

Małgorzata Markowska

Uniwersytet Ekonomiczny we Wrocławiu
e-mail: małgorzata.markowska@ue.wroc.pl

Andrzej Sokółowski¹, Agnieszka Rygiel²

Uniwersytet Ekonomiczny w Krakowie
e-mail: andrzej.sokolowski@uek.krakow.pl; agnieszka.rygiel@uek.krakow.pl

Danuta Strahl

Akademia WSB, Wydział Nauk Stosowanych
e-mail: dstrahl@wsb.edu.pl

APROKSYMACJA WYNIKÓW DYNAMICZNEGO SKALOWANIA WIELOWYMIAROWEGO TRENDAMI NIELINIOWYMI

NONLINEAR TREND APPROXIMATION OF THE RESULTS OF DYNAMIC MULTIDIMENSIONAL SCALING

DOI: 10.15611/pn.2018.507.16

JEL Classification: C14, C23, O11

Streszczenie: Ideą skalowania wielowymiarowego jest rozmieszczenie analizowanych obiektów na płaszczyźnie (rzadziej w przestrzeni trójwymiarowej) w taki sposób, aby odległości tam wyznaczone jak najlepiej oddawały relacje odległości w oryginalnej przestrzeni wielowymiarowej. W *dynamicznym skalowaniu wielowymiarowym* proponujemy takie postępowanie, które uwzględni relacje (w sensie odległości) między punktami w ramach tej samej jednostki czasu, ale również dla różnych jednostek czasu. Kluczowym zabiegiem jest tu przekształcenie kostki danych Y w macierz X oraz zastosowanie tzw. standaryzacji globalnej. Wyniki można przedstawić na jednej płaszczyźnie, ale ciekawszym zabiegiem jest śledzenie ciągu płaszczyzn odpowiadających kolejnym jednostkom czasu oraz przemieszczania się konkretnych punktów. W pracy proponujemy dopasowanie trendów wielomianowych. Przykład empiryczny pokazuje trajektorie rozwoju 28 krajów Unii Europejskiej opisanych pięcioma zmiennymi makroekonomicznymi, w latach 2004-2015.

¹ Publikacja została sfinansowana ze środków przyznanych Wydziałowi Zarządzania Uniwersytetu Ekonomicznego w Krakowie, w ramach dotacji na utrzymanie potencjału badawczego.

² Publikacja została sfinansowana ze środków przyznanych Wydziałowi Finansów i Prawa Uniwersytetu Ekonomicznego w Krakowie, w ramach dotacji na utrzymanie potencjału badawczego.

Słowa kluczowe: skalowanie wielowymiarowe, rozwój gospodarczy, Unia Europejska, aproksymacja trendów nieliniowych, ścieżka rozwoju.

Summary: The main idea of Multidimensional Scaling is the positioning of objects on the plane in such a way that distances on that plane reflect the distances calculated in the original classification space. Most goodness-of-fit measures are based on the sum of squares of differences between the original and 2D matrix, and the most popular measure is STRESS introduced by Kruskal [1964]. In dynamic multidimensional scaling we propose a procedure which takes into account not only the relations between different objects, but also relations of an object with itself at different time points. Data cube should be transformed into spatio-temporal data matrix which is subject to global standardization. The results can be presented on a plane but also in 3D space with one dimension devoted to time variable. We propose to approximate trajectories with polynomial trend functions. The analysis of 28 European Union countries characterized by 5 macroeconomic variables, in 2004-2015 serves as an illustrative example.

Keywords: multidimensional scaling, economic development, European Union, nonlinear trend approximation, development trajectory.

1. Metoda

Skalowanie wielowymiarowe (zob. np. Kruskal i Wish [1978], Schiffman, Reynolds i Young [1981], Cox i Cox [2001], Borg i Groenen [2003], Borg, Groenen i Mair [2013]) to metoda pozwalająca na przedstawienie w mniejszej liczbie wymiarów (zazwyczaj na płaszczyźnie) takiej konfiguracji obiektów, która najlepiej reprezentuje odległości między nimi w oryginalnej przestrzeni danych. Kryterium najczęściej stosowanym w ocenie zgodności tych macierzy jest tzw. STRESS, zaproponowany przez Kruskala [1964].

Celem pracy jest przedstawienie pomysłu wykorzystania trendów nieliniowych do aproksymacji wyników dynamicznego skalowania wielowymiarowego wraz z przykładem o charakterze ilustracyjnym.

Sokołowski, Markowska i Rygiel [2015, 2016] zaproponowali ideę dynamicznego skalowania wielowymiarowego. Dane wyjściowe to kostka danych \mathbf{Y} o wymiarach $(n \text{ obiektów}) \times (m \text{ cech}) \times (T \text{ jednostek czasu})$. Kostka ta jest zamieniana na macierz danych \mathbf{X} o wymiarach $(nT \text{ wierszy/obiekto-okresów}) \times (m \text{ cech})$. Macierz \mathbf{X} poddawana jest standaryzacji. Jest to tzw. standaryzacja globalna (w odróżnieniu od lokalnej), gdzie średnia arytmetyczna i odchylenie standardowe są dla danej cechy liczone z całego okresu T . Przy wykorzystaniu macierzy standaryzowanej \mathbf{X}^* wyznaczana jest macierz odległości \mathbf{D} . Skalowaniem wielowymiarowym wyznaczone są współrzędne punktów (obiekto-okresów) na płaszczyźnie. Dla każdej jednostki czasu można pokazać właściwą jej konfigurację punktów. Z drugiej strony dla każdego obiektu można zilustrować jego ścieżkę przemieszczania się w czasie, wykorzystując współrzędne obiektów na kolejnych diagramach. W cytowanych pracach Sokołowski, Markowska i Rygiel [2015, 2016] zaproponowali aproksymację takich ścieżek za pomocą trendów liniowych. Następnie trendy te można grupować i wyznaczać średnie dla grup obiektów.

W niniejszej pracy proponujemy aproksymować ścieżki za pomocą nieliniowych trendów wielomianowych. Czyli „nieliniowość” dotyczy w zasadzie analizy wyników dynamicznego skalowania wielowymiarowego, a nie samej metody skalowania, która nadal pozostaje w swej klasycznej wersji metrycznej. Podobnie jak przy aproksymacji liniowej, w sytuacji gdy liczba obiektów jest relatywnie duża, warto najpierw te obiekty (lub ich trendy) pogrupować, wyznaczyć średnie współrzędne dla każdej grupy, w każdej jednostce czasu.

2. Przykład ilustracyjny

Jako przykład ilustrujący zaproponowane podejście poddamy analizie 28 państw Unii Europejskiej, w latach 2004-2015, scharakteryzowanych następującymi cechami [World Bank 2017]:

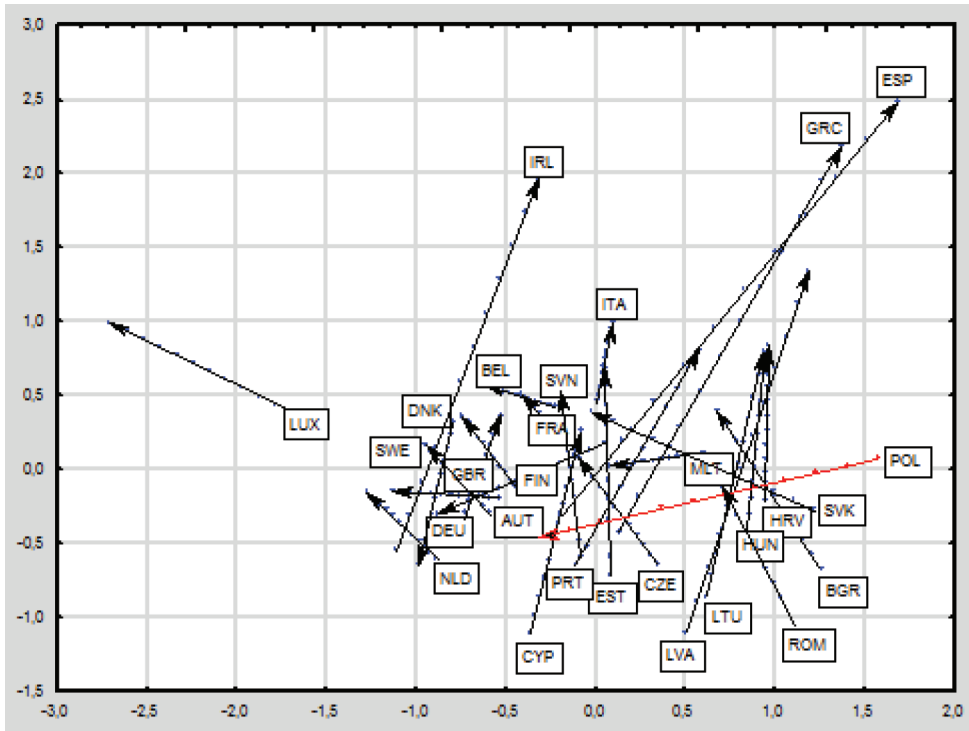
- produkt krajowy brutto *per capita*, z uwzględnieniem parytetu siły nabywczej, w dolarach,
- produkt krajowy brutto na zatrudnionego, w cenach stałych z 1990 r., w dolarach, z uwzględnieniem parytetu siły nabywczej,
- pracujący jako procent populacji w wieku 15+,
- roczny przyrost dochodu narodowego brutto *per capita*, w procentach,
- stopa bezrobocia, w procentach.

Po zastosowaniu procedury skalowania wielowymiarowego w sposób opisany powyżej dla każdego kraju oszacowano proste aproksymujące ścieżkę poruszania się kraju na diagramach skalowania wielowymiarowego odpowiadającym kolejnym latom (rys. 1).

Interpretacja wielu trendów jest utrudniona, może poza obserwacjami ewidentnie odstającymi, jak Luksemburg, Grecja i Hiszpania. W związku z tym proponujemy wstępne grupowanie trendów, a następnie znajdowanie średniej trajektorii dla grupy. W niniejszym opracowaniu zastosowano metodę Warda. Zastosowano ją trzykrotnie dla różnych grup trendów (chodzi o stopień wielomianu) i posługując się kryterium pierwszego wyraźnego skoku w odległości aglomeracyjnej oraz względami merytorycznymi, zdecydowano, że podziałem wynikowym każdorazowo będzie podział krajów na sześć grup. W wyniku aglomeracji trendów liniowych otrzymano następujące grupy krajów:

- A: Grecja, Hiszpania,
- B: Chorwacja, Malta, Polska, Słowacja, Węgry, Rumunia, Litwa, Łotwa, Bułgaria,
- C: Estonia, Czechy, Słowenia, Portugalia, Cypr,
- D: Włochy, Francja, Belgia,
- E: Luksemburg,
- F: Holandia, Dania, Niemcy, Irlandia, Finlandia, Wielka Brytania, Szwecja, Austria.

Na rys. 2 prawa oś pozioma jest osią czasu.



Akronimy: Grecja (GRC), Hiszpania (ESP), Chorwacja (HRV), Malta (MLT), Polska (POL), Słowacja (SVK), Węgry (HUN), Rumunia (ROU), Litwa (LTU), Łotwa (LVA), Bułgaria (BGR), Estonia (EST), Czechy (CZE), Słowenia (SVN), Portugalia (PRT), Cypr (CYP), Włochy (ITA), Francja (FRA), Belgia (BEL), Luksemburg (LUX), Holandia (NLD), Dania (DNK), Niemcy (DEU), Irlandia (IRL), Finlandia (FIN), Wielka Brytania (GBR), Szwecja (SWE), Austria (AUT).

Rys. 1. Liniowe trendy krajów ukazane na płaszczyźnie

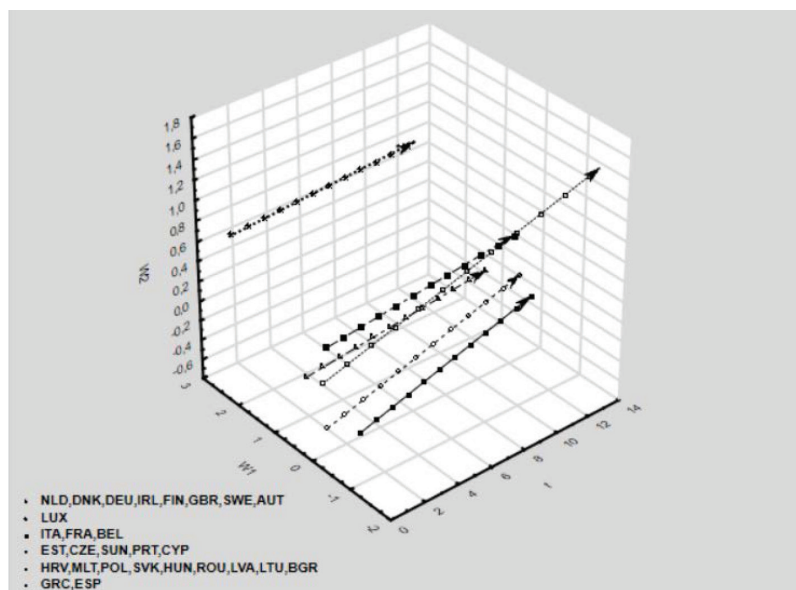
Źródło: opracowanie własne.

Widać, że we wszystkich grupach zanotowano ruch ogólnie w tym samym kierunku. Zdecydowanie wyróżnia się Luksemburg, jeżeli chodzi o punkt wyjścia i poziom. Wiąże się to z wykorzystaniem aż trzech zmiennych związanych z dochodem narodowym, który w Luksemburgu jest niemal dwukrotnie wyższy niż w kolejnym kraju Unii Europejskiej.

W drugim podejściu dla każdego kraju oszacowano trend paraboliczny, nie testując jego istotności. W wyniku grupowania parabol otrzymano te same grupy krajów co w przypadku trendów liniowych.

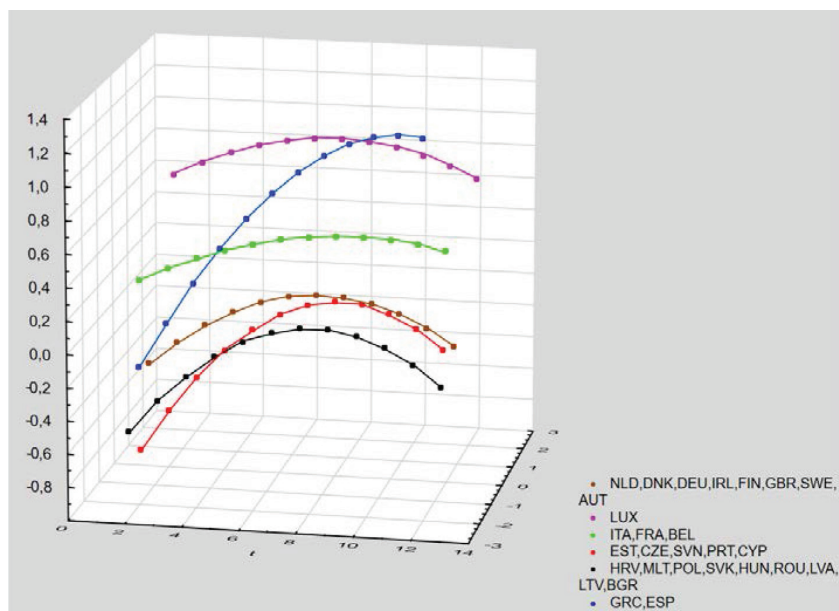
Graficzną ilustrację średnich trendów dla grup przedstawia rys. 3.

Orientacja przestrzenna rys. 3 jest nieco inna niż rys. 2. Tym razem oś czasu jest osią poziomą. Pozwala to zaobserwować wpływ kryzysu ekonomicznego z lat 2007–



Rys. 2. Trendy liniowe średnich dla grup krajów

Źródło: opracowanie własne.



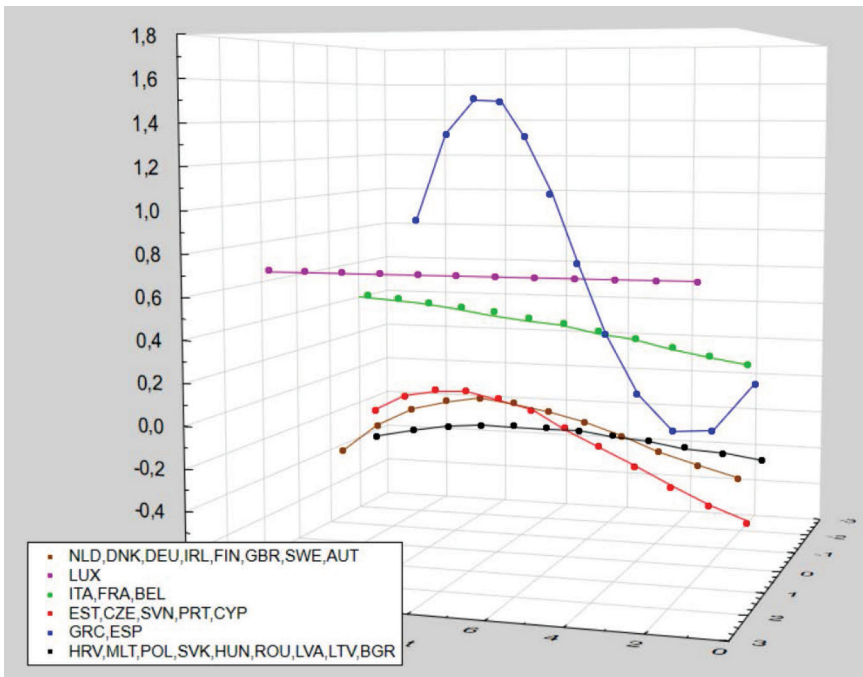
Rys. 3. Trendy paraboliczne dla grup

Źródło: opracowanie własne.

–2008 poprzez przypadające na ten okres wierzchołki parabol. Najwyższa parabola, przebiegająca w płaszczyźnie równoległej do osi czasu, reprezentuje Luksemburg. Parabola reprezentująca Hiszpanię i Grecję przebiega w płaszczyźnie odmiennej od wszystkich pozostałych. Środkowa parabola to grupa obejmująca Włochy, Francję i Belgię.

Trzecie podejście polegało na podejmowaniu prób dopasowania do danych (wyników skalowania wielowymiarowego) dla każdego kraju trendu wielomianowego o stopniu 3. Istotność trendu była weryfikowana testem istotności parametrów strukturalnych, przy poziomie istotności 0,10. W przypadku braku istotności zmniejszano stopień wielomianu. Jeżeli żaden trend (nawet liniowy) nie okazał się istotny statystycznie, wtedy reprezentacją danego wymiaru była średnia arytmetyczna.

Tak jak poprzednio, funkcje trendu dla krajów grupowano, i podobnie jak uprzednio, otrzymano ten sam podział na grupy. Ilustracją wyników jest rys. 4.

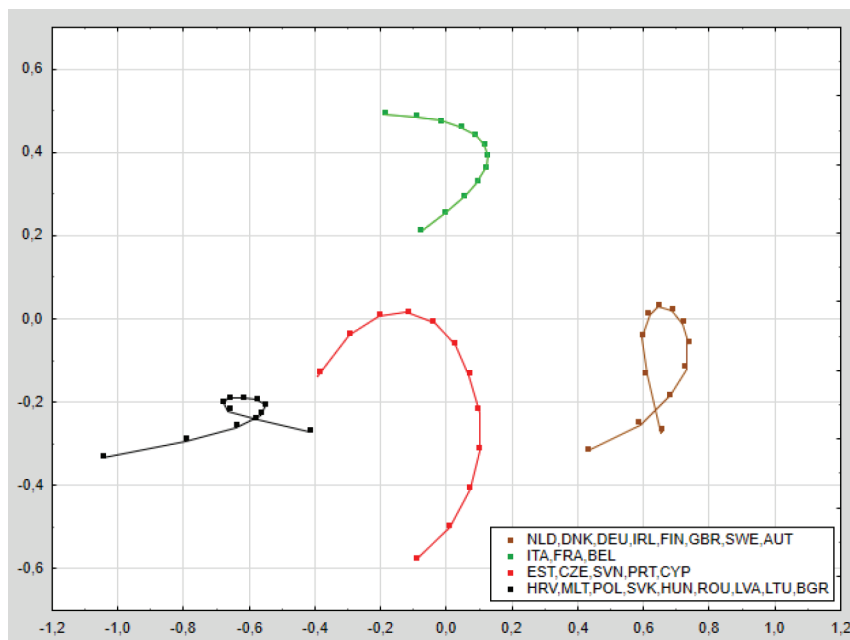


Rys. 4. Statystycznie istotne funkcje trendu dla grup krajów

Źródło: opracowanie własne.

Porządek krzywych jest podobny jak na rys. 3, z tym że najbardziej „skomplikowany” ruch na diagramach skalowania wielowymiarowego dotyczył Hiszpanii i Grecji. Oczywiście wynika to z faktu, że zmienne przyjęte w tym przykładzie ilustracyjnym dotyczą dochodu narodowego i rynku pracy.

Jeżeli zrzutujemy wszystkie punkty na jedną płaszczyznę (zdefiniowaną dwoma wymiarami wyników skalowania wielowymiarowego), to można pokazać przebiegi w przestrzeni dwuwymiarowej. Na rys. 5 zamieszczono krzywe dotyczące tylko czterech grup, z pominięciem Luksemburga, Hiszpanii i Grecji.



Rys. 5. Nieliniowe trendy czterech grup rzutowane na płaszczyznę

Źródło: opracowanie własne.

Jak wskazują współczynniki korelacji cech ze wszystkimi (za cały okres) wymiarami skalowania wielowymiarowego, wymiar reprezentowany przez oś poziomą wyraża poziom dochodu narodowego oraz sytuację na rynku pracy. Oś pionowa reprezentuje tempo przyrostu dochodu narodowego, przy czym występuje tu korelacja ujemna (zachowano tu oryginalne wyniki skalowania wielowymiarowego, choć oczywiście wartości na tej osi można było pomnożyć przez -1), im „wyżej”, tym niższe tempo przyrostu dochodu narodowego.

W przykładzie dotyczy to grupy D: Włochy, Francja i Belgia. Interesująca jest interpretacja zaobserwowanych pętli. Wskazują one – w rozważanym przykładzie – na porządzenie sobie z kryzysem. Najszybciej dokonały tego kraje reprezentowane przez wstążkę z lewej strony – w skład tej grupy wchodzi Polska. Pętla po prawej stronie reprezentuje kraje „starej” Unii Europejskiej.

3. Zakończenie

Zaproponowane podejście przewiduje poszukiwanie prawidłowości w zmianach pozycji obiektów na diagramach skalowania wielowymiarowego, odpowiadających kolejnym jednostkom czasu. Rekomendowana standaryzacja globalna uwzględnia nie tylko relacje między różnymi obiektami, lecz także relacje między położeniem tego samego obiektu w różnych jednostkach czasu. Wydaje się, że warto próbować dopasowywać wielomianowe trendy nieliniowe, jednak o stopniu nie wyższym niż 3. Przy relatywnie znacznej liczbie obserwacji interpretacja krzywych może być utrudniona, dlatego warto najpierw je pogrupować i znaleźć przeciętne krzywe dla grup.

Przedstawiony przykład empiryczny ma wyłącznie charakter ilustracyjny. Wynika jednak z niego obserwacja, że w zróżnicowaniu krzywych charakteryzujących zjawiska makroekonomiczne decydujące znaczenie ma raczej poziom niż dynamika zjawisk. Ponadto widać, że rzutowanie krzywych na płaszczyznę może być ciekawą ilustracją analizowanych zagadnień.

Literatura

- Borg I., Groenen P., 2003, *Modern Multidimensional Scaling. Theory and Applications*, Springer Series in Statistics.
- Borg I., Groenen P.J.F., Mair P., 2013, *Applied Multidimensional Scaling*, Springer, Heidelberg–New York–Dordrecht–London.
- Cox T.F., Cox M.A.A., 2001, *Multidimensional Scaling*, Monographs on Statistics and Applied Probability 88, Second Edition, Chapman & Hall/CRC, Boca Raton–London–New York–Washington D.C.
- Kruskal J.B., 1964, *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis*, Psychometrika, vol. 29, no. 1, s. 1–27.
- Kruskal J.B., Wish M., 1978, *Multidimensional Scaling*, Sage Publications, Newbury Park–London–New Delhi.
- Markowska M., Sokołowski A., Rygiel A., 2015, *Dynamiczne skalowanie wielowymiarowe*, Seminarium naukowe nt.: *Sytuacja społeczno-gospodarcza w Unii Europejskiej w latach kryzysu finansowego – analiza danych regionalnych i międzynarodowych*, Politechnika Rzeszowska, Rzeszów, 9 grudnia 2015.
- Schiffman S.S., Reynolds M.L., Young F.W., 1981, *Introduction to Multidimensional Scaling. Theory, Methods, and Applications*, Emerald Group Publishing Ltd, Bingley.
- Sokołowski A., Markowska M., Rygiel A., 2016, *Predictive Multidimensional Scaling. The Analysis of EU and Turkey Economic Development*, International Conference on Information Complexity and Statistical Modeling in High Dimensions with Applications IC-SMHD-2016, Nevsehir, 18-21 maja 2016 r. (artykuł w druku).
- World Bank 2017, [www.https://data.worldbank.org/\(15.08.2017\)](https://data.worldbank.org/(15.08.2017)).