THE USAGE OF MULTIVARIATE STATISTICAL METHODS FOR CREATION OF TYPOLOGY OF REGIONS

Lucie Sobotková

University of Pardubice, Faculty of Economics and Administration, Institute of Economics

Abstract: *The article deals with searching basic factors of competitiveness of regions whose knowledge enables to increase an efficiency of instruments of regional policy. The finding itself and assessment of mentioned factors of competitiveness is unsophisticated process. Thereby it is suggested a methodics in this article which uses knowledge of factor and cluster analysis. The successful application of these methods makes possible to create a typology of regions. Hereby we can get generalized current status, structure and hypothetical development of a searched region.*

Keywords: *Cluster Analysis, Competitiveness, Factor Analysis, Factors of Competitiveness, Region, Typology of Regions*

1. Introduction

The current regional science and regional policy more and more deals with question of competitiveness of regions. Reaching and continuous increasing of this competitiveness is not only an instrument but also an objective of regional policy of the European Union (and thanks to this effort to coordinate this policy is a content of regional policy of the Czech Republic). Although is conception of regional competitiveness noticed variously among authors, the European Union utilizes it as an instrument for maintenance of cohesion among European regions as well. The both of objectives take effect on one hand to the contrary, they can however support each other as well. The effort of the EU to help to less developed regions supports the economic performance of regions. These regions become a new partner for developed regions and growing competition among regions can support socioeconomic development of all regions concerned.

Evaluating the level of competitiveness of regions and determining of the main determinants of regional competitiveness, it is necessary to deal with methodics of its measurement. Therefore, this article is focused on a given matter thereby it is suggested to use of multivariate statistical methods for evaluation of regional competitiveness. The determination of typology of basic regions is then the output. Through the medium of stated typology it is possible to e. g. determine in which areas it is important to support a region to increase its competitive advantage. The suggested methodics does not only assess economic aspect of regional competitiveness, but also involves in assessment other development conditions of regions.

2. The competitiveness of regions

The conception of regional competitiveness evokes among authors discussions related to its assumption and using. At this point it is possible to remember e. g. the critical approach of KRUGMAN [1994] to using notion “competitiveness” in case of regions. On the other side we can mention supporters of this approach as well (cf. PORTER [2000], MARTIN [2003], VITURKA [2007] etc). This conception is one of basic instruments of regional policy as well. It is necessary to draw the attention that the approaches of supporters of regional competitiveness differentiate. Many authors, who deal with regional competitiveness, agree that this notion it is necessary to catch in a wider context. Regional competitiveness can be often considered as an aggregate of microeconomic competitiveness and the derivation of national competitiveness according to BOROZAN [2008]. On the other hand e. g. CELLINI and SOCI [2002] understand this notion in more complex conception.

Ignoring the question related to competition of regions, we have to explain the meaning of the notion “regional competitiveness”. E. g. PORTER [2000] mentions a big relation between competitiveness and productivity. The growth of competitive advantage is the basis of course. The competitive advantage is however influenced by other factors connected with localization of firms in a certain area or e. g. the possibility of knowledge spillovers [cf. PORTER, 1998]. By contrast BUDD and HIRMIS [2004] call attention to fact that productivity is not ideal measure or indicator of regional competitiveness. The relationship is complex and proceeds via indirect effects. Increasing returns, external economies and endogenous growth effects have greater influence on regional success.

3. Factors of regional competitiveness

Factors of regional competitiveness are in literature defined from the various points of view. It always depends on that fact how authors try to catch this notion. E. g. GARDINER at al. [2004] represents the regional competitiveness in form of pyramidal model. The basis of this pyramid creates sources of competitiveness: economic structure, environment, innovative activity, decision centres, regional accessibility, social structure, skills of workforce, regional culture, research and technical development, SME development, FDI activity, infrastructure and human capital, institutions and human capital. The other three factors, regional performance (measured via GDP), labour productivity, employment rate, connect with the previous ones. The top of this pyramid is created by two target outcomes, which are basically closely interconnected - quality of life, standard of living.

For comparison it is possible to mention factors of regional competitiveness according to KITSON et al. [2004], who displays given factors in form of hexagon: productive capital, human capital, social-institutional capital, culture capital, infrastructure capital, knowledge/creative capital.

Some of above mentioned factors are known as being “soft” factors that have a more indirect competitive impact. Due to this fact, it makes it difficult to measure regional competitiveness. Further we can state other authors dealing with factors of regional competitiveness, e. g. PORTER [1990], SKOKAN [2004] and institutions e. g. BERMANGROUP [2006], ECORYS NEI (see MARTIN, [2003]) etc.

It is evident from given overviews that regional competitiveness is multiconditional (cf. BERMANGROUP [2006]). Great importance is placed on mutual combination of factors that together create “a favourable local environment”.

4. Statistical analysis of indicators of competitiveness

4.1 Factor analysis for determination of regional competitiveness

The determination of factors of competitiveness of regions is not possible to do without analysis of statistical data at regional level. The regression analysis is the suitable method for determining main factors of competitiveness. Its successful application is conditioned by mutual uncorrelation of input data. In addition we can often suppose the multi co-linearity in case of using statistical indicators.

As stated by STANKOVIČOVÁ and VOJTKOVÁ [2007], multi co-linearity can cause high standard errors in the estimated parameters, leading to an unstable regression model. It also complicates the interpretation of results. PACÁKOVÁ et al. [2009] states that duplicity in the analyzed information, that is contained in input indicators, can be a consequence not being able to fulfil assumption. This can lead to significant distortion of results. The stated problems can be solved, e.g. by using multivariate statistical methods, specifically, using factor analysis methods.

Factor analysis is one method of multivariate statistics. The basic aim of factor analysis is to: assess the structure of the relations between the monitored variables, and determine whether the variables can be divided into groups that would significantly eliminate correlation between the variables. This theory is based on the presumption that the interdependences among observed variables are the consequence of the effect of smaller amount of underlying immeasurable variables (so called common factors). This method allows for the recognition and usage (on the basis of interdependences of common factors) of the structure (directly non-observable and immeasurable) of common factors. According to HEBÁK et al. [2007], the factor analysis seeks to derive, create, and understand the common factors (defined as a linear combination of original variables) such as to interpret and clarify the observed dependence. This means that in the final solution, each variable should correlate with a minimum amount of factors.

Here it is important to call attention to frequent critique of this method which concerns to ambiguity of solution, usage of subjectivity and vague interpretation of results [cf. HEBÁK et al., 2007]. The character of factor analysis is rather based on heuristics than on verification of basic data. Its successful usage requires good knowledge of this method and experience with its using. The knowledge of relations among input indicators is very important for analysis as well.

This method results from the set of observable variables (stochastic magnitude) X_j, $j = 1, 2, \ldots, p$, which have multidimensional distribution with p-termed vector of mean values μ_X and with covariance matrix Σ_p with rank p. According to

STANKOVIČOVÁ and VOJTKOVÁ [2007] the general model of factor analysis supposes the existence of q understanding common factors $F_1, F_2, \ldots, F_q$, of which is less than p. These factors enable j-observable stochastic magnitude X_j to express in following equation:

$$X_j = \mu_{Xj} + a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jq}F_q + e_j$$

where:

a_{jk} – factor masses (costs, saturations) which express the influence of k-common factor on variable X_j,

e_j, $j = 1, 2, \ldots, p$ – stochastic (error) components indicate as specific factors.

The model of factor analysis can be written in following matrix note:

$$\mathbf{X} = \boldsymbol{\mu}_\mathbf{X} + \mathbf{AF} + \mathbf{e}, \text{ or } \mathbf{X} - \boldsymbol{\mu}_\mathbf{X} = \mathbf{AF} + \mathbf{e}.$$

where:

A – matrix of factor masses in form of p x q,

F – q-termed vector of common factors,

e – p-termed vector of specific factors,

X – vector of origin measurable variables which are called as indicators.

Factor masses (saturations) a_{jk} represent the regression coefficients between observable variables and insensible factors.

The previous description of factor analysis was illustrated by solution for which the covariance matrix S_p was initial matrix. The factor analysis can be used for solutions which will result from correlation matrix. Moreover according to HEBÁK et al. [2007] the interpretation is often according to correlation matrix the only one possibility because all of the used variables are seldom by analysis in the same measuring units. It is important to use these same units by usage of covariance matrix.

Before the own application of factor analysis it is necessary to evaluate the input data and exclude so called "trivial" factors. According to HEBÁK et al. [2007] we understand those factors which only correlate with one from p pursued variables. If some variable correlates only with one factor and any other variable does not correlate with this factor, this variable is not suitable for factor analysis.

In case of determination of factors in method of factor analysis it is important to set such a number of factors which would mostly interpret the total variance and at the same time would decrease the count of indicators to simplify the interpretation. STANKOVIČOVÁ and VOJTKOVÁ [2007] state that the explained variance should be 90 – 95 % in case of exact sciences and in case of social sciences bigger than 60 – 70 %. Beside this criterion so called "scree plot" can be used which displays number of factors on x-axis and the percent of interpreted variability, i. e. eigenvalues (variance of principal components) of reduced correlation matrix, on y-axis. It is possible to consider as optimal number of factors the value on x-axis; behind it happens to "break" at curve of eigenvalues.

The result of factor analysis is the matrix of factorial masses. This matrix helps to identify the relationship between common factors and identifiers. The first estimate of the factorial masses may not provide a sufficiently clear interpretation of individual factors. STANKOVIČOVÁ and VOJTKOVÁ [2007] indicate that the first solution is not suitable for interpretation. For ease of interpretation, it is necessary to rotate the common factors. The essence of the rotation method is to get as many factorial masses close to zero, as well as a full blast of other masses close to one.

The methods of rotation of factors can be divided into following methods:

- orthogonal (rectangular),
- oblique.

The difference between mentioned methods of rotation consists in solution of matrix of factor masses. The orthogonal transformations lead to solution with un-correlated factors. The elements of matrix of factor masses can be interpreted as regression coefficients of dependence of indicators from factors and also as correlation coefficients among them. The advantage of orthogonal rotations is the fact that these methods change masses of factors but do not change the cumulative percentage of interpreted variability of common factors. On the other hand the methods of oblique rotations lead to obtaining dependent factors.

The best known orthogonal methods are following:

- varimax method,
- quarimax method,
- orthomax method (biquartimax and equamax are modifications of this method).

It is necessary to apply more methods of rotation by usage of factor analysis. The choice decision of appropriate methods of rotation depends on the ability to interpret the resulting factors. The method of varimax is the most widely used method for rotation of factors [cf. HEBÁK et al., 2007].

The next step in case of application of factor analysis consists in denomination of found factors. The procedure mostly consists in that for each of factors is chosen that variable which has the highest value in set of factor masses. If several variables with high, approximately the same masses, exist, then STANKOVIČOVÁ and VOJTKOVÁ [2007] recommend the following:

- for the factor is chosen that variable which is the most considerable representative of the dimension; it can happen that the chosen variable has lower factor mass than is the highest factor mass,
- the next possibility is the using the average among of all of variables which have high and similar identical mass.

The advantage of factor analysis consists in that it is possible to determinate the rank of competitiveness of particular regions. The estimated values of common factors, called factor scores, were used for this procedure. The values of factor scores can be further used as input into other statistical analyses (see e.g. below described cluster analysis). The factor scores represent the estimates of values of insensible values. Methods for their determination relate to regression. The most widely used method for their finding is Bartlett's method (weighted method of least squares) [cf. HEBÁK et al., 2007]. We can obtain the factor scores for each of factor during the application of factor analysis. In case of research of regional competitiveness can be compared regions e.g. according to factors rather of economic character, factors concerning the demography, factors describing the quality of life etc. If we count up the all of factor scores, we can obtain integral factor score. This score expresses the total rate of regional competitiveness.

4.2 The usage of cluster analysis for evaluation of competitiveness of regions

The cluster analysis can be used for decrease of count of surveyed objects (in case of evaluation of competitiveness are directly thought the regions). This method enables to seek regions with similar level of competitiveness thereby the count of surveyed objects is reduced. This fact causes the simplification of whole analysis. Moreover we can explore the main characters of clustered region. This way can be found out the main factors of their competitiveness.

The cluster analysis also enables to classify the input set of objects into several relatively homogeneous clusters. The structure of data set is disclosed in this way, as was indicated above, the particular objects can be classified. After that it is important to find the convenient interpretation as the characteristics of given category. We can successfully apply this method in connection with above mentioned cluster analysis. The cluster analysis namely finds the rate of similarity and dissimilarity of surveyed objects. The similarity of objects is expressed as distance among magnitudes. The following distances are most frequently used:

- the Euclidean distance,
- the squared Euclidean distance,
- the Hamming distance (also the Manhattan distance or the City-block distance),
- the Minkowski distance,
- the Chebyshev distance,
- the Mahalanobis distance.

The other rates of distance except the Mahalanobis distance are dependent on used measuring units. The strong correlation of input data is problematic as well. KUBANOVÁ [2003] states that the existence of strong correlation among input variables influences the size of distances among objects. This correlation influences the result of clustering. The usage of factor scores as input data then enables to solve such a problem.

The basic techniques of clustering can be divided according to system of using classification in following:

- hierarchical clustering approaches,
- non-hierarchical clustering approaches.

We can meet various clustering methods in case of hierarchical approaches. The most widely used methods are: single-linkage clustering, complete-linkage clustering, average linkage clustering, unweighted pair-group centroid, weighted pair-group centroid, Ward's method. We can mention the method of typical points and k-means method from non-hierarchical method.

STANKOVIČOVÁ and VOJTKOVÁ [2007] state that the most common method currently used is Ward's method. HEBÁK et al. [2007], adds that the Ward's method tends to remove small clusters, and to create roughly identical sizes, which is often an acceptable features.

The graphical representation of the hierarchical structure of groups is found in the graph called dendrogram. One dimension of this graph is the examined objects. The second dimension is the distance between objects and clusters.

When applying cluster analysis it is necessary to determine the number of groups of clusters. Dendrogram shows a large variety of groups that are more or less homogeneous. According to STANKOVIČOVÁ and VOJTKOVÁ [2007], the question remains, where to "serve" the tree to obtain the optimal number of clusters. Frequently used heuristic approach determines the number of clusters on the basis of subjective opinion of resolver. In general we can search the biggest "gap" among the separately branches of the tree of clusters (called dendrogram).

The application of cluster analysis, as was mentioned above, enables to create typology of regions. The creating typology of regions becomes a very significant "spring board" for more successful application of regional policy [cf. DOČKAL, 2004]. The typology of regions is used e.g. by executive and decisive organs of states or the European Union. These typologies target the denotation of so called "troubled" regions which require because of their underdevelopment the direct subvention. It is possible to create the typology according to significant factors of competitiveness of regions. E.g. MARTIN [2003] created the typology of regions that describes developing regions and predicted main factors which stand behind the successfully development of these regions. Regions are divided according to this typology into:

- regions as site of export specialization,
- regions as source of increasing returns,
- regions as hub of knowledge.

The usage of single statistic analysis is not sufficient for creating analogical typology of competitiveness. On the basis of multivariate statistical methods is in this paper suggested the methodics of creation of typology. In terms of statistical theory appears as most suitable instrument the usage of combination of factor and cluster analysis. The combination of these methods is necessary in every time. At the moment the combination of factor and cluster analysis appears as suitable instrument.

The described methodics, in this article, for determination of typology of regions can be schematically represented in following Fig. 1.

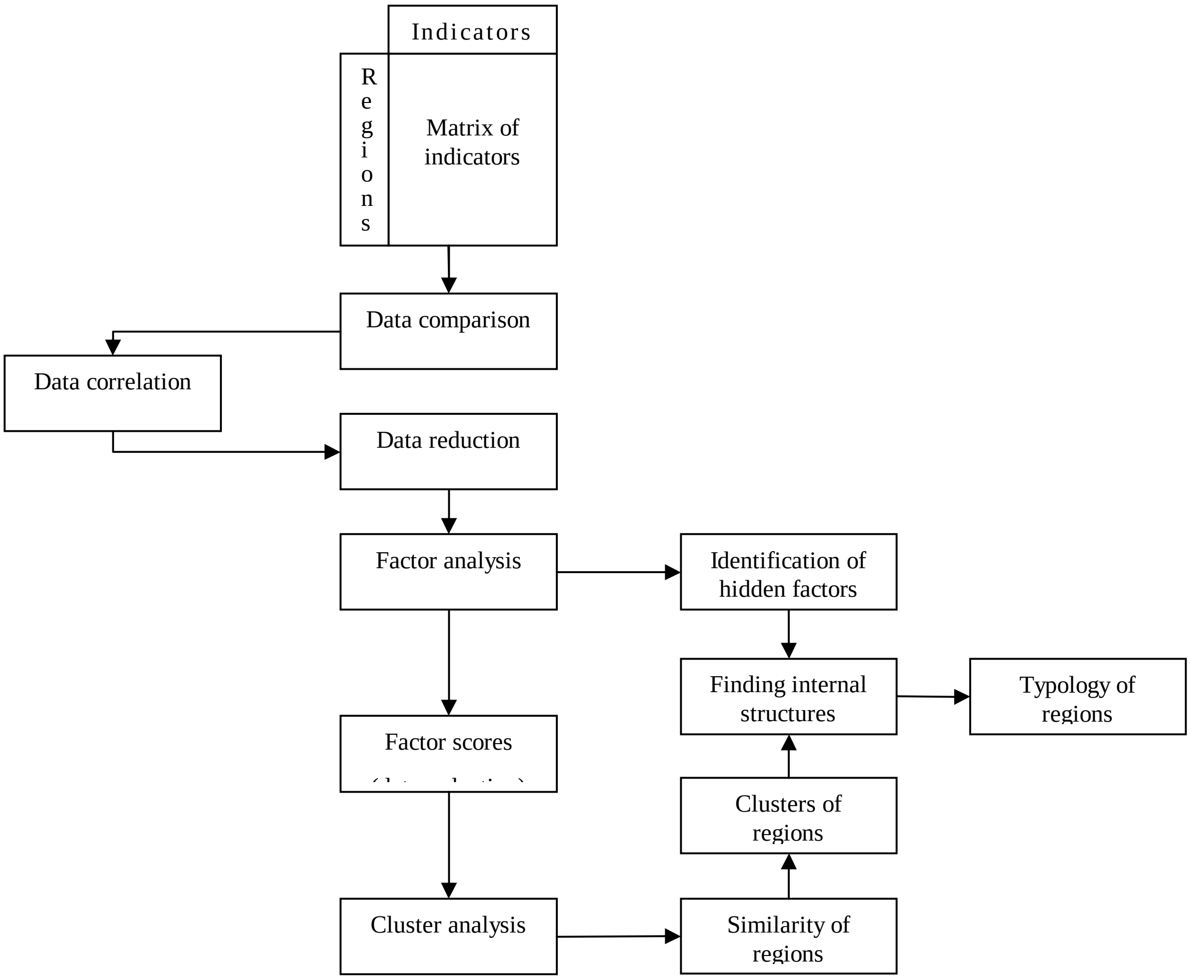

Fig. 3: The scheme of used methodics

Source: own construction

The input data are tested by means of correlation analysis. As a result of this fact the input variables were reduced. Further was applied factor and cluster analysis. The factor analysis helps, as was stated above, to find so called "hidden" factors which have the fundamental impact on competitiveness of region. The other output of factor analysis, so called "factor score", serves as the basis for cluster analysis which enables to identify similar regions. The created clusters of regions are further searched in connection with found hidden factors. This way is revealed the internal structure of regions. The creation of typology of regions, which can be used in terms of regional analysis, is the output of the whole analysis.

5. Conclusion

The question of competitiveness is still popular and actual concept of current regional policy at the level of the whole European Union. The quantity of conferences and specialized publications focused on regional competitiveness is evidence of this statement. Beside the question of holding this notion and its right definition is here solving the question of increasing regional competitiveness. It is not however possible without knowledge its determinants.

Because the regions are considerably varied entities, recognition of main factors of competitiveness is not a simple task. Therefore is in this article proposed a methodics which uses of factor and cluster analysis. On one hand the application of stated methods makes possible to set and generalize basic determinants of regional competitiveness. On the other side it enables to create typology of similar regions. The knowledge of these two outputs of proposed methodics can be used in case of regional policy. That way the efficiency of instruments of regional policy can be increasing.

References:

[1] BERMANGROUP *Regionální hospodářská konkurenceschopnost – příručka pro představitele veřejné správy*. [on-line]. Praha: BermanGroup, 2006. 46 p. [cited 20. 9. 2010]. Available from: <http://www.jic.cz/files/clanky_clanky/regionalni_hospodarska_konkurenceschopnost_118.pdf>.

[2] BOROZAN, D. Regional Competitiveness: Some Conceptual Issues and Policy Implications. In *Interdisciplinary Management Research*, 2008, Vol. 4, p. 50-63.

[3] BUDD, L., HIRMIS, A. K. Conceptual Framework for Regional Competitiveness. In *Regional Studies*, 2004, Vol. 38, No. 9, p. 1015 – 1028. ISSN 0034-3404.

[4] CELLINI, R., SOCI, A. Pop Competitiveness. In *Banca Nazionale del Lavoro Quarterly Review.* 2002. Vol. 55, No. 220, p. 71-101.

[5] DOČKAL, V. Ústřední pojmy regionální politiky EU. In *Středoevropské politické studie*. [on-line]. Brno: Masarykova univerzita., 2004. [cited 25. 8. 2010]. ISSN 1212-7817. <Available from: http://www.cepsr.com/clanek.php?ID=192>.

[6] GARDINER, B., MARTIN, R., TYLER, P. Competitiveness, Productivity and Economic Growth across the European Regions. In *Regional Studies*, 2004. Vol. 38, No. 9, p. 1045 – 1067. ISSN 0034-3404.

[7] HEBÁK, P. a kol. *Vícerozměrné statistické metody* [3]. Praha: INFORMATORIUM, 2007. 271 p. ISBN 978-80-7333-001-9.

[8] KITSON, M., MARTIN, R., TYLER, P. Regional Competitiveness: An Elusive yet Key Concept? In *Regional Studies*, 2004, Vol. 38, No. 9, p. 991 – 999. ISSN 0034-3404.

[9] KRUGMAN, P. Competitiveness: A Dangerous Obsession. In *Foreign Affairs*, 1994, Vol. 73, No. 2, p. 28 – 44. ISSN 0015-7120.

[10] KUBANOVÁ, J. *Statistické metody pro ekonomickou a technickou praxi*. Bratislava: STATIS, 2003. 247 p. ISBN 80-85659-31-X.

[11] MARTIN, R., L. *A Study on the Factors of Regional Competitiveness*. Cambridge: Cambridge Econometrics a Rotterdam: ECORYS-NEI, 2003.

[12] PACÁKOVÁ, V. et al. *Štatistické metódy v ekonómii so zameraním na sociálne analýzy*. Bratislava: Vydavateľstvo EKONÓM, 2009. 208 p. ISBN 978-80-225-2704-0.

[13] PORTER, J. M. *The Competitive Advantage of Nations*. New York: Free Press, 1990. 855 p. ISBN 0-684-84147-9.

[14] PORTER, J. M. *On Competition*. Boston: Harvard Business School Press, 1998. 485 p. ISBN 0-87584-795-1.

[15] PORTER, M. E. Location, competition and economic development: local clusters in a global economy. In *Economic Development Quarterly*, 2000. Vol. 14, No 1, p. 15 – 34. ISSN 0891-2424

[16] SKOKAN, K. *Konkurenceschopnost, inovace a klastry v regionálním rozvoji*. Ostrava: Repronis, 2004. 159 p. ISBN 80-7329-059-6.

[17] STANKOVIČOVÁ, I. VOJTKOVÁ, M. *Viacrozmerné štatistické metódy s aplikáciami*. Bratislava: Iura Edition, 2007. 261 p. ISBN 978-80-8078-152-1.

[18] VITURKA, M. Konkurenceschopnost regionů a možnosti jejího hodnocení. In *Politická ekonomie*, 2007, No. 5, p. 637 – 657. ISSN 0032-3233.

[19] VRTĚNOVÁ, L. *Krátkodobé a dlouhodobé faktory konkurenceschopnosti regionů*. Dissertation. Pardubice: University of Pardubice, 2009. 144 p.

[20] VRTĚNOVÁ, L. *Short-term and Long-term Factors of Competitiveness of Regions*. Pardubice: University of Pardubice, 2009. 32 p. ISBN 978-80-7395-202-0.

Contact address:

Ing. Lucie Sobotková, Ph.D.
University of Pardubice
Faculty of Economics and Administration
Institute of Economics
Studentská 84
532 10 Pardubice
Email: lucie.sobotkova@upce.cz
Phone: +420 466 036 161