STATISTICS IN TRANSITION new series, June 2019
Vol. 20, No. 2, pp. 49–67, DOI 10.21307/stattrans-2019-014

A COMPARATIVE ANALYSIS OF ECONOMIC EFFICIENCY OF MEDIUM-SIZED MANUFACTURING ENTERPRISES IN DISTRICTS OF WIELKOPOLSKA PROVINCE USING THE HYBRID APPROACH WITH METRIC AND INTERVAL-VALUED DATA

Grażyna Dehnel[1], Marek Walesiak[2]

ABSTRACT

The article describes a hybrid approach to evaluating economic efficiency of medium-sized manufacturing enterprises (employing from 50 to 249 people) in districts of Wielkopolska province, using metric and interval-valued data. The hybrid approach combines multidimensional scaling with linear ordering. In the first step, multidimensional scaling is applied to obtain a visual representation of objects in a two-dimensional space. In the next step, a set of objects is ordered linearly based on the distance from the pattern (ideal) object. This approach provides new possibilities for interpreting linearly ordered results of a set of objects. Interval-valued variables characterise the objects of interests more accurately than metric data do. Metric data are atomic, i.e. an observation of each variable is expressed as a single real number. In contrast, an observation of each interval-valued variable is expressed as an interval. The analysis was based on data prepared in a two-stage process. First, a data set of observations was obtained for metric variables describing economic efficiency of medium-sized manufacturing enterprises. These unit-level data were aggregated at district level (LAU 1) and turned into two types of data: metric and interval-valued data. In the analysis of interval-valued data, two approaches are used: symbolic-to-classic, symbolic-to-symbolic. The article describes a comparative analysis of results of the assessment of economic efficiency based on metric and interval-valued data (the results of two approaches). The calculations were made with scripts prepared in the R environment.

Key words: medium-sized enterprise, metric data, interval-valued data, multidimensional scaling, composite measures

JEL: C38, C43, C63, C88, R12

[1] Poznan University of Economics and Business, Department of Statistics, Poznań. E-mail: grazyna.dehnel@ue.poznan.pl. ORCID ID: https://orcid.org/0000-0002-0072-9681.
[2] Wroclaw University of Economics, Department of Econometrics and Computer Science, Jelenia Góra. E-mail: marek.walesiak@ue.wroc.pl. ORCID ID: https://orcid.org/0000-0003-0922-2323.

1. Introduction and motivation

The contribution made to the GDP by small and medium-sized enterprises keeps growing, in contrast to that of large companies. Although the SME sector is dominated by micro enterprises, one cannot ignore the role played by medium-sized companies, employing between 50 and 249 persons (CSO 2017). At present there are nearly 16,000 medium-sized companies in Poland, which accounts for just 0.8% of the entire enterprise sector. This share has remained unchanged for the last 10 years (MED 2017). Medium-sized companies provide more jobs than the small ones (17%). An average medium-sized enterprise employs 104 persons, while the total number of people employed in companies of this category is 1.6 million. Investment outlays in this category account for 33% of the entire enterprise sector, 64% of which are own funds (see Figure 1). Medium-sized enterprises are the most dynamically developing category of companies in terms of the value of exports per one company. They are also characterized by the highest survival rate – 87% of them survive their first year of operation. Medium-sized companies operating for 5 years are likely to survive the next year with a probability of 0.996 (Chaber *et al*. 2017).

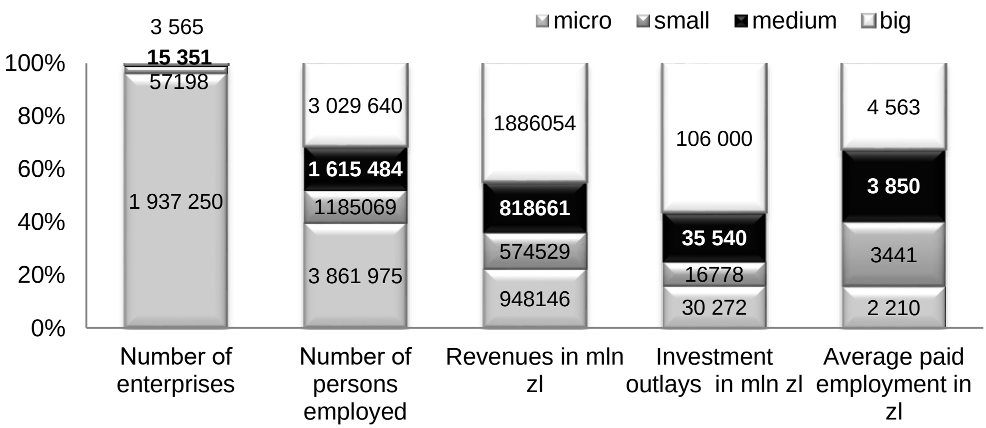

Figure 1. Enterprise characteristics by size class in 2016 (at 31 Dec.)

Source: Based on the CSO study (CSO 2017).

Medium-sized companies are able to compete with large enterprises because they are more flexible and efficient in conducting business activity, are better at controlling costs and take less time to implement innovation and react to changing market requirements.

Taking into account the kind of business activity, one of the most important sections is manufacturing. Looking at the structure of manufacturing companies (see Figure 2), it can be seen that medium-sized enterprises are the smallest group and make up only 3% of all units in this section. People employed by medium-sized manufacturing enterprises account for about 27% of the workforce working in all manufacturing companies. Revenues earned by medium-sized manufacturing enterprises make up 21% of all revenues generated by companies in the manufacturing section.

The empirical study described below is limited to the group of medium-sized manufacturing enterprises, which includes 42% of all medium-sized companies. Those companies employ 44% of the workforce working in this sector. The share of revenues and wages in this group is similar (CSO 2017).

The main objective of the study was to evaluate the economic efficiency of medium-sized manufacturing enterprises in districts of Wielkopolska province. The study was based on metric and interval-valued data and involved a hybrid approach combining multidimensional scaling and linear ordering (Walesiak 2016; Walesiak, Dehnel 2018). Economic efficiency, defined as a relation between effects and investments, in this case, is measured on an operational level using efficiency ratios to assess the company's performance (Kaplan, Cooper 1998; Kaplan 2008; Koliński 2011). Studies of this kind are usually based on a matrix of metric data. The novelty of the present study is the fact that it was based on a table of interval-valued data. In addition, the authors propose an aggregate measure based on the Euclidean Ichino-Yaguchi distance from the pattern object. Interval-valued variables describe objects of interest more accurately than metric data do, which are atomic, meaning that an observation of each variable is expressed as a single real number. In contrast, an observation of each interval-valued variable is expressed as an interval. The following studies (Gioia, Lauro 2006; Brito *et al*. 2015) include real examples of interval-valued data.

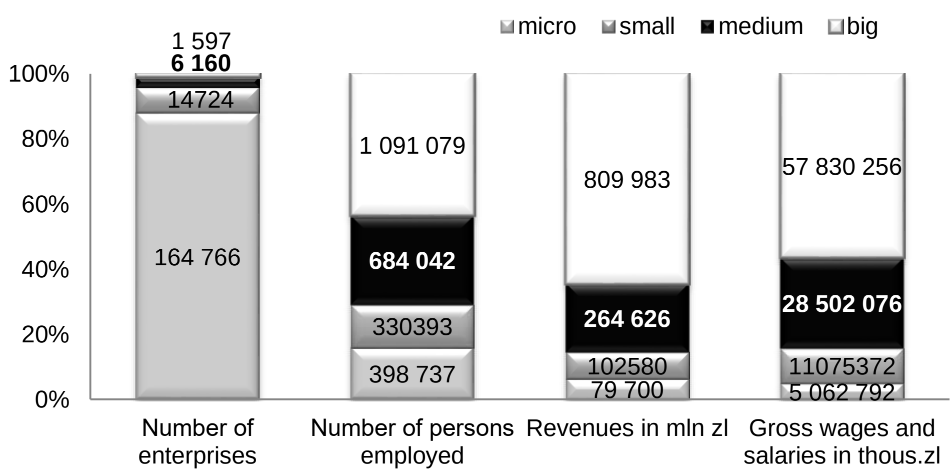

Figure 2. Characteristics of manufacturing enterprises by size class in 2016 (at 31 Dec.)

Source: Based on the CSO study (CSO 2017).

Data for the study were prepared in two steps. The first step involved compiling a set containing metric variables about the economic efficiency of medium-sized manufacturing enterprises; in the second step, the collected data were aggregated at the level of districts, producing metric and interval-valued data. The latter type of data was analysed using two approaches: symbolic-to-classic and symbolic-to-symbolic. Data used in the study come from the DG-1 survey conducted by the Statistical Office in Poznań. The survey is carried out to collect information about basic measures of economic activity in companies (Dehnel 2015). Owing to data availability, the study was conducted for 2012. The official statistics were supplemented by information from the register maintained by the Ministry of Finance.

2. Research methodology

To produce a ranking of medium-sized manufacturing companies operating in districts of Wielkopolska province in terms of economic efficiency, the authors used a hybrid approach, which combines multidimensional scaling (MDS) and linear ordering (Walesiak 2016; Walesiak, Dehnel 2018), which makes it possible to visualize the results of linear ordering. Metric and interval-valued data were used for this purpose. Depending on the type of input and output of multidimensional scaling, three different approaches were used to analyse the data:

a. Classic-to-classic (cc) for metric data,
b. Symbolic-to-classic (sc) for interval-valued data,
c. Symbolic-to-symbolic (ss) for interval-valued data.

The extended analytical procedure (including the above mentioned approaches), accounting separately for metric and interval-valued data, consists of the following steps:

1. Select a complex phenomenon which cannot be measured directly (in this case, it is the economic efficiency of medium-sized manufacturing companies operating in districts of Wielkopolska province).
2. Identify a set of objects of interest and a set of variables that are substantively related to the complex phenomenon. Add a pattern object (upper pole) and an anti-pattern object (lower pole) to the set of objects. Identify preference variables[3] (stimulants, destimulants and nominants).
3. Collect data and construct a data matrix $\mathbf{X} = [x_{ij}]_{nxm}$, (the value of the *j*-th variable for the *i*-th object, $i, k = 1, \dots, n$, $j = 1, \dots, m$) for metric data or a data table $\mathbf{X} = [x_{ij}^l, x_{ij}^u]_{nxm}$ (where $x_{ij}^l \leq x_{ij}^u$) for interval-valued data. The pattern object includes the most favourable variable values, whereas the anti-pattern – the least favourable values of the preference variables (separately for lower and upper bounds of the interval).
4. Normalize variable values and arrange them in the form of a normalized data matrix $\mathbf{Z} = [z_{ij}]_{nxm}$ for metric data or in the form of a normalized data table $\mathbf{Z} =$

3 The idea of a stimulant and a destimulant was introduced by (Hellwig 1972), while that of a nominant in the work by (Borys 1984, p. 118). Definitions can be found, among others, in (Walesiak 2016).

$[z_{ij}^l, z_{ij}^u]_{nxm}$ (where $z_{ij}^l \leq z_{ij}^u$) for interval-valued data. Normalization is used to ensure comparability of variables. This is achieved by removing dimensional units from measurement results and standardizing their orders of magnitude. Interval-valued data require special normalization treatment. The lower and upper bound of the interval of the *j*-th variable for *n* objects are combined into one vector containing 2*n* observations. This approach makes it possible to apply normalization methods used for classic metric data. Metric data were normalized using the data.Normalization function, while interval-valued data – using interval_normalization function, both available in the clusterSim package (Walesiak, Dudek 2018a).

5. In the classic-to-classic approach, select a measure of distance for metric data (Manhattan, Euclidean, Chebyshev, Squared Euclidean, GDM1[4] – see, e.g. Everitt *et al*. 2011, pp. 49-50), calculate distances and create a distance matrix $\boldsymbol{\delta} = [\delta_{ik}(\mathbf{Z})]_{nxn}$ $(i, k = 1, \ldots, n)$.
   For interval-valued data (the symbolic-to-classic approach), select a measure of distance (see Table 1), calculate distances and create a distance matrix $\boldsymbol{\delta} = [\delta_{ik}(\mathbf{Z})]_{nxn}$.

Table 1. Selected distance measures for interval-valued data

Symbol	Name	Distance measure $\delta_{ik}(\mathbf{Z})$
U_2_q1	Ichino-Yaguchi $q = 1, \gamma = 0{,}5$	$\sum_{j=1}^{m} \varphi(z_{ij}, z_{kj})$
U_2_q2	Euclidean Ichino-Yaguchi $q = 2, \gamma = 0{,}5$	$\sqrt{\sum_{j=1}^{m} \varphi(z_{ij}, z_{kj})^2}$
H_q1	Hausdorff $q = 1$	$\sum_{j=1}^{m} [\max(\lvert z_{ij}^l - z_{kj}^l \rvert, \lvert z_{ij}^u - z_{kj}^u \rvert)]$
H_q2	Euclidean Hausdroff $q = 2$	$\left\{\sum_{j=1}^{m} [\max(\lvert z_{ij}^l - z_{kj}^l \rvert, \lvert z_{ij}^u - z_{kj}^u \rvert)]^2\right\}^{1/2}$

$z_{ij} = [z_{ij}^l, z_{ij}^u]$; $\varphi(z_{ij}, z_{kj}) = |z_{ij} \oplus z_{kj}| - |z_{ij} \otimes z_{kj}| + \gamma(2 \cdot |z_{ij} \otimes z_{kj}| - |z_{ij}| - |z_{kj}|)$; $|\quad|$ – interval length; $z_{ij} \oplus z_{kj} = z_{ij} \cup z_{kj}$; $z_{ij} \otimes z_{kj} = z_{ij} \cap z_{kj}$.

Source: Based on works by Billard, Diday 2006; Ichino, Yaguchi 1994.

This step does not apply in the symbolic-to-symbolic approach.

6. In the classic-to-classic and symbolic-to-classic approaches conduct multidimensional scaling (MDS): $f: \delta_{ik}(\mathbf{Z}) \rightarrow d_{ik}(\mathbf{V})$ for all pairs (i, k), where *f* denotes distance mapping from *m*-dimensional space $\delta_{ik}(\mathbf{Z})$ into

4 Cf. Jajuga, Walesiak, Bąk 2003.

corresponding distances $d_{ik}(\mathbf{V})$ in q- dimensional space ($q < m$). To enable graphic presentation of results, q is set to 2. Distances $d_{ik}(\mathbf{V})$ are unknown. The iterative procedure, implemented in the **smacof** algorithm and used to find configuration $\mathbf{V}$ (given q dimensions) and calculate distance matrix $d_{ik}(\mathbf{V})$, is presented in (Borg, Groenen 2005, pp. 204–205).

In the classic-to-classic and symbolic-to-classic approaches, after performing MDS, one obtains a data matrix in 2-dimensional space: $\mathbf{V} = [v_{ij}]_{nxq}$ ($q = 2$). Depending on the location of the pattern and anti-pattern object in the dimensional scaling space $\mathbf{V} = [v_{ij}]_{nx2}$ the coordinate system needs to be rotated by an angle of φ according to the formula:

$$[v'_{ij}]_{nx2} = [v_{ij}]_{nx2} \times D, \quad (1)$$

where: $[v'_{ij}]_{nx2}$ – data matrix in 2-dimensional scaling space after rotating the coordinate system by an angle of φ,

$D = \begin{bmatrix} cos\varphi & -sin\varphi \\ sin\varphi & cos\varphi \end{bmatrix}$ – rotation matrix.

The rotation does not change the arrangement of objects relative to one another but makes it possible to position the set axis connecting the pattern and anti-pattern along the identity line, which improves the visualization of results.

In the symbolic-to-symbolic approach, multidimensional scaling needs to be performed using the I-Scal algorithm. The objective of MDS for interval dissimilarities is to represent the lower and upper bounds of the dissimilarities by minimum and maximum distances between rectangles as well as possible distances in the sense of least-squares (Groenen, Winsberg, Rodriguez, Diday 2006).

Under this approach, after performing MDS, one obtains an interval-valued data table in 2-dimensional space $\mathbf{V} = [v_{ij}^l, v_{ij}^u]_{nxq}$ (where $v_{ij}^l \leq v_{ij}^u$; $q = 2$).

A frequent mistake committed while using MDS results is to evaluate stress mechanically (rejecting an MDS solution because its stress seems “too high”). According to Borg, Groenen, Mair (2013, p. 68; 2018, pp. 85-86) “an MDS solution can be robust and replicable, even if its stress value is high” and “Stress, moreover, is a *summative* index for *all* proximities. It does not inform the user how well a *particular* proximity value is represented in the given MDS space”. In addition we should take into account stress per point measure[5] and Shepard diagram[6] (classic-to-classic and symbolic-to-classic approaches) or the I-Stress per box index and the I-dist diagram (the symbolic-to-symbolic approach).

In this study, we used a solution which enables the selection of an optimal MDS procedure for a given normalization method, distance measure and scaling models (in the classic-to-classic and symbolic-to-classic approaches) and, in the case of the symbolic-to-symbolic approach, according to procedures available in the `mdsOpt` R package (Walesiak, Dudek 2018b).

[5] Cf. Borg and Mair (2017, pp. 31).
[6] Cf. Mair, Borg and Rusch (2016).

7. In all three approaches, MDS results should be presented graphically in a 2-dimensional space and interpreted.

   In the classic-to-classic and symbolic-to-classic approaches, objects are represented as points. Two points, representing the anti-pattern and pattern, are joined by a straight line to form the so-called set axis in the diagram. Isoquants of development (curves of equal development) are drawn from the pattern point. Objects located between the isoquants represent a similar level of development. The same level can be achieved by objects located at different points along the same isoquant of development (due to a different configuration of variable values).
   In the symbolic-to-symbolic approach, objects are represented in the form of rectangles.
8. In the classic-to-classic and symbolic-to-classic approaches, objects should be ordered linearly according to the values of the aggregate measure d_i based on the Euclidean distance from the pattern object (Hellwig 1981):

$$d_i = 1 - \sqrt{\sum_{j=1}^{2}(v_{ij} - v_{+j})^2} \Big/ \sqrt{\sum_{j=1}^{2}(v_{+j} - v_{-j})^2}, \qquad (2)$$

   where: v_{ij} – the *j*-th coordinate for the *i*-th object in the 2-dimensional MDS space, $v_{+j}(v_{-j})$ – the *j*-th coordinate for the pattern (anti-pattern) object in the 2-dimensional MDS space.
   In the symbolic-to-symbolic approach, objects should be ordered according to the values of the aggregate measure d_i based on the Euclidean Ichino-Yaguchi distance (Ichino, Yaguchi 1994) from the pattern object:

$$d_i = 1 - \sqrt{\sum_{j=1}^{2}\varphi\left(v_{ij}, v_{+j}\right)^2} \Big/ \sqrt{\sum_{j=1}^{2}\varphi\left(v_{+j}, v_{-j}\right)^2}, \qquad (3)$$

   where: $v_{ij} = [v_{ij}^l, v_{ij}^u]$; $v_{+j} = [v_{+j}^l, v_{+j}^u]$; $v_{-j} = [v_{-j}^l, v_{-j}^u]$;

   v_{ij}^l and v_{ij}^u – the lower and upper bound of the interval of the *j*-th variable for the *i*-th object in the 2-dimensional MDS space;

   v_{+j}^l and v_{+j}^u (v_{-j}^l i v_{-j}^u) – the lower and upper bound of the interval of the *j*-th variable for the pattern (anti-pattern) object in the 2-dimensional MDS space.

   The values of the aggregate measure d_i given by (2) and (3) belong to the interval $[0; 1]$. The higher the value of d_i, the higher the economic efficiency of medium-sized manufacturing enterprises in the objects (districts). The objects are arranged according to the descending values of the aggregate measure d_i.

3. Results of the Empirical Study

The empirical study uses statistical data about the economic efficiency of medium-sized manufacturing enterprises in districts of Wielkopolska province in 2012. The target data set was prepared in two stages. The first step involved selecting three metric variables (x1 and x2 are stimulants and x3 is a destimulant) describing the economic efficiency of 876 medium-sized manufacturing enterprises:

x1 – return on sales in % (net profit as a percentage of sales revenue).

x2 – sales revenue in thousands PLN per one employee,

x3 – costs in thousands PLN per one employee.

In the second step, the observations were aggregated at the level of districts producing a set of interval-valued data. The economic efficiency of medium-sized manufacturing enterprises operating in 35 districts of Wielkopolska province was measured using three approaches: classic-to-classic, symbolic-to-classic and symbolic-to-symbolic.

In the classic-to-classic approach, the analytical procedure described in the second section was applied to a data matrix containing 35 districts of Wielkopolska province described by the three metric variables. For this purpose, original data for 876 manufacturing enterprises were aggregated at the level of districts by averaging the values of each variable.

In the symbolic-to-classic and symbolic-to-symbolic approaches, the analytical procedure described in the second section was applied to a table containing 35 districts of Wielkopolska province described by the three interval-valued variables. Original data for 876 manufacturing enterprises were aggregated at the level of districts, producing interval-valued data. The lower bound of the interval for each interval-valued variable in each district was given by the first quartile of the entire data set. The upper bound of the interval was obtained by calculating the third quartile.

In the classic-to-classic approach, an optimal scaling procedure was selected after testing combinations of 6 normalization methods (n1, n2, n3, n5, n5a, n12a – see Walesiak, Dudek 2018a), 4 distance measures (Manhattan, Euclidean, Chebyshev, Squared Euclidean, GDM1) and 4 MDS models (ratio, interval, mspline of second and third degree – Borg, Groenen 2005, p. 202) – altogether 120 MDS procedures. As a result of applying the `optSmacofSym_mMDS` function from the **`mdsOpt`** R package (see Walesiak, Dudek 2017; 2018b), the optimal MDS procedure was selected. The procedure uses the normalization method n2 (positional standardization), the mspline 2 scaling model (polynomial of second degree) and the GDM1 distance.

In the symbolic-to-classic approach, an optimal scaling procedure was selected after testing combinations of 6 normalization methods (n1, n2, n3, n5, n5a, n12a), 4 distance measures (Ichino-Yaguchi, Euclidean Ichino-Yaguchi, Hausdorff, Euclidean Hausdorff) and 4 MDS models (ratio, interval, mspline of second and third degree) – altogether 96 MDS procedures. After applying the `optSmacofSymInterval` function from the **`mdsOpt`** R package, the optimal MDS procedure was selected, which involves the normalization method n12a (positional normalization), the mspline 2 scaling model (polynomial of third degree) and the Hausdorff distance.

In the symbolic-to-symbolic approach, an optimal scaling procedure was selected after testing combinations of 6 normalization methods (n1, n2, n3, n5, n5a, n12a) and 2 optimization methods, giving altogether 12 MDS procedures. After applying the `optIscalInterval` function from the **`mdsOpt`** R package, the optimal MDS procedure was selected, which uses the normalization method n1 (standardization) and the MM optimization method (majorization-minimization algorithm).

By taking into account all the three approaches, it was possible to see how assessments of the phenomenon of interest varied when moving from the classic-to-classic approach to more robust ones (symbolic-to-classic, symbolic-to-symbolic). The average value, used in the classic-to-classic approach as the only parameter, which is well known, is strongly affected by outliers. In the other two approaches based on interval-valued data, assessments obtained for districts are not based on average values but account for the variation observed among manufacturing enterprises with respect to the variables of interest. Additional advantage of these approaches is the fact that outliers are excluded from the analysis.

Figures 3, 4 and 5 present MDS results of districts of Wielkopolska province for each approach.

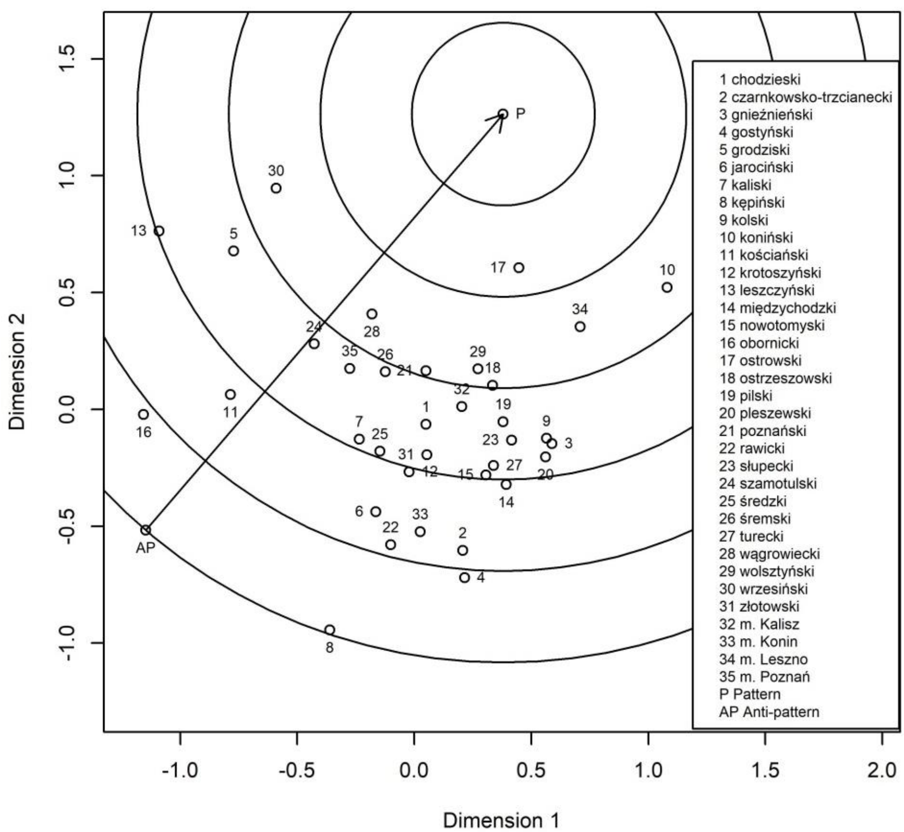

Figure 3. Results of multidimensional scaling of 35 districts of Wielkopolska by economic efficiency of medium-sized manufacturing enterprises in 2012 – the classic-to-classic approach

In the diagram illustrating the classic-to-classic and the classic-to-symbolic approaches, the anti-pattern (AP) and pattern (P) objects were connected by a straight line – the so-called set axis (Figs. 3 and 4). 6 isoquants of development were identified by dividing the set axis into 6 equal parts. The further a given isoquant is located from the pattern object, the less economically efficient are medium-sized companies in districts represented within it.

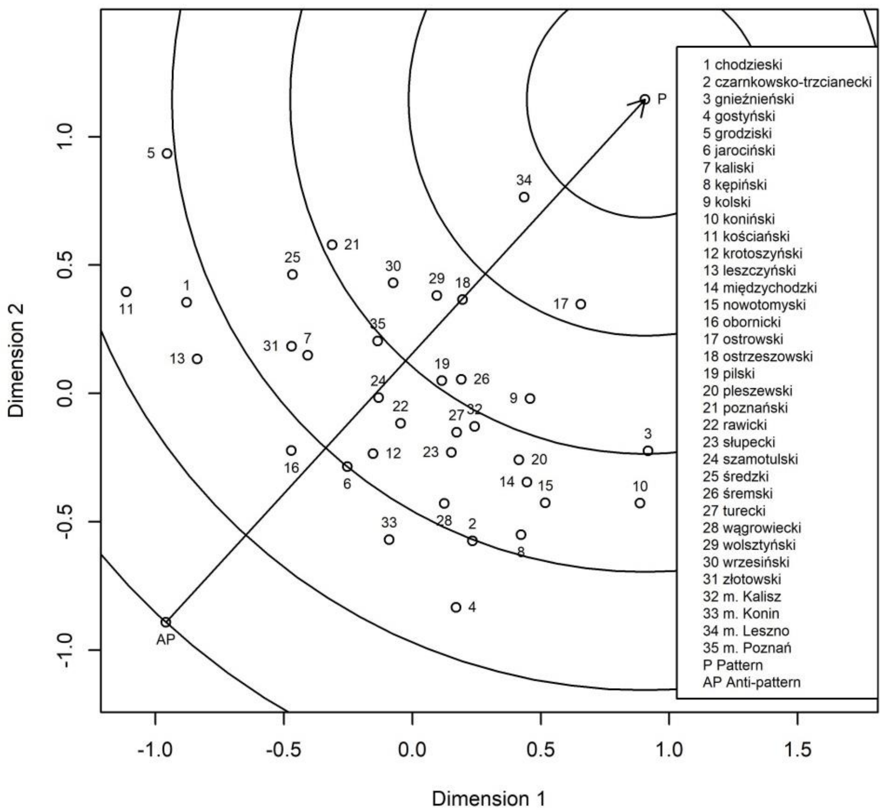

Figure 4. Results of multidimensional scaling of 35 districts of Wielkopolska by economic efficiency of medium-sized manufacturing enterprises in 2012 – the symbolic-to-classic approach

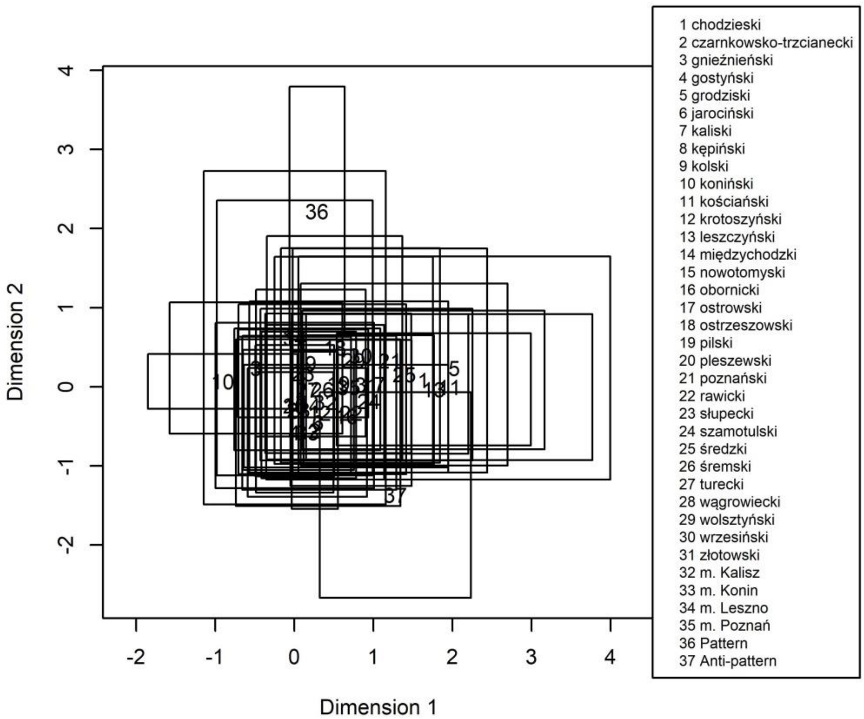

Figure 5. Results of multidimensional scaling of 35 districts of Wielkopolska by economic efficiency of medium-sized manufacturing enterprises in 2012 – the symbolic-to-symbolic approach

By presenting results in this way it is possible to:

- show a graphical ordering of districts in terms of the economic efficiency of manufacturing enterprises measured by three variables according to the values of measure d_i (2),
- distinguish groups of districts with a similar level of economic efficiency (districts between isoquants),
- identify districts characterized by a similar level of economic efficiency, but having a different location on the isoquant of development. Example cases in the classic-to-classic approach include Leszczyński district (13) and Międzychodzki district (14), while in the symbolic-to-classic approach – Grodziski (5) and Jarociński (6) districts. Although the assessment of economic efficiency for these pairs of districts is similar, their respective configurations of values differ.

The visualization of results also reveals that a switch from the classic-to-classic approach to the symbolic-to-classic approach causes a change in the position of objects, and consequently, different assessments of economic efficiency. This is due to the fact that the analysis in the symbolic-to-classic approach is based on the values of the target variables included between the first and third quartile. At least two directions of changes can be observed in the arrangement of objects. Some objects moved along the set axis or relative to it (an object moving closer to, further away from or crossing the set axis). A large majority of the objects (24) moved towards the pattern (higher values of measure d_i), which, in the symbolic-to-classic approach, represents a higher level of economic efficiency of companies. The group of districts with the highest increase in the value of measure d_i includes those in which companies were assessed as least economically efficient in the classic-to-classic approach: Kępiński, Obornicki, Rawicki (Figures 3, 4 and 6, Table 2). A reverse change, i.e. a shift towards the anti-pattern, was observed for 9 districts, which in the classic-to-classic approach received the highest assessment of economic efficiency of companies: Ostrowski, Wrzesiński, Koniński, Wągrowiecki (Figures 3, 4 and 6, Table 2).

Table 2 shows an ordering of 35 districts of Wielkopolska province depending on the economic efficiency of medium-sized manufacturing enterprises in 2012 obtained under the classic-to-classic, symbolic-to-classic and symbolic-to-symbolic approaches.

Table 2. Ranking of 35 districts of Wielkopolska by economic efficiency of medium-sized manufacturing enterprises in 2012

No.	Districts	d_i^{cc}	Rank	d_i^{sc}	Rank	d_i^{ss}	Rank
1	Chodzieski	0.4153	15	0.2941	31	0.3222	22
2	Czarnkowsko-Trzcianecki	0.1967	31	0.3329	28	0.2553	32
3	Gnieźnieński	0.3901	18	0.5050	10	0.4285	9
4	Gostyński	0.1498	33	0.2358	34	0.2452	34
5	Grodziski	0.4504	13	0.3231	29	0.2921	28
6	Jarociński	0.2367	29	0.3352	27	0.2584	31
7	Kaliski	0.3508	22	0.4030	22	0.3757	11
8	Kępiński	0.0063	35	0.3643	25	0.3185	24
9	Kolski	0.3956	17	0.5480	6	0.4757	5
10	Koniński	0.5640	4	0.4303	19	0.3389	18
11	Kościański	0.2843	28	0.2209	35	0.2530	33
12	Krotoszyński	0.3243	27	0.3682	24	0.2947	27
13	Leszczyński	0.3380	25	0.2712	33	0.2772	29
14	Międzychodzki	0.3303	26	0.4325	18	0.3384	19
15	Nowotomyski	0.3453	23	0.4116	21	0.3331	20
16	Obornicki	0.1459	34	0.2979	30	0.2593	30
17	Ostrowski	0.7186	1	0.6964	2	0.5373	1
18	Ostrzeszowski	0.4999	8	0.6169	3	0.5169	3
19	Pilski	0.4331	14	0.5101	9	0.3959	10
20	Pleszewski	0.3745	19	0.4605	14	0.3174	25

Table 2. Ranking of 35 districts of Wielkopolska by economic efficiency of medium-sized manufacturing enterprises in 2012 (cont.)

No.	Districts	d_i^{cc}	Rank	d_i^{sc}	Rank	d_i^{ss}	Rank
21	Poznański	0.5089	7	0.5124	8	0.4360	7
22	Rawicki	0.1879	32	0.4301	20	0.3053	26
23	Słupecki	0.4028	16	0.4340	17	0.4345	8
24	Szamotulski	0.4589	12	0.4356	16	0.3315	21
25	Średzki	0.3435	24	0.4479	15	0.3715	13
26	Śremski	0.4822	9	0.5270	7	0.3660	14
27	Turecki	0.3600	21	0.4608	13	0.3755	12
28	Wągrowiecki	0.5600	5	0.3635	26	0.3186	23
29	Wolsztyński	0.5276	6	0.5971	4	0.4639	6
30	Wrzesiński	0.5642	3	0.5618	5	0.4794	4
31	Złotowski	0.3626	20	0.3926	23	0.3535	16
32	m. Kalisz	0.4599	11	0.4819	12	0.3425	17
33	m. Konin	0.2204	30	0.2823	32	0.2411	35
34	m. Leszno	0.5921	2	0.7804	1	0.5190	2
35	m. Poznań	0.4607	10	0.4927	11	0.3551	15
Parameters		Value		Value		Value	
Mean		0.3840	X	0.4359	X	0.3579	X
Standard deviation		0.1450	X	0.1234	X	0.0821	X
Median		0.3901	X	0.4325	X	0.3389	X
Median absolute deviation		0.1047	X	0.1150	X	0.0694	X

d_i^{cc} – value of measure (2) in the classic-to-classic approach,
d_i^{sc} – value of measure (2) in the symbolic-to-classic approach,
d_i^{ss} – value of measure (3) in the symbolic-to-symbolic approach,

Source: Calculations performed in the R program (R Core Team 2018) and the clusterSim package (Walesiak, Dudek 2018a).

It can be seen that the application of the robust approaches (symbolic-to-classic and symbolic-to-symbolic) results in a different dispersion of objects. The range of d_i values changed from $[0.0063; 0.7186]$ in the classic-to-classic approach to $[0.2411; 0.5373]$ in the symbolic-to-symbolic approach, while the spread of districts expressed in terms of the standard deviation of measure d_i decreased from $S_{d_i} = 0.1450$ in the classic-to-classic approach to $S_{d_i} = 0.0821$ in the symbolic-to-symbolic approach.

The degree of correlation between the values of measure d_i for 35 districts obtained under each approach was measured by the Pearson correlation coefficient. The consistency of rank orders was measured by the Kendall rank correlation coefficient. The results are shown in Table 3.

Table 3. Correlation coefficients (Pearson's *r* and Kendall's *tau*) between the values of measures (2) and (3) obtained under the three approaches

Pearson correlation coefficient				Kendall rank correlation coefficient			
	cc	sc	ss		cc	sc	ss
cc	1.000	0.692	0.702	cc	1.000	0.546	0.543
sc	0.692	1.000	0.911	sc	0.546	1.000	0.741
ss	0.702	0.911	1.000	ss	0.543	0.741	1.000

cc – classic-to-classic approach,
sc – symbolic-to-classic approach,
ss – symbolic-to-symbolic approach.

The highest degree of similarity between rankings of districts (measured by the Kendall rank correlation coefficient) and correlation between districts (measured by the Pearson correlation coefficient) depending on the values of measure d_i is observed for the approaches based on interval-valued data (symbolic-to-classic and symbolic-to-symbolic). The results based on metric data are considerably different from those obtained using interval-valued data. The latter ones are more reliable (since districts were assessed on the basis of intervals of variable values with the exclusion of outliers) than those based on metric data (where districts were assessed on the basis of the mean values of the target variables).

The results of multidimensional scaling of districts of Wielkopolska province obtained under each approach along with the geographical location are presented in a map chart (Figure 6). One can clearly see the impact of Poznań on the neighbouring districts – it functions as a pole of growth (Isard 1960). Districts located further away from Poznań tend to appear lower in the ranking based on measure d_i. The only exceptions are Ostrowski and Ostrzeszowski districts, which, despite their relatively large distance from Poznań, are characterised by very high values of measure d_i regardless of the approach adopted ($d_{17}^{cc} = 0.7186$, $d_{17}^{sc} = 0.6964$, $d_{17}^{ss} = 0.5373$, and $d_{18}^{cc} = 0.4999$, $d_{18}^{sc} = 0.6169$, $d_{18}^{ss} = 0.5169$, respectively). It should be noted that these districts are part of the Kalisko-Ostrowski Industrial District and are important centres of electromechanical and construction industry. Another factor which may be contributing to the high economic efficiency of companies operating in these districts is that fact that they are located in a special economic zone (Kamiennogórska Subzone).

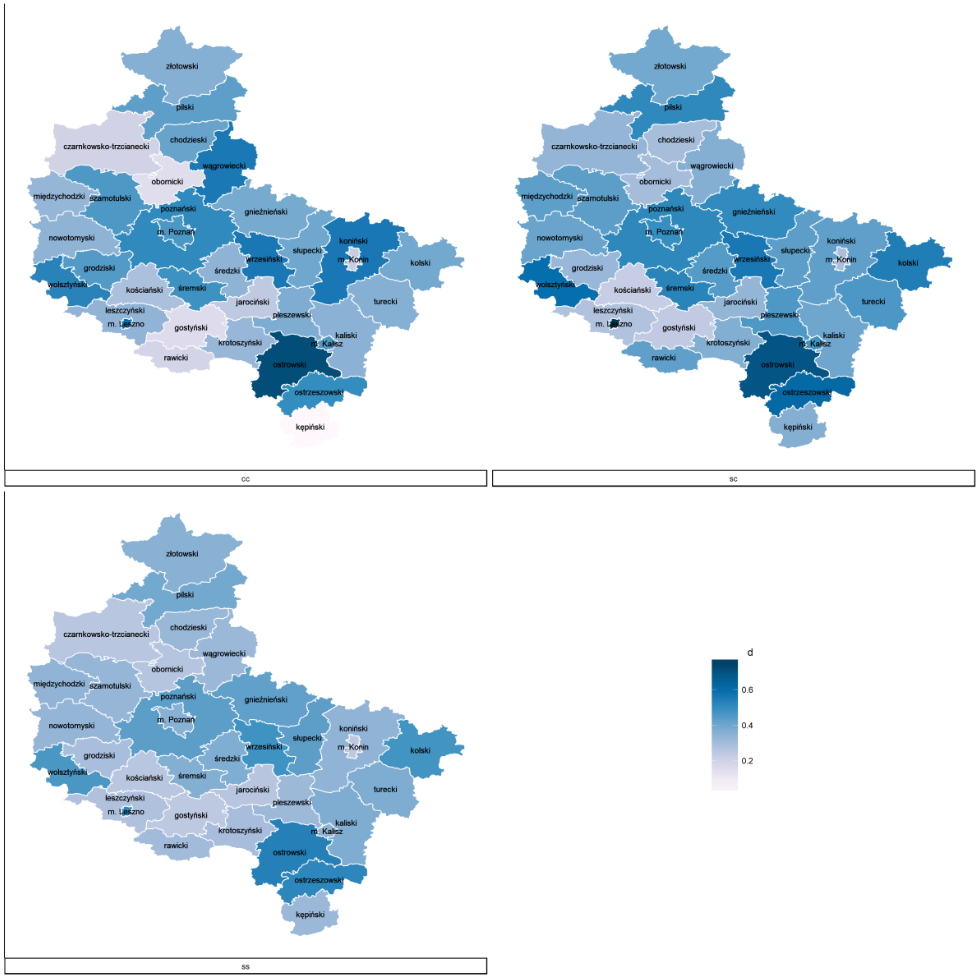

Figure 6. Assessment of districts in terms of economic efficiency of medium-sized manufacturing companies in 2012 in the classic-to-classic (cc), symbolic-to-classic (sc) and symbolic-to-symbolic (ss) approaches

Source: Calculations performed in the R program.

4. Conclusions

The aim of the study was to compare districts of Wielkopolska province in terms of the economic efficiency of medium-sized manufacturing companies, which operated in them in 2012. Variables used in the study are typically used in the financial analysis of economic entities. Assessments were obtained using a hybrid approach combining multidimensional scaling and linear ordering and performed for three types of data set-ups: classic-to-classic, symbolic-to-classic and symbolic-to-symbolic. Thanks to this methodology, it was possible to obtain a graphic presentation of economic efficiency, which is a multidimensional phenomenon, in a 2-dimensional space. In addition, the districts could be ranked according to the economic efficiency of medium-sized manufacturing companies.

By comparing results obtained under three different data set-ups, it was possible to identify changes caused by switching from the classic-to-classic approach to the interval-based approach (interval-valued data). In the two modified approaches, assessments were not based only on the mean values of the target variables describing companies in each district but accounted for the observed variation. Moreover, companies showing outlying values of the financial variables were excluded from the analysis.

The results were used to identify groups of districts with similar levels of economic efficiency and particular districts within the groups with similar and different values of the target variables. The analysis confirmed the impact of Poznań as a pole of growth on the neighbouring districts.

The authors are aware of the limitations resulting from the selected set of variables. However, the main purpose of the study was to present a new methodological approach.

Acknowledgements

The project is financed by the Polish National Science Centre DEC-2015/17/B/HS4/00905.

REFERENCES

BILLARD, L., DIDAY, E., (2006). Symbolic Data Analysis: Conceptual Statistics and Data Mining, John Wiley, Chichester, ISBN: 978-0-470-09016-9.

BORG, I., GROENEN, P. J. F., (2005). Modern Multidimensional Scaling. Theory and Applications. 2nd Edition, Springer Science+Business Media, New York, ISBN: 978-0387-25150-9, URL http://www.springeronline.com/0-387-25150-2.

BORG, I., GROENEN, P. J. F., MAIR, P., (2013). Applied Multidimensional Scaling, Springer, Heidelberg, New York, Dordrecht, London. ISBN 978-3-642-31847-4, URL http://dx.doi.org/10.1007/978-3-642-31848-1.

BORG, I., GROENEN, P. J. F., MAIR, P., (2018). Applied Multidimensional Scaling and Unfolding, Springer, Heidelberg, New York, Dordrecht, London. ISBN 978-3-319-73470-5, URL https://doi.org/10.1007/978-3-319-73471-2.

BORG, I., MAIR, P., (2017). The Choice of initial configurations in multidimensional scaling: local minima, fit, and interpretability, Austrian Journal of Statistics, 46 (2), pp. 19–32,
URL https://doi.org/10.17713/ajs.v46i2.561.

BORYS, T., (1984), Kategoria jakości w statystycznej analizie porównawczej, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 284, Seria:

Monografie i Opracowania nr 23, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław, ISBN: 83-7011-000-0.

BRITO, P., NOIRHOMME-FRAITURE, M., ARROYO, J., (2015). Editorial for special issue on symbolic data analysis, Advances in Data Analysis and Classification, Vol. 9, Issue 1, pp. 1–4, URL https://dx.doi.org/10.1007/s11634-015-0202-1.

CHABER, P., ŁAPIŃSKI, J., NIEĆ, M., ORŁOWSKA, J., ZAKRZEWSKI, R., WIDŁA-DOMARADZKI, Ł., DOMARADZKA, A., (2017). Raport o stanie sektora małych i średnich przedsiębiorstw w Polsce, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, URL https://badania.parp.gov.pl/raport-o-stanie-sektora-msp/stan-sektora-msp-w-polsce.

CSO, (2017). Działalność przedsiębiorstw niefinansowych w 2016 r. (Activity of Non-financial Enterprises in 2016), Central Statistical Office of Poland, Warszawa. URL http://stat.gov.pl/obszary-tematyczne/podmioty-gospodarcze-wyniki-finansowe/przedsiebiorstwa-niefinansowe/dzialalnosc-przedsiebiorstw-niefinansowych-w-2016-r-,2,12.html [Accessed 17 July 2018].

DEHNEL, G., (2015). Robust regression in monthly business survey, Statistics in Transition – new series, Vol. 16, No. 1, pp. 1–16.

EVERITT, B.S., LANDAU, S., LEESE, M., STAHL, D., (2011). Cluster Analysis, Wiley, Chichester, ISBN: 978-0-470-74991-3.

GIOIA, F., LAURO, C. N., (2006). Principal component analysis on interval data, Computational Statistics, 21 (2), pp. 343–363, URL https://doi.org/10.1007/s00180-006-0267-6.

GROENEN, P.J.F. WINSBERG, S., RODRIGUEZ, O., DIDAY, E., (2006), I-Scal: multidimensional scaling of interval dissimilarities, Computational Statistics & Data Analysis, 51 (1), pp. 360–378, URL http://dx.doi.org/10.1016/j.csda.2006.04.003.

HELLWIG, Z., (1972). Procedure of Evaluating High-Level Manpower Data and Typology of Countries by Means of the Taxonomic Method, [In:] Gostkowski Z. (ed.), Towards a system of Human Resources Indicators for Less Developed Countries, Papers Prepared for UNESCO Research Project, Ossolineum, The Polish Academy of Sciences Press, Wrocław, pp. 115–134.

HELLWIG, Z., (1981). Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych. In: Welfe, W. (ed.), Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną, PWE, Warszawa, pp. 46–68, ISBN 83-208-0042-0.

ICHINO, M., YAGUCHI, H., (1994). Generalized Minkowski metrics for mixed feature-type data analysis, IEEE Transactions on Systems, Man, and Cybernetics, 24 (4), pp. 698–708, URL http://dx.doi.org/10.1109/21.286391.

ISARD, W., (1960). Methods of Regional Analysis: An Introduction to Regional Science. Cambridge, Massachusetts: The M.I.T. Press.

JAJUGA, K., WALESIAK, M., (2000). Standardisation of Data Set under Different Measurement Scales, In: Decker, R., Gaul, W., (Eds.), Classification and Information Processing at the Turn of the Millennium, pp. 105–112, Springer-Verlag, Berlin, Heidelberg,
URL http://dx.doi.org/10.1007/978-3-642-57280-7_11.

JAJUGA, K., WALESIAK, M., BĄK, A., (2003). On the General Distance Measure, in Schwaiger, M., Opitz, O., (Eds.), Exploratory Data Analysis in Empirical Research. Berlin, Heidelberg: Springer-Verlag, pp. 104–109,
URL http://dx.doi.org/10.1007/978-3-642-55721-7_12.

KAPLAN, R. S., COOPER R., (1998). Cost & Effect: Using Integrated Cost Systems to Drive Profitability and Performance, Harvard Business School Press, ISBN: 978-0875847887.

KAPLAN, R. S., (2008). Conceptual foundations of the balanced scorecard. In: C. Chapman, A. Hopwood, M. Shields (Eds.), Handbook of Management Accounting Research, Vol. 3, Elsevier, ISBN: 9780080554501.

KOLIŃSKI, A., (2011). Przegląd metod i technik oceny efektywności procesu produkcyjnego, Logistyka, 5, pp. 1083–1091.

MAIR, P., BORG, I., RUSCH, T., (2016), Goodness-of-fit assessment in multidimensional scaling and unfolding, Multivariate Behavioral Research, Vol. 51, No. 6, pp. 772–789,
URL http://dx.doi.org/10.1080/00273171.2016.1235966.

MAIR, P., DE LEEUW, J., BORG, I., GROENEN, P. J. F., (2018). smacof: Multidimensional Scaling. R package ver. 1.10-8,
URL https://CRAN.R-project.org/package=smacof.

MED, (2017). Entrepreneurship in Poland, Ministry of Economic Development, Warsaw, URL https://www.mpit.gov.pl/strony/zadania/analiza-i-ocena-polskiej-gospodarki/przedsiebiorczosc/.

R CORE TEAM, (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org.

WALESIAK, M., (2016). Visualization of linear ordering results for metric data with the application of multidimensional scaling, Ekonometria [Econometrics], 2 (52), pp. 9–21, URL http://dx.doi.org/10.15611/ekt.2016.2.01.

WALESIAK, M., DEHNEL, G., (2018). Evaluation of Economic Efficiency of Small Manufacturing Enterprises in Districts of Wielkopolska Province Using Interval-Valued Symbolic Data and the Hybrid Approach. In M. Papież and S. Śmiech (Eds.), The 12th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings, Foundation of the Cracow University of Economics, Cracow, pp. 563-572,
URL http://dx.doi.org/10.14659/SEMF.2018.01.57.

WALESIAK, M., DUDEK, A., (2017). Selecting the optimal multidimensional scaling procedure for metric data with R environment, Statistics in Transition – new series, 18 (3), pp. 521–540,
URL http://dx.doi.org/10.21307/stattrans-2016-084.

WALESIAK, M., DUDEK, A., (2018a). clusterSim: Searching for Optimal Clustering Procedure for a Data Set. R package, version 0.47-2,
URL https://CRAN.R-project.org/package=clusterSim.

WALESIAK, M., DUDEK, A., (2018b). mdsOpt: Searching for Optimal MDS Procedure for Metric and Interval-valued Symbolic Data, R package, version 0.3-2, URL https://CRAN.R-project.org/package=mdsOpt.