

Andrzej Bąk

Uniwersytet Ekonomiczny we Wrocławiu

e-mail: andrzej.bak@ue.wroc.pl

**STATYSTYCZNE METODY DOBORU ZMIENNYCH
W PORZĄDKOWANIU LINIOWYM**

**STATISTICAL METHODS OF VARIABLES
SELECTION IN LINEAR ORDERING**

DOI: 10.15611/pn.2017.468.03

JEL Classification: C00, C19, C49, C87

Streszczenie: Jednym z najważniejszych problemów w badaniach ekonomicznych prowadzonych za pomocą metod analizy danych (w tym metod porządkowania liniowego) jest dobór zmiennych. Problem ten dotyczy włączenia zmiennych o małym zasobie informacyjnym, pominięcia zmiennych o dużym zasobie informacyjnym lub uwzględnienia zmiennych współliniowych. Może to prowadzić do błędnej oceny badanych zjawisk i podejmowania niewłaściwych decyzji. W literaturze przedmiotu prezentowane są dwa podejścia do zagadnienia doboru zmiennych: merytoryczne i statystyczne. Celem artykułu jest przedstawienie wybranych metod statystycznych doboru zmiennych, które mogą być wykorzystane w badaniach ekonomicznych prowadzonych za pomocą metod porządkowania liniowego oraz ocena ich efektywności w świetle poprawności uzyskiwanych rankingów. W analizie poprawności rankingów zastosowano wybrane mierniki oceny jakości metod porządkowania liniowego. W przeprowadzonych badaniach wykorzystano dane dotyczące rankingów uczelni akademickich oraz pakiet `pllord` i program R.

Słowa kluczowe: porządkowanie liniowe, metody doboru zmiennych, program R.

Summary: One of the major problems in the economic studies conducted by the methods of data analysis (including methods of linear ordering) is a selection of variables. This problem concerns the inclusion of variable low-resource information, omitted variables with high resource information or taking account of collinear variables. This can lead to an incorrect assessment of studied phenomena and making a wrong decision. In the literature there are presented two approaches to the selection of variables: substantive and statistical. The aim of the article is to present some statistical methods of variables selection that can be used in economic studies conducted by the methods of linear ordering and assessment of their effectiveness in the light of the correctness of generated rankings. In the analysis of the correctness of the rankings selected measures of quality assessment of linear ordering methods have been used. This study has used data on the rankings of academic institutions, a `pllord` package and R program.

Keywords: linear ordering, variables selection methods, R program.

1. Wstęp

Metody porządkowania liniowego są stosowane w badaniach ekonomicznych w celu ustalenia kolejności lub klasyfikacji obiektów opisanych przez wiele zmiennych.

Idea porządkowania liniowego obiektów wielowymiarowych opiera się na pojęciu porządkującej relacji binarnej (zwrotnej, antysymetrycznej, przechodniej i spójnej). Z aksjomatów tej relacji wynika, że jest możliwe stwierdzenie, który z dwóch dowolnych obiektów zbioru jest pierwszy (lepszy), a który drugi (gorszy), a także, czy są one identyczne. Przedmiotem porządkowania liniowego mogą być np. takie obiekty, jak kraje (ze względu na poziom rozwoju gospodarczego), przedsiębiorstwa (ze względu na kondycję finansową), produkty (ze względu na walory użytkowe) itp. Takie charakterystyki, jak poziom rozwoju gospodarczego, kondycja finansowa, walory użytkowe, są zmiennymi, których realizacje nie są bezpośrednio mierzalne. Zmienne takie są agregatami, których wartości są generowane przez obserwacje zmiennych diagnostycznych, które są bezpośrednio mierzalne (funkcje agregujące mogą mieć różną postać analityczną). Uzyskane realizacje zmiennej syntetycznej umożliwiają uporządkowanie obiektów wielowymiarowych w sensie relacji preferencji (dominacji).

Celem artykułu jest przedstawienie wybranych metod statystycznych doboru zmiennych, które mogą być wykorzystane w badaniach ekonomicznych prowadzonych za pomocą metod porządkowania liniowego oraz ocena ich efektywności w świetle poprawności uzyskiwanych rankingów. W analizie poprawności rankingów wykorzystane zostały mierniki oceny jakości metod porządkowania liniowego. Podstawą analizy są dane empiryczne dotyczące rankingów polskich szkół wyższych.

Dodatkowym efektem pracy jest implementacja analizowanych metod porządkowania liniowego, mierników oceny jakości tych metod oraz metod doboru zmiennych w pakiecie `pllord` [Bąk 2013, 2016] programu R [R Development Core Team 2016].

2. Metody porządkowania liniowego

Metody porządkowania liniowego, mieszczące się w obrębie wielowymiarowej analizy porównawczej i szerszej taksonomii, są w dużej mierze dorobkiem polskiej myśli statystycznej i ekonometrycznej. Pierwszą propozycję przedstawił Z. Hellwig w pracy [Hellwig 1968]. Publikacja ta zainicjowała intensywne badania w tym zakresie, których efektem były kolejne propozycje metod porządkowania liniowego, zamieszczone m.in. w pracach [Cieślak 1974; Bartosiewicz 1976; Pluta 1976; Strahl 1978; Borys 1978b; Nowak 1984; Walesiak 1993].

Podstawą porządkowania liniowego jest zmienna syntetyczna¹, której wartości są szacowane na podstawie obserwacji zmiennych diagnostycznych opisujących badane obiekty. Zakłada się, że wartości zmiennej syntetycznej, oszacowane za pomocą określonej metody, umożliwiają takie uporządkowanie zbioru obiektów, w którym [Grabiński 1992, s. 135]: (1) każdy obiekt ma przynajmniej jednego sąsiada oraz nie więcej niż dwóch sąsiadów, (2) jeżeli obiekt *a* jest sąsiadem obiektu *b*, to obiekt *b* jest sąsiadem obiektu *a*, (3) istnieją tylko dwa obiekty mające jednego sąsiada.

Zmienna syntetyczna ma charakter zmiennej ukrytej, ponieważ jej realizacje nie są bezpośrednio obserwowane. Realizacje te są natomiast generowane przez obserwacje zmiennych diagnostycznych, które są bezpośrednio mierzalne. Realizacje zmiennej syntetycznej są szacowane za pomocą funkcji agregujących, których postać analityczna może być różna. Rozróżnia się dwie grupy metod, które są stosowane do szacowania wartości zmiennej syntetycznej: metody bezwzorcowe i wzorcowe.

W procedurze porządkowania liniowego wyróżnia się takie etapy postępowania, jak: określenie charakteru zmiennych (stymulanty, nominanty, destymulanty)², wyznaczenie wag zmiennych, normalizacja zmiennych, wyznaczenie współrzędnych wzorca w przypadku agregacji wzorcowej, agregacja bezwzorcowo lub wzorcowa [Grabiński 1984, Pocięcha i in. 1988, Bąk 1999].

3. Zagadnienie doboru zmiennych w porządkowaniu liniowym

Wyróżnia się dwa zasadnicze podejścia w zagadnieniu doboru zmiennych w badaniach ekonomicznych [Nowak 1984]:

- podejście merytoryczne oparte na wiedzy badacza lub ekspertów,
- podejście formalne z zastosowaniem metod statystycznych.

Często łącznie są stosowane obydwie podejścia.

Zmienne opisujące obiekty w porządkowaniu liniowym powinny:

- dobrze dyskryminować obiekty (charakteryzować się dużą zmiennością),
- być słabo skorelowane między sobą,
- być mocno skorelowane ze zmiennymi odrzuconymi.

Do najczęściej stosowanych statystycznych metod doboru zmiennych w porządkowaniu liniowym należą:

- metoda wykorzystująca współczynnik zmienności (klasyczny lub pozycyjny),
- metoda parametryczna klasyfikacji zmiennych Hellwiga,
- metoda odwróconej macierzy korelacji Maliny i Zeliasia,
- analiza czynnikowa,
- metody klasyfikacji (Warda, środków ciężkości).

¹ W literaturze przedmiotu spotkać można inne określenia zmiennej syntetycznej, takie jak np.: zmienna agregatowa, miara syntetyczna, syntetyczna miara rozwoju, taksonomiczny miernik rozwoju, agregatowa miara rozwoju, miara rozwoju gospodarczego.

² Pojęcia zmiennej stymulanty i destymulanty zostały wprowadzone do literatury przedmiotu przez Z. Hellwiga [1968], a pojęcie zmiennej nominanty przez T. Borysa [1978a].

W badaniach empirycznych, których wyniki przedstawiono w artykule, zastosowano metodę parametryczną Hellwiga i metodę odwróconej macierzy korelacji Maliny i Zeliasia.

Algorytm parametrycznej metody Hellwiga [1981]:

- obliczenie macierzy korelacji zmiennych \mathbf{R} :

$$\mathbf{R} = [r_{ij}] = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{bmatrix}_{m \times m}, \quad (1)$$

gdzie: $i, j = 1, 2, \dots, m$; m – liczba zmiennych.

- wyznaczenie wartości progowej współczynnika korelacji (r^*) np. na podstawie wzoru:

$$r^* = \min_i \max_j |r_{ij}|, \quad (2)$$

gdzie: $i, j = 1, 2, \dots, m$; m – liczba zmiennych.

Wartość progową współczynnika (parametr) można także wyznaczyć na podstawie wartości odczytanej z tablic dystrybuanty rozkładu t -Studenta lub też przyjąć w oparciu o arbitralną decyzję:

- obliczenie sumy (lub mediany) bezwzględnych wartości współczynników korelacji każdej kolumny macierzy \mathbf{R} :

$$R_j = \sum_{i=1}^m |r_{ij}|, \quad (3)$$

- wyznaczenie numeru kolumny (k), dla której suma R_j jest największa:

$$R_k = \max_j \{R_j\}, \quad (4)$$

- klasyfikacja zmiennych: zmienna o numerze (k) jest zmienną centralną, a zmienne, dla których $|r_{ik}| > r^*$, są zmiennymi satelitarnymi (tworzą one skupienie zmiennych silnie skorelowanych, a więc powodujących redundancję informacji),
- usunięcie z macierzy \mathbf{R} wierszy i kolumn odpowiadających zmiennym satelitarnym oraz kolumny odpowiadającej zmiennej centralnej,
- procedurę powtarza się do wyczerpania zbioru zmiennych,
- zmienne, które nie znalazły się w żadnym skupieniu, są zmiennymi izolowanymi (tworzą skupienia jednoelementowe),
- do analizy włącza się zmienne centralne i zmienne izolowane (zmienne satelitarne odrzuca się).

Algorytm metody odwróconej macierzy korelacji Maliny i Zeliasia [1997]:

- obliczenie macierzy korelacji zmiennych \mathbf{R} :

$$\mathbf{R} = [r_{ij}] = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{bmatrix}_{m \times m}, \quad (5)$$

gdzie: $i, j = 1, 2, \dots, m$; m – liczba zmiennych;

- obliczenie macierzy odwrotnej \mathbf{R}^{-1} do macierzy korelacji zmiennych \mathbf{R} :

$$\mathbf{R}^{-1} = [\tilde{r}_{ij}] = \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} & \cdots & \tilde{r}_{1m} \\ \tilde{r}_{12} & \tilde{r}_{22} & \cdots & \tilde{r}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{r}_{m1} & \tilde{r}_{m2} & \cdots & \tilde{r}_{mm} \end{bmatrix}_{m \times m}, \quad (6)$$

gdzie: $i, j = 1, 2, \dots, m$; m – liczba zmiennych; elementy diagonalne macierzy \mathbf{R}^{-1} ($\tilde{r}_{11}, \tilde{r}_{22}, \dots, \tilde{r}_{mm}$) przyjmują wartości z przedziału $[1; +\infty]$;

- wyznaczenie wartości progowej \tilde{r}^* dla elementów diagonalnych macierzy \mathbf{R}^{-1} ($\tilde{r}_{11}, \tilde{r}_{22}, \dots, \tilde{r}_{mm}$), np. na poziomie $\tilde{r}^* = 10$ lub $\tilde{r}^* = 100$, lub na innym poziomie przyjętym na podstawie przesłanek merytorycznych;
- usunięcie zmiennych, dla których $|\tilde{r}_{ii}| > \tilde{r}^*$, lub zastąpienie pary zmiennych nową zmienną sztuczną, jeżeli jest to uzasadnione merytorycznie.

4. Mierniki oceny jakości procedur porządkowania liniowego

Poszczególne konfiguracje utworzone z wag zmiennych, metod normalizacji i metod szacowania wartości zmiennej syntetycznej prowadzą na ogół do różnego uporządkowania badanych obiektów. W związku z tym powstaje problem wyboru najlepszej konfiguracji w odniesieniu do empirycznego (lub symulacyjnego) zbioru danych o określonych charakterystykach statystycznych (wektor średnich, macierz kowariancji). Ocenę jakości zmiennej syntetycznej można przeprowadzić na podstawie mierników proponowanych w literaturze przedmiotu z obszaru taksonomii [Grabiński 1984; Grabiński i in. 1989; Bąk 1999, 2015].

Mierniki jakości wykorzystywane do wyboru optymalnej konfiguracji dotyczą w szczególności pomiaru i oceny takich własności procedur porządkowania liniowego, jak:

- zgodność odwzorowania, mierzona wskaźnikiem zróżnicowania odległości między obiektami w przestrzeni zmiennych diagnostycznych oraz w przestrzeni zmiennej syntetycznej,
- korelacja liniowa pomiędzy zmienną syntetyczną a zmiennymi diagnostycznymi, mierzona przeciętnym współczynnikiem „nieokreśloności” oraz współczynnikiem „jednoznaczności” zmiennej syntetycznej,
- korelacja rangowa zmiennej syntetycznej ze zmiennymi diagnostycznymi, mierzona współczynnikiem „nieokreśloności”, współczynnikiem „jednoznaczności”

zmiennej syntetycznej oraz uogólnionym rangowym współczynnikiem rozbieżności,

- zmienność i koncentracja zmiennej syntetycznej, mierzone współczynnikiem obliczonym dla realizacji zmiennej syntetycznej oraz dla pierwszych różnic uporządkowanych niemalejąco wartości zmiennej syntetycznej,
- przeciętna odległość taksonomiczna zmiennej syntetycznej od zmiennych diagnostycznych mierzona na podstawie mierników Hamminga oraz Euklidesa.

W aktualnej wersji pakietu `pllord` programu R uwzględniono następujące mierniki oceny jakości procedur porządkowania liniowego: miernik zgodności odwzorowania, miernik korelacji liniowej zmiennej syntetycznej ze zmiennymi diagnostycznymi, miernik korelacji rangowej zmiennej syntetycznej ze zmiennymi diagnostycznymi, miernik zmienności i koncentracji zmiennej syntetycznej³.

Mierniki te mają charakter cząstkowy o jednoznacznym kierunku preferencji – mniejsze wartości liczbowe każdego miernika wskazują na lepszą procedurę porządkowania liniowego. W związku z tym można przeprowadzić agregację mierników cząstkowych na podstawie wzoru [Seidler i in. 1980]:

$$Q_k = \sqrt{\sum_{l=1}^g g_l^2}, \quad (7)$$

gdzie: Q_k – miernik agregatowy k -tej konfiguracji elementów procedury porządkowania liniowego, g_l – miernik cząstkowy ($l = 1, \dots, g$), g – liczba mierników cząstkowych.

5. Wyniki badań

W badaniach empirycznych analizowano dane zgromadzone na potrzeby rankingów szkół wyższych prowadzonych przez portal „Perspektywy”⁴. Wykorzystano dane do rankingów wyższych uczelni akademickich z lat 2012 i 2016.

Ranking w roku 2012 obejmował 88 uczelni akademickich, które były opisane za pomocą 33 zmiennych reprezentujących 6 obszarów oceny.

Ranking w roku 2016 obejmował 90 uczelni akademickich, które były opisane za pomocą 27 zmiennych reprezentujących 6 obszarów oceny.

Dane dostępne są na stronach internetowych portalu „Perspektywy” odpowiednio dla roku 2012⁵ i dla roku 2016⁶.

W przeprowadzonych analizach uwzględniono 9 miar syntetycznych (agregm), w tym 2 bezwzorcowe (średnia arytmetyczna i średnia harmoniczna) oraz 7 wzorcowe

³ Formuły analityczne tych mierników są zamieszczone w pracach: [Grabiński 1984; Grabiński i in. 1989; Bąk 1999].

⁴ Strona internetowa: <http://www.perspektywy.pl>. Por. także pracę [Jaroeka 2013].

⁵ http://www.perspektywy.pl/ranking/2012/RWU/pdf/tabele_akademickie.pdf (6.09.2016).

⁶ <http://www.perspektywy.pl/RSW2016/pdf/RSW2016-uczelnie-akademickie-Perspektywy.pdf> (6.09.2016).

wych (odległości Hamminga, Euklidesa, Jeffreysa-Matusita, Braya-Curtisa, Clarka, Canberra i kątową).

Obliczenia przeprowadzono z uwzględnieniem wszystkich zmiennych (aggwzm), zmiennych wybranych za pomocą parametrycznej metody Hellwiga (aggmph) i zmiennych wybranych za pomocą metody odwróconej macierzy korelacji Maliny i Zeliasia (aggmmz).

Liczba wybranych zmiennych zależy od wartości progowej współczynnika korelacji (r^*) dla metody Hellwiga i od wartości progowej (\tilde{r}^*) dla metody Maliny i Zeliasia. W przeprowadzonych analizach większa liczba zmiennych prowadziła do mniejszych wartości miernika oceny jakości klasyfikacji (najniższe wartości są obserwowane w przypadku uwzględnienia wszystkich zmiennych). W związku z tym przyjęto takie wartości progowe (r^*) i (\tilde{r}^*), przy których obydwie metody wskazywały taką samą liczbę zmiennych. Wówczas jednak dobrane podzbiory zmiennych były różne.

Syntetyczne wskaźniki jakości porządkowania liniowego uczelni akademickich dla danych z roku 2012 są następujące:

		agregm	aggwzm	aggmph	aggmmz
1	arithmetic mean	26.097	26.744	25.541	
2	harmonic mean	26.097	26.744	25.541	
3	Hamming distance	7.294	12.956	13.023	
4	Euclidean distance	5.618	8.939	9.049	
5	Jeffreys-Matusita distance	4.632	5.954	5.986	
6	Bray-Curtis distance	8.404	15.456	15.445	
7	Clark distance	8.030	14.681	14.697	
8	Canberra distance	6.033	10.766	10.848	
9	angular distance	7.924	15.380	15.391	

Syntetyczne wskaźniki jakości porządkowania liniowego uczelni akademickich dla danych z roku 2016 są następujące:

		agregm	aggwzm	aggmph	aggmmz
1	arithmetic mean	27.154	26.956	27.591	
2	harmonic mean	27.154	26.956	27.591	
3	Hamming distance	10.271	12.680	12.625	
4	Euclidean distance	7.484	9.003	9.045	
5	Jeffreys-Matusita distance	5.270	5.972	6.015	
6	Bray-Curtis distance	12.033	14.888	14.885	
7	Clark distance	11.511	14.249	14.237	
8	Canberra distance	8.488	10.618	10.535	
9	angular distance	11.915	14.887	14.877	

Na podstawie otrzymanych wyników analizy danych można sformułować następujące wnioski:

- jakość rankingów prowadzonych metodami bezwzorcowymi nie jest jednoznacznie zależna od liczby zmiennych (w roku 2012 lepsze oceny uzyskał ranking z uwzględnieniem zmiennych wskazanych metodą Maliny i Zeliasia, a w roku 2016 ranking uwzględniający zmienne wskazane metodą Hellwiga – nie są to jednak różnice znaczące),

- jakość rankingów prowadzonych metodami wzorcowymi jest zawsze lepsza przy uwzględnieniu wszystkich zmiennych, co oznacza, że metody te są bardziej wrażliwe na utratę informacji związaną z pominięciem niektórych zmiennych,
- najlepsze oceny jakości porządkowania w obu rankingach (2012 i 2016) uzyskały metody wzorcowe wykorzystujące odległość Jeffreysa-Matusita i odległość Euklidesa,
- lepsze oceny jakości porządkowania w obu rankingach (2012 i 2016) otrzymano dla podzbiorów zmiennych wybranych metodą parametryczną Hellwiga (przy tej samej liczbie zmiennych), ale nie są to różnice znaczące.

6. Podsumowanie

Metody porządkowania liniowego, zajmujące ważne miejsce w dorobku polskiej taksonomii, znajdują praktyczne zastosowania w wielu dziedzinach badań, w tym w ekonomii. Celem badań empirycznych z wykorzystaniem tych metod jest ustalenie kolejności (porządku) obiektów opisywanych przez zbiór zmiennych. Dobór zmiennych jest istotnym elementem procedury badawczej prowadzonej za pomocą metod porządkowania liniowego, ponieważ ma bezpośredni wpływ na wyniki badań. Różne formalne metody doboru zmiennych prowadzą do różnych podzbiorów zmiennych. W badaniu porównano oceny jakości porządkowania liniowego na podstawie zmiennych wybranych metodami Hellwiga oraz Maliny i Zeliasia. Otrzymane wyniki prowadzą do następujących wniosków:

- liczba zmiennych nie ma jednoznacznego wpływu na ocenę jakości porządkowania metodami bezwzorcowymi,
- redukcja zbioru zmiennych prowadzi do pogorszenia ocen jakości porządkowania metodami wzorcowymi,
- nieco lepsze oceny jakości porządkowania otrzymano w przypadku redukcji zbioru zmiennych metodą Hellwiga.

Główne kierunki dalszych badań:

- porównanie wyników porządkowania na podstawie większej liczby zbiorów danych empirycznych oraz na podstawie zbiorów danych symulacyjnych o określonych rozkładach statystycznych,
- porównanie wyników porządkowania na podstawie innych mierników jakości metod porządkowania liniowego,
- włączenie do badań innych metod doboru zmiennych.

Literatura

- Bartosiewicz S., 1976, *Propozycja metody tworzenia zmiennych syntetycznych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 84, s. 5-9.
- Bąk A., 1999, *Modelowanie symulacyjne wybranych algorytmów wielowymiarowej analizy porównawczej w języku C++*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.

- Bąk A., 2013, *Metody porządkowania liniowego w polskiej taksonomii – pakiet pllord*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 278, s. 54-62.
- Bąk A., 2015, *Zagadnienie wyboru optymalnej procedury porządkowania liniowego w pakiecie pllord*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 384, s. 33-41.
- Bąk A., 2016, *Linear ordering methods – package pllord*. URL: <http://keii.ue.wroc.pl/pllord>.
- Borys T., 1978a, *Metody normowania cech w statystycznych badaniach porównawczych*, Przegląd Statystyczny, z. 2, s. 227-239.
- Borys T., 1978b, *Propozycja agregatowej miary rozwoju obiektów*, Przegląd Statystyczny, z. 3, s. 371-381.
- Cieślak M., 1974, *Taksonomiczna procedura prognozowania rozwoju gospodarczego i określania potrzeb na kadry kwalifikowane*, Przegląd Statystyczny, z. 1, s. 29-39.
- Grabiński T., 1984, *Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie. Seria specjalna: Monografie nr 61.
- Grabiński T., 1992, *Metody taksonometrii*, Akademia Ekonomiczna w Krakowie, Kraków.
- Grabiński T., Wydymus S., Zeliaś A., 1989, *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWN, Warszawa.
- Hellwig Z., 1968, *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, Przegląd Statystyczny, z. 4, s. 307-327.
- Hellwig Z., 1981, *Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych*, [w:] Welfe W. (red.), *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną*, PWE, Warszawa.
- Jarocka M., 2013, *Wpływ metody doboru cech diagnostycznych na wynik porządkowania liniowego na przykładzie rankingu polskich uczelni*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 279, s. 85-94.
- Malina A., Zeliaś A., 1997, *O budowie taksonomicznej miary jakości życia*, Taksonomia, z. 4, s. 238-262.
- Nowak E., 1984, *Problemy doboru zmiennych do modelu ekonometrycznego*, PWN, Warszawa.
- Pluta W., 1976, *Taksonomiczna procedura prowadzenia syntetycznych badań porównawczych za pomocą zmodyfikowanej miary rozwoju gospodarczego*, Przegląd Statystyczny, z. 4, s. 511-517.
- Pociecha J., Podolec B., Sokołowski A., Zając K., 1988, *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, PWN, Warszawa.
- R Development Core Team, 2016, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. URL: <http://cran.r-project.org>.
- Seidler J., Badach A., Molisz W., 1980, *Metody rozwiązywania zadań optymalizacji*, WNT, Warszawa.
- Strahl D., 1978, *Propozycja konstrukcji miary syntetycznej*, Przegląd Statystyczny, z. 2, s. 205-215.
- Walesiak M., 1993, *Statystyczna analiza wielowymiarowa w badaniach marketingowych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 654. Seria: Monografie i Opracowania, nr 101.