PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 476 • 2017

Local and Regional Economy in Theory and Practice ISSN 1899-3192
e-ISSN 2392-0041

Tomasz Bartłomowicz
Wrocław University of Economics
e-mail: tomasz.bartlomowicz@ue.wroc.pl

Katarzyna Cheba
West Pomeranian University of Technology in Szczecin
e-mail: katarzyna.cheba@zut.edu.pl

MULTIDIMENSIONAL COMPARATIVE ANALYSIS OF THE EUROPEAN UNION COUNTRIES IN THE FIELD OF SUSTAINABLE DEVELOPMENT

WIELOWYMIAROWA ANALIZA PORÓWNAWCZA PAŃSTW UNII EUROPEJSKIEJ W OBSZARZE ZRÓWNOWAŻONEGO ROZWOJU

DOI: 10.15611/pn.2017.476.11

Summary: The subject of the study is a comparative analysis of the level of sustainable development of the European Union. The basis of the study was the data collected in the Eurostat database, which the European Commission uses to monitor the implementation of the objectives of EU Sustainable Development Strategy. In order to determine similarities and differences in the level of development of EU countries in the study area, a method of multidimensional scaling was used. The obtained results allowed to identify similar as well as different countries because of their designated sizes. There were also identified countries whose sustainability is the closest to the object of reference and those whose previous development is similar to anti-pattern. In the calculations the environment and packages of R program were used.

Keywords: Eurostat, multidimensional scaling, sustainable development indicators.

Streszczenie: Przedmiotem pracy jest analiza porównawcza poziomu zrównoważonego rozwoju państw Unii Europejskiej. Podstawę badania stanowiły dane zgromadzone w bazie Eurostat, które przez Komisję Europejską wykorzystywane są do monitorowania realizacji celów Strategii Zrównoważonego Rozwoju UE. W celu określenia podobieństw oraz różnic w poziomie rozwoju państw Unii w badanym obszarze wykorzystano metodę skalowania wielowymiarowego. Uzyskane wyniki pozwoliły na zidentyfikowanie krajów podobnych, jak i różniących się ze względu na wyznaczone wymiary. Zidentyfikowano również kraje, których zrównoważony rozwój jest najbliższy obiektowi wzorcowemu, oraz te, których dotychczasowy rozwój zbliżony jest do tzw. antywzorca. W obliczeniach wykorzystane zostało środowisko oraz pakiety programu R.

Słowa kluczowe: Eurostat, skalowanie wielowymiarowe, wskaźniki zrównoważonego rozwoju.

1. Introduction

The study of spatial differentiation, e.g. the level of development, the structure of the objects described with various socio-economic features are quite commonly described in the literature. In this kind of analysis often the methods of multidimensional statistical analysis are used enabling the study of objects and complex phenomena in which the state and behavior also affect many features and factors. A major problem in the study of this type is the complexity of analyzed objects and phenomena as well as the need to simplify the description of multidimensional space by reducing the dimensions. One of the methods to reduce this dimensionality is multidimensional scaling.

The aim of the study was to assess similarities and differences in the level of sustainable development of EU Member States. The method of multidimensional comparative analysis in the form of multidimensional scaling was used for this purpose. It allowed showing in two-dimensional space the relations between EU countries. The interpretation of the relationship enabled dimensions, which had the character of hidden variables, to explain the similarities and differences between the compared objects [Walesiak, Gatnar (eds.) 2009]. The study included all EU Member States. The source data were available in the Eurostat database indicators, which are used by the European Commission to monitor the implementation of the objectives of EU Sustainable Development Strategy (analyzed data from 2014). In the calculations the environment and packages of R program were used.

2. A research tool applied

For the purposes of determining the similarities and differences in the level of sustainable development of the EU countries, in the study the method of multidimensional scaling was used. The theoretical basis of the method and examples of scaling application in marketing research were presented in the works: [Borg, Groenen 2005; Kruskal 1964; Gatnar, Walesiak 2011; Zaborski 2001; Cox, Cox 1994; Green, Rao 1972; Hair et al. 1995]. Presented in the literature scaling objects (e.g. products, services, markets, businesses, households) are described by many variables in the m-dimensional space.[1] The task of multidimensional scaling is to present the relations between the objects in the r-dimensional ($r < m$) and to determine the content of these dimensions. Dimensions r are not directly observable and they are hidden variables, which allows for an explanation of the similarities and differences and then the interpretation of the relationship between the studied objects. Of all the methods of scaling its simplest variant is a classic multidimensional

[1] In the multidimensional scaling a step of selecting variables precede steps of determining the research problem and the selection of a test.

scaling. The starting point here is the design of distance matrix[2] (dissimilarities) between objects in *m*-dimensional space. In a classic scaling directly from the distance matrix it is sought to find such a configuration of points in the *r*-dimension so similar objects would be closer together. In formal terms, this means that for a given set of objects $A = \{A_1, \ldots, A_n\}$ and the distance d_{ik} between objects A_i and A_k (in *m*-dimensional space) such a projection[3] set of objects is looked for in a set of points in the *r*-dimensional space [Borg, Groenen 2005]:

$$d_{ik} \approx \hat{d}_{ik} = f\left(\delta_{ik}\right),$$

where: d_{ik} – is the distance between points and x_i and x_k in *r*-dimensional space,
$\hat{d}_{ik}$ – regression function between d_{ik} and δ_{ik}.

The basic decision problems in a multidimensional scaling includes scaling model selection, a choice of the method of normalization values of the variables and the choice of distance measure. In the case of the variables measured on the metric scales a choice of scaling model includes interval or quotient transformation [Gatnar, Walesiak 2011]. In the case of the method of measuring distance there may be used measure distances for metric data, including, for example Minkowski measures: Manhattan, Euclidean, Chebyshev [Grabinski 1992] and the GDM distance measure (GDM1 for metric data) [Walesiak 2006]. The choice of method of standardization may include: standardization, positional standardization, unitization, positional or with zero minimum unitization, positional normalization and other quotient transformations [Gatnar, Walesiak 2004]. In the scaling a measure of the quality of the resulting mapping is a feature function value STRESS-1. In the simplest version of this function it takes a form [Kruskal 1964]:

$$S^2 = \frac{\sum_{i,k}\left(d_{ik} - \hat{d}_{ik}\right)^2}{\sum_{i,k} d^2{}_{ik}}.$$

It is assumed that the result mapping using STRESS-1 is considered to be very good in the range of 0-2%, good (2-5%), medium (5-10%), low (10-20%) and very low (above 20%).

[2] Among the methods of determining the distance matrix a direct and indirect method is distinguished. Direct methods include e.g. comparing pairs of objects by individual respondents. In indirect methods, the starting point is a matrix of data (the observation of variables), which assumes that the variables are related to the dimensions of the space. Observations on variables are obtained from the secondary data sources.

[3] This means that multidimensional scaling is a kind of technique of data reduction, because its aim is to find a set of points in the space with a small number of dimensions (usually 2- or 3-dimensional), which represent well the configuration of the tested objects in the multidimensional space. A decision concerning the number of *r* dimensions in which the results of multidimensional scaling are presented depends on the researcher.

3. Conducted research

The information basis of studies were indicators (SDI) used by the European Commission to monitor the implementation of the objectives of EU Sustainable Development Strategy. The SDIs have a hierarchic structure whose components are divided into three levels. At the top there are 11 Headline Indicators that are intended to give an overall picture of the progress in terms of the key challenges of the EU SDS. The second level is represented by 31 Operational Indicators that relate to the operational objectives of the Strategy, while on the third, lowest level there are 84 Explanatory Indicators that illustrate the progress of the actions described in the SDS. In addition, the indicators are grouped under 10 themes, including socio-economic development, sustainable consumption and production or sustainable transport. In total, the Eurostat database collected information about 126 indicators describing sustainable development and not all of them are available on the individual level of EU Member States (e.g. in the case of indicators describing the area of natural resources). Additionally, some of them have only informative (explaining) nature (i.e. an indicator describing the modal split of freight transport).

These restrictions meant that the original set of 126 characteristics was reduced to 77 indicators[4] representing different areas of sustainable development, which were the final selection taking into account the statistical criteria. First, the assessment of the coefficients of variation calculated for each variable was carried out. As a criterion for the resignation of the given feature, the coefficient of variation less than or equal to 10% was used. Next a Hellwig's parametric method of selection of features was used [Hellwig 1981]. In this method the critical value of the correlation coefficient equal to 0.5 or higher was implemented.

Finally, among included in the analysis characteristics of EU countries, the following diagnostic features are found (where: S – stimulants and D – destimulants):

x_1 – investment by institutional sectors – total investment [% of GDP] (S),

x_2 – municipal waste generated [kg per capita] (D),

x_3 – livestock density index [livestock units per ha] (S),

x_4 – people at risk of poverty or social exclusion [% and 1000 persons] (D),

x_5 – people living in households with very low work intensity [% and 1000 persons] (D),

x_6 – tertiary educational attainment, age group 30-34 [%] (S),

x_7 – at most lower secondary educational attainment by age [%] (D),

x_8 – individuals' level of computer skills [% of the total number of individuals aged 16 to 74] (S),

x_9 – employment rate of older workers [%] (S),

x_{10} – aggregate replacement ratio [%] (S),

[4] These indicators were analyzed as one group without a division into different areas or levels of monitoring EU Sustainable Development Strategy.

x_{11} – old-age-dependency ratio [per 100 persons] (D),

x_{12} – non-fatal accidents at work by sex (NACE Rev. 2, A, C-N), Standardised incidence rate (D),

x_{13} – electricity generated from renewable sources [% of gross electricity consumption] (S),

x_{14} – share of renewable energy in fuel consumption of transport [%] (S),

x_{15} – combined heat and power generation [% of gross electricity generation] (S),

x_{16} – energy consumption of transport, by mode, road [1000 tonnes of oil equivalent] (D),

x_{17} – CO_2 emissions per inhabitant in the EU [tones] (D).

In this study, apart from all the European Union countries (28 member states), additional objects were included – pattern and anti-pattern respectively, in the form of upper and lower development pole. Defining the coordinates of the upper development pole the best ones were used and in the case of the lower development pole – the less favorable values of the variables. Thus, in the case of an object-model maximum values were used for stimulant and minimum for destimulants. In the case of anti-pattern a definition of coordinates was reversed (minimum values for the stimulants, maximum for destimulants).

Collecting the data for all 30 objects allowed the construction of matrix distance (dissimilarities). Due to the possible use in different scaling: scaling models, standardization of variables and methods of distance measures, which depending upon the selected combination very often lead to different results, in the subsequent step it was decided to determine the optimal procedure for the multidimensional scaling. As proposed by M. Walesiak [Walesiak 2016; Walesiak 2017] and its implementation in R-programme by M. Walesiak and A. Dudek [Walesiak, Dudek 2016] it evaluated the results of scaling of all included combinations of models, methods of standardization and distance measures. The choice of the optimal variant determines the lowest value of the Herfindahl-Hirschman Index HHI [Herfindahl 1950; Hirschman 1964] for a fixed arbitrary maximum level of function STRESS-1.

The survey analyzed data using 80 collected procedures of multidimensional scaling, in particular: 2 scale models[5], 10 methods of normalization of variables[6] and 4 distance measures[7]. The satisfying combination was recognized, the one which achieved the lowest value of the HHI index (assuming the value of the function STRESS-1 < 0.20). This means that in the conducted study the basis of multidimensional scaling were: interval model scaling (MDS interval), the method of normalization of variables in the form of positional normalization (n12a) and the city block distance (Manhattan) as the distance measure.

[5] The study used the scaling interval (interval MDS) and scaling quotient (ratio MDS).

[6] The study used 18 methods of normalization variables (n1-n13) available in the clusterSim R package [Walesiak, Dudek 2015]. Due to generating the same results for some scaling normalization methods (e.g. n1/n6/n12), the study indicates the use of 10 methods of normalization variables.

[7] The study used the following distance measures: Manhattan distance, Euclidean distance, Chebyshev distance and GDM distance (GDM1 for metric data).

4. Study results

An interpretation of the coordinates and dimensions of multidimensional scaling is a difficult task and a clear procedure for its conduct cannot be given [Woźniak, Sikora 2006]. The selected space dimension scaling replaces the original source of information, in this case the indicators describing the sustainable development of EU Member States.

In the paper the results of multidimensional scaling are presented in two-dimensional space, which truly reflects the structure of the EU Member States in the study area. The interpretation of the scaling dimensions was based on the analysis of ratings of correlation coefficients of the variables with coordinates in space scaling. With the first dimension variables are significantly correlated like: x_1 (r = 0.65), x_5 (0.59), x_7 (0.76), x_9 (0.55), x_{14} (0.55), x_{15} (0.55). This dimension describes above all the social aspects of sustainable development.

When analyzing the position of the individual countries to dimension 1 quite a significant gap relative to a pattern among countries of southern Europe (Italy, Spain, Portugal, Greece and Malta) is observed. In recent years these countries have experienced a considerable slowdown covering various areas of socio-economic development. As a direct cause of the deterioration of the situation of the countries in this region in relation to e.g. Eastern European countries (e.g. Czech, Poland and Hungary) the lack of reforms and mistakes in the socio-economic policy and a significantly lower resistance to financial and economic crisis from the years 2007-2008 are mainly shown. The distribution of the EU in relation to dimension 1 also confirms these observations, in particular in the case of the position of Southern European countries (in most cases the furthest away from the model) and the countries of Eastern Europe (located closer to the model than the countries of southern Europe).

At the same time dimension 2 is significantly correlated with the following variables: x_2 (0.78), x_3 (0.62), x_4 (0.57), x_6 (0.57) and x_{17} (0.67). These features represent different areas and may be linked to the economic areas of sustainable development. This dimension can be interpreted by analyzing e.g. the position of countries such as Sweden and Slovenia, which are located very close to each other in relation to this dimension (see Figure 1). Slovenia is the only country located in southern Europe which dealt relatively well with the crisis of 2007-2008 and Sweden is one of the most developed EU Member States. They obtained relatively similar results in the case of indicators describing the area of education (x_6) and economic aspects (generation of waste and CO2 emissions), environmental (x_2, x_{17}). A similar position with respect to this dimension was also observed in the case of countries such as Luxembourg (like Sweden one of the "richest" and most developed countries of the EU) and Belgium (relatively well-developed country located in western Europe). These countries had similar levels of indicators describing the area associated with functioning in the labor market (x_9 and x_{12}). In both dimensions the

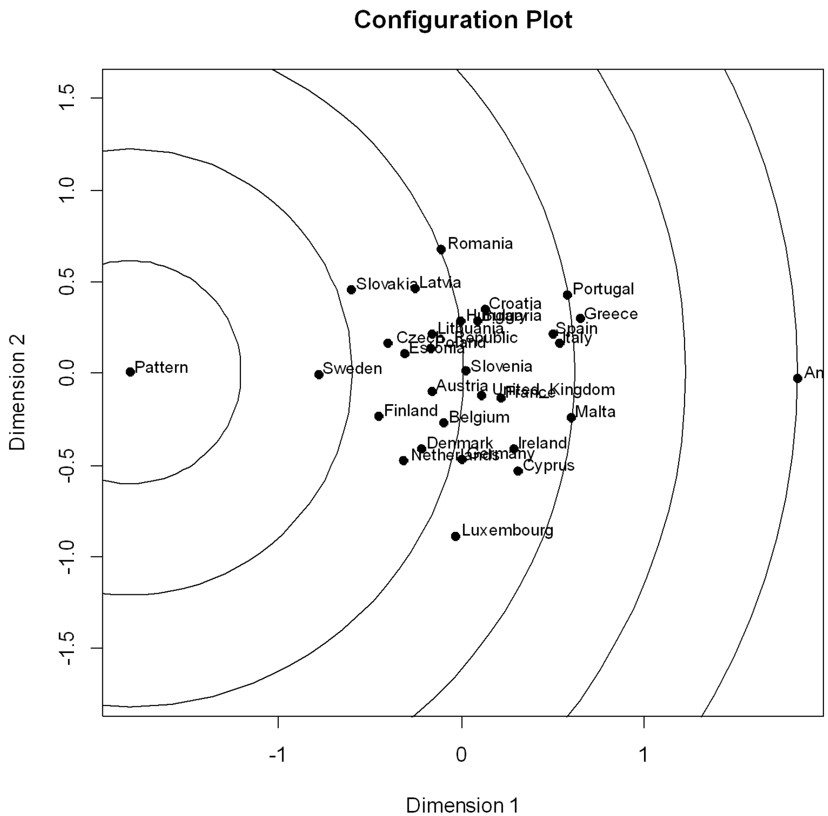

Fig. 1. The results of multidimensional scaling of 30 facilities (28 EU countries, pattern, anti-pattern)

Source: own elaboration using R program.

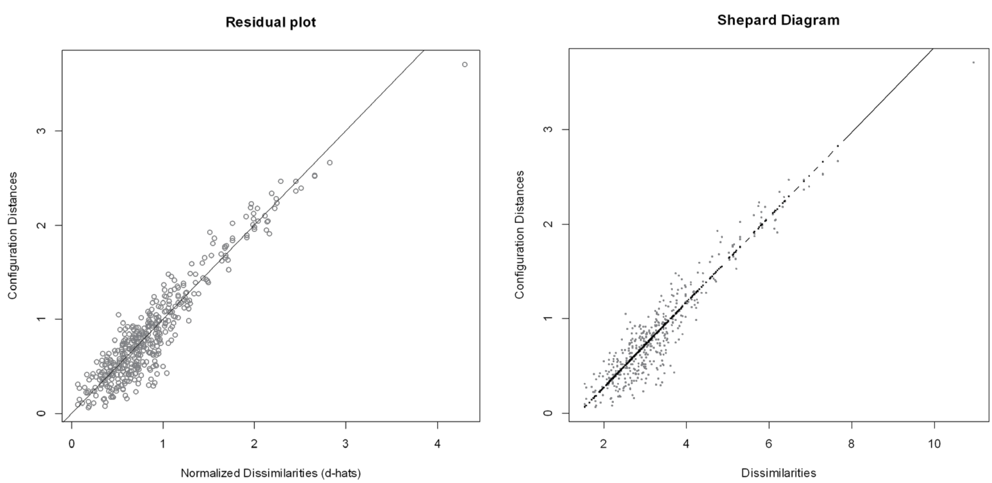

Fig. 2. Graph matching distance (left) and Shepard diagram (right)

Source: own elaboration using R program.

closest located to the model is Sweden, which is by far "ahead of" other EU Member States. In contrast, the closest located to anti-pattern (taking into account both dimensions) are such countries as Malta, Italy and Spain.

In assessing the credibility of the results of multidimensional scaling a graph matching distance Residual Plot and Shepard diagram was used (see Figure 2). The formation of spread points as close to a straight line ($y = x$) confirms the quality of the results of multidimensional scaling. In the case of the conducted research the two charts confirm a good fit distances mapped as a result of multidimensional scaling to the distance observed (real). In particular, the scatter of points on the Shepard diagram confirms the quality of the obtained scaling and a good choice of a 2-dimensional space to present the relations between the studied objects (EU countries).

5. Conclusions

Analyzing individual indicators of the sustainability of the EU does not allow for a clear interpretation of the results. The European Union countries which are included in the study take different positions in terms of the analyzed 17 indicators. The solution in this case is the use of methods of multivariate statistical analysis, which allows the study of objects and complex phenomena in which the state and behavior also affect many features.

In the study a multidimensional scaling method was used, which allowed, among others, to identify the similarities and differences between the analyzed EU countries. As a result of its use countries which development was similar or different to the designated dimensions and those that were located closest in relation to a pattern and anti-pattern were identified. Also relevant is the course of used procedure in the study, which allowed to analyze up to 80 scaling procedures.

References

Borg I., Groenen P., 2005, *Modern Multidimensional Scaling. Theory and Applications*, vol. 2, Springer-Verlag, New York.

Cox T.F., Cox M.A., 1994, *Multidimensional Scaling*, Chapman and Hall, London.

Gatnar G., Walesiak M., 2004, *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych,* Wyd. Akademii Ekonomicznej we Wrocławiu, Wrocław.

Gatnar E., Walesiak M., 2011, *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, C.H. Beck, Warszawa.

Grabinski T. 1992, *Metody taksonometrii,* PWE, Warszawa.

Green P.E., Rao V.R., 1972, *Applied Multidimensional Scaling*, Holt and Winston, New York.

Hair J.F., Anderson R.E., Tatham R.L., Black W.C., 1995, *Multivariate Data Analysis with Readings*, Prentice-Hall, Englewood Cliffs.

Hellwig Z., 1968, *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, Przegląd Statystyczny, no. 4.

Hellwig Z., 1981, *Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych*, [ed:] W. Welfe, *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną*, PWE Warszawa.

Herfindahl O.C., 1950, *Concentration in the U.S. Steel Industry,* unpublished doctoral dissertation, Columbia University.

Hirschman A.O., 1964, *The Paternity of an index*, American Economic Review, vol. 54.

Kruskal J.B., 1964, *Multidemensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis,* Psychometrika, no. 29.

R Development Core Team, 2011, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, URL: http://cran.r-project.org/.

Walesiak M., Gatnar E. (eds.), 2009, *Statystyczna analiza danych z wykorzystaniem programu R*, PWN, Warszawa.

Walesiak M., 2006, *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, Wydawnictwo AE, Wrocław.

Walesiak M., 2016, *Visualization of linear ordering results for metric data with the application of multidimensional scaling*, Ekonometria 2(52), Wyd. UE we Wrocławiu, Wrocław.

Walesiak M., 2017, *Wizualizacja wyników porządkowania liniowego dla danych porządkowych z wykorzystaniem skalowania wielowymiarowego*, Przegląd Statystyczny, iss. 1(2017), Komitet Statystyki i Ekonometrii PAN, Warszawa.

Walesiak M., Dudek A., 2015, clusterSim R package, URL: https://cran.r-project.org/web/packages/clusterSim/.

Walesiak M., Dudek A., 2016, *Wybór optymalnej procedury skalowania wielowymiarowego dla danych metrycznych z wykorzystaniem programu R*, an article presented at the conference "Multivariate Statistical Analysis MSA 2016", Łódź.

Woźniak A., Sikora J., 2006, *Wykorzystanie skalowania wielowymiarowego w analizie potencjału infrastrukturalnego gospodarstw rolnych w wybranych gminach województwa małopolskiego,* Infrastruktura i ekologia terenów wiejskich, no. 3/2.

Zaborski. A., 2001, *Skalowanie wielowymiarowe w badaniach marketingowych*, Wydawnictwo AE, Wrocław.