Dynamic measure of development

Iwona Müller-Frączek[1]

Abstract

The article presents a possible using of the author's method – normalisation with respect to the pattern – in the construction of synthetic measure. When a stimulant (destimulant) is normalized, for each object the share of its distances from the maximum (minimum) in the total distance from the maximum (minimum) of all objects is determined. Such transformation meets the requirements of normalisation - deprives variables their units and unifies their ranges. Normalisation with respect to the pattern has properties suggested in the literature - preserves skewness, kurtosis and the Pearson correlation coefficients. Moreover, although the current data are the sole data used to convert variables, normalized diagnostic variables are comparable across time. This feature gives an advantage of pattern normalisation over other methods in dynamic analysis of complex phenomena.

The article uses normalisation with respect to the pattern in construction of Hellwig's measure of development, in which Euclidean distances from an abstract ideal point are calculated. Since normalized diagnostic variables become destimulants with the minimum value equals 0, the ideal point used to construct a synthetic measure is constant over time. So, the values of modified measures are comparable both across objects and time. One can compare the positions of objects in the rankings as well as the values of the measures themselves (calculate the increments of values, descriptive characteristics, etc.).

Keywords: *synthetic measure, aggregate variable, composite indicator, normalisation, standardisation*

JEL Classification: C19, C38

DOI: 10.14659/SEMF.2018.01.33

1 Introduction

The article concerns the methods of comparing objects due to the level of a complex phenomenon, and more specifically, the linear ordering of these objects. Objects are identified with points in a multidimensional space. To order these points, their one-dimensional projections are constructed. In this way, a synthetic measure of a complex phenomenon is defined. The synthetic measure is also called the aggregate variable or the composite indicator.

The issues raised in the article are quite popular. For a selected complex phenomenon, many different examples of synthetic variables can be found in the literature. For example, in Booysen (2002), the history of creating various composite indicators of socio-economic

[1]Corresponding author: Nicolaus Copernicus University, Faculty of Economic Sciences and Management, Gagarina 13A, 87-100 Toruń, Poland, e-mail: muller@econ.umk.pl.

development of countries has been described. Synthetic variables are created for both science and politics. They are then also widely exploited in journalism.

Due to its simplicity, the synthetic approach to describing qualitative phenomena has many supporters. On the other hand, this type of simplification of complex phenomena is too great for many professionals. The advantages and disadvantages of composite indicators can be found among others in Saltelli (2007).

It is certain that the construction of a synthetic measure should be carried out with due diligence. Rules that should be followed and subsequent stages of construction are presented, for example, in Saisan and Saltelli (2011) or Zeliaś (2002a).

The last stage of constructing a synthetic variable is the aggregation of diagnostic variables. The most common methods are the simplest ones: arithmetic mean and geometric mean. In Poland, Hellwig's method (more advanced) is widespread. The method, called *measure of economic development*, has been described in Hellwig (1968 a, b) and quoted in Fanchette (1971). The measure of development bases on multidimensional Euclidean distances from a pattern - a point in the space whose coordinates are determined by the most favourable observations of diagnostic variables.

The article proposes a modification of the Hellwig's method. It mainly concerns the use of another type of variables normalisation. There are many methods of normalisation (see for example: Milligan and Cooper, 1988; Jajuga and Walesiak, 2000; Pawełek, 2008; Zeliaś, 2002a, 2002b). The article uses *normalisation with respect to the pattern*. This is a new method proposed in Müller-Frączek (2017b). This normalisation has advantages that allow us to construct a dynamic synthetic measure whose values are comparable both across objects and time.

The use of the pattern normalisation in another type of synthetic variable can be found in Müller-Frączek (2017a).

The layout of the article is as follows: Section2 reminds the original Hellwig's construction, Section 3 quotes the concept and main properties of the normalisation with respect to the pattern, Section 4 presents the construction of dynamic measures of objects development, the article ends with conclusions.

2 Hellwig's measure of development

Assume that our goal is to order $n \in \mathbb{N}$objects according to the level of the complex phenomenon. We know a collection of $r \in \mathbb{N}$ diagnostic variables, which characterize this phenomenon. Then the objects are identified with points in a r-dimensional space with coordinates equal to the values of diagnostic variables. Letx_{ip}, $i = 1, \dots, n$,$p = 1, \dots, r$bethe corresponding data matrix.

Assume that diagnostic variables meet both substantive and statistical requirements (for more details see for example Zeliaś, 1982). Among themwe distinguish stimulants and destimulants. Stimulants positively influence the analyzed phenomenon, whereas the influence of destimulants is negative. The set of stimulants is marked by S, while the set of destimulants by D.

In the first step of the original Hellwig's method, diagnostic variables should be standardized to render them comparable. After standardisation they take the form:

$$x_{ip}^{'} = \frac{x_{ip} - \overline{x_p}}{S(x_p)}, \tag{1}$$

where$\overline{x_p}$is the average value of the variable x_p:

$$\overline{x_p} = \frac{1}{n}\sum_{j=1}^{n} x_{jp} , \tag{2}$$

whereas $S(x_p)$is the standard deviation of the variable x_p:

$$S(x_p) = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_{jp} - \overline{x_p})^2} . \tag{3}$$

In the next step of the method, *the patternof economic development*$x^+ = (x_1^+, x_2^+, \dots, x_r^+) \in \mathbb{R}^r$is determined. The pattern is an idealabstract point in the multidimensional space, whose coordinates take the most favourable values of the considered diagnostic variables:

$$x_p^+ = \begin{cases} \max\limits_j x_{jp}^{'} & \text{if} \quad x_p \in S \\ \min\limits_j x_{jp}^{'} & \text{if} \quad x_p \in D . \end{cases} \tag{4}$$

The basis for the construction of Hellwig's development measure are the Euclidean distances between objects and the pattern:

$$d_i^+ = \sqrt{(x_{i1}^{'} - x_{i1}^+)^2 + (x_{i2}^{'} - x_{i2}^+)^2 + \dots + (x_{ir}^{'} - x_{ir}^+)^2} . \tag{5}$$

These distances form a synthetic measure, which is a numerical characteristic of the analyzed qualitative phenomenon.The greateris value d_i^+, the worse is the situation of i-th object.

Since d_i^+ are not normalizedand the direction of the relationship between them and the phenomenon is opposite than expected,*the measure of economic development* of i-th object isgiven by:

$$m_i = 1 - \frac{d_i^+}{\overline{d^+} + S(d^+)}, \tag{6}$$

where $\overline{d^+}$ is theaverage distance between objects and pattern:

$$\overline{d^+} = \frac{1}{n}\sum_{j=1}^{n} d_j^+, \tag{7}$$

while $S(d^+)$ is the standard deviation of these distances:

$$S(d^+) = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(d_j - \overline{d^+}\right)^2}. \tag{8}$$

The denominator of the formula (6) guarantees limitation of development measure. Only in extreme, very rare cases, the values of the measure go beyond the interval $[0,1]$.

In the original, the Hellwig's method was described in a static situation - for a given unit of time. However, in practice it is also used in dynamic researches. In this case, to ensure the comparability of results across time, both observations for all objects and all time units are taken into account when determining the mean (2) and deviation (3) as well as the pattern (4) (see for example Müller-Frączek and Muszyńska, 2016). However, such stochastic approach raises doubts, especially in the case of regional research in which we work with all objects - that is, a population, not a sample (see also Zeliaś, 2002a, 2002b).

An additional disadvantage of the stochastic approach is the necessity of recalculating all results with the appearance of observations for the next unit of time.

3 Normalisation with respect to the pattern

An application of normalisation with respect to the pattern (Müller-Frączek, 2017b) in the construction of synthetic measures can be a solution of the problems indicated at the end of the previous section. A characteristic feature of this method is the comparability of normalized values of variables across time, although a deterministic and not stochastic approach is used for normalisation.

For simplicity, consider one diagnostic variable, which is observed for $n \in \mathbb{N}$objects and $T \in \mathbb{N}$time units. For each$t = 1, \dots, T$let$x^t = (x_1^t, x_2^t, \dots, x_n^t) \in \mathbb{R}^n$be the corresponding data vector.

For each unit of time $t = 1, \dots, T$, we choose the most favourable value of the variable x^t, we call them the *pattern values* (or *patterns* for short):

$$x^{t+} = \begin{cases} \max\limits_i x_i^t & \text{if} \quad x \in S \\ \min\limits_i x_i^t & \text{if} \quad x \in D \,. \end{cases} \tag{9}$$

Note that pattern values change over time.

For $t = 1, \dots, T$ the pattern values are used for normalisation of the variable x^t according to the formula:

$$u_i^{+t} = \frac{|x_i^t - x^{+t}|}{\sum_{j=1}^n |x_j^t - x^{+t}|} = \begin{cases} \dfrac{x^{+t} - x_i^t}{\sum_{j=1}^n \left(x^{+t} - x_j^t\right)} & \text{if} \quad x \in S \\ \dfrac{x_i^t - x^{+t}}{\sum_{j=1}^n \left(x_j^t - x^{+t}\right)} & \text{if} \quad x \in D \,. \end{cases} \tag{10}$$

The transformation (10), called *normalisation with respect to the pattern* or *pattern normalisation* for short, satisfies the requirements for normalisation - it deprives variables their units and unifies their ranges. Furthermore, pattern normalisation has properties suggested in the literature for such type of transformation (compare Walesiak, 2014; Jajuga and Walesiak, 2000): it preserves skewness and kurtosis of distributions, as well as does not change Pearson's linear correlation coefficients between variables (proofs and more others properties can be found at Müller-Frączek, 2017b).The Table 1. presents some descriptive characteristics of variables after normalisation with respect to the pattern, to simplify the notation, the indexes are omitted.

Table 1. Descriptive characteristics of the distribution of variables after pattern normalisation.

Name of characteristic	Value of characteristic
Arithmeticmean	$\overline{u^+} = \dfrac{1}{n}$
Standard deviation	$S(u^+) = \begin{cases} \dfrac{S(x)}{n(x^+ - \bar{x})} & \text{if} \quad x \in S \\ \dfrac{S(x)}{n(\bar{x} - x^+)} & \text{if} \quad x \in D \end{cases}$
Skewness	$A(u^+) = \begin{cases} -A(x) & \text{dla} \quad x \in S \\ A(x) & \text{dla} \quad x \in D \end{cases}$
Kurtosis	$K(u^+) = K(x)$
Pearson correlationcoefficient	$\mathrm{r}(u_1^+, u_2^+) = \begin{cases} \mathrm{r}(x_1, x_2) & \text{if } x_1, x_2 \in S \text{or} x_1, x_2 \in D \\ -\mathrm{r}(x_1, x_2) & \text{otherwise} \end{cases}$

Transformation (10), like standardisation (1), belongs to the group of normalisations, for which the value of the normalized variable for the i-th object is influenced by all the values of the variable. Scaling (or unitarisation), which is very popular method of normalisation in empirical researches, does not have this feature. In this case, only the maximum and minimum influence the values after normalisation.

Transformation (10) is not just a technical operation. After pattern normalisationvariables have a clear interpretation. For i-th object, the value of normalized variable determines the share of its distance from the pattern in the total distance from the pattern of all objects. In the context of constructing synthetic variables, this means that after pattern normalisation diagnostic variables become destimulants, irrespective of their initial nature.

In dynamic studies, the most important advantage of pattern normalisation is the comparability of the values of normalized variables both across objects and time, although only values from the current unit of time are used for transformation. If the value of a normalized variable for the i-th object has increases, then the situation of this object has worsened.

Normalisation with respect to the pattern can be used, among others, for the construction of synthetic measures. An example is the additive synthetic measure presented in Müller-Frączek, 2017a. The present article proposes an application of pattern normalisation in the construction of *dynamic measure of development* based on the Hellwig's concept.

4 Dynamic synthetic measure

As in Section 2, consider a complex phenomenon, which is characterized by a set of $r \in \mathbb{N}$ diagnostic variables. These variables are observed for $n \in \mathbb{N}$ objects in the space and $T \in \mathbb{N}$ units of time. For each$t = 1, \dots, T$letx_{ip}^{t}bethe data matrixof dimension $n \times r$.

The steps of constructing dynamic measure of development are analogous to the original Hellwig's method. At the beginning, all diagnostic variables are brought to comparability, using pattern normalisation instead of standardisation.This transformation causes that all variables become destimulants, and their minimum values are equal to 0. Therefore, the pattern obteined in the next step takes the form $(0,0, \dots ,0) \in \mathbb{R}^{r}$. This pattern does not change over time, this is an important advantage from the point of view of constructing dynamic synthetic measures.

In the next step the multidimensional Euclidean distances between the objects and the pattern d_i^{0t} are determined. They take the form:

$$d_i^{0t} = \sqrt{(u_{i1}^{+t})^2 + (u_{i2}^{+t})^2 + \cdots + (u_{ir}^{+t})^2}\,. \tag{11}$$

Distances (11) are the quantitative descriptions of the objects due to the analyzed qualitative phenomenon.

Because of the comparability of normalized variables, also distances from the pattern are comparable both across objects and time. If the value of d_i^{0t} is greater than $d_j^{0\tau}$, then the situation of the i-th object at the moment t is worse than the situation of the j-th object at the moment τ.Note that, one can compare not only the rankings, but also the values of the synthetic measure d (calculate its increments, descriptive characteristics, etc.).

Next, for $t = 1, \dots, T$, $i = 1, \dots, n$we define *dynamic measures of development*of the i-th object at the moment $t$by the formula:

$$\mu_i^t = 1 - \frac{d_i^{0t}}{\sqrt{r}}. \tag{12}$$

The higher is the value of the measure, the better is the situation of the object.

Similarly to Pluta (1976), the denominator in the formula (12) depends only on the number of variables. Such a form preserves the comparability of measure μ both across objects and time. Additionally, unlike the original method, the values of measures never go beyond the range [0,1].

Conclusions

The article presents the construction of dynamic measure of development, based on Hellwig's concept. The Hellwig's measure of development uses multidimensional Euclidean distances from a pattern - a point in the space whose coordinates are determined by the most favourable observations of diagnostic variables. In the original, the method applies to static situation.The article proposes a dynamic version of Hellwig’s measure.The proposed modification consists primarily in the application of normalisation with respect to the pattern instead of standardisation. The values of the measures are comparable not only across space, but also across time, although only the current observations are used in their determination. In this way, the stochastic approach to normalisation and pattern determination is avoided. Such approach is controversial in regional research, in which we work with the whole population of objects.

In subsequent studies, an attempt will be made to include spatial relationships in the construction of measures, as in Pietrzak (2014).

References

Booysen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research*, *59*(2), 115-151.

Fanchette, S. (1971). Synchronic and diachronic approaches in the UNESCO project on human resources indicators. Wroclaw Taxonomy and Bivariate diachronic analysis. In:*Workshop on Indicators of National Development*, International Social Science Council, European Consortium for Political Research, Lausanne.

Hellwig, Z. (1968a). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny*,*15*(4), 307-326.

Hellwig, Z. (1968b). *Procedure of Evaluating High-level Manpower Data*. UNESCO, Methods and Analysis Unit, Department of Social Sciences.

Jajuga, K.& Walesiak, M. (2000). Standardisation of Data Set under Different Measurement Scales. *Studies in Classification, Data Analysis, and Knowledge Organization Classification and Information Processing at the Turn of the Millennium,* 105-112.

Milligan, G. W.& Cooper, M. C. (1988). A study of standardisation of variables in cluster analysis. *Journal of Classification*,*5*(2), 181-204.

Müller-Frączek, I. (2017a). Propozycja miary syntetycznej. *Przegląd Statystyczny*,*64*(4), 413-428.

Müller-Frączek, I. (2017b). Normalisation with respect to pattern. *Repozytorium UMK.*http://repozytorium.umk.pl/handle/item/4769.

Müller-Frączek, I. & Muszyńska, J. (2016). Implementation of Europe 2020 Strategy: the taxonomic analysis, In: Simberova, I., Milichovsky, F. & Zizlavsky, O. (eds.), *Smart and efficient economy: preparation for the future innovative economy*, University of Technology Faculty of Business and Management, Brno, 346-355.

Pawełek, B. (2008). *Metody normalizacji zmiennych w badaniach porównawczych złożonych zjawisk ekonomicznych*. Wydawnictwo Uniwersytetu Ekonomicznego, Kraków.

Pietrzak, M. (2014). Taksonomiczny miernik rozwoju (TMR) z uwzględnieniem zależności przestrzennych. *Przegląd Statystyczny*, *61*(2), 181-201.

Pluta, W. (1976). Taksonomiczna procedura prowadzenia syntetycznych badań porównawczych za pomocą zmodyfikowanej miary rozwoju gospodarczego. *Przegląd Statystyczny*, *23*(4), 511-518.

Saisana, M. & Saltelli, A. (2011). Rankings and Ratings: Instructions for Use. *Hague Journal on the Rule of Law*,*3*(02), 247-268.

Saltelli, A. (2007). Composite Indicators between Analysis and Advocacy. *Social Indicators Research,81*(1), 65-77.

Walesiak, M. (2014). Przegląd formuł normalizacji wartości zmiennych oraz ich własności w statystycznej analizie wielowymiarowej. *Przegląd Statystyczny, 61*(4), 363-372.

Zeliaś, A. (1982). Kilka uwag o kryteriach doboru zmiennych w modelach ekonometrycznych. *Folia Oeconomica Cracoviensia, 24,* 21-36.

Zeliaś, A. (2002a). Uwagi na temat wyboru metody normowania zmiennych diagnostycznych, In: Kufel, T., Piłatowska, M. (eds.), *Analiza szeregów czasowych na początku XXI wieku*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.

Zeliaś. A. (2002b). Some notes on the selection of normalisation of diagnostic variables. *Statistics in Transition,5*(5), 787-802.