Agris on-line Papers in Economics and Informatics

Volume II Number 4, 2010

Selected approaches of variables weighting in frame of composite indicator analysis

T. Hlavsa

Institute of agriculture economics and information

Abstract

Composite indicators are useful as tool for complex evaluation and aggregation of different variables of regional development. Variables which are aggregated in a composite indicator have first to be weighted. All variables may be given equal weights or they mea be given differing weights which reflect the significance, reliability or other characteristics of the underlying data. The weights given to different variables heavily influence the outcomes of the composite indicator. Aim of this paper is an evaluation of selected methods for weighting of particular variables in frame of composite indicator construction. Evaluation is verified on group of regional economic variables based on Strategy of regional development.

Key words

Composite indicator, region, principal component analysis, expert.

Anotace

Souhrnné indikátory jsou užitečné pro souhrnné a komplexní hodnocení různých ukazatelů regionálního rozvoje. Proměnné, které jsou zahrnuty do souhrnného indikátorů, mohou být ohodnoceny stejnými nebo různými vahami. Váhy pak můžou odrážet významnost, věrohodnost nebo různá specifika podkladových údajů. Přidělené váhy pak mohou výrazně ovlivňovat výsledek hodnocení. Cílem příspěvku je zhodnotit vybrané metody vážení a ověřit, zda některá z testovaných metod je vhodná pro komplexní zhodnocení ekonomik regionů. Analýza je založena na proměnných a datech pocházejících ze Strategie regionálního rozvoje.

Klíčová slova

Souhrnný indikátor, region, analýza hlavních komponent, expert.

Introduction and aim

The structural policy of European Union focuses on regions with declining industries, distant and rural regions. Its target is to reduce their backwardness and provide balanced and sustainable development through development programs and projects. For the identification of less developed or backward regions there are different standards exploited, these are often based on a one-dimensional point of view (e.g. level of 75 % GDP) [3]. As illustrated, GDP is an important indicator; nevertheless it is not the only important indicator for the evaluation of development and the level of regions [8], [11]. The multidimensional point of view on the regional development appears to be more appropriate. The use of multi-criteria framework is a very efficient tool to implement a multi/inter-disciplinary approach [10]. It is asserted through so called composite indicator. Composite indicators – which are synthetic indices of individual indicators – are being developed in a variety of economic performance and policy areas. The proliferation of composite indicators in various policy domains raises questions regarding their accuracy and reliability. Given the seemingly ad hoc nature of their computation, the sensitivity of the results to different weighting and aggregation techniques, and continuing problems of missing data, composite indicators can result in distorted findings on regional performance and incorrect policy prescriptions.

Variables which are aggregated in a composite indicator have first to be weighted. All variables may be given equal weights or they mea be given

differing weights which reflect the significance, reliability or other characteristics of the underlying data. The weights given to different variables heavily influence the outcomes of the composite indicator.

Aim of this paper is an evaluation of selected methods for weighting of particular variables in frame of composite indicator construction. Evaluation is verified on group of regional economic variables. For its achievement there has been set a few partial aims:

A) Selection of weighting's method

B) The valuation of region's position with the regard for results of weighting

Material and methods

The model of the aggregate indicators and the ways of weighting has been applied on chosen indicators of the theme of SRD Economics of regions. The indicators have been chosen on the basis of expert decision, 7 experts participated (4 from the sphere of research, 2 from the sphere of the regional development of regional authorities and 1 from the Ministry for regional development) and on the basis of the statistic methods of the cluster and the correlation analyses. The selection itself is not the content of this article.

These indicators have been chosen:

GDP per capita, share of employed in construction, unemployment rate, average wage, registered job applicants, share of traders and research and development expenditure. The resources of indicators for the years 2007 have been the regional yearbooks of The Czech Statistical office.

The verification of chosen method has been applied on group of 13 regions NUTS3 in the Czech Republic excluding the capital city Prague. The city Prague is featured by specific position compared to other 13 districts, it only consists of city and for period of time before the year 2007 it was restricted from the structural funds. The work is focused on the modeling of multidimensional statistic methods whose analytical apparatus enables complex analyses mutual incidence relevant indicators.

Also has been selected method of construction of composite indicator. Based on [6] it is Ratio-Median method (RMCI), this indicator was defined

$$RM^{CI} = \frac{\sum_{j=1}^{m} y_{ij} \cdot w_j}{\sum_{j=1}^{m} w_j}, \qquad (1)$$

where

$$y_{ij} = \frac{x_{ij}}{\tilde{x}_{.j}} \qquad (2)$$

Note: index i represents region; i = 1, …., 13 and index j variable; j = 1, …, m; where m is number of variables; xij is original variable; $\tilde{x}_{.j}$ is median of the variable; wj is weight of the variable.

Selected approaches for weights wj determination

As is written above, each variable can be given equal or different weight. Different authors such as Freundenberg [4], Saisana [12] or Svatošová [13] have outlined a range of ways of the weights´ determination for the tracked indicators. Also Grupp and Schubert [5] stress to use weights included in composite indicator, but authors mention that composite indicator should be sensitive to weight changes. The multivariate analysis of principal components seems to be appropriate for the exact appraisal of weights on the basis of primary indicators. Weights can be determined also subjectively on the basis of external decision; this approach has been chosen in the field of environment in the work [7].

PCA – principal component analysis seems to be suitable for the identification of factors and analyses of disparity. The method has been thoroughly theoretically illustrated in [1] and [9].

The method is based on covariance matrix or correlation matrix of input variables from whose the set of eigenvectors of this matrix is obtained, that all is done to represent the variance of primary data as well as possible. The target is to find hidden quantities represented as principal components describing the variability and the dependence of variables. In other words, the method has been trying to express the primary variables with the help

of fewer independent fictive variables which can not be directly measured, but might have certain factual interpretation.

The result of analysis of principal components is consecutive components depleting the maximum of remaining variance of set of variables which are mutually independent. The correlation coefficients of primary variables with gained components are usually the base for the interpretation of the principal components. These correlation coefficients are usually described as component weights. With regard to use of the principal components analysis in the evaluation of regional development, it is possible to refer to work [13] in the field of disparities analysis among regions or [12] in the case of construction of variable's weights.

For the selection of suitable method of weighting is one important requirement thought: weights enable differentiate observed variables

Selected approaches are defined:

PCA^{SUM}

$$w_j = \sum_{s=1}^{r} \left| r_{js} \right| \cdot \mathrm{var}_s \qquad (3)$$

PCA^{SELECT}

$$w_j = \left| r_{js} \right| \cdot \mathrm{var}_s \qquad (4)$$

EXP

$$w_j = \frac{pr_j}{k} \qquad (5)$$

Note: wj is weight of the variable, $\left| \mathrm{rjs} \right|$ is absolute value of the correlation coefficient, index j represents variable and index s selected component, vars is share of variance explained by selected component; j = 1, …, m; where m is number of variables; s = 1, … , r; where r is number of selected components; prj is number of assigned preferences, k is number of experts, it express maximal number of preference that could be assigned (in our case 7).

Method PCA^{SUM}

By the share of explained variability of single chosen components (we do not work with all of them, only with those which represent adequately large proportion of primary variability, usually 70 – 90 %) and correlation coefficients of indicators with those components have been determined the weights for each indicator. The weights may take values from 0 to 1 and are expressed in an absolute amount (all of them are stated with the positive sign).

Method PCA^{SELECT}

The second method is also based on the analysis of the main components. The correlation coefficient which is for certain indicator under the chosen components highest is crucial for determination of the weight of each indicator. The proportion of variance, which is explained by the certain component, is also included in the calculation (we do not work with all the components as it was in the case of the method PCA sum, but only with those which represent the sufficient proportion of the total variability, so it is based on the reduced model). The weight can take values in the interval from 0 to 1.

Method EXP

There are weights assigned by subjective opinion of chosen experts of regional development to each indicator in each thematic area. Each thematic area was evaluated by 7 experts, the indicator can theoretically obtain maximum of 7 points and the minimum was 0. The weight can range from 0 to 1, including 0 and 1.

Results and discussion

A) Selection of weighting's method

The composite indicator is possible to calculate in its weighted and non-weighted form. If we knowingly and purposely do not weight the indicators, we automatically allocate the weight 1 to all indicators. Although it is possible to weight them and to allocate higher preferences to chosen indicators which are considered to be more important. The weight can be calculated either accurately or subjectively. We are going to answer the question if and how it is suitable to weight the indicators in our thematic sphere in the following evaluation where there are chosen methods of weighting compared. The weights make sense if

they difference the indicators. If the weights are balanced, it loses sense to include them in the composite indicator. There are three different methods of calculation of weights: According to the PCA^{SUM} method calculated by the formula (3) and also according to PCA^{SELECT} method (4) and EXP method (5). In the thematic sphere Economics of regions where there have been more indicators available to which experts allocated preferences, there was lower variability of weights allocated by experts. In the mentioned sphere there was the variability zero, the experts identically allocated to all important factors-indicators 5 preferences out of 7. However the height of the weight on the basis of expert method is largely influenced by the small number of experts. It makes it impossible the weight to take the values in the interval from 0 to 1 and realistically can take the value of 8 possible heights in the case of our 7 experts. From this point of view the method didn't present to be suitable.

The highest variability of weights in most of the spheres can be traced when using the PCA^{SELECT} method which enables to emphasize differences in the evaluation for the importance of chosen indicators. The steadiest height of weights in most of the spheres is perceptible when using the results of PCA^{SUM} method.

In the thematic sphere Economics of regions there has been lower fluctuation of weights registered in the case of PCA sum method. Weights were ranking within the interval 0,43 to 0,46 excepting the indicator of the share of employed in construction (0,20) and the share of traders (0,26). In the case of PCA select method, there also occurred higher preference of indicators which are generally used for basic description of regions (GDP - 0,40, unemployment - 0,38, number of applicants - 0,36). There has higher differentiation of weights appeared. The variables with the highest weights were the share of employed in construction (0,15) and the share of traders (0,16). It is perceptible that the indicators with markedly lower preferences were separated from the group of indicators with higher weights both when using PCA sum method and PCA select method. Nevertheless, PCA select method enabled to difference the weights and that is why this method is considered to be the more suitable.

B) The valuation of region's position with the regard for results of weighting

This part of paper is engaged in utilization of weight in composite indicator calculation. Enable weight differences among variables? Influence including weights ranking of regions? Evaluation was based also on the same group of economic indicators. In the table 2 are results of composite indicators computed in weighted or non-weighted form. Non-weighted form represents approach, where weights are equal to 1. Weights in weighted form are computed using method PCA^{SELECT}.

Differences in ranking of regions are not large, but some dissimilarity is evident. Minimum of differences is visible on the fringe of ladder, e.g. first and second place and twelfth and thirteenth. On the contrary, the most number of changes was using weights caused roughly in the middle of ladder. According these results can we say, that weights have important role in regions ranking. Including or excluding of regions in or from the group of financial supported regions is very sensitive question. The best ranked regions as a rule are not supported, the worse ranked are supported. But where is the limit? The limit for supported and the rest is anywhere in the middle of the ladder.

Conclusion

There has been a methodical instrument for the evaluation of regional development suggested in this work. It has been verified on selected indicators of the economic regions sphere. The suitable method for the evaluation of position of the regions has been chosen, the method has been modified by author to suit even better the primary requirements. Also there has been possible of engaging the weights to composite indicator considered. The choice of weights can be influenced by special interest groups, it is why is possible to recommend rather exact and objective methods for their assessment. The important base for the determination of the composite indicator is the quantity of data, which is important to gather for all primary indicators. The missing indicators lower the quality of analysis. The method PCA select has been chosen for the calculation, it has enabled to differentiate the indicators the best.

Variable	PCA^{SUM}	PCA^{SELECT}	EXP
GDP per capita	0,43	0,4	0,71
Share of employed in construction	0,2	0,15	0,71
Unemployment rate	0,45	0,38	0,71
Average wage	0,44	0,29	0,71
Registered job applicants	0,45	0,36	0,71
Share of traders	0,26	0,16	0,71
Research and development expenditure	0,46	0,31	0,71
Variation coefficient of weights in %	28,08	34,69	0

Table 1. Results of weights according to used methods.

Region	Value of CI		Ranking of regions based on	
	$RM^{CI}n$	$RM^{CI}w$	$RM^{CI}n$	$RM^{CI}w$
Středočeský	1,60	1,72	1	1
Jihočeský	1,20	1,23	3	3
Plzeňský	1,35	1,45	2	2
Karlovarský	0,83	0,79	13	12
Ústecký	0,83	0,79	12	13
Liberecký	1,02	1,03	7	8
Královéhradecký	1,00	1,04	9	7
Pardubický	1,18	1,22	4	4
Vysočina	1,05	1,07	6	5
Jihomoravský	1,01	1,00	8	9
Olomoucký	0,91	0,88	10	10
Zlínský	1,06	1,07	5	6
Moravskoslezský	0,87	0,84	11	11

Note: RMCIn = composite indicator with weights = 1; RMCIw = composite indicators with weights PCASELECT.
Table 2. Values of composite indicators according to used methods

With the equal weighting approach, there is the risk that certain performance aspects will be double weighted. This is because two or more indicators may be measuring the same behavior. With the different weighting approach, greater weight should be given to components which are considered to be more significant in the context of the particular composite indicator. The relative economic impact of variables could be determined by economic theory or through empirical analysis, particularly by methods based on correlations among the sub-indicators. To be useful for policy, weights need to reflect the relative importance of individual indicators in determining performance outcomes.

Corresponding author:
Ing. Tomáš Hlavsa, Ph.D.
Institute of Agriculture Economics and Information in Prague
Mánesova 75, 120 00 Praha 2
Phone: 00420 222 000 501, e-mail: hlavsa.tomas@uzei.cz

References

[1] ANDĚL, J. Matematická statistika. SNTL, 1978, Praha.

[2] VANĚK, J. – JAROLÍMEK, J. – ŠIMEK, P. Development of communication infrastructure in rural areas of the Czech Republic. Agricultural Economics. 2008, 54, No. 3, s. 129 – 134. ISSN 0139-570X.

[3] EC. Pokyny k regionální podpoře na období 2007 – 2013. Úřední věstník Evropské unie (2006/C 54/08), Brusel, European Commision, [on-line]. c2006, [cit. 2010-05-20]. Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2006:054:0013:0044:CS:PDF

[4] FREUDENBERG, M. Composite Indicators of Country Performance. OECD Science, Technology and Industry Working Papers 2003/16 [on-line]. c2003, [cit. 2010-05-17]. Available from: http://www.olis.oecd.org/olis/2003doc.nsf/linkto/dsti-doc(2003)16.

[5] GRUPP, H. – SCHUBERT, T. Review and new evidence on composite innovation indicators for evaluating national performance. Research Policy, 39, Nr. 1. Amsterdam, 2010. ISSN 0048-7333.

[6] HLAVSA, T. Metodické nástroje hodnocení rozvoje regionů v rámci regionální politiky EU. Disertační práce, Praha 2010.

[7] KUPROVÁ, L. – KAMENICKÝ, J. Multikriteriální postavení krajů ČR v letech 2000 až 2004. Časopis Statistika, Nr. 4. Praha, 2006. ISSN 0322-788x.

[8] MANLY, B., F., J. Multivariate statistical methods. Chapman and Hall, 2005, Laramie, USA. ISBN 1-58488-414-2.

[9] MARTÍN-GUZMÁN, P. – GIL, M. Official statistics for measuring well-being at national and regional level in EU: Some comments from users. Časopis Statistika, Nr. 1. Praha, 2009. ISSN 0322-788x.

[10] MUNDA, G. Social multi-criteria evaluation: Methodological foundations and operational consequences. European Journal of Operational Research, 158, Nr. 3. Amsterdam, 2004. ISSN 0377-2217.

[11] NARDO, M. et al. Handbook on Constructing Composite Indicators. Brusel: Joint Research Centre, European Commision, 2008, ISBN 978-92-64-04345-9, [on-line]. c2008, [cit. 2009-10-17]. Available from: http://browse.oecdbookshop.org/oecd/pdfs/browseit/3008251E.PDF

[12] SAISANA, M. et al. State of the Art Report on Composite Indicators for the Knowledge-based Economy. 2005 [on-line]. c2005, [cit.2009-05-28] Available from: http://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/KEI-WP5-D5.1.pdf

[13] SVATOŠOVÁ, L. Methodological starting points of regional development analyses. Agricultural economics, 51, Nr. 2. Praha, 2005. ISSN 0139-570X.