Multidimensional Modeling of Cohesion Regions

Pavel Petr, Jiří Křupka, and Romana Provazníková

***Abstract*—**The submitted paper focuses on the utilization of the concept of aggregated indicators for expressing the competitiveness of regions. The statistical methods, factor and cluster analysis, have been used for the modeling. In this modeling process real data from the Eurostat statistics database are used. These data contain diverse information about the second level of nomenclature of territorial units for statistics and comparable regions. The result of the factor analysis is the development of a set of new composite indicators that markedly contribute to the increase of competitiveness of a given region. The results of the cluster analysis algorithms (K-means and Self-Organizing maps) are clusters of the selected regions constructed by clustering composite indicators values.

***Keywords*—**Cluster analysis, composite indicators, factor analysis, NUTS 2 region.

I. Introduction

IN the recent years to the front come the question of a complex evaluation of economic, social and environmental activities of states, regions and cities. The most common, the simplest and the most frequently used indicator for the performance of the economy is the Gross Domestic Product (GDP). Various indicators of sustainable development (e.g. Index of Sustainable Economic Welfare), Aggregate or Composite Indicators (CIs), Ecological Footprint, Human Development Index, Happiness Indicators or Life Fulfillment Indicators are cited next to the GDP indicator which expresses the economic activity of a society. A number of indicators used for the measurement and qualification of progress (in various areas) on the national level is becoming to be used on the sub national-regional and municipal level.

Construction [1] of composite (complex, aggregated) indicators for the cohesion regions on Nomenclature of Territorial Units for Statistics (NUTS) 2 level by means of factor analysis (FA) method is proposed in this article. Data reported on the level of these regions provide us with more detailed information and allow to execute the analysis of the economic and social development on a lower than the national level.

Approaches presented so far [2], [3] differentiate especially by the chosen approach to the number of indicators used. Some authors consistently use a larger set of indicators, some work toward some synthesis of indicators. In the first mentioned approach we generally speak about key indicators [4], [5]. These indicators can be used as umbrella indicators since they provide an umbrella to the selected set of indicators. The goal of these indicators is to provide brief and clear information on selected key factors. General public is aware of well known structural indicators that are used by all European Union (EU) countries and presented by Eurostat [6], [7].

To the set of key indicators we can to the contrary use CIs [8], [9], [10], [11], the meaning of which is to provide the synthesis by combining indicators and statistical data concerning various areas. CI is formed when individual indicators are compiled into a single index, on the basis of an underlying model of the multi-dimensional concept that is being measured. A CI measures multi-dimensional concepts – e.g. competitiveness, e-trade or environmental quality, or description of the level of economic well-being, happiness respectively which cannot be captured by a single indicator. Ideally, a CI should be based on a theoretical framework, which allows individual indicators (variables) to be selected, combined and weighted in a manner which reflects the dimensions or structure of the phenomena being measured.

CIs [1] are possible but not the optimal solution, they allow to transform massive amounts of available information into commonly shared findings that can support collective decisions and evaluations, many users (media, politicians and other) favorites composite indicators and linked "rankings" in liking for their "easy handling". On the other side CIs can send misleading policy messages if they are poorly constructed or misinterpreted. Their "big picture" results may invite users (especially policy makers) to draw simplistic analytical or policy conclusions. Instead, CIs must be seen as a starting point for initiating discussion and attracting public interest [8].

The objective of this paper is:

To develop the model on the basis of FA and to interpret the meaning of the set of created "new CIs" from the region competitiveness viewpoint

To analyze NUTS 2 regions in selected countries – Czech Republic (CR), Slovak Republic (SR) and Baltic states (BSs) by means of the cluster analysis (CA) based on created CIs.

II. Problem Formulation

The input data for this analysis are indicators obtained from the Eurostat database [7] for 317 NUTS 2 regions. These are: 271 NUTS 2 regions of 2 EU member states EU-27, 30

This work was supported by the scientific research project of the National Science Foundation of CR under Grant No. 402/08/0849 with title Model of Sustainable Regional Development Management.

statistics level 2 regions of candidate countries (Croatia, the former Yugoslav Republic of Macedonia) and Turkey, and 16 statistics level 2 regions of the European Free Trade Association - Iceland, Liechtenstein, Norway and Switzerland [5].

Eurostat database [7] provides ten main categories of data on the NUTS 2 level regions by the following way: Regional agriculture statistics, Regional demographic statistics, Regional economic accounts – European System of Accounts (ESA95), Regional education statistics, Regional science and technology statistics, Regional health statistics, Regional tourism statistics, Regional transport statistics, Regional labor market statistics, Regional information society statistics. Each of the above mentioned basic categories is characterized by a more detailed set of indicators in a given time series. From the time series the interval from year 2005 to 2008 has been chosen.

The indicators have been constructed on the basis of Eurostat database. We can devote [7] Groups of indicators $G = \{g_1, g_2, \ldots, g_{10}\}$ where g_1 are Regional agriculture statistics, g_2 are Regional demographic statistics, ..., g_{10} are Regional information society statistics. There were 52 time series tables, in other words 52 indicators [5]. There are n indicators (indexes, attributes) a_n (n = 1, 2, …, 52) which are defined for the selected years 2005, 2006, 2007 and 2008. Indicators that had the correlation coefficient equal to 1 were eliminated, for example: Regional GDP (in 10^6 EUR), Regional GDP (in 10^6 purchasing power standards (PPS) and Regional GDP (in PPS per inhabitant in % of EU-27 average) and similar and thus the number of indicators was reduced to 43. The adjusted number of indicators was 43 for 317 NUTS 2 regions, which led to the construction of the matrix $\mathbf{D}(317\times43)$. In the following step indicators that had less than 50% of the value of the given attribute have been eliminated. The resulting number of attributes entering the analysis is 23 (see Table 1). Based on the selected attributes matrix $\mathbf{E}(317\times23)$ has been constructed [5].

The indicators listed in Table 1 form input (manifest) variables of FA and represent five basic groups of indicators which to a larger or lesser extent influence the competitiveness of a given region. These are: Economic performance indicators and standard of living indicators, labor market indicators, regional innovation indicators performance, health indicators – death rate, and tourism [5]. Economic meaning of the selected indicators is in detail described below.

From the group of indicators that demonstrate the economic performance of the region the Regional GDP and Real growth rate of regional GDP indicators have been selected. Regional GDP expressed in PPS per inhabitant eliminates differences in price levels between countries. Calculations on a per inhabitant basis allow for the comparison of economies and regions significantly different in absolute size. GDP is an indicator of the output of a country or a region. GDP per inhabitant in PPS is the key variable for determining the eligibility of NUTS 2 regions in the framework of the EU's

Table 1 Description of selected indicators

Indicators		Description of indicators
x_1	T26_P_Ind	Disposable income of private households – PPS based on final consumption per inhabitant
x_2	T36_P_Ind	Primary income of private households – PPS per inhabitant
x_3	T37_P_Ind	Real growth rate of regional GDP at market prices – percentage change on previous year
x_4	T40_P_Ind	Number of patents applications to the EPO by priority year – per 10^6 of inhabitants
x_5	T30_P_Ind	Number of bed-places in collective tourist accommodation establishments
x_6	T32_P_Ind	Nights spent by total in collective tourist accommodation establishments
x_7	T41_P_Ind	Number of patents applications to the EPO by priority year in the field of high-technology – per 10^6 of inhabitants
x_8	T31_P_Ind	Number of bed-places in hotels and similar establishments
x_9	T60_P_Ind	Death due to all kind of accident – standardized death rate per 10^5 inhabitants, 3 year average
x_{10}	T59_P_Ind	Death due to ischemic heart diseases – standardized death rate per 10^5 inhabitants, 3 year average
x_{11}	T57_P_Ind	All causes of death – standardized death rate per 10^5 inhabitants, 3 year average
x_{12}	T58_P_Ind	Death due to cancer – standardized death rate per 10^5 inhabitants, 3 year average
x_{13}	T61_P_Ind	Death due to transport accidents – standardized death rate per 10^5 inhabitants, 3 year average
x_{14}	T7_P_Ind	Regional employment rate of the age group 15-64 (total)
x_{15}	T54_P_Ind	Regional employment rate of the age 55-64 (total)
x_{16}	T53_P_Ind	Share of long-term unemployment (12 month and more) – percentage of total
x_{17}	T10_P_Ind	Total regional unemployment rate (in %)
x_{18}	T38_P_Ind	Human resources in science and technology – percentage of economic active population
x_{19}	T39_P_Ind	Employment in high-tech sectors – percentage of total employment
x_{20}	T24_P_Ind	Population density – inhabitants per km^2
x_{21}	T77_P_Ind	Air transport of passengers – 10^3 passengers
x_{22}	T5_P_Ind	Regional GDP – PPS per inhabitant
x_{23}	T33_P_Ind	Nights spent by total in hotels and similar establishments

structural policy. The real growth rate of regional GDP at market prices – percentage change on previous year is only recently reported by Eurostat on NUTS 2 level. It should be noted that this indicator is not part of the ESA95 data transmission program and that a harmonized methodology agreed by Member States is not yet available.

Indicators that express, to a certain extent, a household's standard of living are household income before taxation – Primary income (PPS per inhabitant) and income after current taxes on income and wealth paid and current transfer payments received – Disposable income (PPS based on final consumption per inhabitant). In market economies with state redistribution mechanisms, a distinction is made between two stages of income distribution. The primary distribution of income shows the income of private households generated directly from market transactions, in particular the purchase and sale of factors of production. Primary income is the point of departure for the secondary distribution of income, which means the state redistribution mechanism. The density of population expressed as population per km^2 directly correlates with the indicators of economic performance and households incomes. This indicator already includes in itself other indicators reported in the group of regional demographic indicators – both the number of inhabitants indicator and the region total. Generally, regions with the highest population density (exceeding 10^3 inhabitants per km^2) are regions with the capital cities and these regions also generate the highest GDP and achieve above average incomes. By comparison, the EU has a population density of 114 inhabitants per km^2 [6].

The indicators that characterize the labor market are, next to GDP per inhabitant, the most watched macroeconomic indicators on the regional level. They are dependent to a large extent on the GDP development, however they are also influenced by the overall labor productivity, total employment, age and education composition of the inhabitants, and status of transportation infrastructure. In this analysis the following indicators are utilized: Regional unemployment rate, Share of long-term unemployment, Regional employment rate of the age group 15-64 and of the age group 55-64 as a percentage of total employment. All indicators are based on the EU Labor Force Survey. Indicator Regional unemployment rate represents unemployed persons as a percentage of the economically active population (i.e. labor force or sum of employed and unemployed). Unemployed persons comprise persons aged 15-74 who were (all three conditions must be fulfilled simultaneously: without work during the reference week; currently available for work; actively seeking work or who had found a job to start within a period of at most three months). The employed persons are those aged 15-64, who during the reference week did any work for pay, profit or family gain for at least one hour, or were not at work but had a job or business from which they were temporarily absent. The share of long-term unemployment is the share of unemployed persons since twelve months or more in the total number of unemployed persons, expressed as a percentage represents the worst form of unemployment. Regional employment rate of the age group 15-64 and group 55-64 years represents employed persons aged 15-64 (respectively employed persons aged 55-64) as % of the population of the same age group.

The innovation potential of a given region is represented by Human resource quality indicators (for instance Human resources in science and technology (HRST), Employment in high-tech sectors as a percentage of total employment). Current emphasis on innovation as a source of industrial competitiveness has raised awareness of patents. Number of patents applications submitted to the European Patent Office (EPO) and number of submitted patents applications in the field of high-technology patents per million inhabitants of a region are other indicators. Patents reflect a country's inventive activity. Patents also show the country's capacity to exploit knowledge and translate it into potential economic gains. In this context, indicators based on patent statistics are widely used to assess the inventive performance of countries. The number of patent applications or the number of awarded patents both recalculated to 10^6 inhabitants is a relative indicator eliminating the size of a country factor. Indicator Human resources in science and technology gives the percentage of the total labor force in the age group 15-74, that is classified as HRST, i.e. having either successfully completed an education at the third level or is employed in an occupation where such an education is normally required. Patent applications to the EPO by priority year expressed as number of applications per 10^6 of inhabitants. Patent applications are counted according to the year in which they were filed at the EPO and are broken down according to the International Patent Classification (IPC). They are also broken down according to the inventor's place of residence, using fractional counting if multiple inventors or IPC classes are provided to avoid double counting.

From the area of health and health care indicators of death rate have been chosen (in total for all death causes, as well as according to individual causes). Other indicators characterizing health care quality – such as number of physicians, dentists and number of beds have not been taken into consideration due to incomplete data on the relevant NUTS 2 level. Causes of Death data refer to the underlying cause which – according to the World Health Organization – is "the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury" [7]. The indicator is calculated as Standardized death rate per 10^5 inhabitants for 3 years average. Further there are used indicators where the cause of death are civilization diseases such as cancer – Death due to cancer refer to all death caused by a malignant neoplasm a ischemic heart diseases – Death due to ischemic heart diseases (refer to all death caused by reduced blood supply to the heart. Most of these deaths are due to "heart attack"). Further death due to accidents as such – Death due to all kind of accident (transport, drowning, fire, etc.) and accidents due to transport – Death due to transport accidents (refer to all kind of transport – road (car, pedestrian, cyclist), water, air etc.).

Tourism is an important and fast-evolving economic factor in the EU, occupying large numbers of small and medium-sized businesses. Its contribution to growth and employment varies widely across the EU regions. Particularly in rural regions, usually peripheral to the economic centers of their countries, tourism is often one of the main sources of income for the population and a prominent factor in creating and securing an adequate level of employment. Tourism encompasses not only private travel but also business travel. They both make significant demands on transport, accommodation and restaurant services. This area includes data both on accommodation capacity and its utilization. Accommodation capacity is measured by the Number of bed-places in hotels and similar establishments and by Number of bed-places in collective tourist accommodation establishments (on campsites). The central indicator for accommodation services is the number of overnight stays in establishments – Nights spent by total (residents and non-residents) in hotels and similar establishments and Nights spent by total (residents and non-residents) in collective tourist accommodation establishments. Hotels and similar establishments are typified as being arranged in rooms, in number exceeding a specified minimum; as coming under a common management; as providing certain services including room service, daily bed-making and cleaning of sanitary facilities; as grouped in classes and categories according to the facilities and services provided; and as not falling in the category of specialized establishments. A collective tourist accommodation establishment provides overnight lodging for the traveler in a room or some other unit, but the number of places it provides must be greater than a specified minimum for groups of persons exceeding a single family unit and all the places in the establishment must come under a common commercial-type management, even if it is non-profit-making. Indicator that is closely related to the above stated group of indicators is the indicator Air transport of passengers (in 10^3 passengers). The air transport regional data have been calculated using data collected at the airport level in the frame of Commission Regulation No.1358/2003 [7]. They are aggregated at regional level (NUTS 1 and NUTS 2) and also at national level (NUTS 0), excluding double counting within each region. The indicator demonstrates also the state-of-the art of the given region, inhabitants' purchasing power respectively.

III. Model Design Proposal

Data matrix has been constructed for the model – matrix **D** for 317 NUTS 2 regions. Indicators (parameters) have been selected for years 2005 to 2008. For the creation of the model it was essential to create demographic indicators D = $\{d_1, d_2, \ldots, d_5\}$ where d_1 is Name of country/state; d_2 is Code EU; d_3 Name of NUTS 2; d_4 is Country Code; d_5 is NUTS 2 Code. As it was already mentioned, the new matrix **E** for modeling has the dimension **E**(317×23). Basic scheme of modeling is in Fig. 1.

The 23 selected indicators are described in Table 1. Based on the values $v_{n,\text{year}}$ of n-th indicator a_n (n = 1, 2, …, 23) the rates of growth has been calculated growth rate $g_{n,year}$ in the year by the following way:

$$g_{n,year} = \frac{v_{n,year} - v_{n,year-1}}{v_{n,year-1}} \quad (1)$$

where *year* is 2008, 2007, and 2006.

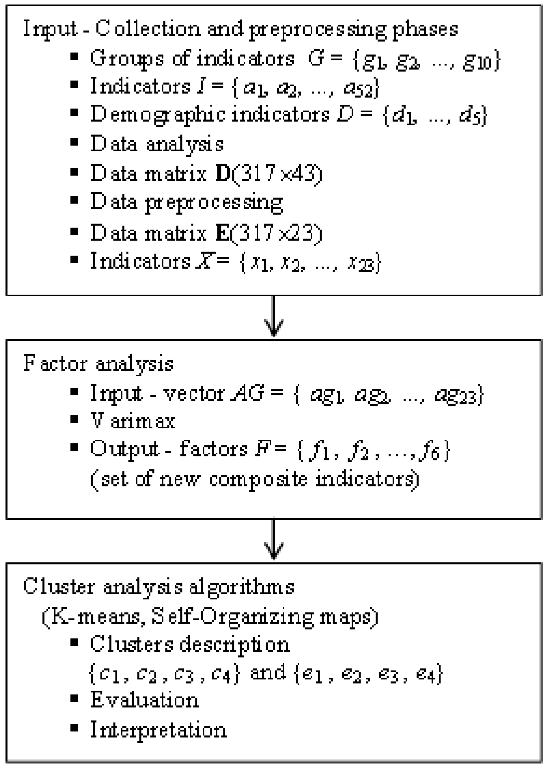


Fig. 1 Basic scheme of modeling

From the growth rate $g_{n,year}$ are set the average growth rate values ag_n of n-th indicator a_n by the following way:

$$ag_n = \frac{\Sigma g_{n,year}}{k} \quad (2)$$

where k = 3 represents the number of *years*.

In this way the input vector of the average growth rate values of the indicators (attributes) for the period 2005 to 2008 has been defined $AG = \{ag_1, ag_2, \ldots, ag_{23}\}$ for algorithm FA.

In this research, to find latent factors, original values for 23 indicators were used as variables to perform FA [12], [13], [14], [15], [16], [17]. We can regard these indicators to be new latent variables. The meaning we assign to these factors is derived from the set of factors which they include. Varimax rotation was used to obtain six factors $F = \{f_1, f_2, \ldots, f_6\}$, which together calculate for 84.9% of the total variance. Factor f_1 named "macroeconomic level of the region", and consists of nine indicators; factor f_2 named "quality of life (mortality)", and consists of five indicators; factor f_3 named "utilization of human potential", and consists of four indicators; factor f_4 named "state-of-the-art of the region" (the quality of human and innovation potential), and consists of two indicators; factor f_5 named “urbanization”, and consists of two indicators; factor f_6 named “tourism”, and consists of a indicator. In the case of the individual factors we can speak of a set of new CIs.

The factor f_1 reflects in the best way the variability of the following indicators (see Table 1 - T26_P_Ind, T36_P_Ind, T37_P_Ind, T40_P_Ind, T30_P_Ind, T32_P_Ind, T41_P_Ind, T31_P_Ind and T5_P_Ind): regional GDP and real growth rate of regional GDP, both indicators characterize, households income – disposable income of households and primary income of households, both indicators of the number of patent applications – number of patent applications and number of high-tech patent applications to the EPO, indicators related to tourism – number of bed-places in collective tourist accommodation establishments and nights spend in these establishments, however also the number of bed-places in hotels. The stated indicators encompass not only private travel but also business travel, which gives tourism an economic dimension. Both sectors mean the same demands on the related service as and represent an important source of incomes for the region’s inhabitants. Expenditure by tourists during their stay at their destination correlates closely with the number of overnight stays. This applies especially to Europe’s island states and island regions, to many coastal regions, particularly in southern Europe, and to the whole Alpine region. Further here enter also indicators related to the level of regional GDP and the GDP rate of increase. The factor expresses both the macroeconomic and the innovation potential of the given region, as well as the standard of living.

Death rate due to various causes indicators enter into factor f_2 (T60_P_Ind, T59_P_Ind, T57_P_Ind, T58_P_Ind and T61_P_Ind) with the same level of the factor burden. Despite the fact that this factor represents only the inhabitant death rate and it is a socio-economic indicator by its character, it can be, to a certain level, related to the health care quality and to inhabitants’ lifestyle (especially with death due to civilization diseases – cancer and heart and cardiovascular system diseases).

Factor f_3 utilization of human potential explains the variability of indicators (T7_P_Ind, T54_P_Ind, T53_P_Ind and T10_P_Ind) characterizing both regional level of unemployment (including long term unemployment) and the unemployment of total work force and the employment of endangered group of inhabitants in the age group 55-64 years. High employment rate and low unemployment rate are again the characteristics of economically strong regions. In the framework of EU some regions still record a double-digit unemployment rate. These are mainly located in the south of Spain, the south of Italy and the eastern regions of Germany. Some regions in SR, Poland and Hungary also recorded unemployment rates above 10% in the observed period.

Variables that characterize the level of employment in research and development and in the high-tech sector do enter into factor f_4. These indicators (T38_P_Ind and T39_P_Ind) indicators jointly with the indicators of achieved education illustrate the quality of human potential that represents an important element in the innovation potential and in the rise of region competitiveness. Also this factor characterizes urbanized regions, mainly regions with capital cities.

The fifth factor f_5 explains the variability of indicators (T24_P_Ind and T77_P_Ind) – population density and the number of passengers in air transportation. This factor is to a certain level specific. It can be expected that the higher the population density, the larger the volume of passengers transported by this means of transport in the region. It is interesting that as the last factor in the model is detached factor f_6 that includes only one indicator (T33_P_Ind) that again deals with tourism (both private or business) – number of nights spent by total (residents and non-residents) in hotels and similar establishments. The importance of incomes from tourism has been already mentioned with the first factor. The explanation why only this factor has been detached must be further studied and it is beyond the extent of this article.

The acquired CIs that were the result of the FA method application [5] have been used by us for a more detailed research into selected regions in CR, SR, and BSs. There are 8 NUTS 2 regions in CR, 4 NUTS 2 regions in SR, BSs are each state one NUTS 2 region – that is together 15 NUTS regions.

Table 2a and 2b illustrate the obtained FA results for the selected regions. The Number represents the NUTS 2 region where 1 is Southeast, 2 is Southwest, 3 is Moravia Silesia, 4 is Prague, 5 is Northeast, 6 is Northwest, 7 is Central Bohemia, 8 is Central Moravia, 9 is Estonia, 10 is Latvia, 11 is Lithuania, 12 is Bratislava region, 13 is Central Slovakia, 14 is East Slovakia, and 15 is West Slovakia. Individual columns show the value of the resulting 6 factors that represent new CIs. If we rank the resulting values inside the individual factors in the descending order (from best to worst) then three groups of regions are appearing – one group of regions comprised of the BSs (Estonia, Latvia and Lithuania) and four CR regions (the capital city Prague, Southeast, Moravia Silesia that together cover the largest Moravia cities – Brno, Ostrava and generates jointly with the capital city Prague the biggest GDP and region Southwest in which Karlovy Vary town is located. This town is major attraction for tourists and generates a lot of revenue from tourism for this region), the group of the last remaining four CR regions – Northeast, Central Bohemia, Central Moravia and Southwest. The last group of regions comprised of the Slovak regions.

With the first factor f_1 “Macro-economic” level (represents economic performance, innovation potential, macroeconomic output and the standard of living of the region) we can see the “worst” group of regions, that reaches negative values, represent all SR regions, then comes the group or regions comprised of four CR regions – Northeast, Central Bohemia, Central Moravia, Southwest – values close to 0.29) and the “best” group (with values around 0.45) includes the BSs and the remaining CR regions (including the capital city Prague, Northwest, Southeast and Moravia Silesia).

Just the opposite is the situation in case of factor f_2 “Quality of life", which includes all causes of death. The group o the best regions (BSs, and CR regions Northwest, Prague and Moravia Silesia) reaches the lowest values of this indicator, the highest values demonstrate the SR regions and in between is the group of the four CR regions (Southwest, Central Moravia, Northeast and Central Bohemia).

Table 2a Factor analysis results for selected regions NUTS 2

Number of regions	Macro economic level	Quality of life	Utilisation of human potential
1	0.45437	0.45606	0.05959
2	0.29769	0.54702	-0.00712
3	0.45417	0.45615	0.05962
4	0.45460	0.45604	0.05949
5	0.29658	0.54739	-0.00673
6	0.45484	0.45589	0.05958
7	0.29680	0.54740	-0.00721
8	0.29697	0.54729	-0.00707
9	0.45442	0.45605	0.05964
10	0.45419	0.45607	0.05997
11	0.45476	0.45586	0.05952
12	-0.07662	0.63961	0.16818
13	-0.52881	0.83591	0.09623
14	-0.35940	0.74201	0.16221
15	-0.23463	0.73103	0.10134

Table 2b Factor analysis results for selected regions NUTS 2

Number of regions	Innovation potential	Urbanisation	Tourism
1	0.18752	-0.16758	0.07701
2	0.50005	0.79341	0.29510
3	0.18780	-0.16746	0.07706
4	0.18768	-0.16742	0.07708
5	0.49933	0.79326	0.29437
6	0.18764	-0.16775	0.07712
7	0.49934	0.79342	0.29414
8	0.49926	0.79335	0.29445
9	0.18745	-0.16749	0.07711
10	0.18759	-0.16761	0.07723
11	0.18783	-0.16755	0.07743
12	0.23426	-0.75151	-5.48102
13	0.42401	0.19919	-5.33888
14	0.11405	-0.76552	-5.56966
15	0.54590	0.20938	-5.26295

More and better demonstration of the FA results for selected factors can be made by using cobweb graph where individual axes represent individual regions.

In Fig. 2 there are shown the values of factors f_1 (represented by the symbol "diamond") and f_2 (represented by the symbol "square"). It is obvious that the worst results in f_1, as well as with f_2 are for Central Slovakia (there is the biggest gap between the values of these two factors). Gradually the gap between the two factors gets smaller (clockwise). For the BSs and the four best CR Regions the values of factor f_2 even overlap with the values of the macroeconomic performance values f_1 in this graph. The most advanced regions from the macroeconomic performance point of view thus show also lower mortality rates than less advanced regions.

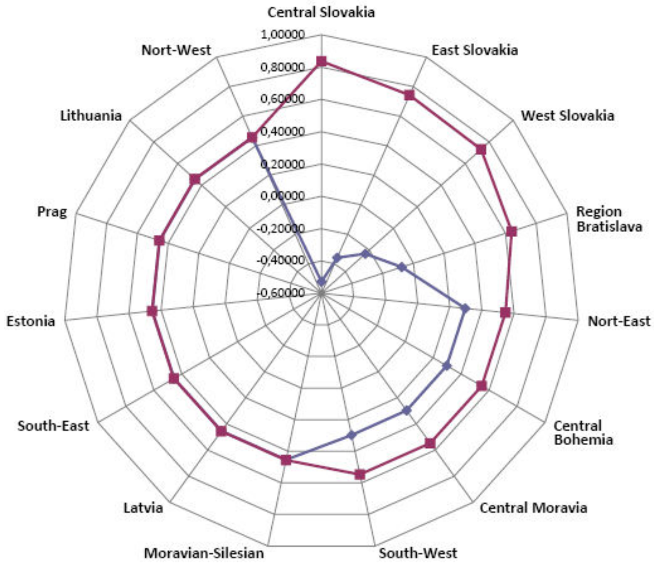


Fig. 2 Cobweb graph of factors f_1 and f_2

With the third factor f_3 "Utilization of human potential" the SR regions reaches the worst values, then come the BSs regions, and the four advanced CR regions and the best values are shown for four CR regions (Southwest, Central Moravia, Northeast, Central Bohemia – negative coefficients work against the value of this factor.

With the fourth and fifth factor f_4 and $f_5$the grouping of regions is different from the previous factors. For factor f_4 named "State-of-the-art of the region", that represents the quality of human potential of the region (level of Human resources employment in Research and Development and in high-tech sectors) we can observe based on the resulting values the following grouping of regions: one group is created by one region East Slovakia with the lowest value 0.114, next the BSs are close to this with their values (0.187) and the four advanced CR regions (Prague, Southeast, Northeast, Moravia Silesia) and also the Bratislava region with the value 0.234 can be grouped here. It is interesting that in the best group with the best factor results (the interval from 0.424 to 0.549) belongs the group of the remaining SR regions and "weaker" CR regions. Paradoxically, the regions including the capital cities do not reach in this factor the highest values, but only the second best values.

In Fig. 3 there is demonstrated mutual dependence of utilization of human capital f_3 factor values (represented by the symbol “diamond”) and employment in high tech f_4 values (represented with symbol “square”). In the group of the best regions the values of both factors move in the same direction, and on the contrary for regions with low utilization of human potential (higher values of factors f_3) the values of the innovation potential f_4 move in the opposite direction (e.g. region East Slovakia). As it has been already stated, it is also interesting that regions including capital city do not reach the highest values in these factors. The scope of this article does not provide enough space to analyze the causes of the stated dependences. The authors of this article shall give more attention to this in their following research works.

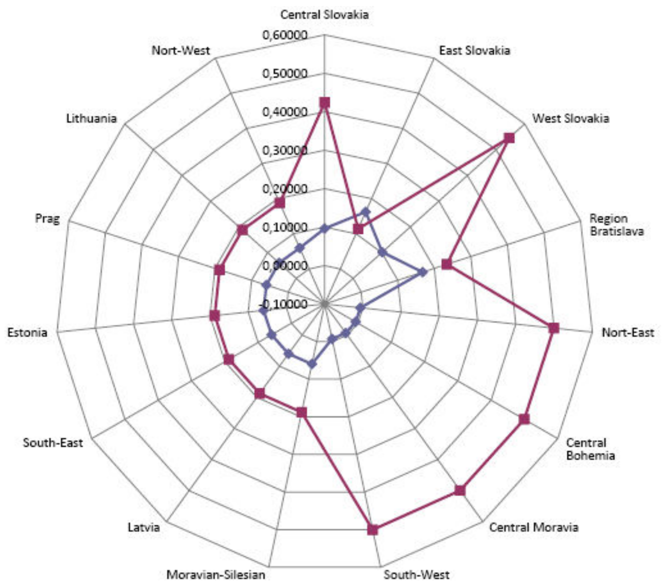


Fig. 3 Cobweb graph of factors f_3 and f_4

For the fifth factor f_5 named “Urbanization” the grouping of regions is different. The most advanced regions work against this factor, thus negative values mean the best result. The group of regions with best values is comprises of Bratislava region and East Slovakia region, second best group of regions are the BSs and the advanced CR regions, the third group comprise of the remaining two SR regions (Central Slovakia, West Slovakia) and the worst results group comprises of the CR regions (Central Moravia, Northeast, Northwest and Central Bohemia).

Regarding the values of the last factor f_6 “Tourism” the distribution of the regions corresponds the distribution into the same groups as is with the first three factors The worst results are reached by the SR regions (negative values), BSs and the four “more advanced” CR regions show medium values of the indicator around 0.77) and the best results are showed by the remaining CR regions.

CA is used for defining clusters of regions based on the value of the individual factors. CA [12], [13], [15], [18], [19], [20] is an exploratory data analysis tool for solving classification problems. The object is sorted into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. The task of clustering is then to divide the set of objects into the disjunctive clusters. The decision making about the object clustering in cluster is realized on the basis of the similarity by application of metric [13], [21]. The basic division of methods is mentioned for instance in [13] and application in [12].

The algorithms K-means and Self-Organizing maps (SOMs) [22], [23] have been used. The number of clusters has been set, based on a previous analysis, at 4. The chosen algorithms allocated based on given factors the given region (object) into a certain segment. The results of the given algorithms are the following: for the first cluster (2 objects witch represents 13.33% of total objects – Central Slovakia and West Slovakia) characterize regions with low macroeconomic level, for the second cluster (7 objects witch represents 46.67% of total objects – Prague, Southeast, Northwest, Moravia Silesia and BSs) is characterized by very similar characteristics of the individual objects and they can be classified as regions with a high level of macroeconomic indicators. The following segment the third cluster (2 objects witch represents 13.33% of total objects – East Slovakia and Bratislava region) are on a very similar macroeconomic level. The last, the forth cluster, is characterized by regions with average macroeconomic indicators and very similar characteristics (4 objects witch represents 26.67% of total objects – Central Moravia, Northeast, Southwest, Central Bohemia).

Means and standard deviation for the clusters c_1, c_2, c_3, and c_4 and factors for K-means algorithm are in Table 3.

Table 3 Cluster analysis results of K-means algorithm

Factors	Clusters			
	c_1 Mean / St.dev.	c_2 Mean / St.dev.	c_3 Mean / St.dev.	c_4 Mean / St.dev.
f_1	-0.382 / 0.208	0.454 / 0.0	-0.218 / 0.2	0.297 / 0.0
f_2	0.783 / 0.074	0.456 / 0.0	0.691 / 0.072	0.547 / 0.0
f_3	0.099 / 0.004	0.06 / 0.0	0.165 / 0.004	-0.007 / 0.0
f_4	0.485 / 0.086	0.188 / 0.0	0.174 / 0.085	0.499 / 0.0
f_5	0.204 / 0.007	-0.168 / 0.0	-0.759 / 0.01	0.793 / 0.0
f_6	-5.301 / 0.054	0.077 / 0.0	-5.525 / 0.063	0.295 / 0.0

On the basis of SOMs means and standard deviation for the clusters e_1 (X = 0, Y = 0), e_2 (X = 0, Y = 2), e_3 (X = 2, Y = 0), and e_4 (X = 2, Y = 2) and factors were computed and they are absolutely same as for K-means algorithm where $e_1 = c_4$, $e_2 = c_2$, $e_3 = c_1$, and $e_4 = c_3$.

IV. Conclusion

From the FA model issues that for the monitored regions, after ranking the values of the CIs, three basic groups of regions are created: SR regions (with the worst values in the majority of factors); the group of the BSs and the four "more advanced" CR regions (capital city Prague, Southeast, Moravia Silesia and Northwest) with the majority of the best values or second best values and the third group of the remaining CR regions (Central Moravia, Northeast, Southeast, Central Bohemia) with the best or the second best value in some factors.

In the following research it shall be useful to concentrate on the joint interconnection of these factors and the inputs into these indicators and to research into, in more detail, their mutual interconnection.

Another grouping of regions has been created based on CA, which produced four clusters. In the analysis there were used two algorithms for the regions segmentation. With using both methods two identical group of regions/objects originated. The results of using of both algorithms show the compactness of the individual regions results. This demonstrates that the similarity of individual regions inside the segments is on a high level and we can talk about their congruity.

Two clusters with the highest count (cluster c_2, e_2 and cluster c_4, e_1) correspond with the group of the best regions and with the second best group of the CR regions. In case of SR the distribution is different – the SR regions have been separated into two clusters (c_1, e_3 and c_3, e_4 cluster). The Bratislava region cluster and the East Slovakia cluster can be explained by the existence of town Košice there. Košice town is the second largest town in SR and it has a major influence on the values of those indicators who were inputs into this model.

REFERENCES

[1] M. Nardo, M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, E. Giovannini, *Handbook on Constructing Composite Indicators Methodology and User Guide*, OECD publishing: Paris, 2008, [Online]. Available: http://browse.oecdbookshop.org/oecd/pdfs/browseit/3008251E.PDF

[2] R. Huggins, H. Izushi, W. Davies, L. Shougui, *World Knowledge Competitiveness Index 2008.* Centre for International Competitiveness, Cardiff School of Management: Cardiff, 2008.

[3] R. Provazníková, J. Křupka, M. Kašparová, "Predictive Modelling on the Regional Level," *TILTAI Bridges*, vol. 39, supplementary issue, pp. 150-158, 2009.

[4] M. Damborský, R. Wokoun, "Location Factors of Small and medium Enterpreneurship under the Economic Condiditions of the Czech Republic," *E+M Economics and Management*, vol. 13, no. 2, pp. 32-43, 2010.

[5] R. Provazníková, P. Petr, J. Křupka, "Analysis of the Regions Based on Multidemensional Methods," *Human Resources – The Main Factor of Regional Development. Journal of Social Sciences*. Klaipeda, Lietuva: Klaipėda University, no. 3, pp. 57-65, 2010.

[6] Eurostat. (2010, July 19). *Eurostat regional yearbook 2009, European Commission: Eurostat Statistical books.* [Online]. Available: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-HA-09-001/EN/KS-HA-09-001-EN.PDF

[7] Eurostat. (2010, June 13). *Regional Statistics*. [Online]. Available: http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/data/main_tables

[8] OECD. (2010, April 05). *The OECD-JRC Handbook on Practices for Developing Composite Indicators*. 2004. [Online]. Available: http://stats.oecd.org/glossary/detail.asp?ID=6278.

[9] OECD. (2010, July 20). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. 2008. [Online]. Available: http://www.oecd.org/dataoecd/37/42/42495745.pdf

[10] R. Čiegis, J. Ramanauskiene, "Sustainable Development Asssesment: The Theoretical and Practical Possibilities," *TILTAI Bridges*, vol. 39, supplementary issue, pp. 86-98, 2009.

[11] C. Lazar, M. Lazar, "Proposal of a Sustainable Development Synthetic Indicator at Local Level," in *Proc. 9th WSEAS Int. Conf. Mathematics and Computers in Business and Economics*, Bucharest, 2008, pp. 74-78.

[12] P. Guidici, *Applied Data Mining: Statistical Methods for Business and Industry*. West Sussex: Wiley, 2003.

[13] J. Han, M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers, 2001.

[14] V. Labudová, M. Vojtková, B. Linda, "Aplikácia viacrozmerných metód pri meraní chudoby," *E+M Economics and Management*, vol. 13, no. 1, pp. 6-22, 2010. (in Slovak)

[15] O. Maimond, L. Rokach, *The Data Mining and Knowledge Discovery Handbook*, New York: Springer, 2005.

[16] D. A. Freedman, *Statistical Models: Theory and Practice*. New York, NY: Cambridge University Press, 2005.

[17] *Electronic Statistics Textbook*, StatSoft, Inc., Tulsa, OK, 2010. [Online]. Available: http://www.statsoft.com/textbook.

[18] N. K. Zivadinovic, K. Dumicic, A. C. Casni, "Multivariate Analysis of Structural Economic Indicators for Croatia and EU 27," in *Proc. 2nd WSEAS Int. Conf. Multivariate Analysis and its Application in Science and Engineering*, Istanbul, 2009, pp. 134-139.

[19] H. Fu, "Cluster Analysis and Association Analysis for the Same Data," in *Proc. 7th WSEAS Int. Conf. Artificial Intelligence, Knowledge Engineering and Data Bases*, Cambridge, 2008, pp. 576-581.

[20] M. Kasparova, J. Krupka, J. Chylkova, "Heavy Metals Contamination Analysis in Selected Czech Localities by Cluster Analysis," *Recent Advances in Environment, Ecosystems and Development*, WSEAS Press: Athens New York, 2009, pp. 240-581.

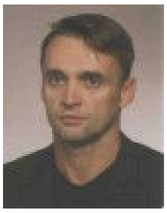
[21] *Clementine® 7.0 User's Guide*, Chicago, IL: SPSS, Inc., 2002.

[22] T. Kohonen, *Self-Organizing Maps*. Berlin Heidelberg: Springer Verlag, 2001.

[23] S. Haykin, *Neural networks: A comprehensive foundation*. 9th India Reprint, Upper Saddle River: Pearson Education, Inc. Pearson Prentice Hall, 2005.

Pavel Petr was born in Pardubice (CR) in 1957. He graduated from the Technical University in Liptovský Mikuláš (SR) in 1981. From 1981 till 1996 he worked in the army and in the Department of Anti-Aircraft Defense of State. Since 1985 he has been working as a lecturer in anti-aircraft systems. There he defended his doctoral thesis in 1987 and he habilitated in 1996. Since 1996 he is working as associated professor at Institute of System Engineering and Informatics and and as a vice-dean of Faculty of Economics and Administration at the University of Pardubice (CR).

Assoc. Prof. Petr published a number of papers concerning genetic programming, data mining and cluster analysis.



Jiří Křupka was born in Prostějov (CR) in 1962. He graduated from the Military Technical University in Liptovský Mikuláš (SR) in 1985. From 1985 till 1990 he worked in the Department of Technical Support System's and Automation in the Air Defense. From 1990 till 2004 he worked as a lecturer, a senior lecturer, and vice-dean for education at the Faculty of Air Defense at the Military Academy in Liptovský Mikuláš. There he finished his doctoral thesis in 1995 and habilitated in 1997. Since 2004 he is working as associated professor and head of Institute of System Engineering and Informatics, Faculty of Economics and Administration University of Pardubice (CR).

Assoc. Prof. Křupka has published parts of book and a number of papers concerning with fuzzy decision, fuzzy control, case based reasoning, and

rough set theory. Nowadays he is focusing on modeling of environmental and social systems.

Romana Provazníková is currently working at the position of associated professor at Institute of Economy and vice-dean at Faculty of Economics and Administration at University of Pardubice (CR). She graduated at 1986 at Prague School of Economy, the PhD theses she defended at Faculty of Economics and Public Administration at Masaryk University Brno in 2001. She has been habilitated at the Economic Faculty at Matej Bel's University at Banská Bystrica (SR) in 2008.

Her research topics are macroeconomic management under the fiscal federalism conditions, public finance and the fiscal imbalance and local government finance.

Assoc. Prof. Provaznikova published a number of papers on these topics.