EFFECT OF ADOPTED DIAGNOSTIC VARIABLE WEIGHING SYSTEM ON THE RESULTS OF INVESTIGATION OF SOCIOECONOMIC DEVELOPMENT OF VOIVODSHIPS IN POLAND

RAFAŁ CZYŻYCKI

University of Szczecin, POLAND
e-mail: rafal.czyzycki@wzieu.pl

ABSTRACT   The regional development is certain, complex and multidimensional statistics, ambiguously defined, and although it is frequently used, both in theory and practice, its measurement method remains inconclusive. The proper assessment of the region development level is mostly performed by the methods of linear ordering, whose basic tool is a synthetic measure, being a function aggregating partial information, contained in individual attributes (measures). The results of the research are determined at the same time not only the final list of diagnostic variables, but also the choice of the distance measure and aggregation formulas. As there are many formulas for variable standardisation, methods for weight determination, methods for averaging of standardised values, methods for establishment of reference object coordinates and formulas for distance calculation, lots of different aggregate measures.

However, the literature frequently encourages to, in the absence of unambiguous indications for varied meaning and role of individual features, tacitly assume that all diagnostic variables are of the same weight, in general, there are two methods of obtaining weights for individual diagnostic variables: the weights are established by the experts' method (*a priori* method) or with the use of the computation algorithms based on information included in the primary (raw) data (*a posteriori* method). This study will be conducted on the basis of statistical methods for obtaining weights for individual diagnostic variables (*a posteriori* weights).

Introduction

When investigating the level of socioeconomic development of specific regions, it needs to be borne in mind that it is some kind of multi-dimensional characteristics which cannot be measured directly, but which is described by a range of diagnostic variables which should be related to that term. The very term "regional development" can be defined in different ways. Regional development can be treated as a process of positive changes in the

quantitative growth and qualitative progress occurring in the country (region), i.e. a supralocal socio-territorial system characterised by specific features of space, structure of economy and social bond arising from the common national identity (Strahl, 1998, p. 29), but it can also be identified with constant growth of economic potential in regions and systematic and constant improvement in their competitiveness as well as improved quality of life of the inhabitants Chądzyński, Nowakowska, Przygodzki, 2007). Regional development occurs also when a region exports goods or services, develops its infrastructure, uses rich and diverse resources, when innovative entrepreneurs and competent and active institutions of public authorities supporting those processes operate there (Flejterski, 2007, p. 227).

Regions are ranked from "the best" to "the worst" on the basis of the value describing individual regions (synthetic variable, synthetic development measure), obtained from the function aggregating information included in the adopted diagnostic variables. The main stages of the procedure aiming to obtain a synthetic variable include (Kukuła, 2000, pp. 42–76; Walesiak, 1996, pp. 125–130):

- – selection of diagnostic variables related in substance to the paramount criterion according to which the objects (voivodships) will be arranged,
- – depriving variable base units of values and unification of orders of magnitude to lead them to comparability,
- – weighing of standardised diagnostic features, i.e. assignment of weights to individual variables, which weights determine their meaning for the general criterion compared to other features,
- – selection of the aggregating formula and determination of the value of the synthetic development measure on its basis.

So far, as regards methodical aspects of investigation of socioeconomic development of regions, the subject of interest was, inter alia, the effect of selection of diagnostic variables (Czyżycki, 2018a), effect of selection of the standardising formula (Czyżycki, 2012, pp. 15–23) or aggregating formula (Czyżycki, 2018b) on the obtained results of region ranking. This article aims to determine the effect and merit of the adopted diagnostic variable weighing system on the final results of such investigations.

In general, there are two methods of obtaining weights for individual diagnostic variables (Kao, 2010, p. 1779): direct and indirect. In the former method, the values of weights for individual variables are obtained by interviewing experts, on the basis of conducted surveys, etc. even before the stage of collecting data on shaping of individual variables in the studied objects. Therefore, they are sometimes referred to as *a priori* weights. In the latter method, the weights are obtained directly from analysis of data obtained beforehand (*a posteriori* weight); this is why they seem more compelling from the substantive point of view. As opposed to the weights determined by experts, which can be constant in further analyses of the same problem, *a posteriori* weights have to be determined anew in each and every analysis (Sokołowski, 1985, p. 48). However, the literature frequently encourages to, in the absence of unambiguous indications for varied meaning and role of individual features, tacitly assume that all diagnostic variables are of the same weight (Kukuła, 2000, p. 64).

This study will be conducted on the basis of statistical methods for obtaining weights for individual diagnostic variables (*a posteriori* weights).

Research methodology

Bearing in mind that the list of variables included in the investigation of socioeconomic development of regions is key, it should, therefore, be discussed among experts and regarded as the best list representing the analysed

issue. A wide review of the literature in this respect was performed by R. Klóska (2015, pp. 99–108) who, on the basis of in-depth studies, offered 18 indicators which can be employed to assess regional development in three dimensions:

a) from a social perspective: number of infant deaths per 1,000 live births (X_1), at-risk-of-poverty rate (X_2 – %), number of students of the tertiary education institutions per 10,000 inhabitants (X_3), registered unemployment rate (X_4 – %), number of road fatalities per 100,000 inhabitants (X_5) and total water consumption for the needs of the national economy and population (hm^3) per 10,000 inhabitants (X_6);

b) from an economic perspective: GDP (current prices) per capita (X_7), share of economy entities' financial outlays in the total outlays on research and development activities (X_8 – %), number of newly registered national economy entities in the private sector per 10,000 inhabitants (X_9), number of employees per 1,000 inhabitants (X_{10}) and total investments (current prices) per capita in PLN (X_{11});

c) from an environmental perspective: percentage of the population using the sewage treatment plants (X_{12} – %), forest cover (X_{13} – %), recycling of packaging waste (X_{14} – %), share of devastated and degraded lands requiring reclamation in the total area (X_{15} – %), share of waste (excluding municipal waste) recovered in the amount of waste generated during a year (X_{16} – %), share of the electricity generation from renewable energy sources in the total electricity generation (X_{17} – %) and electricity consumption per 1 million PLN GDP (X_{18} – GWh).

Employing statistical website strateg.stat.gov.pl, information for all voivodships in Poland was obtained as to the shaping of eighteen diagnostic variables proposed in the article. At the same time, it was assumed that the socioeconomic development would be analysed at the end of 2016 and in the case of thirteen variables, the shaping of their values in individual voivodships in that year was assumed for studying, but in the case of Gross Domestic Product per capita (X_7), share of expenditures for R&D activity financed from the enterprise sector in the expenditures for R&D activity in total (X_8) and the consumption of electric energy per PLN 1 million of GDP (X_{18}), the study included their value in 2015, in the case of packaging waste (X_{14}) – in 2014 and as regards the share of waste (except for municipal waste) subject to recycling in the amount of waste generated throughout the year (X_{16}), the level of 2013 was adopted. For the above five variables, the indicated years were the last for which, at the moment of conducting the investigation, the Central Statistical Office provided information concerning the shaping of the values of those variables in individual voivodships.

The values of weights for individual diagnostic variables (w_j) will be determined based on the following methods:

1. Method of uniform degree of magnitude of individual variables:

$$w_j^1 = \frac{1}{m} \tag{1}$$

where m means the number of diagnostic variables (m = 18).

2. Method based on standard deviation of standardised variables (Diakoulaki, Mavrotas, Papayannakis, 1995, p. 766):

$$w_j^2 = \frac{\sigma_j}{\sum_{k=1}^{m} \sigma_k} \tag{2}$$

Variables shaped in range [0, 1] are obtained:
– in the case of stimulants, according to formula:

$$z_{ij} = \frac{x_{ij} - \min\limits_i x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}} \tag{3}$$

– in the case of destimulants:

$$z_{ij} = \frac{\max\limits_i x_{ij} - x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}} \tag{4}$$

3. Methods based on primary variable variability indices (Śledzik, 2012, p. 258):

$$w_j^3 = \frac{V_j}{\sum_{k=1}^m V_k} \tag{5}$$

4. Method based on entropy of variables consisting in the following (Wang, Luo, 2010, p. 8):
Step 1: data normalization:
– in the case of stimulants, according to formula:

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}}} \tag{6}$$

– in the case of destimulants:

$$z_{ij} = \frac{\left(x_{ij}\right)^{-1}}{\sqrt{\sum_{i=1}^m \left(x_{ij}\right)^{-1}}} \tag{7}$$

Step 2: determination of the value of entropy (E_j) and degree of differentiation (d_j):

$$E_j = -\frac{1}{\ln(m)} \times \sum_{i=1}^m z_{ij} \ln z_{ij} \tag{8}$$

$$d_j = 1 - E_j \tag{9}$$

Step3: determination of weights:

$$w_j^4 = \frac{d_j}{\sum_{k=1}^m d_k} \tag{10}$$

CRITIC method (*Criteria Importance through Inter-criteria Correlation method*) (Diakoulaki, Mavrotas, Papayannakis, 1995, pp. 764–765), which requires standardisation of variables in accordance with formulas (3) and (4), determination of standard deviation σ_j for each standardised variable and linear correlation indices r_{ij} among all variables. Values of individual weights are determined on the basis of formula (11):

$$w_j^5 = \frac{C_j}{\sum_{k=1}^n C_k} \tag{11}$$

where

$$C_j = \sigma_j \times \sum_{k=1}^{m} (1 - r_{ik}) \qquad (12)$$

Method of minimisation of distance from the standard (Ma, Fan, Huang, 1999) which consists in determination of such values of weights for individual variables that will minimise the distance defined as follows:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left(z_0^+ - z_{ij} \right) w_j^2 \to \min \qquad (13)$$

where

$$z_0^+ = \begin{cases} \max_i z_{ij} & \text{in the case of stimulants} \\ \min_i z_{ij} & \text{in the case of destimulants} \end{cases} \qquad (14)$$

and z_{ij} are determined by way of formulas (3) and (4).

The solution of assumption (13) is to determine the values of individual weights based on the following formula:

$$w_j^6 = \frac{\dfrac{1}{\sum_{i=1}^{m} \left(z_{ij} - z_0^+ \right)^2}}{\sum_{k=1}^{m} \left(\dfrac{1}{\sum_{i=1}^{m} \left(z_{ij} - z_0^+ \right)^2} \right)} \qquad (15)$$

The values obtained by way of methods 1–6 meet two basic assumptions related with the weights of diagnostic variables, namely: positivity ($w_j > 0$) and summation to units $\left(\sum_{j=1}^{m} w_j = 1 \right)$.

Effect of selection of method for determination of weights of individual diagnostic variables on the results of investigation of development of regions will be assessed with the TOPSIS method (*technique for order preference by similarity to ideal solution*) (Hwang, Yoon, 1981, pp. 130–132) according to which the ranking of objects is created based on values C_i determined according to the following formula:

$$C_i = \frac{d_i^-}{d_i^+ + d_i^-} \qquad (16)$$

where

$$d_i^+ = \sqrt{\sum_{j=1}^{n} \left(z_{ij} \times w_j - z_0^+ \right)^2}, \; i = 1, 2, ..., m$$

$$d_i^- = \sqrt{\sum_{j=1}^{n} \left(z_{ij} \times w_j - z_0^- \right)^2}, \; i = 1, 2, ..., m \qquad (17)$$

$$z_0^+ = \begin{cases} \max_i z_{ij} \times w_j & \text{in the case of stimulants} \\ \min_i z_{ij} \times w_j & \text{in the case of destimulants} \end{cases}$$

$$z_0^- = \begin{cases} \min_i z_{ij} \times w_j & \text{in the case of stimulants} \\ \max_i z_{ij} \times w_j & \text{in the case of destimulants} \end{cases}$$

(18)

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}$$

(19)

Study results

Analysing the values of individual weights obtained from the formulas, which assume the possibility of occurrence of their varied level, it can be observed that the most important variable (i.e. the variable having the strongest weight) of all analysed diagnostic variables was either variable X_{14} (in the case of w_3, w_4 and w_6) or X_{15} (in the case of w_2 and w_5). On the other hand, the lowest weight was associated with variables X_6 (w_6), X_7 (w_5), X_{11} (w_2) or X_{13} (w_3 and w_4). Nevertheless, it can be generally stated that the magnitude of individual diagnostic variables changed depending on the adopted weight determination method. Detailed information on the shaping of the values of obtained weight indices depending on the adopted method is illustrated in Table 1.

Table 1. Values of weight indices for individual diagnostic variables

	w_j^1	w_j^2	w_j^3	w_j^4	w_j^5	w_j^6
X_1	0.0556	0.0596	0.0280	0.0136	0.0645	0.0509
X_2	0.0556	0.0569	0.0449	0.0302	0.0601	0.0536
X_3	0.0556	0.0507	0.0360	0.0217	0.0595	0.0395
X_4	0.0556	0.0642	0.0411	0.0245	0.0725	0.0401
X_5	0.0556	0.0527	0.0228	0.0071	0.0537	0.0269
X_6	0.0556	0.0562	0.1559	0.0979	0.0519	0.0167
X_7	0.0556	0.0483	0.0337	0.0156	0.0436	0.0931
X_8	0.0556	0.0520	0.0470	0.0342	0.0525	0.0576
X_9	0.0556	0.0630	0.0243	0.0089	0.0576	0.0583
X_{10}	0.0556	0.0557	0.0198	0.0059	0.0523	0.0641
X_{11}	0.0556	0.0463	0.0304	0.0132	0.0439	0.0761
X_{12}	0.0556	0.0485	0.0311	0.0138	0.0470	0.0771
X_{13}	0.0556	0.0519	0.0133	0.0027	0.0495	0.0338
X_{14}	0.0556	0.0516	0.2083	0.3937	0.0456	0.1224
X_{15}	0.0556	0.0684	0.0664	0.0559	0.0760	0.0380
X_{16}	0.0556	0.0602	0.0466	0.0371	0.0581	0.0307
X_{17}	0.0556	0.0600	0.1221	0.2129	0.0543	0.0869
X_{18}	0.0556	0.0537	0.0284	0.0111	0.0575	0.0341

Source: own work.

The obtaining of various values of individual weight indices automatically translates into obtaining of various results of voivodship rankings in terms of socioeconomic development. In each case, the Mazovian Voivodship takes the first place in the ranking, but the West-Pomeranian Voivodship, depending on the adopted methods for weighing the diagnostic variables measured in the study, may take the second, third, tenth or even the last, sixteenth, place in the ranking. Table 2 shows details on the shaping of TOPSIS values (C_i) and the voivodship ranking place determined on its basis.

Table 2. Shaping of TOPSIS (C_i) values and the place taken in the ranking (R_i) by individual voivodships, depending on the method for determination of value of weights for individual diagnostic variables

	C_1	R_1	C_2	R_2	C_3	R_3	C_4	R_4	C_5	R_5	C_6	R_6
Lower Silesia (LS)	0.439	7	0.381	9	0.383	7	0.196	7	0.229	9	0.229	9
Kuyavian-Pomeranin (KP)	0.435	8	0.312	13	0.365	13	0.165	14	0.187	12	0.187	12
Lublin (LU)	0.411	13	0.277	16	0.356	14	0.161	16	0.157	15	0.157	15
Lubusz(LB)	0.419	12	0.412	5	0.388	5	0.194	8	0.187	13	0.187	13
Łódź (ŁD)	0.449	6	0.294	15	0.378	8	0.180	11	0.199	10	0.199	10
Lesser Poland (LP)	0.486	3	0.381	8	0.371	9	0.180	10	0.241	8	0.241	8
Masovian (MS)	0.750	1	0.627	1	0.802	1	0.898	1	0.848	1	0.848	1
Opole (OP)	0.411	14	0.331	12	0.369	11	0.176	12	0.179	14	0.179	14
Subcarpathian (SC)	0.467	4	0.400	6	0.385	6	0.188	9	0.244	7	0.244	7
Podlasie (PL)	0.397	15	0.303	14	0.370	10	0.170	13	0.138	16	0.138	16
Pomeranian(PM)	0.501	2	0.413	4	0.393	4	0.214	5	0.273	5	0.273	5
Silesian (SL)	0.430	9	0.358	11	0.366	12	0.164	15	0.190	11	0.190	11
Świętokrzyskie (ŚK)	0.335	16	0.397	7	0.397	3	0.483	2	0.428	2	0.428	2
Warmian-Masurian (WM)	0.424	11	0.375	10	0.398	2	0.245	4	0.249	6	0.249	6
Greater Poland (GP)	0.453	5	0.444	3	0.314	15	0.202	6	0.278	4	0.278	4
Western Pomeranian (WP)	0.425	10	0.446	2	0.301	16	0.278	3	0.341	3	0.341	3

Source: own work.

The degree of conformity of obtained rankings can be assessed by way of the Kendall's tau coefficient (τ). This coefficient assumes values from the range <–1, 1>, where value 1 means full conformity, value 0 means lack of conformity in arrangement and value –1 means their full contradiction. In order to verify the hypothesis about the conformity of obtained rankings, the test of significance of the Kendall's tau coefficient is conducted, for which the testing statistics are defined as follows:

$$Z_\tau = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \tag{20}$$

for $n > 10$, it has asymptotically a regular distribution (Abdi, 2007) (for the investigation of socioeconomic development of voivodships in Poland, $n = 16$).

A statistically significant convergence of obtained rankings occurs when the values of weights are determined:
– with method of uniform degree of magnitude of individual variables (w_j^1), the CRITIC method (w_j^5) and the method of minimisation of distance from the standard (w_j^6),

– with method based on standard deviation of standardised variables (w_j^2) and the following methods: method of entropy (w_j^4), CRITIC method (w_j^5) and the method of minimisation of distance from the standard (w_j^6),

– with method based on the primary variable variability index (w_j^3) and entropy (w_j^4),

– methods of entropy (w_j^4), CRITIC (w_j^5) and minimisation of distance from the standard (w_j^6) (cf. Table 3).

Table 3. Values of the Kendall's tau coefficient, examining the conformity of the obtained rankings (values over the main diagonal) along with the p-value for the test assessing the statistical significance of the obtained Kendall's tau coefficient (under the main diagonal)

	w_j^1	w_j^2	w_j^3	w_j^4	w_j^5	w_j^6
w_j^1	x	1.8009	0.7204	1.1706	2.2511	2.2511
w_j^2	0.0717	x	1.2606	3.3317	3.5118	3.5118
w_j^3	0.4713	0.2074	x	2.7914	1.7109	1.7109
w_j^4	0.2418	8.63E-04	0.0052	x	3.9620	3.9620
w_j^5	0.0244	4.45E-04	0.0871	7.43E-05	x	5.4027
w_j^6	0.0244	4.45E-04	0.0871	7.43E-05	6.56E-08	x

Source: own work.

The graphic comparison of results of arranging individual voivodships in 2016 indicates a certain dependency of the obtained place in the ranking on the manner of weighing the magnitude in the investigation of individual diagnostic variables (cf. Figure 1).
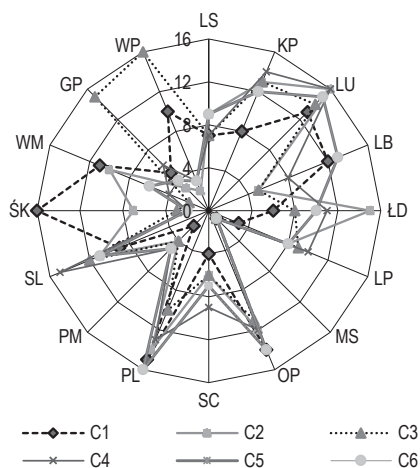
Figure 1. Places of voivodships of Poland based on the level of socioeconomic development in 2016

Source: own work.

Conclusions

On the basis of conducted analyses, it seems legitimate to put forward the following conclusions:

1. In the case of investigation of regional development with methods being part of a multi-dimensional comparative analysis, the assumption of a certain diagnostic variable weighing system may significantly influence the obtained final results of the investigation.

2. It seems debatable to indicate one specific method for determination of the values of weights as the best method, but it also needs to be borne in mind that the assumption of equal magnitude of all diagnostic variables, which occurs in most such investigations, is equal to selecting a specific method – method of uniform degree of magnitude of individual variables described by formula (1).

3. It seems valid to propose that such studies not only present the final results, but also describe the study methods in detail. It will allow for exact repetition of the conducted analyses and possible checking of their correctness and for, which seems more important, explanation of any differences in rankings of objects obtained by various researchers even if they adopt the same diagnostic variables.

References

Abdi, H. (2007). The Kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics. Sage*, *Thousand Oaks*, *CA*, 508–510.

Chądzyński, J., Nowakowska, A., Przygodzki, Z. (2007). *Region i jego rozwój w warunkach globalizacji*. Warszawa: CeDeWu. PL Wydawnictwa Fachowe.

Czyżycki, R. (2012). Badanie rozwoju społeczno-gospodarczego województw – wpływ metodyki badań na uzyskane wyniki. *Zeszyty Naukowe Wyższej Szkoły Bankowej w Poznaniu*, *42*, 15–22.

Czyżycki, R. (2018a). Determinanty rozwoju społeczno-ekonomicznego regionów w Polsce. *Ekonomiczne Problemy Usług* (w druku).

Czyżycki, R. (2018b). Metody bezwzorcowe w badaniu rozwoju regionów - wpływ formuły agregacyjnej na uzyskane wyniki badań. *Progress in Economic Sciences*, *5* (w druku).

Diakoulaki, D., Mavrotas, G., Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The critic method. *Computers and Operations Research*, *22* (7), 763–770. DOI: 10.1016/0305-0548(94)00059-H.

Flejterski, S. (2007). Zastosowanie kategorii kapitał do analizy rozwoju regionalnego i lokalnego. *Zeszyty Naukowe Uniwersytetu Szczecińskiego*, *471*.

Hwang, C.-L., Yoon, K. (1981). Methods for Multiple Attribute Decision Making. In: *Multiple Attribute Decision Making: Methods and Applications A State-of-the-Art Survey*. Berlin Heidelberg: Springer-Verlag. DOI: 10.1007/978-3-642-48318-9.

Kao, C. (2010). Weight determination for consistently ranking alternatives in multiple criteria decision analysis. *Applied Mathematical Modelling*, *34* (7), 1779–1787.

Klóska, R. (2015). *Innowacyjność jako determinanta rozwoju regionalnego w Polsce*. Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

Kukuła, K. (2000). *Metoda unitaryzacji zerowanej*. Warszawa: Wydawnicwto Naukowe PWN.

Ma, J., Fan, Z.P., Huang, L.H. (1999). A subjective and objective integrated approach to determine attribute weights. *European Journal of Operational Research*, *112* (2), 397–404. DOI: 10.1016/S0377-2217(98)00141-6.

Sokołowski, A. (1985). Wybrane zagadnienia pomiaru i ważenia cech w taksonomii. *Zeszyty Naukowe/Akademia Ekonomiczna w Krakowie*, *203* (Prace z zakresu statystyki), 41–53.

Strahl, D. (1998). *Taksonomia struktur w badaniach regionalnych*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.

Śledzik, K. (2012). Wielowymiarowa ocena pozycji konkurencyjnej krajowych banków giełdowych za okres 2007–2011. *Zarządzanie i Finanse*, *4* (2), 257–269.

Walesiak, M. (1996). *Metody analizy danych marketingowych*. Warszawa: Wydawnictwo Naukowe PWN.

Wang, Y.-M., Luo, Y. (2010). Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making. *Mathematical and Computer Modelling*, *51* (1–2), 1–12.