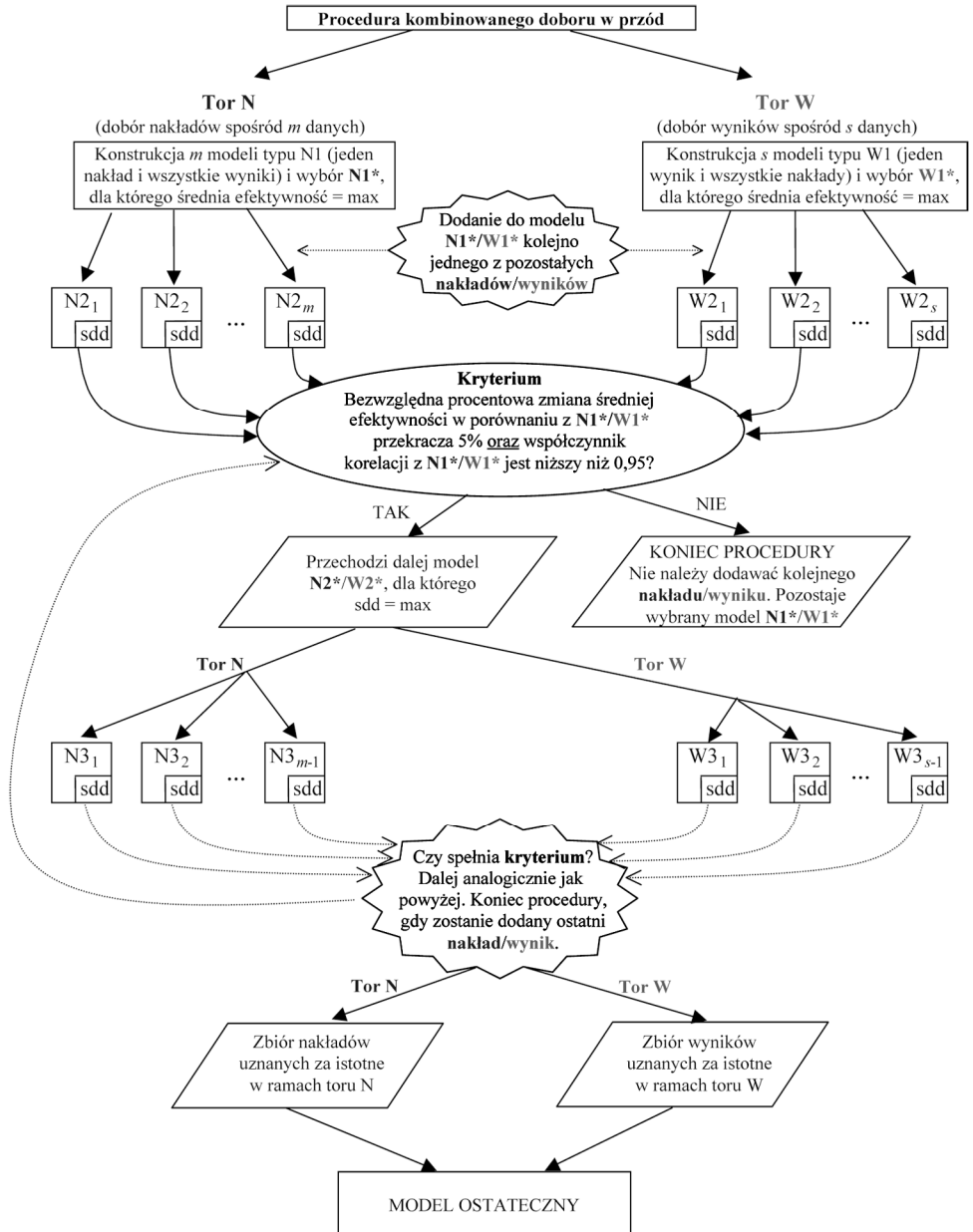
Warto zwrócić uwagę, iż w praktyce może pojawić się więcej przypadków, w których oba powyższe kryteria będą spełnione. W takiej sytuacji należy wybrać tę gałąź algorytmu, w której zmiana efektywności była największa, a współczynnik korelacji najniższy. Jeśli jednak wybór nie jest jednoznaczny, sugeruje się zbadanie

<sup>6</sup> Kryterium to jest modyfikacją warunków zastosowanych przez Serrano-Cincę, Mar-Molinero i Chaparro-García [2002] oraz Wagner i Shimshaka [2007].

<sup>7</sup> Pomysł wykorzystania współczynnika korelacji (choć inaczej zastosowanego) zaczerpnięto z pracy Normana i Stokera [1991].

<sup>8</sup> Tor W (czyli dobór wyników) oczywiście przebiega analogicznie, na podstawie tych samych kryteriów.

<sup>9</sup> Wartości graniczne przyjmuje się arbitralnie na podstawie symulacji empirycznych.



**Rys. 1.** Algorytm procedury kombinowanego doboru w przód

Źródło: opracowanie własne.

wskaźników *sdd* (czyli wskaźników tzw. skłonności do dodania) dla poszczególnych modeli z dodanymi „zmiennymi-kandydatkami”. Wskaźnik *sdd* powstaje przez uśrednienie dwóch wskaźników cząstkowych:

- pierwszym jest  $\Delta^* = \frac{\Delta_k}{\Delta_{\max}}$ , gdzie  $\Delta_k$  to bezwzględna wartość procentowej

zmiany średniej efektywności uzyskanej w badaniu po wprowadzeniu analizowanej w danym momencie *k*-tej zmiennej w stosunku do średniej efektywności bez uwzględnienia tej zmiennej, a  $\Delta_{\max}$  to największa (co do wartości bezwzględnej) procentowa zmiana średniej efektywności w badaniu po wprowadzeniu nowej zmiennej (spośród *K* zmiennych) w stosunku do średniej efektywności bez tej zmiennej, czyli  $\Delta_{\max} = \max_k \{\Delta_k\}$ ;

- drugim jest  $r^* = \frac{r_{\min}}{r_k}$ , przy czym  $r_k$  to wartość bezwzględna współczynnika

korelacji między rezultatami analizy po dodaniu badanej w danym momencie *k*-tej zmiennej a rezultatami bez tej zmiennej. Z kolei  $r_{\min}$  to najniższy (co do wartości bezwzględnej) współczynnik korelacji między rezultatami analizy po dodaniu nowej zmiennej a rezultatami bez tej zmiennej, czyli  $r_{\min} = \min_k \{r_k\}$ .

Wskaźnik „skłonności do dodania” ma postać:

$$sdd = \frac{\Delta^* + r^*}{2}. \quad (1)$$

Oblicza się go dla każdej ze „zmiennych-kandydatek”. Do dalszej analizy wprowadza się tę zmienną, dla której wskaźnik *sdd* ma najwyższą wartość. Model z taką zmienną oznaczany jest jako N2\*.

**Etap III:** Do wybranego modelu N2\* dodaje się kolejno jeden z pozostałych jeszcze nakładów. Powstaje w ten sposób *m* – 2 modeli typu N3. Decyzję o tym, który nakład można dodać do modelu, podejmuje się analogicznie jak w etapie II.

W kolejnych etapach konstruuje się modele typu N4, N5 itd. i sprawdza, które z pozostałych nakładów należy włączyć do modelu. Tor N kończy się, gdy dodano ostatni nakład lub kiedy dla żadnego rozszerzenia listy nakładów nie notuje się istotnych zmian w rezultatach badania efektywności w porównaniu z modelem z poprzedniego etapu.

Tor W przebiega analogicznie. Zaczyna się od prostego modelu (jeden wynik i pełna lista nakładów), a potem kolejno dołącza się wyniki – do momentu, gdy dodanie następnego wyniku nie spowoduje istotnej zmiany rezultatów badania efektywności obiektów.

W ostatnim etapie procedury łączy się informacje uzyskane w analizach przeprowadzonych dla torów N oraz W. Model ostateczny uwzględnia zbiór zmiennych

będący kombinacją obu torów – nakłady wyodrębnione przez tor N oraz wyniki wskazane przez tor W.

### 3. Przykład zastosowania procedury kombinowanego doboru w przód

Badaniu poddana została grupa 45 obiektów. Każdy z nich opisany był przez intuicyjnie wybrany zbiór siedmiu zmiennych<sup>10</sup>, z których cztery to nakłady (oznaczane przez  $x_i$ ), a trzy to wyniki (oznaczane przez  $y_i$ ). W przykładzie wybrano model SBM-I (zorientowany na nakłady model SBM bez nadefektywności<sup>11</sup>).

W celu sprawdzenia, czy istnieje możliwość zredukowania liczby zmiennych, zastosowano procedurę kombinowanego doboru w przód. Algorytm procedury składa się z etapów przedstawionych na rys. 1. Poniżej omówiono tylko kolejne etapy doboru nakładów (a więc tor N). Dobór wyników przebiegał analogicznie.

#### Tor N

**Etap I:** Procedura zaczyna się od przeprowadzenia analizy efektywności na podstawie modeli oznaczonych jako N1, które zawierają pojedynczy nakład oraz cały zestaw wyników. W przykładzie są cztery takie proste modele. Dla każdego z nich ustala się średnią efektywność obiektów. Rezultaty podano w tab. 1.

**Tabela 1.** Etap I – wybór pierwszego nakładu

| Model N1 zawiera zmienne $y_1, y_2, y_3$ oraz jedną z: | Średnia efektywność modelu N1 |
|--|-------------------------------|
| $x_1$  | 0,292                         |
| $x_2$  | 0,545                         |
| $x_3$  | 0,500                         |
| $x_4$  | <b>0,567</b>                  |

Źródło: opracowanie własne.

Średnia efektywność jest najwyższa dla modelu zawierającego nakład  $x_4$ , a więc ta zmienna wprowadzana jest do modelu. Model ten oznaczono jako N1\*.

**Etap II:** Badana jest efektywność modeli (oznaczonych jako N2), zawierających zestaw zmiennych ustalonych w etapie I oraz jeden nakład spośród tych, które pozostały.

Rezultaty analiz przeprowadzonych w tym kroku porównuje się z efektami etapu poprzedniego, a więc ustala się zmianę średniej efektywności wybranego w etapie I modelu N1\* w stosunku do średniej efektywności z poszczególnych modeli N2 oraz

<sup>10</sup> Prezentowany przykład oparty jest na danych rzeczywistych, wykorzystywanych w empirycznej części niepublikowanej rozprawy doktorskiej autorki.

<sup>11</sup> Warto tu zaznaczyć, że zaleca się stosowanie procedury doboru na podstawie modeli bez nadefektywności, gdyż istotne są przede wszystkim zmiany poziomu efektywności obiektów nieefektywnych. Obiekty efektywne z dużą nadefektywnością sztucznie zawyżałyby poziom średniej efektywności w całej badanej grupie.



oblicza się współczynniki korelacji między rezultatami z modelu N1\* a rezultatami z poszczególnych modeli N2. W tabeli 2 zamieszczono efekty tych porównań, a także podano wskaźniki „skłonności do dodania”<sup>12</sup>.

**Tabela 2.** Etap II – wybór drugiego nakładu

| Model N2:                       | Średnia efektywność modelu N2 | $\Delta_k$ – % zmiana średniej efektywności (w stosunku do modelu N1*) | $r_k$ – wsp. korelacji (z wynikami modelu N1*) | Wskaźniki cząstkowe |       | <i>sdd</i> |
|---------------------------------|-------------------------------|--|--|---------------------|-------|------------|
|                                 |                               |  |  | $\Delta^*$          | $r^*$ |            |
| $x_4 \setminus y_1 y_2 y_3 x_1$ | 0,662                         | 16,71%   | 0,730  | 0,621               | 0,739 | 0,680      |
| $x_4 \setminus y_1 y_2 y_3 x_2$ | <b>0,720</b>                  | <b>26,91%</b>  | <b>0,539</b>                                   | 1,0                 | 1,0   | <b>1,0</b> |
| $x_4 \setminus y_1 y_2 y_3 x_3$ | 0,704                         | 24,25%   | 0,700  | 0,901               | 0,770 | 0,836      |

Źródło: opracowanie własne.

Jak wspomniano wcześniej, wartości progowe średniej zmiany efektywności ( $\Delta_{kryt}$ ) oraz współczynnika korelacji ( $r_{kryt}$ ), od których uznaje się zmianę rezultatów badania efektywności za istotną, ustala się na podstawie symulacji empirycznych. W przykładzie przyjęto, że  $\Delta_{kryt} = 5\%$ , a  $r_{kryt} = 0,95$ .

Z informacji zawartych w tab. 2 wynika, iż zmienną, która ma najsilniejszy wpływ na rezultaty analizy DEA, jest zmienna  $x_2$ . Procentowa zmiana średniego poziomu efektywności w analizie uwzględniającej model z jednym nakładem ( $x_4$ ) w stosunku do średniej efektywności w analizie z dwoma nakładami ( $x_4$  oraz  $x_2$ ) wynosi prawie 27%, a więc przekracza założone 5%. Natomiast współczynnik korelacji między rezultatami tych analiz (bez zmiennej  $x_2$  i ze zmienną  $x_2$ ) wynosi w przybliżeniu 0,54 (jest więc niższy niż przyjęty poziom 0,95).

Warto zwrócić uwagę, że dodanie nakładu  $x_1$  czy  $x_3$  również wywołuje zmiany efektywności, które można uznać za istotne<sup>13</sup>. Wyodrębnienie właściwej zmiennej ułatwia obliczanie wskaźnika *sdd* („skłonności do dodania”), który w tab. 2 dla zmiennej  $x_2$  ma wartość najwyższą, równą 1,0. Ostatecznie więc wybrany w tym etapie model (z dodaną zmienną  $x_2$ ) oznacza się jako N2\*.

**Etap III:** Do modelu N2\* dodaje się kolejno jeden z pozostałych jeszcze nakładów. Powstaje w ten sposób szereg modeli N3, z których każdy zawiera zestaw nakładów wyodrębniony w kroku poprzednim oraz jeden z pozostałych (czyli  $x_1$  lub  $x_3$ ). Efekty podsumowano w tab. 3.

Jak widać, kolejną zmienną, którą powinno się dodać do modelu, jest zmienna  $x_1$ . Nakład  $x_3$  również spełnia warunki dotyczące wprowadzania zmiennych do modelu, jednakże obliczony dla niego wskaźnik „skłonności do dodania” jest niższy niż *sdd* dla nakładu  $x_1$ . Model z uwzględnionym nakładem  $x_1$  oznacza się jako N3\*.

<sup>12</sup> Obliczone według wzoru (1).

<sup>13</sup> Jednak, jak zaznaczono wcześniej, do modelu wprowadza się tę zmienną, której dodanie wywołuje najwyższą zmianę średniej efektywności i przy której współczynnik korelacji jest najniższy.

**Tabela 3.** Etap III – wybór trzeciego nakładu

| Model N3:            | Średnia efektywność modelu N3 | $\Delta_k$<br>(w stosunku do modelu N2*) | $r_k$<br>(z wynikami modelu N2*) | Wskaźniki cząstkowe |       | <i>sdd</i>   |
|----------------------|-------------------------------|--|----------------------------------|---------------------|-------|--------------|
|                      |                               |  |                                  | $\Delta^*$          | $r^*$ |              |
| $x_2x_4y_1y_2y_3x_1$ | <b>0,876</b>                  | <b>21,72%</b>                            | <b>0,627</b>                     | 1,0                 | 0,954 | <b>0,977</b> |
| $x_2x_4y_1y_2y_3x_3$ | 0,823                         | 14,34%                                   | 0,598                            | 0,660               | 1,0   | 0,830        |

Źródło: opracowanie własne.

**Etap IV:** Do zestawu wyodrębnionego w etapie poprzednim dodaje się kolejno pozostałe nakłady. W przykładzie pozostał już tylko jeden nakład, stąd krok ten polega tylko na sprawdzeniu, czy zmienna  $x_3$  jest zmienną istotną. Wyniki podano w tab. 4.

**Tabela 4.** Etap IV – wybór czwartego nakładu

| Model N4:               | Średnia efektywność modelu N4 | $\Delta_k$<br>(w stosunku do modelu N3*) | $r_k$<br>(z wynikami modelu N3*) | Wskaźniki cząstkowe |       | <i>sdd</i> |
|-------------------------|-------------------------------|--|----------------------------------|---------------------|-------|------------|
|                         |                               |  |                                  | $\Delta^*$          | $k^*$ |            |
| $x_1x_2x_4y_1y_2y_3x_3$ | 0,907                         | <b>3,53%</b>                             | 0,765                            | ---                 | ---   | ---        |

Źródło: opracowanie własne.

Okazuje się, iż zmienna  $x_3$  nie jest istotna<sup>14</sup>. Tor N kończy się. Spośród intuicyjnie wybranych czterech nakładów do modelu powinny wejść:  $x_1$ ,  $x_2$  oraz  $x_4$ .

### Tor W oraz etap końcowy

Powyżej przedstawiono dobór nakładów (czyli tor N). W sposób analogiczny przeprowadza się analizę istotności poszczególnych wyników. Procedura zaczyna się od prostych modeli (typu W1) z tą różnicą, że tym razem stałym zestawem zmiennych są wszystkie początkowe nakłady (a więc  $x_1$ ,  $x_2$ ,  $x_3$  oraz  $x_4$ ), a iteracyjnie zostają dobrane kolejne wyniki. Finalnie przeprowadzony w przykładzie tor W wskazał, że do modelu powinny wejść dwa spośród trzech początkowych wyników<sup>15</sup>. Są to:  $y_2$  oraz  $y_3$ .

Ostatecznie do modelu wybiera się zmienne, które zostały wskazane w obu torach procedury. Oznacza to, iż w badaniu efektywności powinna zostać wykorzystana lista:  $x_1$ ,  $x_2$ ,  $x_4$  oraz  $y_2$ ,  $y_3$ .

<sup>14</sup> Wprawdzie współczynnik korelacji jest niższy niż 0,95, jednak średnia zmiana efektywności nie przekroczyła minimalnego poziomu 5%.

<sup>15</sup> Jak wspomniano, tor W przebiega tak samo jak tor N, nie zaprezentowano więc jego poszczególnych etapów.

## 4. Podsumowanie

Opracowanie miało na celu propozycję nowej metody doboru zmiennych do modeli typu DEA. Działanie metody zilustrowano na przykładzie modelu SBM-I. Warto jednak podkreślić, iż omawiana procedura kombinowanego doboru w przód, ze względu na swoją konstrukcję, jest niezwykle uniwersalna i może być stosowana również przy innych modelach typu DEA (co ważne, tematyka badania nie ma tu znaczenia) – pod jednym tylko warunkiem, że pozwalają one na uzyskanie jedno-wymiarowej miary efektywności<sup>16</sup>.

## Literatura

- Guzik B., *Prosta metoda doboru zestawu nakładów w modelach DEA*, „Przegląd Statystyczny” 2009, nr 1.
- Hellwig Z., *Problem optymalnego wyboru predyktant*, „Przegląd Statystyczny” 1969, nr 3-4.
- Norman M., Stoker B., *Data Envelopment Analysis: The Assessment of Performance*, John Wiley & Sons Inc., 1991.
- Serrano-Cinca C., Mar-Molinero C., Chaparro-Garcia F., *On model selection in data envelopment analysis: a multivariate statistical approach*, Discussion Paper Series, “Management” nr M02-7, University of Southampton, 2002.
- Wagner J.M., Shimshak D.G., *Stepwise selection of variables in data envelopment analysis: procedures and managerial perspectives*, “European Journal of Operational Research” 2007, nr 180.

### **PROPOSAL OF A NEW METHOD FOR VARIABLE SELECTION IN DEA MODELS (COMBINED FORWARD STEPWISE SELECTION METHOD)**

**Summary:** The article presents a proposal of a new method for variable selection in Data Envelopment Analysis (DEA). The method is based on stepwise adding of variables to a simple (one output/one input) model. Two conditions are used to recognize which variable should be added to the model: relative change of a mean efficiency (which is due to incorporation of a new variable) and correlation coefficient between the results of efficiency analysis before and after incorporating a new variable.

**Keywords:** DEA, efficiency, stepwise selection of variables.

---

<sup>16</sup> Oznacza to, iż procedura ta nie znajdzie bezpośredniego zastosowania np. dla addytywnego modelu DEA.