Central European Review of Economic Issues
EKONOMICKÁ REVUE

Regional classification: The case of the Visegrad Four

Eva POLEDNÍKOVÁ*

Department of European Integration, Faculty of Economics, VŠB-Technical University of Ostrava, Sokolská 33, 701 21 Ostrava, Czech Republic.

Abstract

Since the biggest enlargement of the European Union in 2004, the EU cohesion policy has made a significant contribution to spreading growth and prosperity across the European territory. Despite this, the cohesion policy is still confronted with persistent economic, social and territorial disparities among countries and regions and weakened economic growth and competitiveness. The aim of the paper is to evaluate regional disparities in the case of the Visegrad Four (V4) countries in the year 2010 and to propose an alternative method of regional classification that could be helpful for the efficient allocation of European funds. In the paper, an analysis of the disparities in the V4 NUTS 2 regions is undertaken on the basis of cluster analysis. The three determined clusters confirm that NUTS 2 regions with capital cities (Praha, Bratislavský kraj, Mazowieckie and Közép-Magyarország) still occupy the dominant positions in comparison with other regions in the V4. Significant disparities between clusters are visible, especially regarding the economic and innovative performance and territorial cohesion.

Keywords

Classification, cluster, cluster analysis, regional disparities, Visegrad Four.

JEL Classification: C38, R11, Y1

---

* eva.polednikova@vsb.cz
This research paper was supported by the Faculty of Economics, VŠB-Technical University of Ostrava within the Student Grant Competition, project registration number SP2013/173, and within the Education for Competitiveness Operational Programme, project number CZ.1.07/2.3.00/20.0296.

ISSN 1212-3951 (Print), 1805-9481 (Online)
ER-CEREI, Volume 14: 25–37 (2014).
doi: 10.7327/cerei.2014.03.03

Regional classification: The case of the Visegrad Four

Eva POLEDNÍKOVÁ

1. Introduction

The economic, social and territorial disparities in the level of regional performance are a major obstacle to the balanced and harmonious development of the regions, but also of the territory as a whole. The admission of new member states to the European Union (EU) in 2004 and 2007 has been associated with an increase in regional disparities that has negatively affected the EU's competitiveness and internal cohesion (Mendez et al., 2013).

To promote harmonious and balanced development of the European territory, the elimination of disparities is considered as the primary objective of the EU's development activities. The EU cohesion policy plays a key role in regional development and EU funding (see Leonardi, 2005; Molle, 2007; Mendez et al., 2013). In the programming period 2007–2013, the EU cohesion policy seeks to eliminate regional differences through the support of regional growth, innovation and job creation. In practice, the policy is implemented by multi-annual development programmes co-financed by various types of financial instruments, notably the structural funds (the European Regional Development Fund – ERDF, the European Social Fund – ESF) and the Cohesion Fund (Molle, 2007). The current EU cohesion policy, which remains in effect until 2013, distributes funds to the total amount of 347 billion EUR among the European regions according to three objectives (Convergence, Regional competitiveness and employment, and European territorial cooperation) and the relative wealth of regions (measured by the gross domestic product per capita compared with the EU average). Almost 82% of the total funds focus on less prosperous regions within the *Convergence* objective. The Convergence objective covers NUTS 2 regions[1] with a gross domestic product per capita in purchasing power parity (GDP per capita in PPS) less than 75% of the average GDP of the EU-25. This objective also includes the category of regions called *phasing out*. These regions used to be eligible for funding under the Convergence objective, but now they are above the 75% threshold. Due to the statistical effect of the EU enlargement to 25 countries, these regions will receive transitional support until 2013. Another 16% of the total funds are intended for all the regions of the EU that are not eligible for funding from the Convergence objective. The Regional competitiveness and employment objective also covers another category of regions called *phasing in*. These regions, which used to be covered under the convergence criteria but are now above the 75% threshold even within the EU-15, receive *phasing-in* support. Less than 3% of the funds are available for cross-border, transnational and interregional cooperation under the European territorial cooperation objective (European Commission, 2010).

The level of regional disparities in the EU countries are still actual and important topics of many discussions and regional research studies, at the European level, e.g. Vorauer (1997), Wishlade and Yuill (1997), Soares et al. (2003), Felsenstein and Portnov (2005), Bacarić-Rašić (2006), Campo et al. (2008) and Zivadinovic et al. (2009). In recent years, the attention has been focused on the measurement of the regional disparities that impede the well-balanced development and strengthening of cohesion in *new* EU countries. For example, Viturka et al. (2009) research the regional disparities in 10 new EU member countries, Matlovič et al. (2008) deal with the regional disparities in Slovakia, Goshin et al. (2008) analyse the regional disparities in Romania and the regional differences in the Visegrad Four countries are analysed by Kutscherauer et al. (2010), Tuleja (2010) and Tvrdoň and Skokan (2011).

In the paper, the problem of cohesion policy and regional disparities in the Visegrad Four countries is considered. The Visegrad Four (the Czech Republic, Hungary, Poland and Slovakia) belong to the central European states, of which the economic development of the last 10 years has been strongly linked to European funding. Although the regional disparities have been reduced in the V4 with contributions from the EU cohesion policy, disparities still persist, especially between regions of capital cities and regions that are

[1] The Nomenclature of Territorial Units for Statistics (NUTS) serves as a reference for the collection, development and harmonization of EU regional statistics, socio-economic analyses of the regions and the framing of the EU regional policies (for the purpose of appraising eligibility for aid from the structural funds).

more distant from capital cities. The level of regional disparities has a significant impact on V4 eligibility for EU funding in the next period, 2014–2020.

The first goal of this paper is to evaluate the regional disparities in the case of the Visegrad Four countries in the year 2010. The results should contribute to the acceptance or rejection of the hypothesis that more developed NUTS 2 regions with capital cities (Praha, Bratislavský kraj, Mazowieckie and Közép-Magyarország) have a persistent significant socio-economic position that differs from the other regions and will be grouped into one homogeneous cluster. The second goal of the paper is to propose, with the example of the Visegrad Four, an alternative method of regional classification that could be helpful for the efficient allocation of European funds.

The rest of this paper is organized as follows. The basic concepts of the cohesion policy and regional disparities in the European Union are introduced in Section 2. In Section 3, the theoretical background of cluster analysis as a multivariate statistical method is discussed. In Section 4, an empirical case of regional disparity analysis in the V4 countries is illustrated. Moreover, the classification of regions based on cluster analysis and the European Commission (EC) classification of regions based on the GDP per capita are discussed. In Section 5, the main conclusions and remarks are provided.

The cluster analysis is performed through the statistical software PASW Statistics 18 and the table processor Microsoft Office Excel 2007. The European Statistical Office (Eurostat) serves as a basic database of the available and comparable regional indicators.

2. EU cohesion policy and the concept of regional disparities

The EU cohesion policy has been a force for change over the last 10 years, making a significant contribution to convergence and growth in the EU. The cohesion policy has directly created over 1 million jobs, invested in training to improve the employability of over 10 million people, co-financed the construction of over 2 000 km of motorway and 4 000 km of railway and set up at least 800 000 small and medium-sized enterprises (SMEs) (European Commission, 2011). Undoubtedly, without the cohesion policy, the disparities would be greater. Nevertheless, the lasting economic and social effects of the financial and economic crisis, the demand for innovation arising from increased global challenges and the imperative to increase competitiveness call for an ambitious reform of the policy for the next seven years.

2.1 EU cohesion policy after 2013

The future cohesion policy should continue to play a critical role in the task of delivering smart, sustainable and inclusive growth, while promoting harmonious development of the Union and its regions by reducing the regional disparities. On 6 October 2011, the European Commission adopted a draft legislative package that will frame the EU cohesion policy for the period 2014–2020. The EC proposed a number of important changes to the way in which the cohesion policy is designed and implemented, namely: concentrating on the Europe 2020 strategy's priorities of smart, sustainable and inclusive growth; rewarding performance; supporting integrated programming; focusing on results – monitoring progress towards the agreed objectives; reinforcing territorial cohesion; and simplifying delivery (European Commission, 2011). The European Commission decided that the cohesion policy should remain an essential element of the next financial package and underlined its pivotal role in delivering the Europe 2020 strategy. The total proposed budget of the EU cohesion policy for the period 2014–2020 will be 325.1 billion EUR (according to the proposal of the Multiannual Financial Framework 2014–2020 submitted in February 2013). After 2013, the cohesion policy will pursue two objectives: *Investment for growth and jobs* and *European territorial cooperation*. According to the EC, every European region may benefit from the support of the ERDF and ESF and thus three categories of regions have been newly proposed. A classification of regions as less developed, transitional and more developed regions will exist in order to ensure concentration of the funds within the first objective, for which the allocation is more than 313.1 billion EUR. The level of GDP per capita remains the main criteria for eligibility for regional funding. A definition of regions' eligibility with regard to each category is provided in Table 1.

2.2 Regional disparities in the context of the EU cohesion policy

The existence of disparities between regions, including their elimination, is one of the main aspects of the EU cohesion policy. In this context, we distinguish three types of regional disparities: economic, social and territorial. *Economic disparities* represent different levels of economic convergence of countries and regions that can be measured by economic indicators. *Social disparities* are related to how people perceive the spatially differentiated quality of life, standard of living or social inequality and they are mostly measured by the indicators of the labour market. *Territorial disparities* reflect the strong inequalities in the EU competitiveness factors. Territory inequality is expressed by the significant differences in the economic performance, geographical potential and transport and technical infrastructure, capacity for innovations or quality of the environment (Molle, 2007).

In the European concept, the level of disparities can be regarded as a measure of cohesion. According to Molle (2007), cohesion can be expressed as such a level of differences between countries, regions or groups that are politically and socially tolerable. Based on the typology of disparities, three dimensions of cohesion are recognized, i.e. economic, social and territorial. *Economic cohesion* evaluates economic convergence and can be expressed as disparities reducing the development levels of countries and regions by economic indicators. *Social cohesion* tends to achieve objectives in employment and unemployment, education level, social exclusion of different groups and demographic trends. *Territorial cohesion* is a supplementary term to economic and social cohesion. This concept develops economic and social cohesion by transferring the basic objective of the EU, i.e. balanced and sustainable development, into a territorial context (Kutscherauer et al., 2010).

In the EU, the disparity trends and cohesion situation of all the EU member states and their regions are evaluated within the *Reports on Economic, Social and Territorial Cohesion* (European Commission, 2007, 2010), published by the EC every three years. The most frequent indicators that are used for expressing the regional disparities and level of cohesion are provided in Table 2.

2.2 Selected approaches to regional disparity evaluation

The qualitative analysis of disparities provides important information about the key problematic issues in the region on one hand and its development potential on the other hand. The attitude of researchers towards the measurement and evaluation of regional disparities is not uniform. Most of the existing approaches to regional disparity evaluation use several disparity indicators that are processed by different *mathematical and statistical methods*. The aim is usually to obtain one comprehensive index that represents each of the analysed territories. Most of the regional economic inequalities are measured by a variety of indices based on the indicator of GDP – the coefficient of variation and the Hoover Concentration Index, the Herfindahl index, the Geographic Concentration Index and the Theil index (see e.g. Tvrdoň, 2012). A highly innovative approach to regional disparity analysis is presented by Viturka (2010). From the point of view of low calculation difficulty, a high informative level and the applicability of the results in practice, these mathematical and statistical methods are often used to measure disparities (Kutscherauer et al., 2010): the point method, traffic light method (scaling), method of average (standard) deviation, method of standardized variables and method of distance from the imaginary point.

From the perspective of practical utilization, the traffic light method can be applied in the phase of identification and quantification of the variables (see e.g. Melecký and Skokan, 2011). The point method, method of standardized variables and method of distance from the imaginary point are often used in an integrated approach based on the calculation of a synthetic index of disparities (see e.g. Tuleja, 2010; Poledníková, 2012; Svatošová and Boháčková, 2012) or a weighted synthetic index of disparities (Melecký, 2012). One of the possible methods that can be useful in the process of regional disparity evaluation is cluster analysis (CA); see e.g. Rovan and Sambt (2003), Soares et al. (2003) and Zivadinovic et al. (2009).

Table 1 Geographical coverage of EU cohesion policy support 2014–2020

Category of regions	Definition of eligibility	Eligibility for V4 regions
Less developed regions	NUTS 2 regions with a GDP per capita less than 75% of the average GDP of the EU-27	Střední Čechy, Jihozápad, Jihovýchod, Severovýchod, Střední Morava, Severozápad, Moravskolezsko, Közép-Dunántúl, Dél-Dunántúl, Észak-Magyarország, Dél-Alföld, Nyugat-Dunántúl, Észak-Alföld, Łódzkie, Małopolskie, Śląskie, Lubelskie, Podkarpackie, Świętokrzyskie, Podlaskie, Wielkopolskie, Zachodniopomorskie, Lubuskie, Dolnośląskie, Opolskie, Kujawsko-Pomorskie, Warmińsko-Mazurskie, Pomorskie, Východné Slovensko, Stredné Slovensko a Západné Slovensko
Transition regions	NUTS 2 regions with a GDP per capita between 75% and 90% of the EU-27 average	
More developed regions	NUTS 2 regions with a GDP per capita above 90% of the average GDP of the EU-27	Praha, Közép-Magyarország, Bratislavský kraj, Mazowieckie

Source: European Commission (2011), Ministerstwo Rozwoju Regionalneho (2013), own modification

As presented, there is neither a uniform approach to regional disparity analysis nor a comprehensive index for disparity evaluation at the European and national level. This paper is thus a response to the multidimensional problems of regional disparities and presents an alternative method for their evaluation.

3. Formulation of a model for regional disparity evaluation and the classification of regions

Regional disparity evaluation and the classification of regions will be undertaken simultaneously by cluster analysis. Cluster analysis represents one of the multivariate statistical methods for classifying objects into homogeneous clusters. The objects in each cluster are similar to each other in some ways and dissimilar to those in other clusters.

3.1 Theoretical basis of cluster analysis

Cluster analysis is a major technique for classifying a large amount of information into meaningful sub-groups, called clusters, which are more manageable than individual data. It allows the identification of homogenous groups of objects and the determination of what in our sample belongs to which group. In this term, a cluster means a group of relatively homogeneous cases or observations (Burns and Burns, 2008). The objects in a specific cluster share many characteristics, but are very dissimilar to objects that do not belong to the cluster. The aim of cluster analysis is to minimize the variability within clusters and maximize the variability between clusters. Cluster analysis examines the inter-relationships between variables, making no distinction between dependent and independent variables.

Cluster analysis is a method for quantifying the structural characteristics of a set of observations and has strong mathematical properties but no statistical foundations. The requirements of normality, linearity and homoskedasticity that are so important in other techniques have little bearing on cluster analysis. Nevertheless, there are two assumptions – *representativeness of the sample* and *multicollinearity* among the variables (Hair et al., 2009). The variables are examined for substantial multicollinearity and, if found, the variables have to be reduced. The cluster solution can also be distorted by outliers.

There are several clustering procedures for forming the groups of objects. The most popular procedures represent the hierarchical methods (agglomerative and divisive) and non-hierarchical methods. Each of the procedures follows a different approach to grouping the most similar objects into a cluster and to determining each object's cluster membership (Mooi and Sarstedt, 2011).

Table 2 Selected indicators of regional disparities

Type of disparities	Indicator	Abbreviation
Economic disparities	GDP per head (purchasing power standard per inhabitant)	GDP
	Disposable income of households (purchasing power standard per inhabitant)	DI
	Labour productivity (% GDP per person employed in PPS, EU27 = 100)	LP
	Gross fixed capital formation (purchasing power standard per inhabitant)	GFCF
	Gross domestic expenditure on research and development (GERD) (% of GDP)	GERD
	Patent applications to the European Patent Office (EPO) (number per million of inhabitants)	EPO
	Human resources in science and technology (% of active population)	HRTS
	Employment in technology and knowledge-intensive sectors (% of active population)	ETKI
Social disparities	Employment rate (% of population aged 15–64)	ER15to64
	Employment rate of older workers (% of population aged 55–64)	ER55to64
	Employment rate of woman (% of woman population aged 15–64)	ERw15to64
	Unemployment rate (% of labour force aged 15–64)	UR15to64
	Unemployment rate of youth (% of labour force aged 15–24)	URy15to24
	Long-term unemployment (% of labour force)	LtUR
Territorial disparities	Collective tourist accommodation establishments (number)	TE
	Tourism intensity (number)	TI
	Density of railways (km/1000 km^2)	DR
	Hospital beds (number per thousand inhabitants)	HB
	People killed in road accidents (number of deaths per million inhabitants)	PKR
	Infant mortality rates (%)	IMR

Source: European Commission (2007, 2010), Eurostat (2012), own elaboration

Hierarchical cluster analysis

Hierarchical cluster analysis uses dissimilarities such as distances between objects when forming the clusters. The most common way of computing the distances between objects in a multidimensional space is to compute *Euclidean distances*, an extension of Pythagoras' theorem. The *squared Euclidean distanc*e is used more often than the simple Euclidean distance in order to place progressively greater weight on objects that are further apart (Mooi and Sarstedt, 2011). There are also other frequently used alternative distance measures, e.g. *the Manhattan metric*, *the Chebychev distance* or *the Mahalanobis distance*. After the computation of the distance measure, the clustering algorithm has to be selected. There are several agglomerative procedures and they can be distinguished by the way in which they define the distance from a newly formed cluster to a certain object or to other clusters in the solution. The most popular agglomerative clustering procedures include (Mooi and Sarstedt, 2011): *methods of single linkage (nearest neighbour)*, *complete linkage (furthest neighbour)*, *average linkage*, *centroid*, *Ward's* and *median*. The last step of cluster analysis is the graphical representation of the distance at which clusters are combined (figure dendogram) and the selection of the cluster solution that best represents the data sample.

Non-hierarchical cluster analysis

Non-hierarchical clustering differs from hierarchical clustering in several ways. Non-hierarchical clustering methods have the advantage of being able to *optimize* the cluster solution better by reassigning observations until the maximum homogeneity within clusters is achieved. The number of clusters is determined from the hierarchical results. A primary benefit of non-hierarchical cluster methods is the ability to develop a cluster solution later in the process that is not based on the clusters formed earlier (in hierarchical methods, the cluster solution is directly based on the combination of two clusters formed earlier in the process). Non-hierarchical methods are generally preferred for their *fine-tuning* of an existing cluster solution from a hierarchical process (Hair et al., 2009).

The most common and popular clustering algorithm is called *k-means*,[2] where k is the number of required clusters. The k-means clustering algorithm assigns cases to clusters based on the smallest amount of distance between the cluster mean and the case. The first step in k-means clustering is to find the *k* centres (seed points) that are the initial starting point for each cluster. Centres can be selected by two methods – generation of the sample by the cluster software (i.e. random selection) or specification by the researcher (Hair et al., 2009). After the initial cluster centres have been selected, each case is assigned to the closest cluster, based on its distance from the cluster centres. This is performed iteratively. After all of the cases have been assigned to clusters, the cluster centres are recomputed, based on all of the cases in the cluster. Case assignment is performed again, using these updated cluster centres. The process continues until the homogeneity within the clusters cannot be increased by further movement of the cases between clusters.

3.2 Model for the evaluation of the regional disparities in the V4 and the classification of regions

For the cluster analysis, 20 indicators available for 35 NUTS 2 regions[3] in the V4 countries were selected for the year 2010. The selection of appropriate regional indicators results from the concept of regional disparity evaluation in the EU. These indicators represent the most frequently used indicators of economic, social and territorial disparities in the *Reports on Economic, Social and Territorial Cohesion* (European Commission 2007, 2010), which evaluate the trends of the disparities and cohesion in the EU member states and regions. However, the determination of appropriate and comparable regional statistics has faced significant problems of limited availability at the required territorial level and time series. From this reason, the most recent data exist for the year 2010. The indicators that are available in the Eurostat database at the level NUTS 2 V4 regions are presented in Table 2.

Because of the multicollinearity, it was necessary to remove five indicators[4] from the follow-up analysis: the disposable income of households, labour productivity, human resources in science and technology, unemployment rate and unemployment rate of youth. Because of the different scales of the clustering variables, the variables are standardized by Z-score (by SPSS). The final input matrix for the cluster analysis is created by five economic indicators, four social indicators and six territorial indicators in the year 2010.

After the preparation of the data, hierarchical cluster analysis is used to determine the optimum cluster solution in the case of the V4 regions. The agglomerative clustering process is based on Ward's method, applying the squared Euclidean distance as the dis-

[2] Alternative algorithms are k-medoids and fuzzy clustering.

[3] The Czech Republic: 8 NUTS 2 regions, Hungary: 7 NUTS 2 regions, Slovakia: 4 NUTS 2 regions, Poland: 16 NUTS 2 regions.

[4] The Pearson correlation coefficient achieved a value above 0.8.

tance (similarity) measure. The squared Euclidean distance is used because it is the basis for Ward's method. It can be expressed by the following formula (Řezánková, 2009):

$$D_{ES}\left(x_i, x_j\right) = \sum_{l=1}^{m}\left(x_{il} - x_{jl}\right)^2, \qquad (1)$$

where x_{il} is the value of the l-th attribute of the i-th object and x_{jl} is the value of the l-th attribute of the j-th object. By finding the similarities of objects (regions), the matrix of distance called the *Proximity Matrix* is obtained. The greater the value of the distance, the smaller the degree of similarity between the objects.

Ward's method is selected as the clustering algorithm because it is generally a very efficient method and uses an analysis of variance approach to evaluate the distances between clusters. The results obtained by different hierarchical clustering methods are often very different, due to the interaction space between the objects. In this case, Ward's method appears to be appropriate, since it extends the space between cases by the formation of compact clusters with a large number of cases. The information about the hierarchical clustering process produces the *Agglomeration Schedule* table, in which the column *Coefficients* is very important. The coefficients help to decide how many clusters are optimal for the representation of the data. In this case, the cluster formation should be stopped when the increase in the coefficients between two adjacent steps is *large* (Meloun et al., 2005). The distance at which clusters are combined is presented by a *dendogram*. Finally, the extracted clusters are interpreted using the *profile of clusters* based on the mean value of the standardized variable. The profile of clustering variables for the cluster solution helps to define the characteristics of the cluster.

On the basis of the hierarchical cluster solution, the non-hierarchical clustering process is executed. The algorithm selected is the *optimizing algorithm* in SPPS that finds k cases that are well separated and uses these values as initial cluster centres, allowing the reassignment of observations among clusters until a minimum level of heterogeneity is reached. The clusters are described by the distances of the final cluster centres and by ANOVA F-tests (analysis of variance). The F statistics from one-way ANOVAs examine whether there are statistically significant differences between clusters for each of the clustering variables. The independent variable is cluster membership and the dependent variables are clustering variables.

4. Evaluation of the regional disparities in the V4 and the classification of regions

In the first step, the analysis of regional disparities in the V4 is based on the agglomerative hierarchical clustering process. According to the determined clusters, a classification of regions is proposed and discussed. Subsequently, the results of hierarchical clustering are verified by the non-hierarchical clustering process.

4.1 Empirical results of the cluster analysis

According to the *Proximity Matrix*, the greatest differences exist between the Czech region Praha and the Slovak region Západné Slovensko in the year 2010 (178). The lowest distance is recorded between two Polish regions, Warmińsko-Mazurskie and Opolskie (2.8). From the analysis of distances at which clusters are joined (see the dendrogram presented in Figure 2 in the Appendix), it can be considered that an optimal solution must fall within *three to five* clusters. As stated above, the increase in the coefficients in the Proximity Matrix represents the important information for the determination of the cluster solution. In this case, the largest increase in coefficients is recorded between the second and the third cluster, so the best interpretation of the data ensures a *three-cluster solution*. Moreover, for the purpose of comparison, the classification of regions resulting from the presented approach with three categories proposed by the EC (see Table 1), the focus will be on the three-cluster solution.

Cluster I represents only the Praha region. *Cluster II* is created by three regions of capital cities, Mazowieckie, Közép-Magyarország and Bratislavský kraj, and by five Czech regions, Střední Čechy, Jihozápad, Jihovýchod, Severovýchod and Střední Morava. *Cluster III* comprises two Czech regions – Severozápad and Moravskolezsko, Hungarian regions – Közép-Dunántúl, Dél-Dunántúl, Észak-Magyarország, Dél-Alföld, Nyugat-Dunántúl and Észak-Alföld, Polish regions – Łódzkie, Małopolskie, Śląskie, Lubelskie, Podkarpackie, Świętokrzyskie, Podlaskie, Wielkopolskie, Zachodniopomorskie, Lubuskie, Dolnośląskie, Opolskie, Kujawsko-Pomorskie, Warmińsko-Mazurskie and Pomorskie – and three Slovak regions – Východné Slovensko, Stredné Slovensko and Západné Slovensko. The gradual clustering of the V4 regions in the year 2010 is presented in the dendogram in Figure 2 in the Appendix.

From the classification of regions into clusters, we can state that the clusters correspond to the three categories defined by the EC for the cohesion policy 2014–2020 (see Table 1). Cluster I corresponds to the category *more developed regions*, cluster II corresponds to the *transitional regions* and cluster III corresponds to the *less developed regions*. To obtain a precise interpretation of the clusters' characteristics and a comparison of the regions' classification, we construct the profile of each cluster that is based on *the mean value of the standardized indicators (variable).*

Three defined clusters confirm that the socio-economic situation of the metropolitan NUTS 2 regions is different from the other V4 NUTS 2 regions; therefore, these regions tend to be naturally grouped into one cluster. As Figure 1 shows, the Praha region has an important position, being classified in Cluster I. The region achieves the highest average value of the economic indicators. It is also characterized by a high-quality structure of the labour force (a high share of human resources in science and technology), high support of research and development and a flexible labour market (all the indicators achieve the most favourable development). According to the cluster analysis, the Praha region is considered to be the most developed region. It corresponds to the fact that Praha has been classified by the European Commission as a *more developed region*.

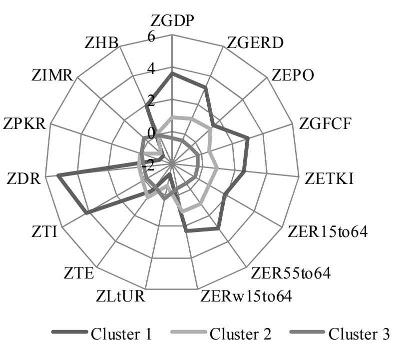

Figure 1 Clusters' profile, 2010 (mean of the standardized variables)

On the contrary, Cluster III is characterized by the worst results in the economic, social and territorial segment. The disparities between Cluster I and Cluster III are significant in the territorial indicators and economic performance (especially in the indicators tourism intensity, density of railways, GDP per capita and GERD). Cluster III can be regarded as the least developed. As in the first case, the classification of regions based on multivariate analysis reflects the EC classification. All the regions in this cluster are classified as *less developed regions*.

Cluster II includes capital regions – Közép-Magyarország, Bratislavský kraj and Mazowieckie – and highly developed Czech regions, which can be seen as a positive trend in reducing the disparities and increasing the convergence between the capital regions themselves and between the Czech regions. As shown in Figure 2, this cluster is characterized by average values of the indicators. In this case, the classification that takes account of the economic, social and territorial development of regions is not in accordance with the EC classification based only on the GDP per capita criterion. For the period 2014–2020, the regions Közép-Magyarország, Bratislavský kraj and Mazowieckie have been defined by the EC as *more developed regions* and the Czech regions Střední Čechy, Jihozápad, Jihovýchod, Severovýchod and Střední Morava have been considered as *less developed regions*. However, these regions should rather be considered as *transition regions* according to the multidimensional evaluation of their development. The allocation of financial resources based on a threshold corresponding to 75% of the European average GDP per capita can already cause inefficient funding and can negatively affect the poorest areas. It seems that in the case of the V4, the less developed regions would not be negatively affected by the EC classification. The opposite situation can be exactly observed in the case of NUTS 2 regions Közép-Magyarország, Bratislavský kraj and Mazowieckie. The region Mazowieckie, as the first Polish region, will leave the group of the poorest EU regions because it exceeds the 75% per capita GDP threshold for EU regions. This means that the region should be classified as a *more developed region* in the period 2014–2020. However, the rule of GDP per capita does not reflect three main aspects: the existence of social and territorial differences in the region, internal economic, social and territorial differences that are considerable at the level of NUTS 3 regions and the dominant position of the capital city, Warsaw, which lies in the Mazowieckie region and statistically affects the level of development of the whole region. According to the author's approach, the Mazowieckie region rather corresponds to a *transition region*. The better explanatory ability of the classification based on multivariate analysis would also confirm the fact that the European Commission finally included the Mazowieckie region in *safety nets* (regions in this category will receive an allocation under the structural funds equal to at least two-thirds of their 2007–2013 allocation).

4.2 Empirical results of the non-hierarchical cluster analysis

The final classification of regions according to the three-cluster solution determined in the hierarchical process is shown in Table 4 in the Appendix. There are no differences between the hierarchical and the non-hierarchical results for the year 2010.

The similarity (dissimilarity) between clusters provides another output of the k-means clustering process. Table 3 shows the differences between the final cluster centres based on the Euclidean distances.

Greater distances between clusters mean that there are greater dissimilarities. As shown in Table 3, the highest dissimilarity is confirmed between Cluster I and Cluster III. On the contrary, the lowest differences are found between Cluster II and Cluster III.

Table 3 Distances between the final cluster centres (2010)

Cluster	1	2	3
1		8.738	11.652
2	8.738		4.235
3	11.652	4.235	

Subsequently, the ANOVA table indicates which variables contribute the most to the cluster solution. Variables with large mean square errors provide the least help in differentiating between clusters. According to Table 5 in the Appendix, the variables that are not as helpful as the other variables in forming and differentiating clusters are especially territorial indicators: collective tourist accommodation establishments (ZTE), people killed in road accidents (ZPKR), hospital beds (ZHB) and also long-term unemployment (ZLtUR).

The non-hierarchical cluster analysis confirmed the clusters' membership as well as the classification of regions given by the hierarchical cluster analysis.

5. Conclusion

The results of the cluster analysis show that NUTS 2 regions with capital cities (Praha, Bratislavský kraj, Mazowieckie and Közép-Magyarország) still have a significant and different socio-economic position from the other regions in the V4. Significant disparities between Cluster I and Cluster III are visible, especially regarding the economic and innovative performance and territorial cohesion.

The comparison of the three-cluster classification with the three categories proposed by the EC shows that each of three categories of the EC classification comprises different groups of regions with greater heterogeneity than in the case of classification based on the multivariate method. The comparison of the two approaches with the regional classification implies that one-dimensional evaluation is insufficient for characterizing the dissimilarity among regions and for designing the solutions to the different groups of regions with regard to their different needs within the EU territory. The usage of multivariate classification for purposes of the cohesion policy could clarify the association of financial resources and policies with the problems of regions identified through different indicators of regional disparities.

The author of the paper takes into account that the informative level of the cluster analysis is influenced by the characteristic of the data file (e.g. the occurrence of outliers and the correlation of the variables), by the selected number and type of the indicators, as well as by the selected clustering technique, distance criterion and algorithm (method) of the clustering. The author also takes into consideration that the V4 regions represent only a small sample of the European regions. Therefore, it seems an interesting task to conduct further analysis aiming to compare the results with those from other different EU countries and/or different territorial units (NUTS 3).

References

BURNS, R.B., BURNS, R.A. (2008). *Business Research Methods and Statistics Using SPSS*. London: Sage.

BAKARIĆ-RAŠIĆ, I. (2006). Uncovering regional disparities – the use of factor and cluster analysis. *Croatian Economic Survey* 9: 11–34.

CAMPO, C., MONTEIRO, C.M.F., SOARES, O.J. (2008). The European regional policy and the socio-economic diversity of European regions: A multivariate analysis. *European Journal of Operational Research* 187(2): 600–612. http://dx.doi.org/10.1016/j.ejor.2007.03.024.

EUROPEAN COMMISSION (2011). *Cohesion Policy 2014–2020. Investing in Growth and Job*. Luxembourg: Publications Office of the European Union. doi:10.2776/43732

EUROPEAN COMMISSION (2010). *Fifth Report on Economic, Social and Territorial Cohesion. Investing in Europe's future*. Luxembourg: Publications Office of the European Union.

EUROPEAN COMMISSION (2007). *Growing Regions, Growing Europe. Fourth Report on Economic and Social Cohesion*. Luxembourg: Office for Official Publications of the European Communities.

FELSENSTEIN, D., PORTNOV, B.A. (2005). *Regional Disparities in Small Countries*. Berlin: Springer. http://dx.doi.org/10.1007/3-540-27639-4

GOSHIN, Z., CONSTANTIN, D., ROMAN, M., ILEANU, B. (2008). The current state and dynamics of regional disparities in Romania. *The Journal of the Romania Regional Science Association* 2: 80–105

HAIR, J. F., BLACK, W.C. et al. (2009). *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.

KUTSCHERAUER, A. et al. (2010). *Regionální disparity: Disparity v regionálním rozvoji země, jejich pojetí, identifikace a hodnocení*. Ostrava: VŠB-TUO.

LEONARDI, R. (2005). *Cohesion Policy in the European Union: The Building Europe*. New York: Palgrave Macmillan.

MATLOVIČ, R., KLAMÁR, R., MATLOVIČOVÁ, K. (2008). Development of regional disparities in Slovakia at the beginning of the 21st in the light of selected indicators. *Regionální studia* (2): 2–12.

MELECKÝ, L. (2012). Assessment of cohesion in selected European Union countries by construction of synthetic indices of disparities. In: *Recent Research in Business and Economics – Proceedings of the 4th World Conference on Applied Economics, Business and Development (AEBD 2012)*. Porto: WSEAS Press, 211–216.

MELECKÝ, L., SKOKAN, K. (2011). EU cohesion and its evaluation in the case of Visegrad Four countries. In: *Liberec Economic Forum 2011*. Liberec: Technical University of Liberec, 314–326.

MELOUN, M., MILITKÝ, J., HILL M. (2005). *Počítačová analýza vícerozměrných dat v příkladech*. Praha: Academia.

MENDEZ, C., WISHLADE, F., BACHTLER, J. (2013). *EU Cohesion Policy and European Integration. The Dynamics of EU Budget and Regional Policy Reform*. USA: Ashgate.

MOLLE, W. (2007). *European Cohesion Policy*. London: Routledge.

MOOI, E., SARSTEDT, M. (2011). *A Concise Guide to Market Research. The Process, Data, and Methods Using IBM SPSS Statistics*. Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-12541-6

POLEDNÍKOVÁ, E. (2012). Hodnocení disparit v kontextu soudržnosti zemí Visegrádské čtyřky. In: *MEKON 2012, 14th Intenational Conference*. Ostrava: VŠB-TUO, 1–13.

ROVAN, J., SAMBT, J. (2003). Socio-economic differences among slovenian municipalities: A cluster analysis approach. In: Ferligoj, A., Mrvar, A., (Eds.): *Developments in Applied Statistics*. Ljubljana: Fakulteta za družbene vede, 265–278.

ŘEZÁNKOVÁ, H. (2009). *Shluková analýza dat.* Praha: Professional Publishing.

SOARES, O. J., MARGUES, M.M.L., MONTEIRO, C.M.F. (2003). A multivariate methodology to uncover regional disparities: A contribution to improve European Union and governmental decision. *European Journal of Operational Research* 145(1): 121–135. http://dx.doi.org/10.1016/S0377-2217(02)00146-7

SVATOŠOVÁ, L., BOHÁČKOVÁ, I. (2012). Metodologické přístupy k hodnocení regionálních disparit. In: *Sborník příspěvků, XV. Mezinárodní kolokvium o regionálních vědách*. Brno: Masarykova univerzita, 11–18.

TULEJA, P. (2010). Praktická aplikace metod hodnocení regionálních disparit. *Acta Academica Karviniensia* (1): 496–509.

TVRDOŇ, M. (2012). Cohesion, convergence and regional disparities: The case of European Union. *WSEAS Transactions on Business and Economics* 9(2): 89–99.

TVRDOŇ, M., SKOKAN, K. (2011). Regional disparities and the ways of their measurement: The case of the Visegrad Four countries. *Technological and Economic Development of Economy* 17(3): 501–518. http://dx.doi.org/10.3846/20294913.2011.603485

VITURKA, M. (2010). Regionální disparity a jejich hodnocení v kontextu regionální politiky. *Geografie* 115(2): 131–143.

VITURKA, M., ŽÍTEK, V., KLÍMOVÁ, V., TONEV, P. (2009). Regional analysis of new EU member states in the context of cohesion policy. *Review of Economic Perspectives* 9(2): 71–90. http://dx.doi.org/10.2478/v10135-009-0001-8

VORAUER, K. (1997). *Europäische Regional Politik Regionale Diparitäten*. Passau: Münchener Geographische Hefte.

ZIVADINOVIC, N.K., DUMICIC, K., CASNI, C.A. (2009). Multivariate analysis of structural economic indicators for Croatia and EU 27. In: *Proceedings of the 2nd WSEAS International Conference on Multivariate Analysis and its Application in Science and Engineering*. Istanbul: WSEAS Press, 134–139.

Additional sources

EUROSTAT (2012). *Regional Statistics*. [On-line], accessed at 30. 12. 2012. Available from: <http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes>.

MINISTERSTVO ROZWOJU REGIONALNEHO (2013). *Programowanie perspektywy finansowej 2014–2020. Umowa Partnerstwa*. [On-line], accessed at 20. 07. 2013. Available from: <http://www.mrr.gov.pl/fundusze/Fundusze_Europejskie_2014_2020/Programowanie_2014_2020/Umowa_partnerstwa/Documents/UP_19_07_2013.pdf>.

TULEJA, P. (2008). *Metody měření regionálních disparit v územním rozvoji České republiky*. [On-line], accessed at 02. 01. 2012. Available from: <http://disparity.vsb.cz/dokumenty2/RD_0803.pdf>.

WISHLADE, F., YUILL, D. (1997). Measuring disparities for area designation purposes: Issues for the European Union. *Regional and Industrial Policy Research Paper,* No. 24. [On-line], accessed at 10. 08. 2012. Available from: <http://www.eprc.strath.ac.uk/eprc/Documents/PDF_files/R24MeasDispforAreaDesigPurposes.pdf>.

Appendix

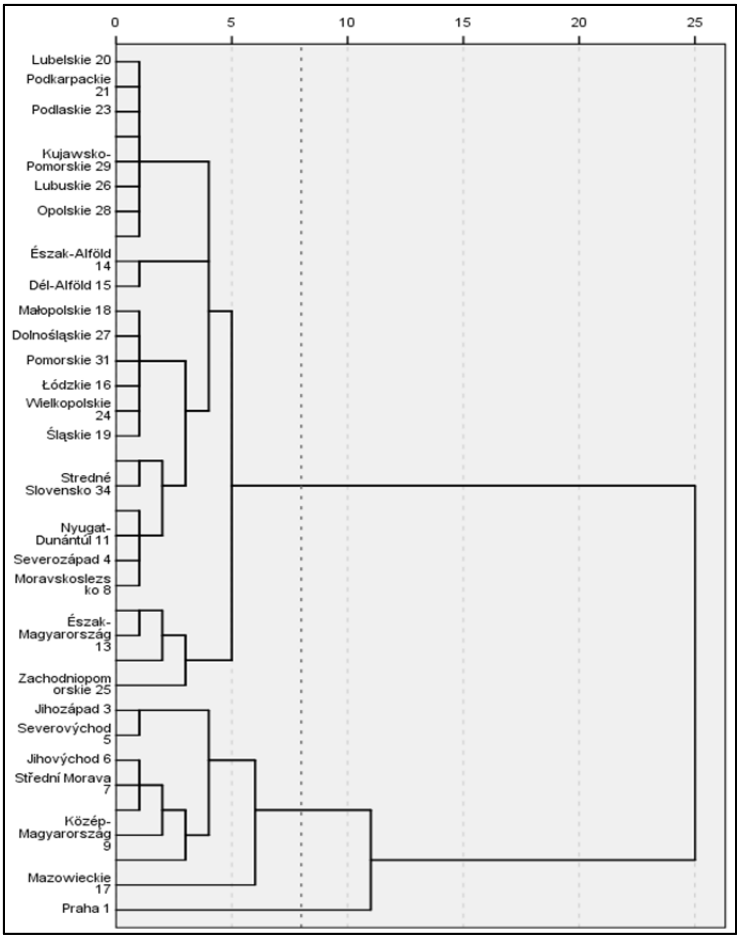

Figure 2 Dendogram, 2010

Table 4 Comparison of cluster membership based on Ward´s method and K-means (2010)

Code	Region	2010	
		Ward´s method	K-means
CZ01	Praha	1	1
CZ02	Střední Čechy	2	2
CZ03	Jihozápad	2	2
CZ04	Severozápad	3	3
CZ05	Severovýchod	2	2
CZ06	Jihovýchod	2	2
CZ07	Střední Morava	2	2
CZ08	Moravskoslezsko	3	3
HU10	Közép-Magyarország	2	2
HU21	Közép-Dunántúl	3	3
HU22	Nyugat-Dunántúl	3	3
HU23	Dél-Dunántúl	3	3
HU31	Észak-Magyarország	3	3
HU32	Észak-Alföld	3	3
HU33	Dél-Alföld	3	3
PL11	Łódzkie	3	3
PL12	Mazowieckie	2	2
PL21	Małopolskie	3	3
PL22	Śląskie	3	3
PL31	Lubelskie	3	3
PL32	Podkarpackie	3	3
PL33	Świętokrzyskie	3	3
PL34	Podlaskie	3	3
PL41	Wielkopolskie	3	3
PL42	Zachodniopomorskie	3	3
PL43	Lubuskie	3	3
PL51	Dolnośląskie	3	3
PL52	Opolskie	3	3
PL61	Kujawsko-Pomorskie	3	3
PL62	Warmińsko-Mazurskie	3	3
PL63	Pomorskie	3	3
SK01	Bratislavský kraj	2	2
SK02	Západné Slovensko	3	3
SK03	Stredné Slovensko	3	3
SK04	Východné Slovensko	3	3

Table 5 Anova (2010)

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
ZGDP	11.140	2	0.366	32	30.413	0.000
ZGERD	11.662	2	0.334	32	34.952	0.000
ZEPO	9.935	2	0.442	32	22.501	0.000
ZGFCF	6.688	2	0.645	32	10.376	0.000
ZETKI	7.794	2	0.575	32	13.546	0.000
ZER15to64	4.519	2	0.780	32	5.793	0.007
ZER55to64	11.633	2	0.335	32	34.684	0.000
ZERw15to64	9.550	2	0.466	32	20.510	0.000
ZLtUR	3.139	2	0.866	32	3.623	0.038
ZTE	2.167	2	0.927	32	2.337	0.113
ZTI	9.631	2	0.461	32	20.911	0.000
ZDR	14.157	2	0.178	32	79.681	0.000
ZPKR	0.761	2	1.015	32	0.750	0.481
ZIMR	7.579	2	0.589	32	12.872	0.000
ZHB	2.083	2	0.932	32	2.235	0.123