

# A PROPOSAL OF DATA – DRIVEN METHOD FOR DETERMINING THE WEIGHTS OF COMPOSITE INDICATORS

**Kinga Kądziołka**

WSB University, Dąbrowa Górnicza, Poland

e-mail: kinga.kadziolka@vp.pl

ORCID: 0000-0001-9506-3044

© 2021 Kinga Kądziołka

*This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License.*

*To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>*

*Quote as:* Kądziołka, K. (2021). A proposal of data – driven method for determining the weights of composite indicators. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, 25(1).

DOI: 10.15611/eada.2021.1.03

JEL Classification: C02, C38, C63

---

**Abstract:** The paper proposed a simulation method for determining the weights of components of taxonomic measures. The method takes into account the degree of similarity of the final ranking to other rankings and other properties, e.g. the clustering ability of the measure. The analyses were performed on publicly available data published by the General Statistic Office, concerning selected characteristics of the labour market in Poland at the level of subregions. The results obtained by the proposed method depend on the initial set of weights vectors. Due to the fact that the proposed method does not provide an invariant solution for a given data set, the stability of the rankings obtained using this method was assessed. There was high consistency in the orderings of objects obtained in the consecutive repetitions of the procedure.

**Keywords:** taxonomic measure, composite indicator, weighting schema, semi-standard deviation, labour market, Spearman's correlation coefficient.

---

## 1. Introduction

The synthetic variable (known also as taxonomic measure or composite indicator) is the result of an appropriate transformation of the group of diagnostic variables (Wydymus, 1984, p. 188). Taxonomic measures allow for the comparison and linear ordering of objects described by many different characteristics. There is not a single universal method for constructing taxonomic measures. Some of these methods were presented by, among others Kukuła and Luty (2018). The standarised sum method was used, due to its simplicity.

The analysed taxonomic measures took the form:

$$TM_i = \sum_{j=1}^m w_j z_{ij},$$

where  $w_j$  is the weight of  $j$ -th diagnostic variable,  $\sum_{j=1}^m w_j = 1$ ,  $w_j > 0$ ,  $j = 1, \dots, m$ ,  $z_{ij}$  – value of  $j$ -th variable (in form of stimulant<sup>1</sup> and after normalization) for  $i$ -th object,  $i = 1, \dots, n$ .

The weights of the diagnostic variables reflect their relative importance. The methods for determining the weights of the diagnostic variables can be categorized into three groups: experts-based weighting, equal weighting and statistics-based weighting. Gan et al., analysed literature to answer the question of what are the most commonly used methods for weighting and aggregation diagnostic variables. According to their study, the most often adopted was the equal weighting method (Gan et al., 2017, p. 492). The existing literature offers many quantitative methods to determine the weights of the composite indicators, such as: principal component analysis, factor analysis, multiple linear regression, mathematical programming (Becker, Saisana, Paruolo, and Vandecasteele, 2017; Greco, Ishizka, Tasiou, and Torrisi 2019; Zhou, Ang, and Poh, 2007).

In this paper, a simulation method for determining the weights is proposed. The idea of the method is to create ranking of objects being similar to the rankings obtained with other analysed taxonomic measures. Spearman's correlation coefficient was used to assess the similarity of rankings. The proposed method consists of four steps:

1. Generate randomly  $k$  vectors of weights and determine  $k$  taxonomic measures based on the generated vectors of weights.

2. Determine for values of each of constructed taxonomic measures the mean and semi-standard deviation of Spearman's correlation coefficients with the values of other analysed taxonomic measures. The author used semi-standard deviation that incorporated only the negative deviations from the mean value.<sup>2</sup> Deviations above the target are a positive phenomenon. The higher the value of Spearman's correlation coefficient, the more similar the rankings.

3. Determine a subset (denoted as  $D$ ) of the constructed measures such that for each taxonomic measure belonging to this subset there is no other taxonomic measure (among the initial set of  $k$  measures) with the higher mean of Spearman's correlation coefficients and lower or the same semi-standard deviation, or with the same mean of Spearman's correlation coefficients and lower semi-standard deviation.

4. Select the final taxonomic measure from the set  $D$  based on the adopted criterion. The five criteria of the selection of the final taxonomic measure are compared.

---

<sup>1</sup> A stimulant is a variable whose higher value indicates a higher level of the analysed phenomenon. A destimulant is a variable having a negative impact on this phenomenon (Młodak, 2006, p. 33).

<sup>2</sup> The formula for its calculation is presented in Just and Śmiglak-Krajewska (2013, p. 61).

The proposed method is presented based on the example of multidimensional comparative analysis of the labour market data at subregional level. The analysed data are publicly available on the website of General Statistic Office (GUS). All the calculations were conducted using R software.

## 2. Characteristics of the analysed data

In this study the taxonomic measure was constructed in order to assess the situation on the labour market in Poland at subregional level in 2018 (Table 3). The following four diagnostic variables were chosen to construct the taxonomic measure:

- registered unemployment rate ( $x_1$ ),
- people registered as unemployed for a period lasting longer than 1 year (% of overall unemployed; the so-called long-term unemployment rate) ( $x_2$ ),
- participation of unemployed persons in the age group of 18-24 years in the total number of people of this age ( $x_3$ ),
- participation of unemployed persons with at most lower secondary education in the total number of unemployed ( $x_4$ ).

The diagnostic variables were chosen arbitrarily. Their choice was motivated, among others, by data availability. Table 1 presents examples of the sets of variables used by other authors for the multidimensional assessment of the situation on the labour market in Poland.

**Table 1.** Diagnostic variables used in other studies

Author	Components of taxonomic measures
1	2
E. Badach, M. Ziolo (2015, p. 26)	Unemployment rate, long-term unemployment rate, participation of unemployed persons over 55 years in the total number of unemployed, participation of unemployed persons of 18-24 years in the total number of unemployed, participation of unemployed persons with higher education in the total number of unemployed, participation of persons with disabilities in the total number of unemployed, people registered as unemployed per one job offer.
M. Gawrycka, A. Szymczak (2013, p. 77)	Labour productivity, employment rate, unemployment rate, tax burdens, investment expenditure for research and development, labour force participation, life-long learning of adults, gross enrolment rate.
E. Sojka (2013, p. 35, 2014, p. 104)	Participation of unemployed persons of 18-24 years in the total number of unemployed, persons without internship or with internship not exceeding 1 year in the total number of unemployed, long-term unemployment rate, people registered as unemployed per one job offer, participation of unemployed persons with higher education in the total number of unemployed, participation of people working in the private sector in the total number of working people, participation of people working in services in the total number of working people, gross earnings in relation to the regional average (Silesia region = 100).

Table 1, cont.

1	2
M.B. Pietrzak (2016, p. 50)	The monthly average gross salary, new registered national economy entities per every 10 thousand of working age population, business investment expenditure per 1000 of working age population, unemployment rate.
A. Tatarczak, O. Boichuk (2018, p. 375)	Participation of unemployed persons of 15-24 years in the total number of unemployed, participation of unemployed persons without internship in the total number of unemployed, participation of unemployed persons with higher education in the total number of unemployed, job vacancy rate, the monthly average gross salary in relation to the national average.
E. Rollnik-Sadowska, M. Jarocka, E. Glińska (2018, p. 440-441)	People registered as unemployed for a period lasting longer than 1 year (% of overall unemployed), average monthly number of people registered as unemployed per one job offer, unemployment rate, average monthly gross earnings in relation to the national average, new registered entities per every 10 thousand of working age population, business investment expenditure per one working age person, national economy entities per one thousand working age citizens, employment rate.

Source: own elaboration.

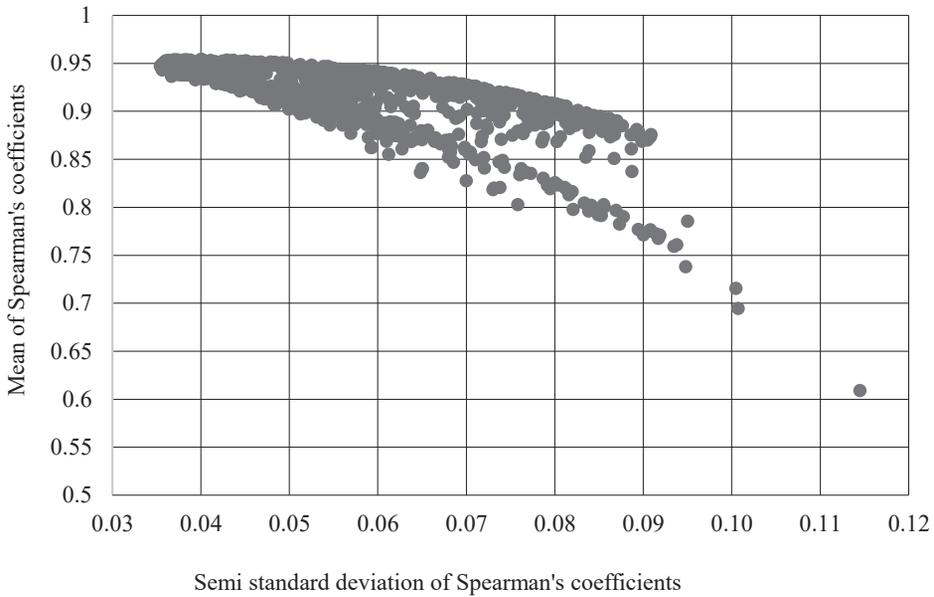
In this paper, in addition to the registered unemployment rate, the variables concerning long-term unemployment and unemployment among young people and people with low education were used, as these are particularly dangerous phenomena on the labour market, which may lead to an increase in the crime rate (Kądziołka, 2015, p. 72).

In the analysed case, all diagnostic variables were destimulants. They were converted into stimulants according to the formula:  $z_{ij} = \frac{\max_i\{x_{ij}\} - x_{ij}}{\max_i\{x_{ij}\} - \min_i\{x_{ij}\}}$ ,  $i = 1, \dots, n, j = 1, \dots, m$ .

### 3. Application of the proposed method

In the analysed case, there were randomly generated 1000 vectors of weights  $(w_1^r, \dots, w_4^r)$ ,  $r = 1, \dots, 1000$ , from whose one can choose a representative, with which the final taxonomic measure is constructed and the linear ordering of subregions performed. Firstly,  $k = 1000$  taxonomic measures are constructed:  $TM_i^r = \sum_{j=1}^m w_j^r z_{ij}$ ,  $i = 1, \dots, n; r = 1, \dots, k, m = 4$ , associated with the generated weights. From the set of these taxonomic measures, one is chosen to be the final solution of the linear ordering of objects (here: subregions). Figure 1 presents the scatterplot of the semi-standard deviation and the mean of Spearman's correlation coefficients for the constructed taxonomic measures.

Based on the values of taxonomic measures, the author created rankings of objects. The subregions were ordered from the best to the worst according to the

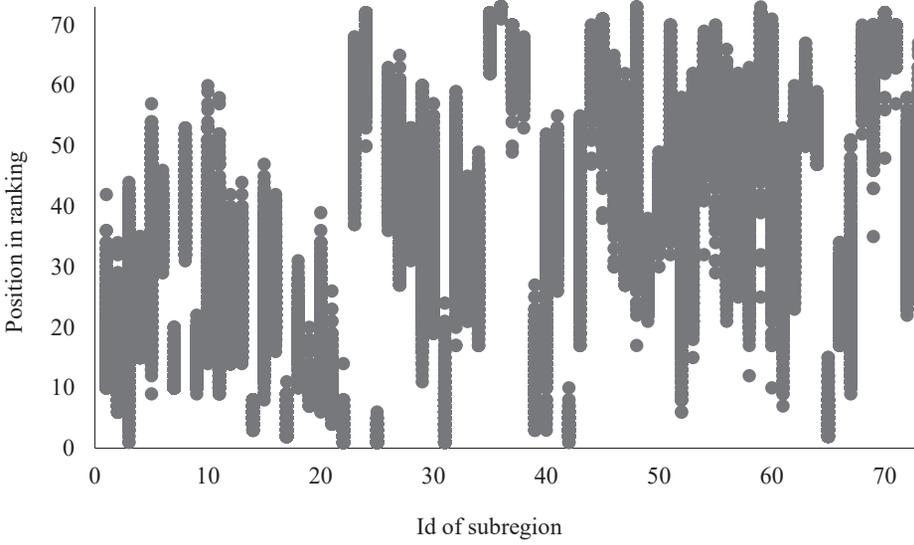


**Fig. 1.** Semi-standard deviation and mean of Spearman’s correlation coefficients

Source: own elaboration.

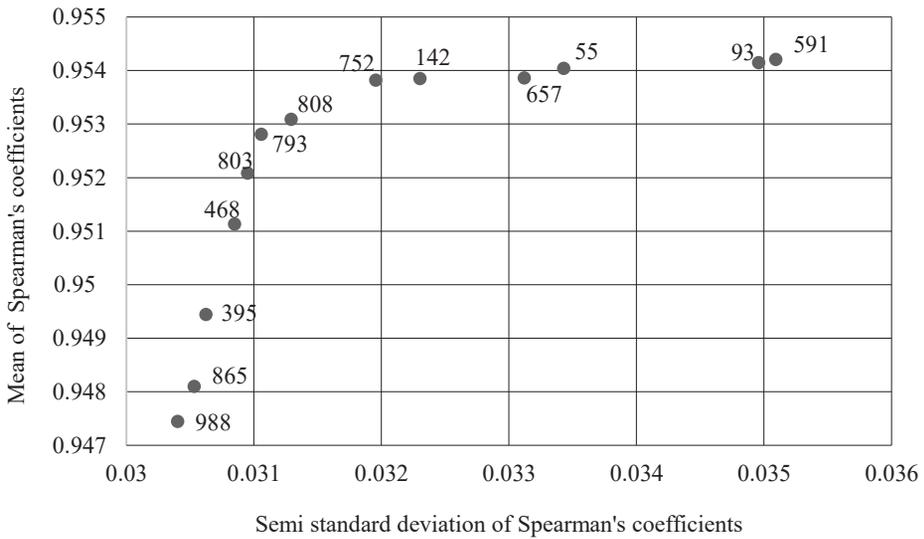
values of the taxonomic measures. Figure 2 shows positions of individual subregions in rankings obtained for the analysed 1000 taxonomic measures. Position 1 is the object with the highest value of the taxonomic measure (the best subregion) and position 73 the object with the lowest value of the measure (the worst subregion). It can be seen that for the majority of subregions there were large differences in positions according to the particular rankings.

Next the author created a subset of taxonomic measures containing those for which there exists neither a measure with higher mean of Spearman’s correlation coefficients, and lower or the same semi-standard deviation of Spearman’s correlation coefficients nor a measure with the same mean of Spearman’s correlation coefficients and lower semi-standard deviation. In this case there were 13 such taxonomic measures (see Figure 3). There is a similarity to the determination of an efficient frontier of investment portfolios. The labels in Figure 3 contain the identification number (Id) of the appropriate taxonomic measure. Figure 4 shows the structure of the weights of the taxonomic measures belonging to the reduced set, and Figure 5 the positions of the subregions in the rankings for the taxonomic measures from the reduced set of measures. It can be seen that the range of positions for individual subregions has decreased.



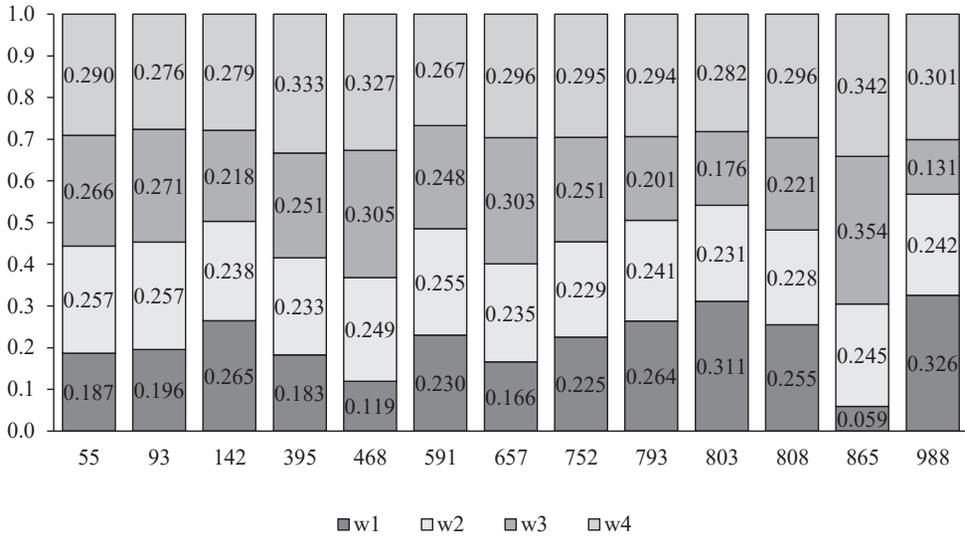
**Fig. 2.** Positions of subregions in rankings according to the values of the taxonomic measures

Source: own elaboration.



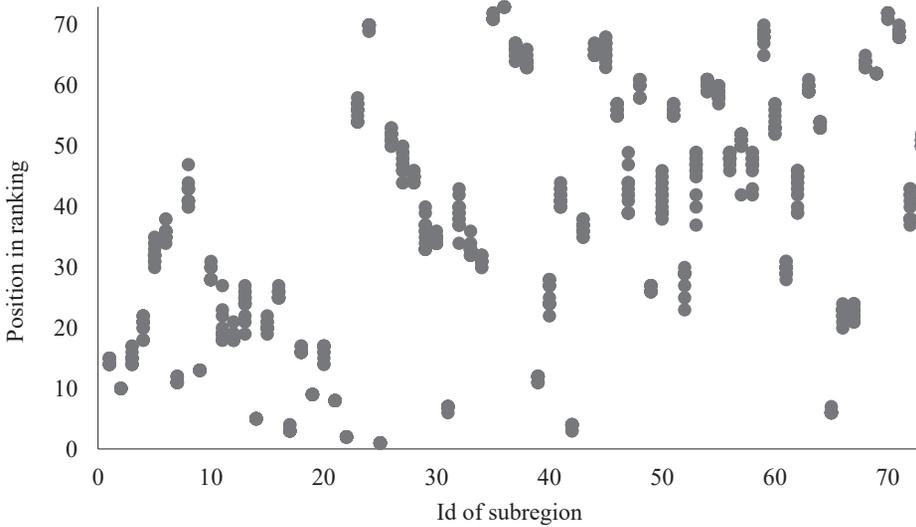
**Fig. 3.** Reduced set of taxonomic measures

Source: own elaboration.



**Fig. 4.** Vectors of weights of taxonomic measures belonging to the reduced set

Source: own elaboration.



**Fig. 5.** Positions of subregions according to values of TMs from the reduced set

Source: own elaboration.

#### 4. The choice of the final representative

The results obtained with the use of the five methods for selecting the final measure were compared.

In the first case, the measure characterized by the minimal value of the semi-standard deviation of Spearman's correlation coefficients was selected.

In the second case, the measure characterized by the maximal mean of Spearman's correlation coefficients was selected.

In the third case, the measure characterized by the maximal value of Sokołowski's discrimination coefficient<sup>3</sup> was selected. Sokołowski's coefficient is determined according to the formula:  $G = 1 - \sum_{i=1}^{n-1} \min\left\{\frac{TM_i - TM_{i+1}}{R}, \frac{1}{n-1}\right\}$ ,  $R = \max_i\{TM_i\} - \min_i\{TM_i\}$ , where  $TM_i$ ,  $TM_{i+1}$  – values of taxonomic measure  $TM$  in descending order. The higher the value of Sokołowski's coefficient, the higher the ability of the taxonomic measure to create clusters of similar objects (Roszkowska and Lasakevic, 2014, p. 46).

In the fourth case, the measure was selected for which the sum of the distances to other measures (in two-dimensional space, see Figure 3) was minimal.

In the last case, the measure for which the mean value of indicators of the similarity of weights structures was maximal, was selected. The indicator of the similarity of two structures was determined according to the following formula:  $P_{ij}^* = \sum_{k=1}^r \min(p_{ik}, p_{jk}) / \sum_{k=1}^r \max(p_{ik}, p_{jk})$ , where:  $i, j$  – numbers (Id) of objects,  $k$  – number (Id) of the component of the structure,  $p_{ik}$  – share of  $k$ -th component in the structure of  $i$ -th object,  $p_{jk}$  – share of  $k$ -th component in the structure of  $j$ -th object. The higher the value of this indicator, the more similar structures of objects. When the value of the indicator equals 1, the structures are identical (Sobczyk, 2010, p. 181).

Table 2 presents the results of the selection of the final measure according to various criteria. The results differ from each other as different selection methods may lead to different outcomes.

Table 3 presents the rankings of the subregions according to the measures selected using methods 1 to 5 (Table 2). The rankings of the subregions in the first ten positions are identical. In each of the five final rankings, Szczecin was the best subregion, while the Włocławski subregion was the worst. Table 4 shows the values of Spearman's correlation coefficients for the five analysed TMs. There was a high consistency in the linear orderings of the subregions according to the values of the analysed measures.

---

<sup>3</sup> Taxonomic measures were used to create rankings of objects and identify groups of objects with similar values of the measure. This is the reason for considering in this article the discrimination ability of taxonomic measures.

**Table 2.** Selected weight vectors according to different methods

Id	Method for selecting the final representative	Id (number) of selected weight vector
1	Minimal value of the semi-standard deviation of Spearman's correlation coefficients	988
2	Maximal value of the mean of Spearman's correlation coefficients	591
3	Maximal value of Sokołowski's coefficient	468
4	Minimal sum of distances to other taxonomic measures	808
5	Maximal mean of similarity indicators of weight structures	752

Source: own elaboration.

**Table 3.** Rankings of the subregions according to the taxonomic measures listed in Table 2

Id of subregion	Subregion	TM 468	TM 591	TM 752	TM 808	TM 988
1	2	3	4	5	6	7
25	miasto Szczecin	1	1	1	1	1
22	miasto Poznań	2	2	2	2	2
17	kaliski	3	3	3	3	3
42	trójmiejski	4	4	4	4	4
14	tyski	5	5	5	5	5
65	miasto Warszawa	6	6	6	6	6
31	miasto Wrocław	7	7	7	7	7
21	poznański	8	8	8	8	8
19	leszczyński	9	9	9	9	9
2	miasto Kraków	10	10	10	10	10
39	gdański	11	11	12	12	12
7	bielski	12	12	11	11	11
9	częstochowski	13	13	13	13	13
3	nowosądecki	14	17	15	15	14
1	krakowski	15	15	14	14	15
18	koniński	16	16	16	16	16
20	pilski	17	14	17	17	17
4	oświęcimski	18	22	20	21	21
12	rybnicki	19	19	18	18	18
15	gorzowski	20	20	21	20	20
66	warszawski wschodni	21	23	23	23	24
13	sosnowiecki	22	26	24	24	27
11	katowicki	23	18	19	19	19

Table 3, cont.

1	2	3	4	5	6	7
67	warszawski zachodni	24	21	22	22	23
16	zielonogórski	25	25	25	25	25
49	piotrkowski	26	27	26	26	26
52	kielecki	27	29	29	29	29
40	śląpski	28	24	27	27	22
61	tarnobrzesci	29	31	30	30	28
10	gliwicki	30	28	28	28	30
5	tarnowski	31	35	32	32	31
34	bydgosko-toruński	32	30	31	31	32
29	wałbrzyski	33	33	36	37	40
33	opolski	34	32	33	33	33
30	wrocławski	35	34	34	34	36
43	chojnicki	36	37	37	36	35
32	nyski	37	38	39	40	43
6	nowotarski	38	36	35	35	34
62	białostocki	39	46	44	42	44
72	siedlecki	40	43	38	38	38
8	bytomski	41	41	41	43	47
47	łódzki	42	39	43	44	49
41	starogardzki	43	40	42	41	41
50	sieradzki	44	42	40	39	39
28	legnicko-głogowski	45	45	45	45	45
58	krośnieński	46	48	48	47	48
27	jeleniogórski	47	44	46	49	50
53	sandomiersko-jędrzejowski	48	47	47	46	37
56	lubelski	49	49	49	48	46
26	szczeciński	50	51	52	52	53
57	puławski	51	52	51	51	42
73	żyrardowski	52	50	50	50	51
60	rzeszowski	53	57	53	53	52
64	suwalski	54	53	54	54	54
46	olsztyński	55	55	57	57	56
23	koszaliński	56	54	56	56	58
51	skierniewicki	57	56	55	55	55
55	chełmsko-zamojski	58	60	59	58	57
54	białski	59	61	61	61	60

1	2	3	4	5	6	7
63	łomżyński	60	59	60	59	59
48	miasto Łódź	61	58	58	60	61
69	ostrołęcki	62	62	62	62	62
68	ciechanowski	63	64	64	63	63
45	ełcki	64	66	67	67	68
38	świecki	65	63	63	64	64
37	inowrocławski	66	67	66	66	67
44	elbląski	67	65	65	65	66
59	przemyski	68	69	69	68	65
71	płocki	69	68	68	69	69
24	szczecinecko-pyrzycki	70	70	70	70	70
70	radomski	71	72	72	72	72
35	grudziądzki	72	71	71	71	71
36	włocławski	73	73	73	73	73

Source: own elaboration.

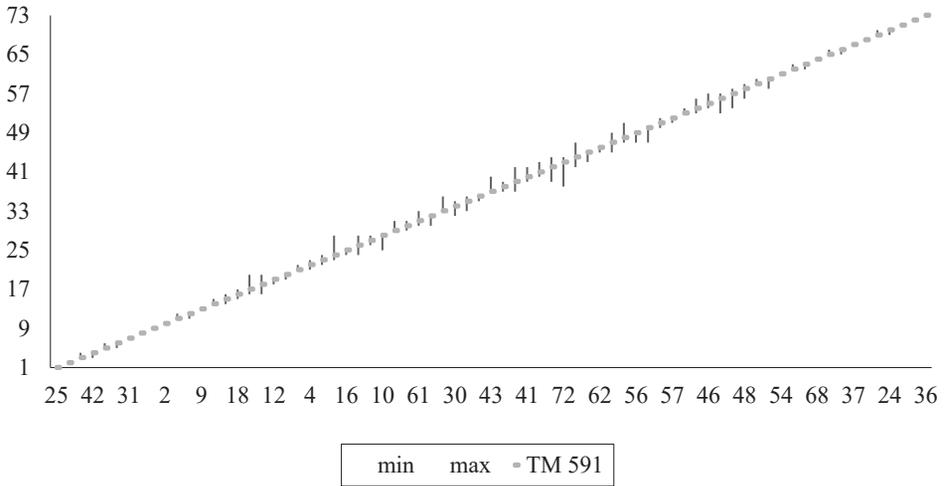
**Table 4.** Spearman's correlation coefficients for different final representatives

	TM 468	TM 591	TM 752	TM 808	TM 988
TM 468	1.0000	0.9953	0.9972	0.9971	0.9903
TM 591	0.9953	1.0000	0.9977	0.9966	0.9897
TM 752	0.9972	0.9977	1.0000	0.9994	0.9936
TM 808	0.9971	0.9966	0.9994	1.0000	0.9952
TM 988	0.9903	0.9897	0.9936	0.9952	1.0000

Source: own elaboration.

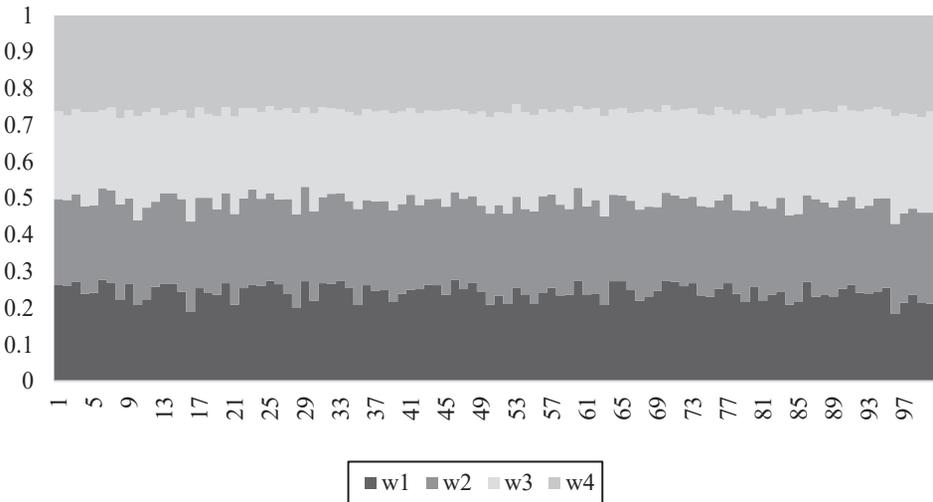
## 5. Stability of results

The weights obtained by the use of the proposed method depend on the initial set of  $k$  weights vectors. To assess the stability of the obtained rankings, the procedure was repeated 100 times and the obtained results were compared. The final criteria in the fourth step of the method was the maximal mean of Spearman's correlation coefficients. Figure 6 presents the range of the subregions' positions obtained in 100 final rankings. The positions obtained according to the values of TM 591 were used as a benchmark (Tables 2 and 3). There was a high consistency in the linear ordering of the subregions. Next, the matrix was determined containing Spearman's correlation coefficients for the values of the obtained taxonomic measures; it contained 100 rows and 100 columns. The minimal value of these coefficients was 0.9967. This result confirms the high consistency of the linear orderings of the subregions.



**Fig. 6.** Range of positions of the subregions in the 100 final rankings

Source: own elaboration.



**Fig. 7.** Structure of weights of the 100 final representatives

Source: own elaboration.

Next, the author analysed the stability of the weights' structures of the 100 final representatives. Figure 7 shows the structures of the weights of the measures. The matrix containing the indicators of the structures' similarity was determined;

it contained 100 rows and 100 columns. The minimal value of these indicators was 0.8165. This result confirms the high similarity of the weights' structures.

## 6. Conclusion

The existing literature offers many different methods for determining the weights of the components of taxonomic measures. Different methods can lead to different rankings. The paper proposed a simulation method for determining the weights of the diagnostic variables. The proposed method makes it possible to find a local solution (a vector of weights and the taxonomic measure depending on the initial set of weights). The accuracy of the obtained solution depends on the number of the initial weights vectors that are randomly generated. The higher the number of these initial weights vectors, the more accurate the obtained solution.

## References

- Badach, E., and Ziolo, M. (2015). Badanie właściwości populacji bezrobotnych na lokalnych rynkach pracy województwa małopolskiego z wykorzystaniem taksonomii wrocławskiej. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, XVII/3, 25-30.
- Becker, W., Saisana, M., Paruolo, P., and Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, (80), 12-22.
- Gan, X. et al. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, (81), 491-502
- Gawrycka, M., and Szymczak, A. (2013). Zróżnicowanie rynków pracy na przykładzie krajów Grupy Wyszehradzkiej. *Studia Ekonomiczne Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (160), 74-82.
- Greco, S., Ishizka, A., Tasiou, M., and Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation and robustness. *Soc Indic Res*, (141), 61-94.
- GUS. (n.d.). Retrieved from [www.stat.gov.pl](http://www.stat.gov.pl)
- Just, M., and Śmiglak-Krajewska, M. (2013). Pomiar zmienności cen na rynku ziarna roślin strączkowych uprawianych w Polsce oraz rynku śruty sojowej. *Zeszyty Naukowe SGGW w Warszawie – Problemy Rolnictwa Światowego*, 13(28), 1, 58-69.
- Kądziołka, K. (2015). Bezrobocie, ubóstwo i przestępczość w Polsce. Analiza zależności na poziomie województw. *Studia Ekonomiczne Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (242), 72-84.
- Kukuła, K., and Luty, L. (2018). O wyborze metody porządkowania liniowego do oceny gospodarki odpadami w Polsce w ujęciu przestrzennym. *Problemy Rolnictwa Światowego*, 18, XXXIII/2, 183-192.
- Młodak, A. (2006). *Analiza taksonomiczna w statystyce regionalnej*. Warszawa: Difin.
- Pietrzak, M. B. (2016). Zastosowanie przestrzennego taksonomicznego miernika rozwoju (pTMR) w analizie rynku pracy w Polsce. *Studia Ekonomiczne Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (291), 47-58.
- Rollnik-Sadowska, E., Jaročka, M., and Glińska, E. (2018). Taxonomic analysis of diversity of the local labour market in Poland. *Society, Integration, Education, Proceedings of the International Scientific Conference*, VI, 437-448.

- Roszkowska, E., and Lasakevic, R. (2017). Taksonomiczna analiza zróżnicowania rozwoju rynku pracy na Litwie w latach 2004-2014. *Optimum. Studia Ekonomiczne*, 6(90), 42-61.
- Sojka, E. (2013). Analiza sytuacji na lokalnym rynku pracy z wykorzystaniem zmiennej syntetycznej. *Studia Ekonomiczne Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (160), 33-43.
- Sojka, E. (2014). Analiza porównawcza powiatowych rynków pracy województwa śląskiego. *Studia Ekonomiczne Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (181), 102-113.
- Sobczyk, M. (2010). *Statystyka opisowa*. Warszawa: Wydawnictwo C.H. Beck.
- Tatarczak, A., Boichuk, O. (2018). The multivariate techniques in the evaluation of unemployment analysis of Polish regions. *Oeconomia Copernicana*, 9(3), 361-380.
- Wydymus, S. (1984). Zmienne syntetyczne w modelowaniu ekonometrycznym. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, (1), 187-200.
- Zhou, P., Ang, B. W., and Poh, K. L. (2007). A mathematical programming approach to constructing composite indicators. *Ecological Economics*, (62), 291-297.

## PROPOZYCJA METODY WYZNACZANIA WAG SKŁADOWYCH MIERNIKÓW TAKSONOMICZNYCH

**Streszczenie:** W pracy zaproponowano symulacyjną metodę wyznaczania wag składowych miernika taksonomicznego. Uwzględnia ona zarówno stopień podobieństwa rankingów uzyskanego na podstawie wartości skonstruowanego finalnego miernika taksonomicznego do innych rankingów, jak również inne jego własności, np. zdolność do grupowania obiektów. Metodę zastosowano do wielowymiarowej oceny sytuacji na rynku pracy w Polsce na poziomie podregionów. Uzyskiwane za pomocą proponowanej metody rezultaty zależą m.in. od początkowego zestawu wygenerowanych losowo wektorów wag. W związku z tym, że proponowana metoda nie daje jednoznacznego rozwiązania (jedynego wektora wag, który byłby niezmienny dla danego zbioru danych), dokonano oceny stabilności uzyskiwanych rankingów. Występowała duża zgodność uporządkowania obiektów z wykorzystaniem mierników taksonomicznych uzyskanych w ramach kolejnych powtórzeń procedury.

**Słowa kluczowe:** miernik taksonomiczny, wagi zmiennych diagnostycznych, semi-odchylenie standardowe, współczynnik korelacji Spearmana, rynek pracy.