THE APPLICATION OF DIFFERENT SYNTHETIC MEASURES FOR STANDARD OF LIVING MEASUREMENT

[a] MARTA KUC

Nicolaus Copernicus University, Ul. Gagarina 13a, 87-100 Toruń, Poland
email: [a]martakuc@doktorant.umk.pl,

Abstract: The synthetic variable is one of the methods to examine differences in the standard of living among countries. The synthetic variable allows to replace the whole set of variables into one aggregated variable. This variable is the basis for organizing and grouping countries in terms of standard of living. In the literature one can find various aggregation methods. The purpose of this article is to show, how the application of chosen method of creating the synthetic measure affects the final result of ordering and grouping objects.

Keywords: synthetic measure, comparative analysis, standard of living.

1 Introduction

Nowadays the standard of living has an increasing role in the European Union integration process. Without doubts there is a need to analyze the standard of living issue because it is a source for defining goals and measuring the effectiveness of social policy. The standard of living is a multidimensional and interdisciplinary category thus it is hard to define and quantified it in a direct manner. In this paper the definition proposed by Bywalec and Wydmus [1992] has been used – by standard of living we can understand the level of wealth, comfort, material goods and necessities available to a certain socioeconomic class in a certain geographic area.

The synthetic variable is one of the methods that can be used to describe changes in the standard of living. It allows to indentify and measure the spatial differentiation among given objects. In this article four different methods of creating the synthetic variable were presented. The study was carried out for European Union countries with the exception on Cyprus, Malta and Luxemburg. The empirical material was taken from databases published by Eurostat, Euromonitor and the World Health Organization.

To construct the synthetic variables 35 diagnostic variables have been used (see table 1). All those variables according to formal and essential conditions are crucial to describe the studied phenomenon. Synthetic variable allows to transform the units described by many variables into the one-dimensional space. Results of analysis allowed to compare the effects of ordering and grouping countries received by using different aggregation method.

2 Different proposals of synthetic variable construction

There are variety of methods for creating a synthetic variable, for instance: Hellwig [1968], Cieślik [1974], Strahl [1978], Grabiński [1992], Grabiński, Wydmus and Zeliaś [1993], Zeliaś and Malina [1997] and so on. In this paper, four chosen methods have been briefly described.

2.1 Taxonomic measure of development proposed by Hellwig

In this method we are looking for an "ideal object" which is described as:

$$y_{oj} = \max_i y_{ij};\ j \in S$$

$$y_{oj} = \min_i y_{ij};\ j \in D$$

The ideal coordinates of object are the maximum value if a variable is stimulant (the higher the value of analyzed variable, the better the studied issue is evaluated) or the minimum value if a variable is destimulant (the lower value of analyzed variable, the better the studied issue is evaluated).

Table. 1. The set of variables

Symbol	Variables
	1. Population
$X_{1,1}$	Total fertility rate
$X_{1,2}$	Demographic dependency ratio of elderly people (in %)
	2. Labour market and job security
$X_{2,1}$	Unemployment rate (in %)
$X_{2,2}$	Number of deaths due to accident at work per 100 000 inhabitants
	3. Health and social care
$X_{3,1}$	Number of deaths due to cancer per 100 000 inhabitants
$X_{3,2}$	Number of deaths due to diabetes per 100 000 inhabitants
$X_{3,3}$	Number of new AIDS cases per 100 000 inhabitants
$X_{3,4}$	Number of doctors per 100 000 inhabitants
$X_{3,5}$	Number of nurses per 100 000 inhabitants
$X_{3,6}$	Number of hospital beds per 100 000 inhabitants
$X_{3,7}$	Obese population (BMI 30kg/sq m or more) as a percentage of population aged 15+
	4. Education
$X_{4,1}$	Number of university students per 1000 inhabitants
$X_{4,2}$	Number of academic teachers per 1 student
	5. Recreation, culture and leisure time
$X_{5,1}$	Annual cinema trips per capita
$X_{5,2}$	Number of hotels per 1000 inhabitants
	6. Living conditions
$X_{6,1}$	Number of newly built dwellings per 1000 households
	7. Transport and communication
$X_{7,1}$	Number of newly registered cars per 1000 inhabitants
$X_{7,2}$	Length of expressways in km per 1 sq km of land
$X_{7,3}$	Proportion of paved roads as a percentage of total road network
$X_{7,4}$	Density of road network in km per 1 sq km of land
$X_{7,5}$	Length of public railway network operated per 1000 sq km of land
$X_{7,6}$	Number of mobile phones subscribers per 100 inhabitants
$X_{7,7}$	Railway passenger traffic in million passenger-km per 1000 inhabitants
$X_{7,8}$	Airline passenger traffic in millions of passenger-km per 1000 inhabitants
	8. Social security
$X_{8,1}$	Number of suicides and self-harm per 100 thousand inhabitants
$X_{8,2}$	Number of divorces per 1000 inhabitants
$X_{8,3}$	Number of crimes per 100 000 inhabitants
	9. Population incomes and expenditures
$X_{9,1}$	Annual average rate of inflation (in %)
$X_{9,2}$	Gross domestic product per capita in USD
$X_{9,3}$	Household saving as % of disposable income
	10. Degradation and protection of the environment
$X_{10,1}$	Sulfur oxides emissions in kg per capita
$X_{10,2}$	Nitrogen oxide emissions in kg per capita
$X_{10,3}$	Carbon monoxide emissions in kg per capita
$X_{10,4}$	Nationally protected areas as a percentage of land
$X_{10,5}$	Forest land as a percentage of land

Source: Author's own study

When we have the coordinates of the model object we are looking for the distance between this object and given countries using Euclidean distance:

$$c_{io} = \left[\sum_{j=1}^{n} \left(z_{oj} - z_{ij} \right) \right]^{1/2}$$

where:

c_{io} - Euclidean distance between country z_{ij} and an "ideal object" z_{oj} ,

z_{oj} - coordinates of an "ideal object",

z_{ij} - normalized value of the j-th variable on object i , according to formula:

$$z_{ij} = \frac{x_{ij} - \overline{x_j}}{s_j}$$

where:

x_{ij} – real value of the j-th variable on object i ,

$\overline{x_j}$ - mean value of the j-th variable,

s_j - standard deviation of j-th variable.

The synthetic variable is calculated using formula:

$$d_i = 1 - \frac{c_{io}}{c_o}$$

where
d_i - taxonomic measure of development proposed by Hellwig
c_{io} - Euclidean distance between country z_{ij} and an "ideal object" z_{oj} ,
c_o - critical distance between objects and an "ideal object" z_{oj} ,

$$c_o = \overline{c_o} + 2s_d$$

$$\overline{c_o} = \frac{1}{n}\sum_{j=1}^{n} c_{io}$$

$$s_d = \left[\frac{1}{n}\sum_{j=1}^{n}\left(c_{io} - \overline{c_o}\right)^2\right]^{1/2}$$

2.2 Absolute measure of development proposed by Cieślik

In Cieślik's method we are not looking for any ideal object. Countries are ordered on the basis of the value of attributes and average variation between objects, expressed by the standard deviation. To build the synthetic variable we use the formula:

$$d_i = \sum_{j=1}^{n} x'_{ij}$$

where:
d_i - absolute measure of development proposed by Cieślik,
x'_{ij} – normalized value of the j-th variable on object i .

$$x'_{ij} = \frac{x_{ij}}{s_j}$$

2.3 Synthetic variable proposed by Strahl

Also in Strahl's method we are not looking for any ideal object. The synthetic variable is built using following formula:

$$d_i = \sum_{j=1}^{n} x'_{ij} = \sum_{j=1}^{n} x'^{S}_{ij} + \sum_{j=1}^{n} x'^{D}_{ij}$$

where:
d_i - synthetic variable proposed by Strahl,
x'_{ij} – normalized value of the j-th variable on object i ,

$$x'^{S}_{ij} = \frac{x_{ij}}{\max x_{ij}}$$ if a variable is a stimulant,

$$x'^{D}_{ij} = \frac{\min x_{ij}}{x_{ij}}$$ if a variable is a destimulant.

2.4 Synthetic variable proposed by Zeliaś and Malina

The matrix of standardized diagnostic variables is the basis for the construction of a synthetic variable according to the formula:

$$d_i = \frac{1}{p}\sum_{q=1}^{p} z_{iq}$$

where:
d_i – synthetic variable value for country i ,
z_{iq} – synthetic variable value for country i calculated on the base of variables belonging to group q ,
p – number of groups.

Creation of a synthetic variable proceeded as follows:
- calculation the synthetic variable for a given group as a mean of the standardized variables,
- construction the synthetic measure according to given formula as a mean of synthetic variables calculated for each group.

3 The standard of living measurement

Synthetic varaibles describing the standard of living in the European Union countries were calculated using all methods mentioned in section 2. Table 2 shows values of synthetic variables for each method and position of each country according to achived standard of living. Calculated values of synthetic variable describing the standard of living in chosen European Union countries are the basis to order these countries from best to worst in terms of the studied phenomenon. The grades were given to each country, in such a way that a rank 1 represents the country with the highest value of the synthetic variable, and a rank 24 represents the country with the lowest value of the variable.

Table. 2. Values of synthetic variables for European Union countries and position of each country according to achived standard of living.

Country	Method proposed by							
	Hellwig		Cieślik		Strahl		Zelias & Malina	
	d_i	rank	d_i	rank	d_i	rank	d_i	rank
AT	0,704	3	49,943	4	0,711	3	0,614	3
BE	0,651	14	44,558	15	0,586	15	0,528	13
BG	0,616	22	38,403	24	0,493	24	0,432	24
DK	0,674	9	49,116	5	0,662	7	0,575	7
EE	0,645	15	44,172	16	0,568	17	0,478	18
FI	0,665	11	48,875	6	0,640	12	0,541	12
FR	0,670	10	47,955	10	0,650	10	0,594	5
GR	0,683	6	48,574	8	0,673	6	0,571	8
ES	0,715	2	48,494	9	0,692	4	0,613	4
NL	0,696	4	54,786	2	0,724	2	0,624	2
IE	0,853	1	62,887	1	0,888	1	0,773	1
LT	0,622	21	41,279	19	0,509	22	0,437	23
LV	0,614	24	39,632	22	0,502	23	0,443	22
DE	0,676	8	47,348	11	0,654	8	0,570	9
PL	0,636	19	43,595	17	0,570	16	0,503	17
PT	0,642	17	39,553	23	0,530	20	0,470	20
CZ	0,653	13	45,180	14	0,600	14	0,521	14
RO	0,616	23	40,687	21	0,531	19	0,472	19
SK	0,644	16	45,347	13	0,608	13	0,517	15
SL	0,638	18	41,361	18	0,564	18	0,516	16
SE	0,663	12	48,602	7	0,644	11	0,551	11
HU	0,624	20	40,878	20	0,518	21	0,450	21
UK	0,690	5	50,210	3	0,689	5	0,583	6
IT	0,682	7	46,219	12	0,651	9	0,567	10

Source: Author's own study.

On the basis of previous considerations, the classification of EU countries has been made. To create the synthetic measure the variable interval was built using mean $\overline{d}$ and the standard deviation s_d . Groups were formed as follows:
- group I (high quality of life): $d_i > \overline{d} + s_d$,
- group II (medium quality of life): $\overline{d} + s_d > d_i > \overline{d}$,
- group III (low quality of life): $\overline{d} > d_i > \overline{d} - s_d$,
- group IV(the lowest quality of life): $d_i < \overline{d} - s_d$

The results of grouping countries are presented in table 3 and figure 1.

Table 3.Countries clustering.

Group no.	Method proposed by			
	Hellwig	Cieślik	Strahl	Zeliaś and Malina
Group I	IE	IE, AT, NL	IE, NL	IE, NL
Group II	AT, DK, FR, GR, ES, NL, DE, UK, IT	DK, FR, GR, ES, DE, UK, IT, FI, SE	AT, DK, FR, GR, ES,DE, UK, IT, FI, SE	AT, DK, FR, GR, ES, DE, UK, IT, FI, SE
Group III	BE, EE, FI, LT, PL, PT, CZ, SK, SI, SE, HU	BE, EE, PL, PT, CZ, SK, SI, RO	BE, EE, PL, PT, CZ, SK, SI, RO	BE, EE, PL, CZ, SK, SI, RO, LT, HU
Group IV	BG, LV, RO	LT, HU, BG, LV	LT, HU, BG, LV	PT, BG, LV

Source: Author's own study.

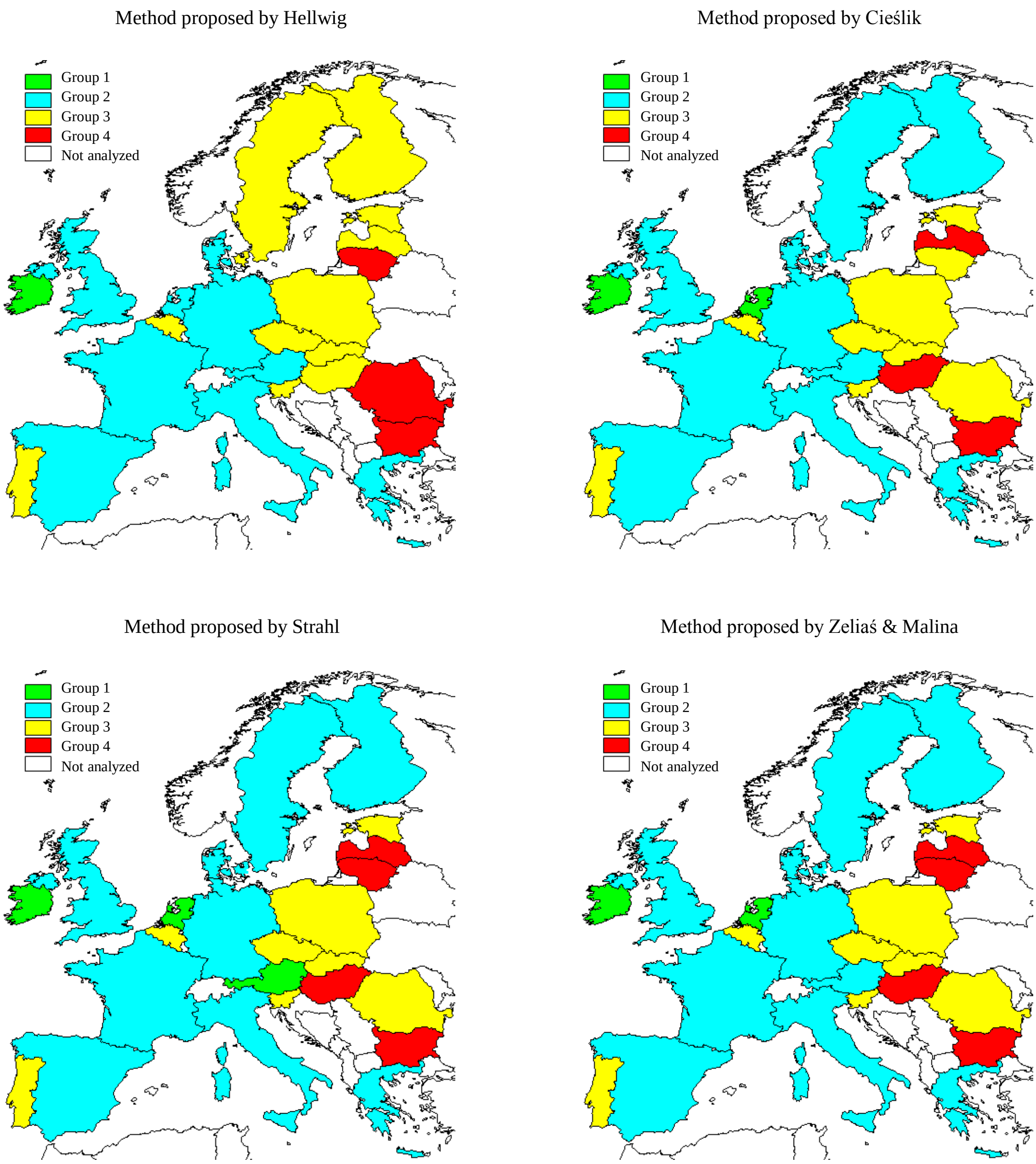

Figure 1. Countries clustering using different aggregation methods.
Source: Author's own study.

4 The comparison of presented methods

4.1 Spearman rho

In order to verify concordance between two linear orders the Spearman rho was calculated [B. Monjeardet], according to formula:

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)}$$

where:
r – Spearman rho,
$d_i = r_{1i} - r_{2i}$,
r_{1i} – rank of i th unit in the first ranking,
r_{2i} – rank of i th unit in the second ranking,
n – number of units.

Table 4. Values of Spearman rho comparing the ordering of the European Union countries.

	Hellwig	Cieślik	Strahl	Zeliaś & Malina
Hellwig	-----	0,900	0.968	0,951
Cieślik		----	0,942	0,926
Strahl			----	0,979
Zeliaś & Malina				----

Source: Author's own study

Afterwards the statistical significance of the correlation coefficient has been examined using t-student statistic. For all the values presented in table 4 there is a relationship $t_{(n-2)} > t_{\alpha}$ so the null hypothesis is rejected, therefore Spearman rho is statistically significant for α = 0.01. This means that there is a high correlation between rankings of EU countries. As we can see in table 4 the highest correlation of results we achieve using methods proposed by Strahl and Zeliaś & Malina. The biggest differences are between method proposed by Hellwig and Cieślik. No matter which method was used Ireland was at the first plane in the ranking. Country with the lowest standard of living according to three methods was Bulgaria.

4.2 Cramer's V

Earlier analysis shows that results of clustering in all methods are very similar. However we should conduct a detailed analysis of accuracy in results obtained by these four methods. For this purpose, contingency tables were constructed and values of Cramer's V were calculated [Grabiński, Wydmus, Zeliaś, 1983].

$$v = \sqrt{\frac{\chi^2}{m \cdot \min\{(r-1),(s-1)\}}}$$

where:
m – number of objects,
r and s – size of the contingency table,
χ^2 - value of chi-square statistics, calculated as:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(m_{ij} - \hat{m}_{ij}\right)^2}{\hat{m}_{ij}}$$

where:

m_{ij} - values calculated inside the contingency table,

$\hat{m}_{ij}$ – theoretical values, calculated as:

$$\hat{m}_{ij} = \frac{m_{i.}m_{.j}}{m_{ij}}$$

where:
$m_{i.}m.j$ - boundary elements of the contingency table.

Cramer's V is defined between $\langle 0,1 \rangle$, wherein the coefficient is closer to unity, the higher accuracy of the results. Table 5 contains the calculated values of Cramer's V for the classification results by all methods. Cramer's V calculated according to the above mentioned formula is compared to the threshold value V*= 0.485. If the calculated value of Cramer's V exceeds the threshold value, this means that there is agreement in the classification between two methods. Analyzing the results from the table 5 we can see a high accuracy of the results.

Table 5. Cramer's V value between results obtained using different aggregation methods.

	Hellwig	Cieślik	Strahl	Zeliaś & Malina
Hellwig	-----	0,698	0.619	0,665
Cieślik		----	0,753	0,893
Strahl			----	0,877
Zeliaś & Malina				----

Source: Author's own study

As we can see at table 5 the biggest differences in clustering countries can be observed when we compare method proposed by Hellwig and Cieślik. Hellwig's method is the one with the lowest correlation with other methods. The average Cramer's V value for Hellwig's method is only 0,661- it is still bigger than V*=0,485 but the agreement in the classification is not so high.

5 Conclusions

- Based on the Spearman rho we can see that all aggregation methods gives very similar results. Thus their usage do not significantly affect the resutls of ordering countries due to the standard of living.
- The differences in ranking usually oscillate between +/- 2-3 position.
- Based on Cramer's V value we can see that presented aggregation methods gives similar clusteris results. The lowest correlation is observed in Hellwig's method. In Hellwig's method Ireland was the only country clasiffied in group I, however in onther methods to the first group belongs also Netherlands. Using Hellwig's method Finland was classified in group III, but other methods place Finland in gropu II. Also Romania has lower position in Hellwig's clustering than in ohter methods. Even thought the usage of different aggregation methods do not significantly affect the result of clustering countries due to the standard of living.
- Considering the standard of living measurement the highest position in the entire set belongs to Ireland – this was confirmed by all clasiffications. The lowest standard of living in Bulgaria was confirmed by three of the four clasiffications.
- It seems that the best method to construt the synthetic varaible for the standard of living measurement is the method proposed by Zeliaś and Malina. It was shown in this paper that this method has the highest avarage correlation with other methods and it is also relatively easy to calcuatle and interpret.

Literature:

1. Bywalec, Cz., Wydmus, S.: *Poziom życia ludności Polski w porównaniu z krajami Europejskiej Wspólnoty Gospodarczej. Ekonomista* nr 5-6. ISSN 0013-3205.
2. Grabiński, T.: *Metody taksonometrii.* Kraków: Wydawnictwo Akademii Ekonomicznej, 1992.
3. Grabiński, T., Wydmus, S., Zeliaś, A.: *Metody prognozowania rozwoju społeczno-gospodarczego*. Kraków: Wydawnictwo Akademii Ekonomicznej, 1993. ISBN 83-20802-73-3.
4. Hellwig, Z.: *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom rozwoju oraz zasady i strukturę wykwalifikowanych kadr, Przegląd Statystyczny*, nr 4, Warszawa: Wydawnictwo Naukowe PWN, 1968. ISSN 0033-2372.
5. Młodak, A.: *Analiza taksonomiczna w statystyce regionalnej.* Warszawa: Difin, 2006. ISBN 83-72516-05-7.
6. Monjardet, B.: *Concordance between Two Linear Orders: The Spearman and Kendall Coefficients Revisited,* "Journal of Classification" vol. 14. DOI: 10.1007/s003579900013.
7. Strahl, D.: *Propozycja konstrukcji miary syntetycznej*, „Przegląd Statystyczny". Warszawa: Wydawnictwo Naukowe PWN, 1978. ISSN 0033-2372.
8. Zeliaś, A., Malina, A.: *O budowie taksonomicznej miary jakości życia*., „Taksonomia", z. 4. Wrocław, 1994. ISBN 83-7011-138-6.
9. Zeliaś, A.: *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*. Kraków: Wydawnictwo AE, 2002. ISBN: 83-7252-065-8.
10. Zeliaś, A.: *Poziom życia w Polsce i krajach Unii Europejskiej,* Warszawa: Wydawnictwo Naukowe PWE, 2004. ISBN: 83-208-1526-6.

Primary Paper Section: A

Secondary Paper Section: H, O