Evaluation of economic efficiency of small manufacturing enterprises in districts of Wielkopolska province using interval-valued symbolic data and the hybrid approach

Marek Walesiak [1], Grażyna Dehnel[2]

Abstract

The article describes a hybrid approach to evaluating economic efficiency of small manufacturing enterprises (employing from 10 to 49 people) in districts of Wielkopolska province. The analysis was based on data prepared in a two-stage process. First, a dataset of 2,162 observations was obtained for three metric variables describing economic efficiency of small manufacturing enterprises. These unit-level data were aggregated at district level and turned into interval-valued symbolic data. Economic efficiency of small manufacturing enterprises was evaluated using a hybrid approach. In the first step, multidimensional scaling (see Borg and Groenen, 2005; Mair et al., 2017) 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 Euclidean distance from the pattern (ideal) object. The proposed approach provides new possibilities for interpreting linearly ordered results of a set of objects.

Keywords: *small enterprises, interval-valued symbolic variables, multidimensional scaling ,composite measures*

JEL Classification: C38, C43, C63

DOI: 10.14659/SEMF.2018.01.58

1 Introduction and motivation

The SME sector plays an important role in the development of the Polish economy, and the group of small businesses (employing between 10 and 49 people) is of particular interest in this respect. What makes small companies noteworthy is their ability to compete even with the largest enterprises thanks to a strict control of costs, their flexibility, which enables them to react quickly to changing market requirements and the ability to implement innovative solutions relatively quickly. At present about 57,000 small businesses are active in Poland, most of which tend to operate locally. In terms of industrial classification, manufacturing is one of the most important and also most numerous category of activity in this group: manufacturing companies account for 26% of all small businesses, generate 20% of total revenue, and provide 30% of jobs in this sector (Główny Urząd Statystyczny, 2017).

[1]Corresponding author: Wrocław University of Economics, Department of Econometrics and Computer Science, ul. Nowowiejska 3, 58-500 Jelenia Góra, Poland, marek.walesiak@ue.wroc.pl.

[2] Poznan University of Economics and Business, Department of Statistics, al. Niepodległości 10, 61-875 Poznań, Poland, g.dehnel@ue.poznan.pl.

Economic efficiency is defined as a relationship between effects and outlays, which, in this case, is measured on an operational level using efficiency ratios to assess the company's performance (Jaki, 2012; Koliński, 2011).

The article describes a study designed to evaluate the economic efficiency of small manufacturing enterprises in districts of Wielkopolska province. The evaluation was conducted using a hybrid approach combining multidimensional scaling (MDS) and linear ordering. Studies of this type are typically based on a classical data matrix. The novelty of the present study is the fact that is based on interval-valued symbolic data obtained in a two-stage process. Interval-valued variables describe objects of interest more precisely than classical metric variables. For classical metric data, an observation on the *j*-th variable for the *i*-th object in a data matrix is expressed as one real number. In contrast, for symbolic interval-valued data, observations on each variable are expressed as intervals $x_{ij} = [x_{ij}^l, x_{ij}^u](x_{ij}^l \leq x_{ij}^u)$, where x_{ij}^l denotes the lower bound and x_{ij}^u the upper bound of the interval. Studies by (Gioia and Lauro, 2006; Brito et al.,2015) provide different examples of data that in real life are of interval type. The empirical study described in this article was based on official statistics from a survey of small businesses. The survey is carried out to collect information about basic measures of economic activity in companies (Dehnel, 2015). The reference period for survey data was 2012. Another data source was the register maintained by the Ministry of Finance.

2 Research methodology

A two-step hybrid approach, presented by Walesiak (2016), enables the visualisation of linear ordering results. In this study it was adopted to order analysed objects. A research procedure that takes into account the specificity of interval-valued symbolic variables includes steps:

1. Select the research problem.

2. Select objects and interval-valued variables substantively related to the research problem. A pattern object (upper pole) and an anti-pattern object (lower pole) are added to the set of objects. Preference variables (stimulants, destimulants and nominants) are included among the interval-valued variables. Definitions of these variables are available in the study e.g. (Walesiak, 2016). Nominants are transformed into stimulants.

3. Collect data and construct data table $\mathbf{X} = [x_{ij}]_{nxm}$ for $x_{ij} = [x_{ij}^l, x_{ij}^u]$, where $x_{ij}^l \leq x_{ij}^u$, $i, k = 1, \ldots, n$ and $j = 1, \ldots, m$. 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. Select the variable normalization method and the construction of normalized data table $\mathbf{Z} = [z_{ij}]_{nxm}$ for $z_{ij} = [z_{ij}^l, z_{ij}^u]$, where $z_{ij}^l \leq z_{ij}^u$(z_{ij} –normalized observation). Interval-valued symbolic data require a special normalization approach. The lower and upper bound of the interval of the *j*-th variable for *n* objects are combined into one vector containing *2n* observations. This approach makes it possible to apply normalization methods used for classical metric data. The data were normalized using the interval_normalization function from the cluster Simpackage (Walesiak and Dudek, 2017a).

5. Select the distance measure for interval-valued data (4 distance measures were taken into account: Ichino-Yaguchi, Euclidean Ichino-Yaguchi, Hausdorff, Euclidean Hausdorff – see Billard and Diday, 2006; Ichino and Yaguchi, 1994)and construct a distance matrix in *m*-dimensional space $\boldsymbol{\delta} = [\delta_{ik}(\mathbf{Z})]_{nxn}$ for $i, k = 1, \ldots, n$.

6. Perform multidimensional scaling (MDS): $f: \delta_{ik}(\mathbf{Z}) \rightarrow d_{ik}(\mathbf{V})$ for all pairs (i, k)– mapping distances in *m*-dimensional space $\delta_{ik}(\mathbf{Z})$ into corresponding distances $d_{ik}(\mathbf{V})$ in *q*-dimensional space ($q < m$) by a representation function *f*. The distances $d_{ik}(\mathbf{V})$ are always unknown. That is, MDS must find a configuration $\mathbf{V}$ of predetermined dimensions q on which the distances are computed. To enable graphic presentation of linear ordering results $q = 2$. Iterative procedure in the smac of algorithm is presented in the study by (Borg and Groenen, 2005).The solution allowing the choice of an optimal MDS procedure was used to account for the methods used to normalize the variables, the distance measure for interval-valued variables and scaling models, according to the procedure available in mdsOpt package (Walesiak and Dudek, 2017b, 2017c), which applies the smac of Sym function from the smacof package (Mair et al. 2017; De Leeuw and Mair, 2009).

7. Finally, after applying the optimal multidimensional scaling procedure, a data matrix in two-dimensional space $\mathbf{V} = [v_{ij}]_{nxq}$($q$ equals 2) is generated.

8. Depending on the position of the pattern and anti-pattern in two-dimensional space $\mathbf{V} = [v_{ij}]_{nx2}$ the coordinate system needs to be rotated by an angle of φ according to the formula $[v'_{ij}]_{nx2} = [v_{ij}]_{nx2} \times D$ ($[v'_{ij}]_{nx2}$ – data matrix in two-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).

9. Graphic presentation and interpretation of the results in a two-dimensional space. 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).

10. Objects are ordered linearly using an aggregated measure (composite indicator) d_i based on the Euclidean distance from the pattern object (Hellwig, 1981):

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

where: v_{ij} – j-th coordinate for the i-th object in a two-dimensional MDS space, v_{+j} (v_{-j}) – j-th coordinate for the pattern object (anti-pattern) in the 2-dimensional MDS space.

Values of aggregated measure d_i are included in the interval $[0; 1]$. The higher the value of d_i, the higher the economic efficiency of small manufacturing enterprises in districts. Target objects are ranked according to the descending values of the aggregated measure (1).

3 Empirical results

The empirical study uses statistical data about the economic efficiency of small manufacturing enterprises in districts of Wielkopolska province in 2012. The target dataset was prepared in two stages. First, a dataset of 2,162 observations was compiled with three metric variables describing economic efficiency of small manufacturing enterprises (employing 10-49 people): x1 – return on sales in % (net profit as a percentage of sales revenue); x2 – sales revenue in thousand PLN per one employee; x3 – costs in thousand PLN per one employee.

The study did not cover more variables due to unavailability of data. Variables x1 and x2 are stimulants and x3 is a destimulant. In the second step, the observations were aggregated at the level of districts of Wielkopolska province, producing a set of symbolic interval-valued data. The lower bound of the interval for each symbolic interval-valued variable in each district was given by the first quartile of the entire dataset. The upper bound of the interval was obtained by calculating the third quartile. Table 1 presents interval-valued symbolic data describing the economic efficiency districts of Wielkopolska province.

The selection of an optimal scaling procedure was made after testing combinations of ten normalization methods (n1, n2, n3, n5, n5a, n8, n9, n9a, n11, n12a – see Walesiak and Dudek, 2017a; Jajuga and Walesiak, 2000), four distance measures for interval-valued data (Ichino-Yaguchi, Euclidean Ichino-Yaguchi, Hausdorff, Euclidean Hausdorff) and four MDS models (ratio, interval, mspline of second and third degree – see Borg and Groenen, 2005) – altogether 160 MDS procedures. MDS was performed for each procedure separately. Next, the procedures were arranged in ascending order taking into account values of the *Stress*-1, which measures goodness-of-fit (see e.g. Borg et al. 2013). The percentage shares of objects in the value of *Stress*-1 (*spp* – stress per point) measure, was used to calculate the HHI index (Herfindahl, 1950; Hirschman, 1964):$HHI_p = \sum_{i=1}^{n} spp_{pi}^2$. The HHI_p index takes values in the interval$[10{,}000/n\,; 10{,}000]$. From the perspective of MDS the lowest value of the HHI_p index is desirable. Of the acceptable MDS procedures, for which$Stress\text{-}1_p \leq$ mid-range($Stress$-1), we selected one which meets the condition $\min_p\{HHI_p\}$. It wasprocedure 95: normalization n5 (normalization in range [–1; 1]); mspline MDS model of second degree; Euclidean Ichino-Yaguchi distance.

Table 1. Interval-valued data for three variables describing the economic efficiency of small manufacturing enterprises in districts of Wielkopolska province in 2012.

No.	District	x1	x2	x3
1	chodzieski	[1.86, 10.36]	[85, 265.21]	[80.06, 247.58]
2	czarnkowsko-trzcianecki	[1.17, 15.49]	[92.04, 215.08]	[79.24, 208.82]
3	gnieźnieński	[1.44, 12.49]	[67.61, 198.79]	[65.35, 175.54]
4	gostyński	[2.28, 12.01]	[65.48, 205.99]	[59.82, 168.22]
5	grodziski	[2.3, 9.72]	[129.02, 341.85]	[112.96, 323.22]
6	jarociński	[2.08, 17.09]	[64.92, 153.01]	[53.69, 135.89]
7	kaliski	[1.07, 5.77]	[104.95, 394.23]	[112.33, 358.85]
8	kępiński	[1.87, 8.9]	[76.8, 161.89]	[73.4, 150.87]
9	kolski	[1.48, 7.91]	[75.86, 437.76]	[73.17, 433.45]
10	koniński	[1.72, 7.97]	[99.07, 267.03]	[89.85, 246.49]
11	kościański	[2.41, 14.53]	[98.37, 217.48]	[87.7, 195.97]
12	krotoszyński	[1.83, 10.67]	[81.67, 181.89]	[73.61, 153.8]
13	leszczyński	[1.09, 9.45]	[100.63, 197.59]	[95.72, 191.07]
14	międzychodzki	[2.29, 9.96]	[71.29, 178.49]	[67.49, 172.63]
15	nowotomyski	[3.56, 12.94]	[73.95, 250.52]	[71.61, 219.97]
16	obornicki	[0.63, 8.03]	[91.14, 197.87]	[88.30, 196.24]
17	ostrowski	[2.05,11.46]	[72.59, 217]	[67.2, 186.12]
18	ostrzeszowski	[1.83, 9.12]	[79.34, 270.57]	[70.31, 261.16]
19	pilski	[1.03, 12.08]	[82.57, 227]	[75.69, 194.45]
20	pleszewski	[1.39, 14.4]	[62.39, 178.2]	[63.67, 173.23]

21	poznański	[1.53, 12.6]	[104.53, 262.46]	[96.56, 246.48]
22	rawicki	[2.78, 8.27]	[69.74, 183.93]	[61.31, 171.96]
23	słupecki	[1.34, 10.09]	[74.83, 333.98]	[64.62, 327.98]
24	szamotulski	[0.01, 11.5]	[129.28, 267.24]	[120.39, 253.81]
25	średzki	[1.5, 10.31]	[136.71, 275.28]	[125.13, 269.17]
26	śremski	[0.69, 9.82]	[55.12, 281.88]	[55.46, 273.49]
27	turecki	[1.91, 10.64]	[71.71, 136.18]	[69.26, 127.26]
28	wągrowiecki	[2.23, 7.04]	[77.38, 231.82]	[74.72, 227.62]
29	wolsztyński	[2.17, 10.03]	[124.61, 493.23]	[107.75, 457.19]
30	wrzesiński	[1.21, 9.76]	[71.62, 187.06]	[68.09, 180.97]
31	złotowski	[2.39, 7.78]	[88.54, 263.99]	[82.16, 249.96]
32	m. Kalisz	[2.17, 10.3]	[47.08, 187.11]	[56.67, 196.88]
33	m. Konin	[3.73, 13.63]	[67.23, 210.7]	[60.84, 170.29]
34	m. Leszno	[3.29, 11.25]	[74.92, 266.16]	[70.53, 227.62]
35	m. Poznań	[0.77, 12.8]	[94.52, 268.76]	[77.55, 251.88]
P	Pattern	[3.73, 17.09]	[136.71, 493.23]	[53.69, 127.26]
AP	Anti-pattern	[0.01, 5.77]	[47.08, 136.18]	[125.13, 457.19]

Fig. 1 (left panel) shows the Shepard diagram which confirms the correctness of the selected MDS model. The right panel (Stress plot) shows that the MDS configuration represents all proximities almost equally well. Finally, after applying the optimal MDS procedure a data matrix in a two-dimensional space was obtained. Fig. 2 presents results of MDS of 37 objects (35 districts, the pattern and anti-pattern), in terms of the economic efficiency of small manufacturing enterprises.The coordinate system was rotated by an angle $\varphi = 0.9\pi$. The anti-pattern (AP) and pattern (P) were connected by a straight line to form the so-called set axis. Six isoquants of development were defined by dividing the set axis into 6 equal parts. Next, the values of the composite measure (1) were calculated. Table 2 presents the ordering of 35 districts in terms of the economic efficiency of small manufacturing enterprises, in descending order of values of (1). The calculations were performed using R (R Core Team, 2017).

By presenting results in this way it is possible to:

– order districts by the economic efficiency of small manufacturing enterprises measured by three variables based on values of measure (1) and present them graphically in Fig. 2,

– distinguish classes of districts (districts between isoquants) sharing a similar level of economic efficiency (see Fig. 2),

– identify districts characterized by a similar level of economic efficiency, but having a different location on the isoquant of development (see Fig. 2). For example, Kępiński District (8) and Kolski District (9) have a similar level of economic efficiency, but are located at different points on the isoquant of development and in different parts of the province (see Fig. 3). A similar situation occurs for Wolsztyński District (29) and Turecki District (27): while these districts achieved a similar level of development, they were characterized by quite different configurations of variable values.

Table 2. Ordering of districts of Wielkopolska province according to the economic efficiency of small manufacturing enterprises in 2012.

Rank	District	No.	d_i	Rank	District	No.	d_i
1	czarnkowsko-trzcianecki	2	0.5956	19	śremski	26	0.4305
2	kościański	11	0.5589	20	ostrzeszowski	18	0.4158
3	nowotomyski	15	0.5569	21	krotoszyński	12	0.4013
4	m. Poznań	35	0.5490	22	m. Kalisz	32	0.3911
5	szamotulski	24	0.5426	23	międzychodzki	14	0.3810
6	poznański	21	0.5376	24	wrzesiński	30	0.3807
7	m. Konin	33	0.5247	25	leszczyński	13	0.3772
8	m. Leszno	34	0.4963	26	koniński	10	0.3741
9	średzki	25	0.4865	27	złotowski	31	0.3634
10	pleszewski	20	0.4848	28	wolsztyński	29	0.3616
11	pilski	19	0.4808	29	turecki	27	0.3551
12	jarociński	6	0.4767	30	kępiński	8	0.3340
13	gnieźnieński	3	0.4648	31	kolski	9	0.3258
14	grodziski	5	0.4596	32	wągrowiecki	28	0.3252
15	chodzieski	1	0.4536	33	obornicki	16	0.3189
16	gostyński	4	0.4521	34	rawicki	22	0.3157
17	ostrowski	17	0.4517	35	kaliski	7	0.2775
18	słupecki	23	0.4315				
Mean		0.4324		Median		0.4315	
Standard deviation		0.0818		Median absolute deviation		0.0960	

The results of the MDS of districts combined with information about their geographical location seem to confirm the assumptions of the theory of growth poles. This is true not only for its original application, which was limited to economic entities, but also in the current interpretation, which accounts for the spatial dimension (Isard, 1960). One can clearly see the impact of Poznań, which acts as a growth pole, on the neighbouring districts (see Fig. 3). Districts located further away from Poznań are found lower in the ranking based on measure (1), except for two districts – obornicki and wągrowiecki – for which the value of measure d_i was very low (0.32 and 0.33 respectively). This discrepancy can be explained in a number of ways. For one things, these two districts are characterised by high unemployment rate, a relatively small number of working persons; moreover, their inhabitants mainly work in Poznań. According to a study of commuting flows, the largest number of commuters working in Poznań come from districts located north of the city (Główny Urząd Statystyczny, 2014).

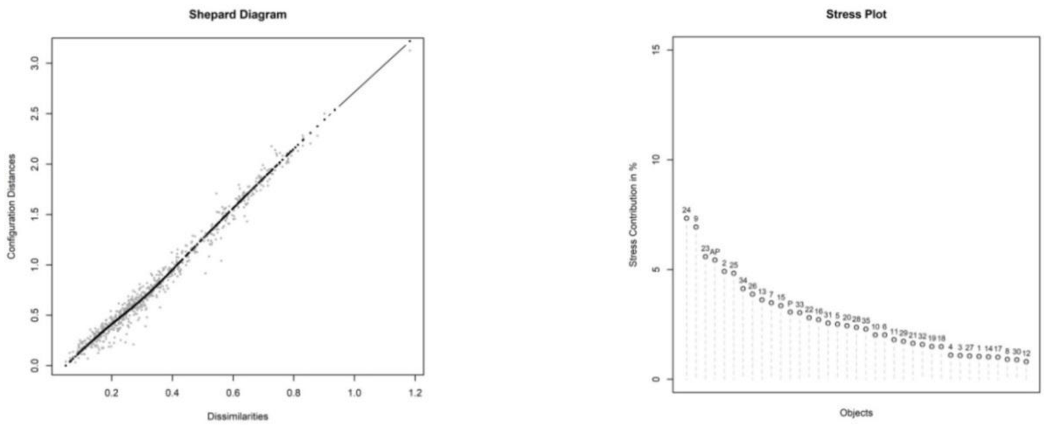

Fig. 1. Shepard diagram and Stress plot.

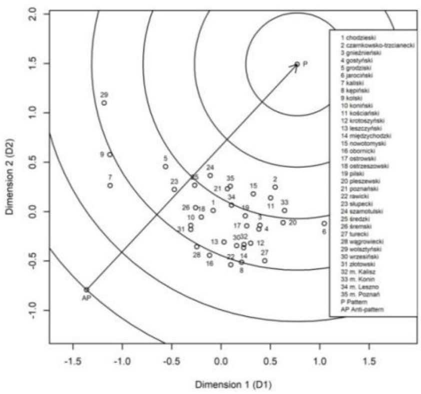

Fig. 2. Results of MDS of 35districtsof Wielkopolska byeconomicefficiency of small manufacturing enterprises.

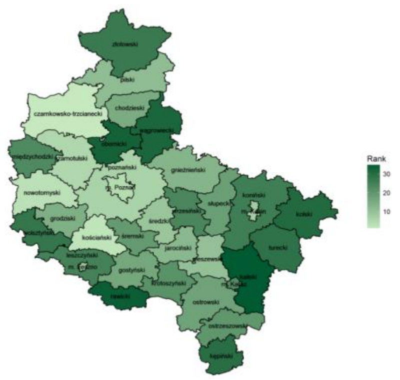

Fig.3. Spatial distribution of districts of Wielkopolska based on the ranking Table 2.

Conclusions

The study described above was an attempt to compare districts of Wielkopolska province in terms of the economic efficiency of small manufacturing companies.The authors used a hybrid approach combining multidimensional scaling and linear ordering.The empirical study was based on interval-valued symbolic data. Districts were evaluated according to the economic efficiency of small manufacturing companies measured by three variables. Thanks to the methodological approach used in the study, it was possible to present the results of linear ordering graphically in a two-dimensional space. In this way districts could be arranged in terms of the economic efficiency of small manufacturing companies and divided into groups sharing a similar level of economic efficiency.The graphical presentation also facilitated the identification of groups of similar districts characterised by similar values of the target variables and those with a different configuration of variable values.The empirical results presented on a map confirm the influence of Poznan as a growth pole on the neighbouring districts. The authors are aware that the results depend on the kind of variables taken into account but the main emphasis of the study was to implement a particular 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*. Chichester: John Wiley.

Borg, I. & Groenen, P. J. F. (2005). *Modern Multidimensional Scaling. Theory and Applications*. New York: Springer Science+Business Media.

Borg, I., Groenen, P. J. F. & Mair, P. (2013). *Applied Multidimensional Scaling*. Heidelberg, New York, Dordrecht, London: Springer.

Brito, P., Noirhomme-Fraiture, M. & Arroyo, J. (2015). Editorial for special issue on symbolic data analysis. *Advanced in Data Analysis and Classification*, *9*(1), 1-4.

Dehnel, G. (2015). Robust regression in monthly business survey. *Statistics in Transition new series*, *16*(1), 1-16.

De Leeuw, J. & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, *31*(3), 1-30.

Gioia, F. & Lauro, C. N. (2006). Principal component analysis on interval data.

Computational Statistics, *21*(2), 343-363.

Główny Urząd Statystyczny.(2014). *Dojazdy do pracy*. Warszawa: GUS.

Główny Urząd Statystyczny. (2017). *Działalność przedsiębiorstw niefinansowych w 2015 roku*. Warszawa: GUS.

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*(pp. 46-68). Warszawa: PWE.

Herfindahl, O. C. (1950). *Concentration in the US steel industry*. Unpublished PhD. Dissertation, Columbia University.

Hirschman, A. O. (1964). The paternity of an index. *Am Econ Rev.*, *54*(5), 761-762.

Ichino, M. & Yaguchi, H. (1994). Generalized Minkowski metrics for mixed feature-type data analysis. *IEEE Transactions on Systems, Man, and Cybernetics*, *24*(4), 698-708.

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). Berlin, Heidelberg: Springer-Verlag.

Jaki, A. (2012). *Mechanizmy procesu zarządzania wartością przedsiębiorstwa*. Zeszyty Naukowe UE w Krakowie. Seria Specjalna, Monografie (215).

Koliński, A. (2011). Przegląd metod i technik oceny efektywności procesu produkcyjnego. *Logistyka*, 5, 1083-1091.

Mair, P., De Leeuw, J., Borg, I. & Groenen, P. J. F. (2017). smacof: Multidimensional Scaling. *R package*, version 1.9-6. http://CRAN.R-project.org/package=smacof

R Core Team. (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. http://www.R-project.org

Walesiak, M. (2016). Visualization of linear ordering results for metric data with the application of multidimensional scaling. *Ekonometria*, *2*(52), 9-21.

Walesiak, M. & Dudek, A. (2017a). clusterSim: Searching for Optimal Clustering Procedure for a Data Set. *R package*, version 0.47-1. http://CRAN.R-project.org/package=clusterSim

Walesiak, M. & Dudek, A. (2017b). mdsOpt: Searching for Optimal MDS Procedure for Metric Data. *R package*, version 0.2-2. http://CRAN.R-project.org/package=mdsOpt

Walesiak, M. & Dudek, A. (2017c). Selecting the optimal multidimensional scaling procedure for metric data with R environment. *Statistics in Transition new series*, *18*(3), 521-540.