Chapter 2

Classification of the Regions

Jana Hančlová

Department of Mathematical Methods in Economics
Faculty of Economics, Technical University of Ostrava
Sokolská 33, 701 21 Ostrava

jana.hanclova@vsb.cz

Lubor Tvrdý

Department of Regional Economics
Faculty of Economics, Technical University of Ostrava
Sokolská 33, 701 21 Ostrava

lubor.tvrdy@vsb.cz

KEY WORDS

Classification, qualitative and quantitative approach, expert estimation, cluster analysis, factor analysis, REDIF – methodology, regional data

SUMMARY

This chapter deals with the basic approaches to the region classification both in general and in specific cases. The chapter is divided into four parts.

The first part defines basic terms of classification, identification, typology etc. There are also mentioned the possibilities of different classification approaches (quantitative or qualitative, static or dynamic).

The second part deals with a definition of the region and new administrative arrangement of the Czech Republic (CZ NUTS). There is also a basic information of regional data sources (including quality of indicators) in particularly in the Czech Republic.

The third part explains the nature and purpose of cluster analysis as a methodological background of the regional classification.

Next part of this chapter presents three empirical studies that are based on research background of the authors in the field of economics, sociology, demography and regional science. The first study evaluates 14 regions in the Czech Republic in the year 1997, 2000 and 2002 on the basis of two extracted factors using the factor analysis of 18 socio-economic regional indicators. Five significant clusters are identified by means of the hierarchical cluster analysis. The second study includes multi-period classification of the districts (NUTS 5) in the Czech Republic using unemployment rate and wages from the year 1996 through 2000 using absolute or relative differences. The last empirical study is to compare a regional competitiveness level and territorial capital of six selected pilot IRON CURTAIN reference areas by REDIF (REgional DIFferential) method. The attention is devoted to the comparison of the quantitative and qualitative methodological approach towards the classification.

1 Theoretical Background

1.1 Classification

The classification of objects ranks among the fundamental cognitive instruments. Through the classification a man searches for the order in the chaos. This need for the order or an arrangement, if you like, of the reality, can be found even in the case of the natural nations.

By classification, we mean the basic set of objects dissociation leading to the creation of the classes system – the so-called classification system. It is necessary to distinguish between the terms “classification” and “identification” which are usually confused (unified). *Identification* is the assignment of the individual objects into classes which have been created prior the actual assignment; it means that during identification we need to know the classification system. If this system has not been created yet or is inappropriate, then the classification needs to be done first.

The classification system has to fulfill the condition of completeness, i.e. the classification has to be exhaustive so that each object from the basic set could be classifiable; and the condition of definiteness, i.e. each object may be associated only with one class, hence the classes do not overlap.

Another term that is often wrongly used as a synonym for the term “classification” is a typology. *Typology* means the categorization of a set of objects using the *type* as a certain model. In contrast to the classification system, typology does not have to fulfill the condition of definiteness, hence it allows for the existence of mixed (overlapping) types which occur in the reality. In order to fulfill the condition of completeness, in typology it is possible to make a logical construction when an artificial group of objects is labeled as “the others” or “uncategorized”. Making typologies is closer to the practical use of classification methods, e.g. types of society development and suchlike. When it comes to the comprehension of variability principles and traits development at investigated objects, it is possible to generalize the typology and thus create a classification system.

The existence of a certain order in nature, respectively in matter, is the basic condition for the creation of classifications. Hence, the purpose of classifications is not only the heterogeneity simplification of reality and the arrangement of the reality for the cognition but they should also lead to finding the hidden elementary structures that exist independently on our conscience. The Linn's hierarchic taxonomy system and the Mendeleyev's periodic table of elements serve as an example of these classification systems.

1.2 Qualitative and Quantitative Classification

Quantitative methods proceed from an assumption that the objects properties can be measured by means of arguments. The term *argument* denotes the character of an object which is perceivable by senses. The number codes are being assigned to the individual states of an argument – the so-called *values* of an argument. An argument has to manifest the following properties:

- Distinguishability (ability to take on at least two values),
- Completeness (every possible state of an object has its defined value of the argument),
- Definiteness (any two different values of an argument may not correspond to the same state).

In some cases the argument values may even have the character of language descriptions which can be modeled by means of the "fuzzy" sets, for example.

Functionality and the meaningfulness are the elementary verificatory conditions for accepting a classification system. Besides that, the classification should explain even the value variability of other arguments which the classification was not created from but which it is relevant to. The classification system stability in time represents another verificatory condition.

Qualitative methods have a great advantage in comparison with the ones mentioned above in the sense that a scientist is creating the classification during the collection of data. Hence the classification is not only a mere "dry" ex-post created construct but also it is being constantly confronted with the reality and is constantly being modified by the expanding experience. In this sense, the process of classification takes place simultaneously with identification. It is possible to compare the cognition process to the centripetal motion along a vertical spiral. In the middle of the spiral there should be situated the "classification system" which corresponds to the real structure of processes and its vertical dimension represents the increase in cognition. With some simplification it can be said that the sense of the entire qualitative research is to find a structure and a regular pattern in the obtained data which can be called the creation of a classification system from these data.

Both of these above-mentioned approaches have their advantages as well as disadvantages which are summarized in Table 1.1.

Table 1.1: Comparison of qualitative and quantitative methods

	Quantitative methods	Qualitative methods
Advantages	1. objectiveness 2. higher reliability 3. comparability 4. simplicity	1. high validity 2. interaction between a researcher and the reality 3. acceptation and interpretation of all the objects including the extremes
Disadvantages	1. problem with measureability (namely in case if the soft factors) 2. lower validity 3. exclusion of the extreme values (loss of information)	1. the influence of the view subjectivity 2. lower reliability 3. demanding as to the skills and experience of a researcher

When creating classification systems it is to certain extent necessary to proceed from both of the two approaches. An author of the quantitative classification has to be able to interpret it and this interpretation should be meaningful even for the rest of the professional public. The author has to use his/her expert knowledge for the interpretation, i.e. the use of the qualitative approach. Even the orthodox followers of the numerical taxonomy dating back to the 60s and 70s of the 20th century have come to this conclusion. Their conclusions had even more defects than so-criticized conclusions of the classic taxonomy because they were very unnatural even though otherwise they were totally objective, see (Lukasová - Šarmanová, 1985, pp. 13 – 14). On the other hand, it is convenient to operationalize and test the qualitative classification at the quantitative level as well.

1.3 Static and dynamic classification

At general level, classification systems should be totally time-independent. It holds provided that the classification proceeds from the law that is especially defined in the context of sciences, e.g. the periodical law of elements. Nevertheless, these classification systems occur just randomly in the field of the social sciences and therefore the condition of time-independence is loosened in this field. Although the suggested classification system has to have certain stability in time, the development variability in time is acceptable, especially in such a case where it can be clearly explained.

If we suggest a classification system by means of quantitative methods for two different time periods, a shift in the profile of the individual classes occurs, more specifically of their centroids. The membership of objects of these classes changes, too. The problem of classification system instability in time is also possible to eliminate by suggesting a single classification system for more time periods. There exist two possible approaches:

On one side, it is possible to suggest a classification system in which the objects can move between the individual classes. If the centroids of these classes are stable during the entire studied time period (i.e. the coordinates of the centroids do not change), it concerns the static classification system.

On the other side, it is possible to suggest a classification system proceeding from development of the objects within the whole time period. In this case, each object belongs to exactly one cluster (centroid). Thus we create the objects classification system. In this classification system the centroid coordinates express a development (a change), hence the system can be understood as the dynamic classification system.

Classification systems for more time periods:

- static;
- dynamic;
    - one variable (time series clustering);
    - two or more variables (generalization).

The system created on the basis of one argument, which is called *time series clustering*, is a special case of the dynamic system. This approach has become popular during recent time, particularly in the field of data mining. When constructing the dynamic classification system it is possible to use the absolute as well the relative form of expressing the *proximity.* The absolute form, e.g. the Euclidean distance, identifies the mutual position of objects toward each other. On contrary, the relative form, e.g. the Pearson's coefficient of correlation, compares the differences in the objects development character. It is possible to assemble the dynamic classification system even for more arguments describing the development of objects.

1.4 Conclusion of the section

In this chapter, different possibilities of classification approaches have been presented. Choice of methodology depends on the particular problem; a combination of several approaches often gives the best results. Nevertheless, we should stick on the classification also under the resulting approach different from the one under which it was created. This complementarity property is best shown at the case of qualitative and quantitative approach. The individual approaches are described only at general level. For better understanding, some examples of the above-mentioned classification systems are presented in subchapter 4.

2 Regions and Data

2.1 Definitions of basic terms

A *region* is a territorial unit in which it is possible to define the elementary social-economic system and to accurately describe this system with the help of one or more arguments. The region can be described on the basis of the relations, i.e. we merge the territories which possess strong spatial interrelations, on the basis of homogeneity, i.e. we merge the homogeneous territories according to the selected arguments or on the basis of functionality. A social-economic system does not find itself in an abstract space but it is defined in a specific environment with which it is in an interaction.

A *natural region* is characterized by the cultural-social elements (historical development of the local fellowships) as well as by the geographical elements (e.g. soil fertility, altitude, climatic conditions). The strong economic, social and cultural internal relations are typical for these regions. Furthermore, the existence of one or more centers and gravity areas is also typical for the natural region. Particularly, the gravity areas are defined by the extent of commuting to the work but usually for more complex definition we work with the extent of commuting to the places where services or, recreation activities are located.

An *administrative region* is an area for the public administration control system which, however, often differs from the natural social-geographical region.

By *regional differences* we mean the distances between regions in an abstract metric space. This space may be described either by a single descriptor of a region or by the whole set of descriptors, both static and dynamic.

2.2 Information sources

In order to get the needed information about the regions, the "Statute Roll of the Czech Statistical Office from April 27th 1999 with the effect from January 1st 2000" (see the Collection of laws 1999, Part 33) has been accepted as the so-called CZ-NUTS classification (*La Nomenclature des Unités Territoriales Statistiques*). This classification and the administrative demarcation partially overlap each other and the individual levels are defined as follows:

- NUTS 5 – municipal areas,
- NUTS 4 - districts (76 districts + Prague),
- NUTS 3 – higher administratively independent territorial units – regions
  (13 regions + Prague),
- NUTS 2 – cohesions regions (8 units),
- NUTS 1 – the territory of the Czech Republic,
- NUTS 0 – the state of the Czech Republic.

Regarding the data sources when carrying out the social-economic analyses and the classifications of regions, it is useful to emphasize that for some problems a finer classification according to the basic residential units (particularly the urban districts) can be used; for the details see The Register of the added districts (RSO).

The basic information sources used for regional classifications:

- Surveys carried out by the Czech Statistical Office (ČSÚ) or the statistical services of the section ministries or, possibly, by other commissioned subjects, e.g.:
  - General Headquarters of customs charges,
  - Trade Office,
  - Financial Office,
  - Land registry Office,
  - The Czech Hydro-meteorological Institution,
  - The Czech Ecological Office,
  - River-basin Control etc.
- LANDSAT – Satellite / Remote Sensing data documenting the Land Use Change

- Former surveys that have already been carried out within solving the projects, e.g. environment and the attitudes toward the local policy, regional differences in market prices on the accommodation market, the European research of values etc. More detailed information on the above-mentioned surveys can be found in the sociological data archive which is administered by the Sociological Institution of the Czech Academy of Science (the web page is: http://archiv.soc.cas.cz/czech/index.html).
- One's own survey at companies, experts and local inhabitants

Some data are processed within the Regional information systems (RIS, http://www.risy.cz/), or pertinently in the advanced Integrated regional information system (IRIS, http://www.iriscrr.cz/) (see Ježek, 2001). It mainly concerns the data coming from the regional statistics basic database of the Czech Statistical Office:

- Database **KROK** (region, district),
- Database **MOS/MIS** (urban and municipal statistics / urban information system of the burghs),
- **RES** – the **R**egistry of **E**conomic **S**ubjects.

2.3 Regional data and their classification

If we deal with a regional data analysis, the first problem we immediately come across is grounded in defining the structure of the data matrix since there the cross-sectional and time data get combined here. Here it is necessary to define the purpose of the basic analysis, to clear and structure the data. The use and splitting of the data matrix according to either time or a region is extremely useful for making the work with the regional data easier (Split File).

When dealing with the regions classification, the basic problem is the insufficiency of data at the required level. Therefore it is often necessary to use the methods of data regionalization on the higher-level. The aggregated indicators are most often recalculated according to their value of the population on the lower-level. In the case of the economic indicators the employment structure is used for the recalculation, e.g. during regionalization of GDP productivity in the sectors and unemployment is recalculated and expressed *per an employee* and, subsequently, aggregated according to the employment in the region. In order to obtain a more accurate estimate, these values can be weighted by the weights being the productivity in the individual regions. The wage, more specifically the wage difference is used to estimate the productivity in the region.

The basic problem concerning the regional data quality resides in spatial validity and reliability, namely in case of smaller regions (i.e. at the level of NUTS 3).

The problem with the spatial validity resides firstly in overlapping of the local activities beyond the border of the region, namely in the form of commuting to work and places offering services and also in business activities. Therefore it is necessary to survey the methodology of collecting and coding the data. For example, in companies there are two approaches of the wage costs surveillance: the corporate method and the workplace method. In the former, the wage resources are classified with regard to the residence of a corporation and in the latter, they are classified with regard to the location of a workplace (the Labor Report 3-01). If the average monthly wage is being calculated using these values, then the difference e.g. in the district of Karviná was approx. CZK 1,000 in 2001. The reason for this fact is that the residence of OKD, a.s. company is situated in the district of Ostrava. Unfortunately, the Register has been invalidated by this fact.

The problem of data spatial reliability exists namely in the case of small regions (at the level of NUTS5). The time series of these regions manifest high variability and time predictions are difficult. Therefore it is necessary to stabilize these time series either by the moving averages or by the methods of exponential smoothing. It is also possible to use the annual or perennial averages. Another approach to the time series stabilization resides in their spatial aggregation. This approach is, on one hand, convenient regarding the improvement in the possibility of predictability of these regions development but, on the other hand, there occurs a suppression of regional differences accompanied by the information loss due to the reduction of the degree of freedom.

The choice of the regions size for the actual analysis should respect their relative withdrawnness and the representation of a sample in the case of the argument for a given area (e.g. the number of observations depending on the number of inhabitants).

The regional indicators can be classified into a number of categories with regard to the area which they cover within the region. The simplest classification has three categories: social-demographic, economic and environmental. It is typically proceeded from this classification and it is modified with regard to the purpose of analysis. For instance, in order to observe the regional difference in the Czech Republic the following categories are used, see (MMR, 2000):

- Overall characteristics of a region
- Economic potential
- Human potential
- Technical facilities of a territory
- Environment

As the second example, there are listed the categories that are used when measuring the territorial competitiveness as defined by the LEADER methodology, see (LEADER, 1999):

- Physical resources
- Environmental conservation

- Human resources
- Cultural identity of the region
- Know-how and skills
- Governance
- Institutional capacity
- Activities and business firms
- Markets and external relations
- Economic structure
- Image and perception of region
- Social welfare

2.4 Conclusion of the section

The core problem of regional classification at the chosen spatial level is availability of data of sufficient spatial validity and reliability. This requirement applies specially to quantitative data, but it needs to be taken into consideration also in case of qualitative information. In order to enhance the data reliability, yearly averages or data smoothed by the moving averages and/or exponential smoothing can be used. The problem of validity occurs particularly at the NUTS5 level, sometimes also at the NUTS4 level. The majority of data sources in the Czech Republic cover the level of NUTS3. However, the significant suppression of the regional differences occurs at this level.

3 Methodological background for the regional classification

3.1 Classification technique for the regions

The most commonly used technique for the region classification is cluster analysis. The cluster analysis groups individuals or objects into clusters so that objects in the same cluster are more similar to one another than they are to objects outside the cluster. The aim is to maximize the homogeneity of objects within the clusters while also maximizing the heterogeneity between the clusters (Paelinck-Nijkamp, 1975, p. 170). We neglect their spatial links by including hierarchical structure of centers. Classification methods for the regions are based on cross-section analysis and they presume non-differentiated abstract space where the regions are described by several selected descriptors.

This part explains the nature and purpose of cluster analysis for classification of the regions.

Cluster analysis, like the other multivariate techniques can be viewed from the four-stage approach:

- setting of a research problem,
- selecting and analyzing suitable clustering variables,
- selecting of clustering algorithm,
- interpreting and validating the clusters.

3.2 Setting of a research problem

Setting of a research problem includes a set of objectives and selection of clustering variables. The primary goal of cluster analysis is achieved any of three objectives:

- exploratory purposes and the formation of a taxonomy – an empirically based classification of objects,
- data simplification - with a defined structure, the observations can be grouped for further analysis,

- identification of relationships – with the clusters defined and the underlying structure of the data represented in clusters.

In any application, the objectives of cluster analysis cannot be separated from the selection of variables used to be characterize the objects (regions). This selection must be done with regard to theoretical and conceptual as well as practical considerations. The research must realize the importance of including only those variables that characterize the objects being clustered and relate specifically to the objectives of the cluster analysis.

3.3 Selecting and analyzing of suitable clustering variables

With the objectives defined and variables selected, our research should address three questions in the second stage for our model: detecting outliers, selecting similarity measures, and standardizing variables. Cluster analysis is very sensitive to the inclusion of irrelevant variables but also to *outliers*. The outliers do not represent the general population or under-sample actual group(s) in the population that causes an under-representation of the groups(s) in the sample. In both cases, the outliers distort the true structure and make the derived clusters unrepresentative of the true population structure. For this reason, a preliminary screening for outliers by box-plots is always necessary.

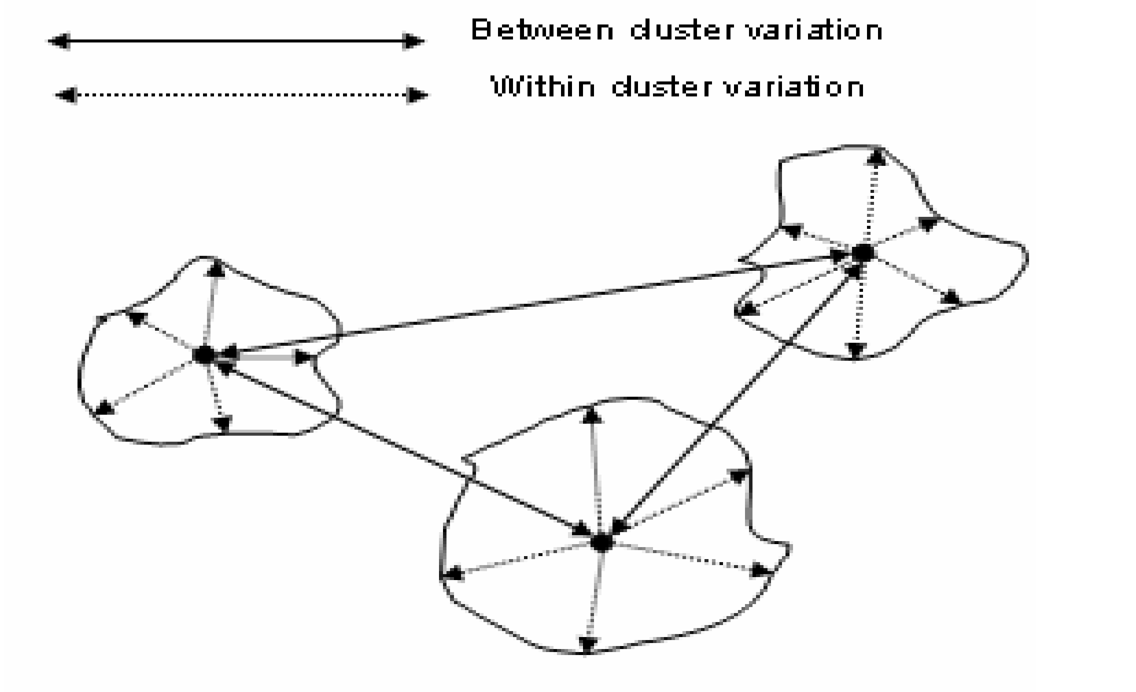
The concept of *similarity* is fundamental to cluster analysis. The inter-object similarity is a measure of correspondence, or resemblance, between objects to be clustered. Three methods dominate the applications of cluster analysis: correlation measures, distance measures (the most commonly used is the Euclidean distance), and association measures. Both the correlation and distance measures require metric data, whereas the association measures can be used for non-metric data.

Next problem is an impact of *unstandardized data values*. A problem faced by all the distance measures of non-standardized data involves the inconsistencies between cluster solutions when the scale of the variables is changed. The most common form of standardization is a conversion of each variable to standard scores (also known as *Z* scores) by subtracting the mean and dividing it by the standard deviation for each variable. This is also an option in the menu offer of the SPSS software.

3.4 Selecting of clustering algorithm

Last but one stage in our four-stage model building is selecting of a clustering algorithm, identifying clusters and assessing the overall fit. With the variables selected and the similarity matrix calculated, the partitioning process begins. We must first select a cluster algorithm for creating clusters and then make the decision on how many clusters to be identified. The essential criterion of the algorithm is that we want to maximize the differences between clusters relative to the variation within the clusters, as shown in the following figure 3.1.

Figure 3.1. Cluster diagram showing between- and within-cluster variation



The ratio of the between-cluster variation to the average within-cluster variation is then similar (but not identical) to the F ratio in analysis of variance. Usual clustering algorithms can be classified into two general categories (Hair et al. 1998, p. 493):

– hierarchical,

– non-hierarchical.

Hierarchical procedures involve construction of a hierarchy of a treelike structure using two types of hierarchical clustering procedures - agglomerative and divisive.

In *the agglomerative methods,* each object or observation starts out as a cluster. In subsequent steps, the two closest clusters are unified into a new aggregate cluster, thus reducing the number of clusters by one in each step. An important characteristic of hierarchical procedures is that the results at an earlier stage are always nested within the results at a later stage, creating a similarity to a tree. This process can be graphically described by a dendrogram. Five popular agglomerative algorithms used to develop clusters are single linkage, complete linkage, average linkage, Ward's Method, and centroid method.

The *single-linkage* procedure is based on the minimum distance. It identifies two objects separated by the first cluster. Then the next-shortest distance is found, and either a third object joins the first two to form a cluster, or a new two-member cluster is created. The process continues until all objects are unified in one cluster.

The *complete linkage* procedure is similar to single one except that the cluster criterion is based on the maximum distance.

The average linkage method starts out in the same way as that of single linkage or complete linkage, but the cluster criterion is the average distance one and unifies all individuals in one cluster. Such procedure does not depend on extreme values and also tend to be biased toward creation of the clusters with approximately the same variance.

In the *Ward's method* the distance between two clusters is the sum of squares between the two clusters summed over all variables. At each stage of the clustering procedure, the within-cluster sum of squares is minimized over all partitions obtainable by unifying two clusters from the previous stage. This procedure tends to combine clusters with a small number of observations. It is also biased toward the creation of the clusters with approximately the same number of observations.

In the *centroid method* the distance between two clusters is the distance (usually squared Euclidean) between their centroids. Cluster centroids are the mean values of the observations of the variables in the cluster. The advantage of this method is that it is less affected by outliers than the other hierarchical methods.

When the clustering process proceeds in the direction opposite to agglomerative methods, it is referred to as a *divisive method.* In this method, we begin with only one large cluster containing all the objects. In succeeding steps, the objects that are most dissimilar are split off and made into smaller clusters. This process continues until each observation is an individual cluster.

In contrast to hierarchical methods, non-hierarchical procedures (K-means clustering) do not involve the treelike construction process. Instead, they assign objects into clusters once the number of clusters to be formed is specified. There are several different approaches for selecting cluster seeds and assigning objects. There exist three main approaches for assigning individual observations to one of the clusters: sequential threshold method, parallel threshold method and optimization methods.

3.5 Interpreting and validating the clusters

The last stage in the four-stage model building approach is interpreting and validating the clusters.

The *interpretation stage* involves examining each cluster in terms of the cluster variable to name or assign a label accurately describing the nature of the clusters.

Validation of the cluster solution attempts to assure that the cluster solution is representative of the general population and thus it can be generalized to others objects and is stable over time. The most direct approach in this sense is to analyze separate samples, to compare the clusters and to asses some correspondence between the results. This approach is often impractical because of time or cost constraints or because of the unavailability of objects for multiple cluster analyses. In this case a common approach is to split the sample into two groups. Each is analyzed separately, and the results are compared.

Cluster analysis provides an empirical and objective method for performing one of the most inherent tasks of human classification. Whether for purposes of simplification, exploration or confirmation, cluster analysis is a potent analytical tool that has a wide range of applications. But with this technique comes a responsibility on the researcher to apply the underlying principles appropriately.

4 Case studies for the regional classification

4.1 Clusters of regions in the Czech Republic according to the regional disparities

4.1.1 Introduction

The regional economic models developed in the sixties of the twentieth century might be characterized as a search for systematic and quantitative representations of spatial economic system. Much emphasis was placed on the definition and specification of the components and interactions among the systems. From the middle of the 1970s onward, efforts have been made to design integrated spatial economic models that are suitable for evaluating actual regional trends by means of the whole spectrum of regional objectives or side-conditions.

Some of these models are multidisciplinary or even interdisciplinary in nature, incorporating also demographic, environmental and social variables. They also help us to focus our attention on a multiregional system rather than on a single-region one. In the third generation, the regional economic models have a clearer multiregional orientation than they had before because of the two principal reasons, see (Bolton, 1980), (Glickman, 1982). From a theoretical point of view, the design of a single-region model ignores the links between the region in question and the other regions. From a viewpoint of policy, multiregional economic models appear to respond needs of the decision-maker better than the single-region ones.

This study deals with the evaluation of 14 regions in the Czech Republic in the years 1997, 2000 and 2002 on the basis of 18 socio-economic indicators. The result of the factor analysis is the extraction of two rotated factors. We classify these regions by the help of hierarchical cluster analysis, particularly five significant clusters that slightly change in the time dimension have been identified. The methodology used in this work applied multivariate statistical methods, namely factor and cluster analysis.

Similar studies, however, for different variables, regions and years, have been conducted for the US and the UK by (Ozimek, 1993) and (Openshaw, 1995), for Portugal (Soares et al., 2002). The last mentioned paper is concerned with 275 Portugal municipalities analyzed by 33 local socio-economic indicators concerning demographic, economic, health, education, employment and cultural characteristics reported to the year 1995. The results lead to the identification of nine axes of socio-economic factors and separating of the Portuguese territory into four groups with different degrees of development.

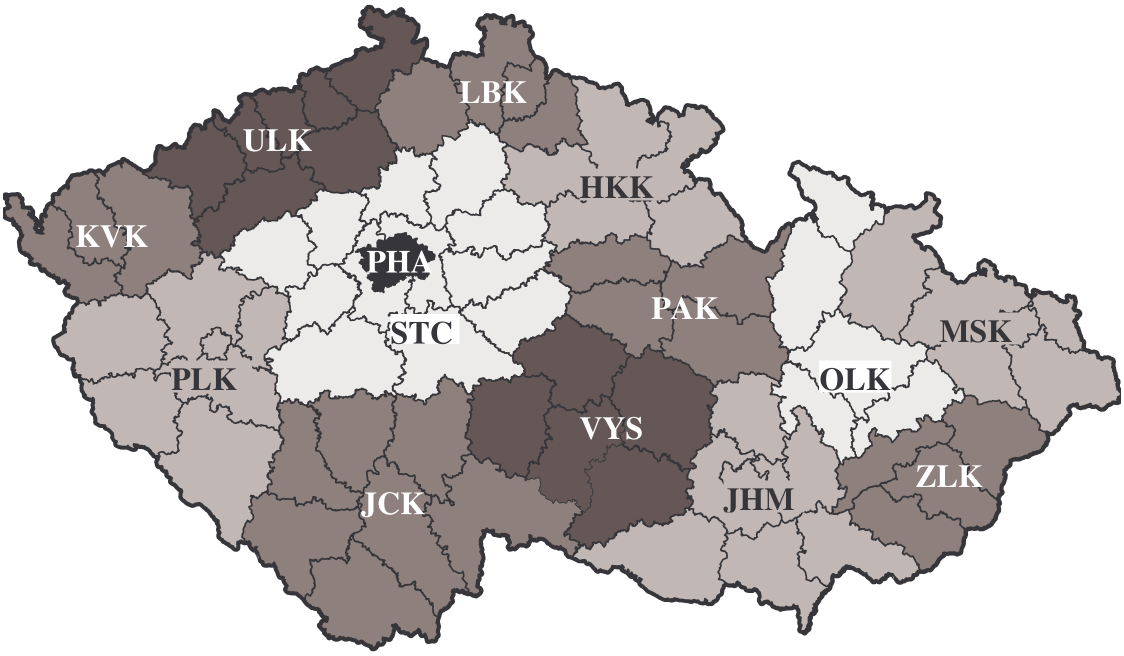
The structure of this contribution is adapted to the above mentioned purpose. The first part deals with the socio-economic regional indicators, the second part deals with some application of the factor analysis for extracting latent factors. The most important is the third part where the mentioned regions in the Czech Republic are classified into 5 groups on the basis of disparities in 1997, 2000 and 2002. In the final evaluation positive as well as negative results are summarized and possibilities of further focus of research or empirical recommendations for performance of regional economy are presented.

4.1.2 Data – important regional socio-economic indicators

The investigated territory - the Czech Republic was divided into *14 higher territorial self-governing units* with the effect from January 1st, 2000 (see Fig. 4.1). The following regions are the objects of the regional policy: the capital Prague (PHA), the region of Central Bohemia (STC), the region of Southern Bohemia (JHC), the Pilsen region (PLK), the Karlovarský region (KVK), the Ústecký region (ULK), the Liberecký region (LBK), the Královéhradecký region (HKK), the Pardubický region (PAK), the Vysočina region (VYS), the region of Southern Moravia (JHM), the Olomoucký region (OLK), the Zlínský region (ZLK), the Moravia-Silesia region (MSK).

The variables used in this work consist of 18 regional socio-economic indicators published by the Czech Statistical Office (see www.czo.cz) in the year 1997, 2000 and 2002. Their code, description and type can be found in Table 4.1. It is important to notice that there exists a greater number of indicators, however, many of them are unavailable or with some methodological changes during the time period.

Figure 4.1. NUTS3 Regions of the Czech Republic



We consider five types of indicators – demographic, economic, employment, environment and health indicators. In the next part of this study we give reasons for reduction of 18 variables into 8 indicators. The descriptive statistics for these 8 selected regional indicators are depicted in Table 4.2, they reflect some huge differences in the 14 regions in 1997, 2000 and 2002. For example, the average monthly gross wages increase from mean value 10 316 CZK in 1997 up to 14 508 CZK in 2002. Figure 4.2 provides some evidence of the growing deviation through the 14 regions in the monitored years, i.e. it presents raising regional disparities (1 147 CZK in 1997, 1673 CZK in 2002). We can also indicate some outliers for the capital city of Prague and with the very high level of wages through the monitored years.

Table 4.1. Description of variables and respective codes

	Code	Descriptor	Type of indicator
1	**OLD-AGE**	Age index = the number of citizen who are over 64 years of age per 100 children at age of 0 through 14)	**DEMOGRAPHIC**
2	**NATURAL**	Natural increase/decrease per 1 000 inhabitants	
3	**MIGRATION**	Net migration increase/decrease per 1 000 inhabitants	
4	**BALANCE**	Balance of migration per 1 000 inhabitants Total increase per 1 000 inhabitants	
5	**URBAN**	Proportion of urban citizens %	
6	**BIRTHS**	Live births per 1 000 inhabitants	
7	**DEATHS**	Deaths per 1 000 inhabitants	
8	**WAGE**	Average monthly gross wage	**ECONOMIC**
9	**WAGE-IND**	Average monthly gross wage in industry	
10	**GDP**	Share of the region in GDP of the CR, % (CR = 100)	
11	**INCOME**	Disposable income per inhabitants	
12	**UNEMPLOY-MENT**	Registered unemployment rate	**EMPLOY-MENT**
13	**APPLICANTS-VACANCIES**	Registered job applicants/ Vacancies	
14	**S_DWEL-LINGS**	Started dwellings per 10 000 inhabitants	**ENVIRONMENT**
15	**C_DWEL-LINGS**	Completed dwellings per 10 000 inhabitants	
16	**CAR**	Registered personal motor cars per 1000 inhabitants	
17	**CRIMES**	Reported crimes per 1000 inhabitants Offences per 1000 inhabitants	
18	**INCAPACI-TY**	Average percentage of incapacity for work	**HEALTH**

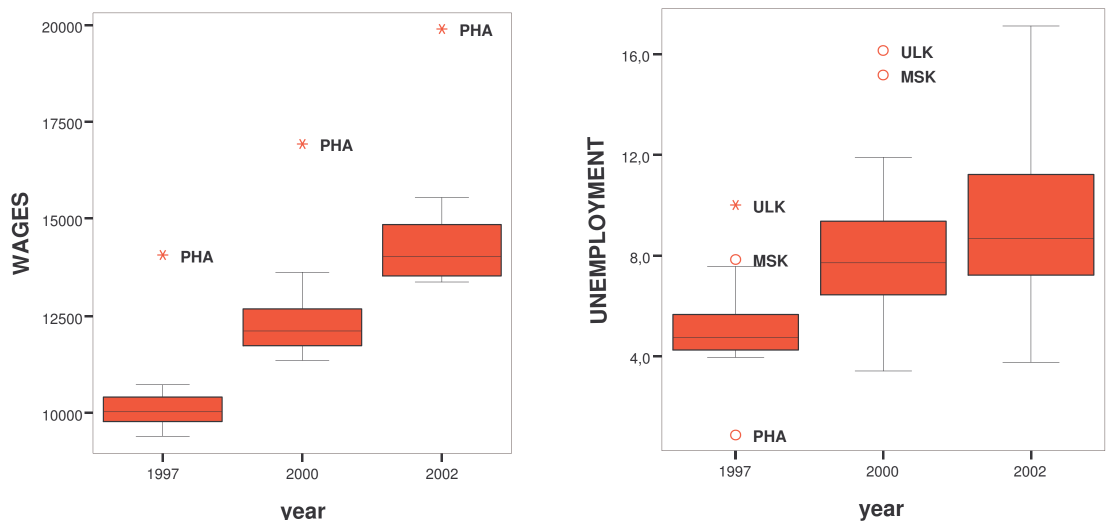
Further interesting indicator is form a regional labor market - registered unemployment rate. Mean values of this indicator increase again, particularly there is an evident shock between 1997/2000 from mean value 5,2% up to 8,5%. There are increasing regional differences at the same time (measured by variance). The Ústecký and Moravia-Silesia regions are located among the regions with very high level of the unemployment rate (outliers). On the opposite side there are another outliers – Prague with the value of registered unemployment of 0,9% (in 1997).

Analogically we can analyze also the other indicators.

Table 4.2. Descriptive statistics for 8 selected socio-economic regional indicators

	year	Minimum	Maximum	Mean	Std. Deviation
OLD_AGE	1997	62,9	110,9	76,929	11,8458
	2000	70,6	120,6	84,395	12,5442
	2002	74,4	123,4	88,151	12,3138
NATURAL	1997	-4,2	-,5	-1,943	1,0610
	2000	-3,4	-,4	-1,562	,7706
	2002	-3,1	,3	-1,320	,7619
WAGES	1997	9383	14073	10315,79	1146,465
	2000	11346	16923	12515,93	1397,357
	2002	13373	19897	14508,29	1672,667
GDP	1997	2,7	21,5	7,129	4,8789
	2000	2,4	24,8	7,143	5,6213
	2002	2,3	24,9	7,143	5,6614
UNEMPLOYMENT	1997	,9	10,0	5,241	2,1408
	2000	3,4	16,1	8,489	3,5887
	2002	3,7	17,1	9,595	3,6172
CAR	1997	265,0	444,0	325,630	48,4585
	2000	270,3	452,0	331,288	48,8917
	2002	287,0	479,9	353,389	51,2368
CRIMES	1997	18,2	86,7	35,004	16,7175
	2000	17,9	90,4	34,566	17,4047
	2002	17,3	88,8	33,359	17,3291
INCAPACITY	1997	4,8	7,4	6,366	,6710
	2000	5,0	7,8	6,617	,7252
	2002	5,3	8,0	6,984	,6937

Figure 4.2. Boxplots for wage and unemployment variables



4.1.3 Identification of the underlying socio-economic factors

Factor analysis is used to look for a small number of socio-economic dimensions that adequately summarize information contained in the original set of variables. This analysis is a class of multivariate statistical methods aimed at investigating dimensions or constructs assumed to underline a set of interdependent variables (Hair et al., 1998). Exploratory analysis summarizes information contained in the set of original variables into a smaller group of factors - linear combinations of sets of original highly correlated variables.

Factor analysis involves 4 steps (Norusis, 1990):

Firstly, based on the correlation matrix for all variables, the appropriateness of the factor model has to be evaluated.

Secondly, it is necessary to decide which factor model should be used, the number of factors that should be extracted, and to assess how well the model fits the original data.

Thirdly, the choice of the rotation method to make factors more interpretable needs to be made.

Finally, the computed factor scores can be used in other statistical analyses – regional classification. This methodology has been applied in the present research in this part.

Evaluating the appropriateness of factor analysis means whether the input regional variables are significantly and sufficiently correlated with each other so that their number can be reduced by applying the factor analytical model. This can be done:

- with a visual inspection of the correlation matrix for all variables,
- by computing some statistics, including the Bartlett test of sphericity and the Keiser-Meyer-Olkin measure of sampling adequacy,
- by a visual inspection of the inverse correlation matrix.

We provide the above mentioned investigation with results for 8 selected indicators through years 1997, 2000 and 2002.

The correlation matrices can be found in Table 4.3. All but two variables have significantly at least one correlation coefficient with an absolute value larger than 0.3, the value that is usually suggested as the minimum value for including the variable in consideration. The following exceptions occur: unemployment-GDP (signif. 0. 136) in the year 1997; unemployment rate – natural, wages, GDP (signif. 0.098, 0.130, 0.249) in 2000; and unemployment rate-natural, wages, GDP, crimes (signif. 0.164, 0.069, 0.168, 0.126) in 2002.

The Bartlett test of sphericity is a statistical test for presence of correlations among the variables. The significance level gives the result of the test. Very small values (less than 0.05) indicate that there are probably significant relationships among variables. Values higher than 0.1 may indicate that our data are not suitable for factor analysis.

The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy is a statistic which indicates the proportion of variance in our variables which might be caused by underlying factors. High values (close to 1.0) generally indicate that a FA may be useful with our data. If the value is less than 0.5, the results of factor analysis probably won't be very useful.

Table 4.3. Correlation matrices for 8 selected variables

Correlation Matrix[b]

Correlation

	Z_OLD_AGE	Z_NATURAL	Z_WAGES	Z_GDP	Z_UNEMPLOYMENT	Z_CAR	Z_CRIMES	Z_INCAPACITY
Z_OLD_AGE	1,000	-,881	,779	,723	-,718	,785	,676	-,656
Z_NATURAL	-,881	1,000	-,666	-,615	,536	-,688	-,557	,530
Z_WAGES	,779	-,666	1,000	,928	-,492	,714	,939	-,642
Z_GDP	,723	-,615	,928	1,000	**-,315**	,518	,845	-,450
Z_UNEMPLOYMENT	-,718	,536	-,492	**-,315**	1,000	-,733	-,392	,553
Z_CAR	,785	-,688	,714	,518	-,733	1,000	,684	-,841
Z_CRIMES	,676	-,557	,939	,845	-,392	,684	1,000	-,669
Z_INCAPACITY	-,656	,530	-,642	-,450	,553	-,841	-,669	1,000

b. year = 1997, (-,315 sign. 0,136)

Correlation Matrix[b]

Correlation

	Z_OLD_AGE	Z_NATURAL	Z_WAGES	Z_GDP	Z_UNEMPLOYMENT	Z_CAR	Z_CRIMES	Z_INCAPACITY
Z_OLD_AGE	1,000	-,892	,800	,800	-,622	,798	,720	-,634
Z_NATURAL	-,892	1,000	-,758	-,793	,367	-,608	-,650	,483
Z_WAGES	,800	-,758	1,000	,938	-,323	,766	,949	-,656
Z_GDP	,800	-,793	,938	1,000	-,198	,615	,894	-,530
Z_UNEMPLOYMENT	-,622	**,367**	**-,323**	**-,198**	1,000	-,721	-,280	,521
Z_CAR	,798	-,608	,766	,615	-,721	1,000	,715	-,846
Z_CRIMES	,720	-,650	,949	,894	**-,280**	,715	1,000	-,686
Z_INCAPACITY	-,634	,483	-,656	-,530	,521	-,846	-,686	1,000

b. year = 2000

Correlation Matrix[b]

Correlation

	Z_OLD_AGE	Z_NATURAL	Z_WAGES	Z_GDP	Z_UNEMPLOYMENT	Z_CAR	Z_CRIMES	Z_INCAPACITY
Z_OLD_AGE	1,000	-,833	,794	,794	-,687	,752	,667	-,647
Z_NATURAL	-,833	1,000	-,769	-,805	,283	-,565	-,607	,483
Z_WAGES	,794	-,769	1,000	,947	-,417	,778	,926	-,707
Z_GDP	,794	-,805	,947	1,000	-,277	,628	,868	-,620
Z_UNEMPLOYMENT	-,687	**,283**	**-,417**	**-,277**	1,000	-,775	-,329	,555
Z_CAR	,752	-,565	,778	,628	-,775	1,000	,713	-,818
Z_CRIMES	,667	-,607	,926	,868	**-,329**	,713	1,000	-,706
Z_INCAPACITY	-,647	,483	-,707	-,620	,555	-,818	-,706	1,000

b. year = 2002

Table 4.4 summarizes results for both the Bartlett test and KMO Measure of Sampling Adequacy test in 1997, 2000 and 2002.

Table 4.4 Results for both the Bartlett and the KMO tests

		1997	2000	2002
KMO Measure of Sampling Adequacy		0,751	0,765	0,681
Bartlett test of sphericity	Chi-square	105,99	117,87	121,65
	df	28	28	28
	Sign.	0,000	0,000	0,000

The Barlett test with a respective value of 106; 118 and 122 and associated probability less than 0.05 in the selected years 1997, 2000 and 2002 suggests that the data structure is adequate to be subjected to factor analysis.

There are also KMO Measures of Sampling Adequacy with the values of 0.751; 0.765 and 0.681, which are also acceptable.

The last approach consists of investigation of the inverse correlation matrices. The inverse matrices contain the partial covariances and correlations. They can give an indication of correlations which aren't due to the common factors. Small values indicate that our variables are relatively free of unrepaired correlations. Most or all values off the diagonal should be small (close to 0). Each value on the diagonal of the anti-image correlation matrix shows the Measure of Sampling Adequacy for the respective item. Values less than 0.5 may indicate variables do not seem to fit with the structure of the other variables. Consider dropping such variables from our analysis. The Table 4.5 includes the inverse correlation matrices for our 8 selected socio-economic indicators. The results from this table suggest that the unemployment rate variable should be omitted, particularly in the year 2002 because the lowest Measure of Sampling Adequacy (MSA) for this variable is 0.422. All other MSA values belong to the intervals 0.612-0.872 (in 1997); 0.683-0.805 (in 2000) and 0.422 – 0.861 (in 2002).

Table 4.5. Inverse correlation matrices for 8 selected variables

Anti-image Matrices[b]

Anti-image Correlation

	Z_OLD_AGE	Z_NATURAL	Z_WAGES	Z_GDP	Z_UNEMPLOYMENT	Z_CAR	Z_CRIMES	Z_INCAPACITY
Z_OLD_AGE	**,730[a]**	,761	,284	-,530	,650	-,019	-,076	,292
Z_NATURAL	,761	**,726[a]**	,260	-,282	,480	,222	-,205	,254
Z_WAGES	,284	,260	**,732[a]**	-,786	,431	-,177	-,675	,108
Z_GDP	-,530	-,282	-,786	**,688[a]**	-,500	,211	,182	-,274
Z_UNEMPLOYMENT	,650	,480	,431	-,500	**,612[a]**	,322	-,304	,265
Z_CAR	-,019	,222	-,177	,211	,322	**,872[a]**	-,070	,541
Z_CRIMES	-,076	-,205	-,675	,182	-,304	-,070	**,838[a]**	,168
Z_INCAPACITY	,292	,254	,108	-,274	,265	,541	,168	**,816[a]**

a. Measures of Sampling Adequacy(MSA)

b. year = 1997

Anti-image Matrices[b]

Anti-image Correlation

	Z_OLD_AGE	Z_NATURAL	Z_WAGES	Z_GDP	Z_UNEMPLOYMENT	Z_CAR	Z_CRIMES	Z_INCAPACITY
Z_OLD_AGE	**,750[a]**	,741	,295	-,540	,564	-,298	-,060	-,052
Z_NATURAL	,741	**,773[a]**	,338	-,134	,286	-,252	-,300	-,115
Z_WAGES	,295	,338	**,741[a]**	-,574	-,065	-,620	-,638	-,341
Z_GDP	-,540	-,134	-,574	**,805[a]**	-,336	,345	-,117	-2,521E-05
Z_UNEMPLOYMENT	,564	,286	-,065	-,336	**,683[a]**	,387	,078	,202
Z_CAR	-,298	-,252	-,620	,345	,387	**,740[a]**	,278	,653
Z_CRIMES	-,060	-,300	-,638	-,117	,078	,278	**,824[a]**	,443
Z_INCAPACITY	-,052	-,115	-,341	,000	,202	,653	,443	**,779[a]**

a. Measures of Sampling Adequacy(MSA)

b. year = 2000

Anti-image Matrices[b]

Anti-image Correlation

	Z_OLD_AGE	Z_NATURAL	Z_WAGES	Z_GDP	Z_UNEMPLOYMENT	Z_CAR	Z_CRIMES	Z_INCAPACITY
Z_OLD_AGE	**,615[a]**	,855	,388	-,591	,903	,449	-,343	,215
Z_NATURAL	,855	**,573[a]**	,400	-,337	,805	,480	-,499	,227
Z_WAGES	,388	,400	**,749[a]**	-,757	,326	-,221	-,609	-,118
Z_GDP	-,591	-,337	-,757	**,741[a]**	-,502	,065	,168	,046
Z_UNEMPLOYMENT	,903	,805	,326	-,502	**,422[a]**	,671	-,407	,238
Z_CAR	,449	,480	-,221	,065	,671	**,741[a]**	-,199	,550
Z_CRIMES	-,343	-,499	-,609	,168	-,407	-,199	**,782[a]**	,112
Z_INCAPACITY	,215	,227	-,118	,046	,238	,550	,112	**,861[a]**

a. Measures of Sampling Adequacy(MSA)

b. year = 2002

Evaluating the appropriateness of factor analysis using above mentioned approaches can be summarized: the data structure for 8 selected regional indicators is adequate to be subjected to factor analysis for years 1997, 2000 and 2002.

The second stage of factor analysis is a selection of a suitable method for factor extraction and making decision on the number of factors to be extracted.

We extracted our factors using principal component analysis to check for stability and robustness of the solution in accordance with (Kline, 1994).

Usually, the number of factors to extract should be decided by four criteria:

- eigenvalue criterion,
- percentage of variance criterion,
- scree test criterion, and finally,
- interpretability of the factor structure.

By the *eigenvalue criterion* we consider that all factors with eigenvalues greater than 1 should be retained. The reason for using the eigenvalue criterion is that any factor should explain at least the variance of a single variable. From Table 4.6, we can see that 2 factors should be retained for 3 selected years.

By the *percentage of variance criterion* one should extract all factors that explain at least 60% (approximately) of the variance of the original variables. Although no absolute cut-off point has been adopted for all data, this figure is normally accepted as satisfactory in the Social Sciences. From Table 4.6, the minimum number of factors that should be retained is also 2 for each year because we can meet the necessary threshold of variance criterion with the respective values of 84%; 86% and 86% through our years.

The *scree test criterion* is given by the point at which the curve begins to become horizontal. In Figure 4.3 for the first year 1997, by this criterion we select 2 factors. You can see also results for the years 2000 and 2002, there.

Table 4.6. Results of principal component analysis

Total Variance [a]

Component	Extraction Sums of Squared			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	**5,671**	70,887	**70,887**	3,366	42,073	42,073
2	**1,051**	13,133	**84,019**	3,356	41,946	84,019

Extraction Method: Principal Component

a. year = 1997

Total Variance [a]

Compone	Extraction Sums of Squared			Rotation Sums of Squared		
	Total	% of Varianc	Cumulative	Total	% of Varianc	Cumulative
1	**5,732**	71,64	**71,649**	4,302	53,77	53,77
2	**1,173**	14,66	**86,316**	2,603	32,54	86,31

Extraction Method: Principal Component

a. year = 2000

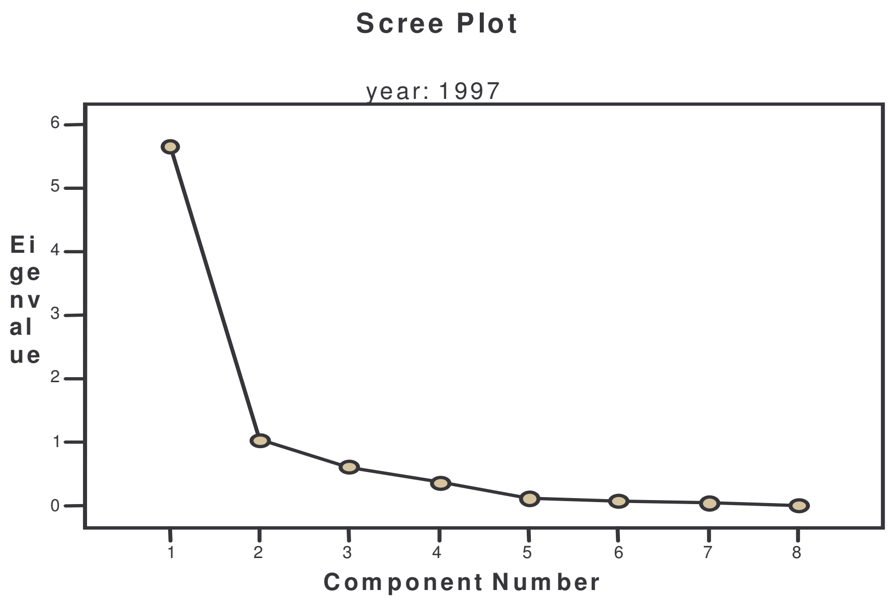
Total Variance [a]

Compone	Extraction Sums of Squared			Rotation Sums of Squared		
	Total	% of Variance	Cumulative	Total	% of Variance	Cumulative
1	**5,756**	71,951	**71,951**	4,187	52,332	52,332
2	**1,122**	14,023	**85,974**	2,691	33,641	85,974

Extraction Method: Principal Component

a. year = 2002

Figure 4.3. Scree plot for principal components analysis in 1997



After 2 candidate factors have been determined through our criteria, the final number of factors has to pass the *interpredicability test.* In practice, the ability to interpret and assign some meaning to the factors, acts as an extremely important criterion in determining the final number of factors to extract.

We used the Varimax rotation to provide a more interpretable structure of factors. The Varimax rotation, which imposes an orthogonal structure on the data, should always be used when the resulting factor scores are to be analyzed by the other statistical procedures. The rotated factor matrices are shown in Table 4.7. We can assess the quality of the resulting factor matrices and then naming the factors. In an ideal solution each variable would only load in one factor with a score of 1, and would not load any other factors. However, in practice, factor loadings greater than 0.3 are considered significant; whereas loadings greater than 0.5 are considered very significant (Hair et al., 1998). The results in Table 4.7 reveal that all variables have at least one factor loading greater than 0.646 (in absolute value -0.646, -0.685, -0.665), whereas the great majority of the variables have very high loadings of only one factor.

Table 4.7. Varimax rotated factor matrixes.

	Component 1	Component 2
Z_GDP	,948	
Z_WAGES	,892	,413
Z_CRIMES	,879	
Z_UNEMPLOYMENT		-,896
Z_CAR	,401	,851
Z_OLD_AGE	,566	,742
Z_INCAPACITY		-,731
Z_NATURAL	-,508	-,646

b. year = 1997

	Component 1	Component 2
Z_GD	,970	
Z_WAGE	,928	
Z_CRIME	,896	
Z_NATUR	-,796	
Z_OLD_A	,732	,576
Z_UNEMPLOYME		-,946
Z_CA	,538	,795
Z_INCAPACI	-,488	-,685

b. year = 2000

	Component 1	Component 2
Z_GD	,956	
Z_WAG	,903	
Z_CRIME	,850	
Z_NATUR	-,844	
Z_OLD_A	,695	,592
Z_UNEMPLOYM		-,962
Z_CA	,505	,812
Z_INCAPACI	-,527	-,665

b. year = 2002

A good factor solution should also account for between 50% and 70% of the amount of variance of each individual variable. Our 2 – factor structure found explains between 65-95%, 69-97%, 67-96% of the variance of each variable. This highlights the very good quality of the results of the factor analytic model.

The last stage of the factor analysis is naming the factors.

Factors interpretation for *the year 1997*:

The first factor presents ECONOMIC EFFICIENCY within a market environment (it has high positive loadings on the GDP, wages and reported crimes).

The second factor was labeled LABOR MARKET development and has negative high loadings on unemployment rate. We also noticed the significant positive loadings on the registered personal cars, demographic indicators (positive loading for the age index and negative for the natural decrease) and also the significant negative loadings for the average percentage of incapacity for work.

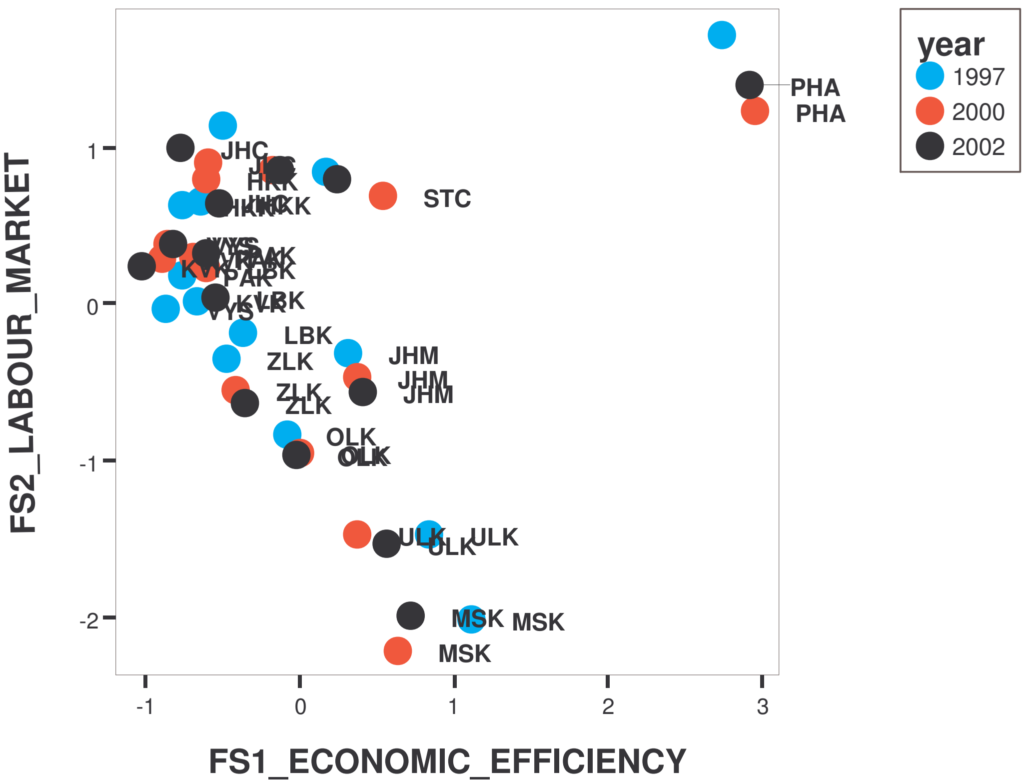
Factors interpretation for *the years 2000 and 2002*:

The first factor is also economic efficiency including influence of demographic mobility.

The second factor represents a labor market within impacting standard of living.

Further we can illustrate a position of the region using two extracted factors in 1997, 2000 and 2002 and we can also investigate movements of the regions during these years (see Figure 4.3).

Figure 4.4. The position of the regions on the basis of 2 extracted factors and monitoring development in the years 1997, 2000 and 2002



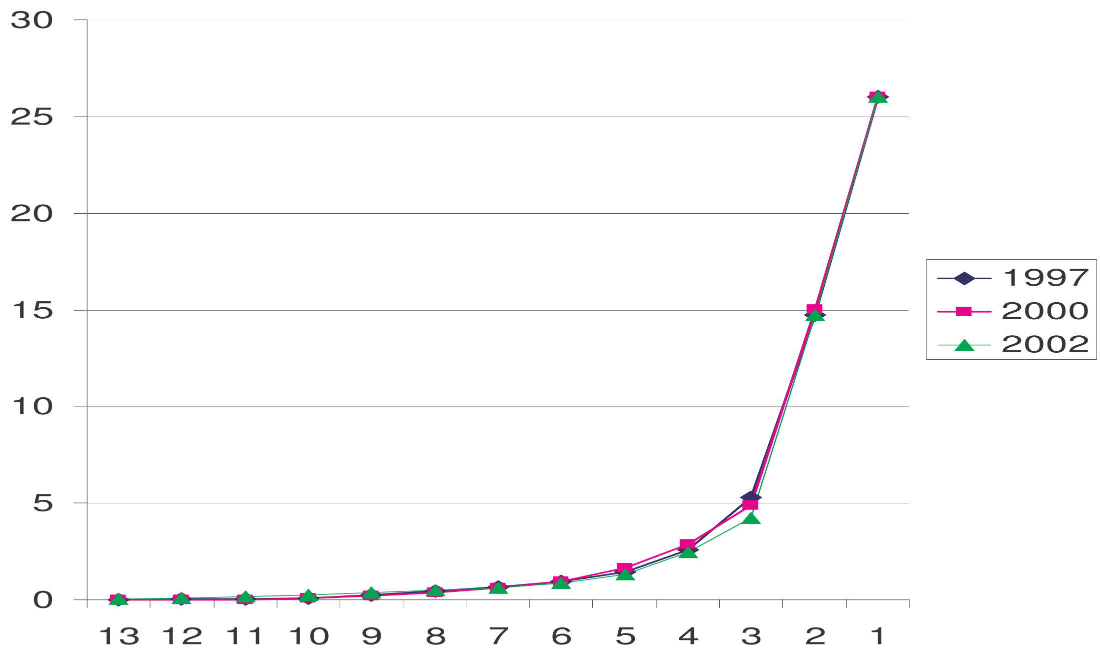
Cluster analysis can be used to look for groups of regions with similar of socio-economic development using extracted factors.

4.1.4 Classifying the regions in the Czech Republic

Cluster analysis was applied by the steps of the above mentioned methodology. The method of Cluster analysis was run by using the standardized two extracted factors. A range of hierarchical cluster procedures using various distance measures was used in the search for homogeneous groups of regions. The results from the majority of computations have suggested a 5-clusters solution, with an appropriate number of regions in each cluster.

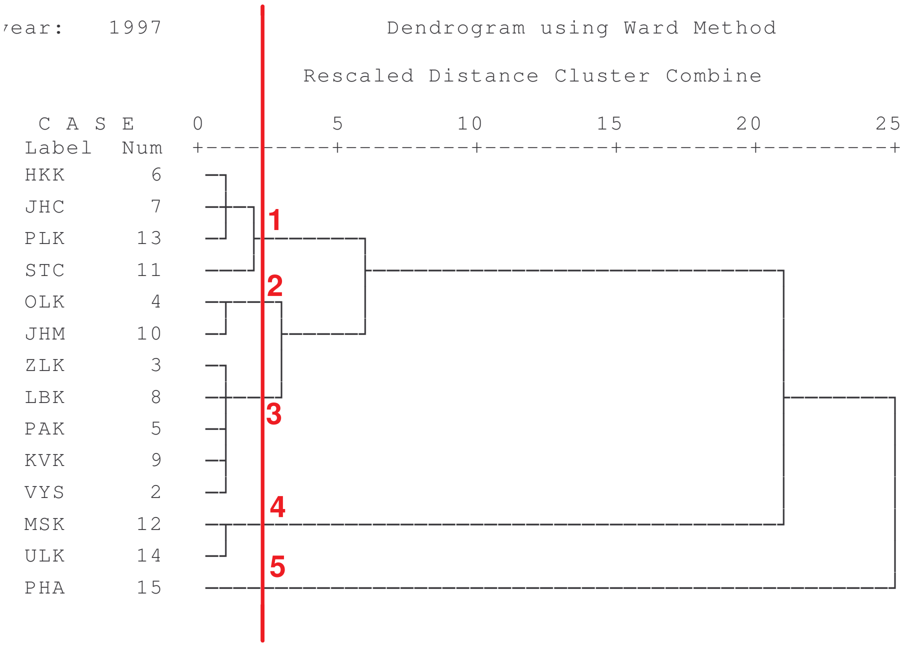
A graphical representation of this situation is shown in Figure 4.5. We selected the final cluster solution using the agglomeration coefficient. We looked for a major increase in the value, similar to the scree test in factor analysis. The clustering agglomeration coefficients show pretty large increases in going from four to five clusters at each year.

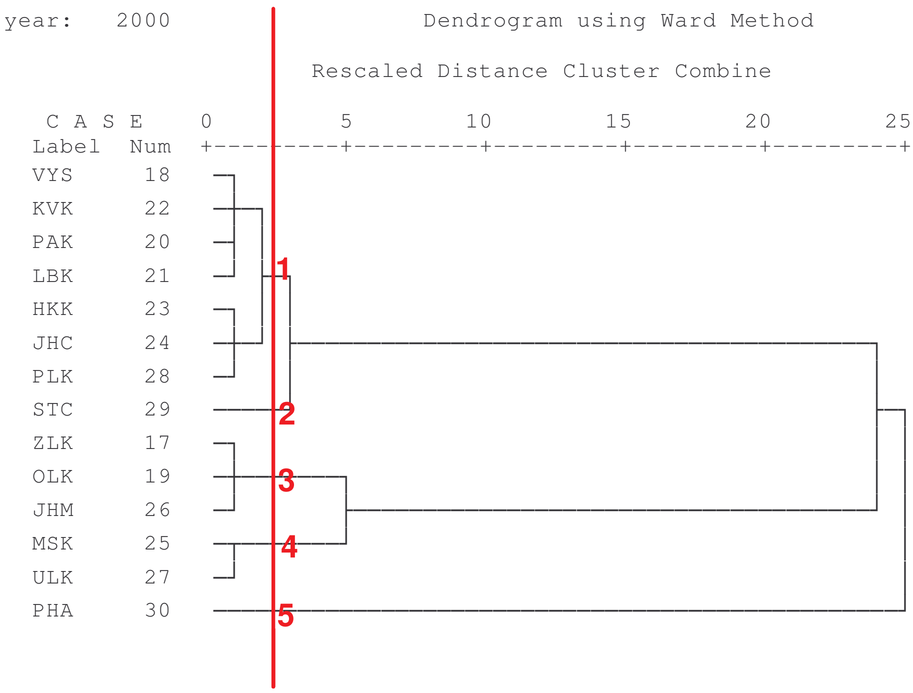
Figure 4.5 Graphical presentation of the agglomeration coefficients

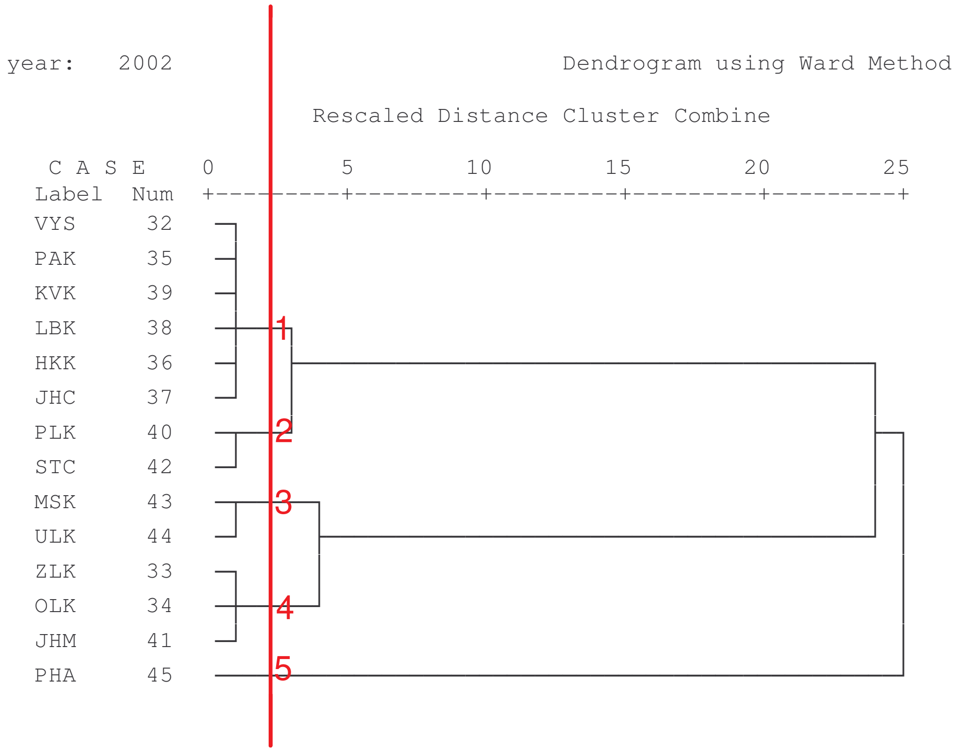


Here, we used hierarchical cluster procedure using the Euclidean distance. For developing clusters the best solution given by Ward's method was used, see Figure 4.6.

Figure 4.6. Dendograms (Ward method)







A graphical representation of the five clusters is shown in Figure 4.7.

Cluster 5, named "Prague" includes only the capital - city of Prague. The cluster is characterized by a very good economic prosperity and no problems with labor market. This cluster contains outliers – Prague.

Cluster 4 – includes MSK (Moravia-Silesia), and ULK. The regions in this cluster have relatively good economic efficiency but the biggest labor market problems with very high level of unemployment rate.

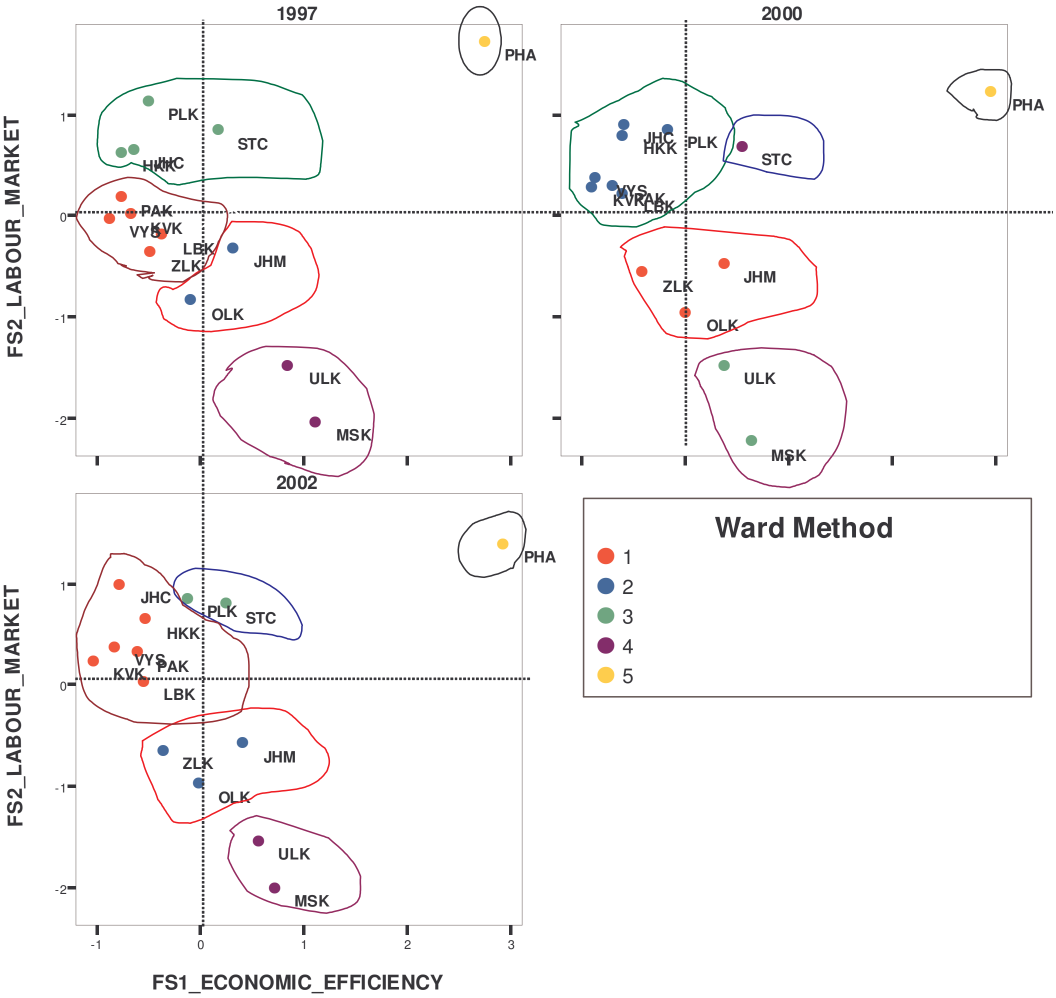
Cluster 3 – includes STC (Central Bohemia), and PLK (Pilsen region for 1997 and 2002). These regions have an average level of economic efficiency and above average situation on the labor market in comparison with the Czech Republic.

Cluster 2 - includes OLK, JHM (Southern Moravia), and ZLK in principle. This cluster represents the regions with the average level of economic efficiency (similar to Cluster 3), but below the average labor market deficiency.

Cluster 1 – consists of regions VYS, PAK, LBK, and KVK. This cluster is characterized by existing economic problems and with no serious problems on the labor market.

The other regions, particularly HKK and JHC (Southern Bohemia), are unstable during the years 1997, 2000 and 2002.

Figure 4.7. Clusters of the regions in the Czech Republic showing socio-economic levels of development



4.1.5 Conclusions of the empirical study

The first conclusion of this work is that the used multivariate techniques have been successful in allowing identify:

the main 2 axes of socio-economic characterization (the factors):

factor 1 = economic efficiency,

factor 2 = labor market development.

The 5 region clusters in the Czech Republic with various degrees of development.

The second conclusion is that the number of observations (regions) is suitable for extending the structure of a set of the available regional descriptors, in particular, from the point of view of verification of our results.

The third - final conclusion, and perhaps the most important, is that the classification scheme resulting from this research makes evident that NUTS3 regions are treated as heterogeneous form the point of view of socio-economic development and also need various approaches for the regional development policy.

4.2 Multi-period Classification of the Regions

4.2.1 Static Classification System

This type of classifications is based on the assumption that a classification system, which is being looked for, does not develop in the course of time. Therefore, when suggesting a classification system, we can drop the time dimension. However, the dimension will later be used during the consequential analysis of transition of the regions among the individual classes. The following classification will be clarified using the following two key regional descriptors:

the rate of unemployment (*iu*), and

the wage (*w*).

As to the rate of unemployment, the annual average is used and the data for wages are taken from the Labor Report 3-01. The five-period (i.e. five-year) data are used, more specifically from the year 1996 through 2000. These data coming from the individual years have been standardized using *Z*-score.

Table 4.8: Centroids of statistical classification

	Very high *w* and low *iu*	Lower *w* and low *iu*	Higher *w* and high *iu*	Lower *w* and high *iu*	Higher *w* and low *iu*
Zscore(*w*)	4,785	-,383	,731	-,425	1,476
Zscore(*iu*)	-1,679	-,601	1,893	,533	-1,127

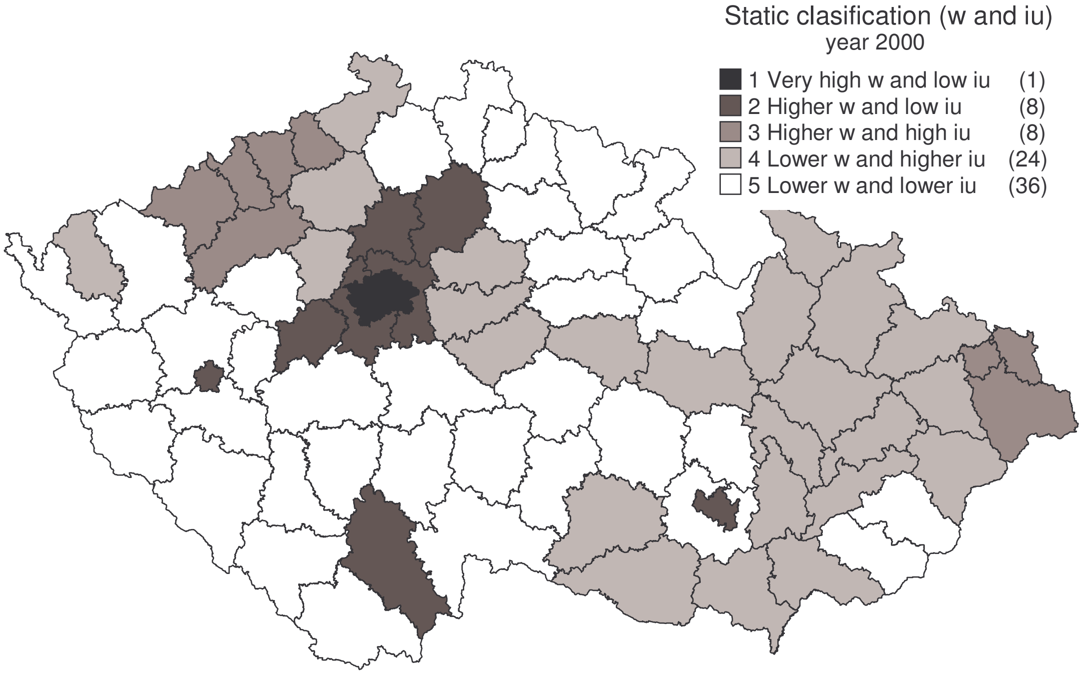
The category with a very high wage *w* and a low rate of unemployment *iu* is found only in Prague. In the period of interest the difference in the labor market character between Prague and the other districts show a sharp increase. The difference is represented especially by wage.

Into the category with higher *w* and lower *iu*, we rank the satellites of Prague - the districts of Prague – West (PZ) and of Prague – East (PY), the neighboring districts of Mělník, Beroun and naturally Mladá Boleslav. Further, there are other districts that also belong to the category: Brno, Plzeň and České Budějovice. These territories can be regarded as very prosperous with sufficient demand for labor and a qualified workforce. The favorable status of these territories is magnified by the presence of foreign capital.

The category of regions with higher *w* and high *iu* is formed by the districts situated in Ostrava-Karviná coal-mining area (OV, KI, FM), Northern Bohemia coal-mining area (MO, CV, TP, UL) and Louny. With the exception of the district of Louny these districts are typical territories with mining, metallurgical and, in case of Ostrava, mechanical industry. It concerns typically structurally-handicapped districts. Higher levels of wages are caused by the character of production activities (compensation differences, prosperity and export possibilities of companies or the power of labor unions in big corporations). They are territories with long-term problems on the labor market. The district of Louny is classified to this category because of strong spatial binds with the above-mentioned districts. More than 10% of economically active inhabitants commute to work located outside the region they live in. Particularly, they commute to Most and Chomutov district.

The category with lower *w* and low *iu* and the one with lower *w* and higher *iu* differ from each other in higher rate of unemployment manifested by Moravian districts. Both of these categories comprise the peripheral districts located near the state border and at the Bohemia-Moravia border. It concerns the district with low proportion of employment in the sector of services.

Figure 4.8: Cartogram of classification with regard to wage and rate of unemployment



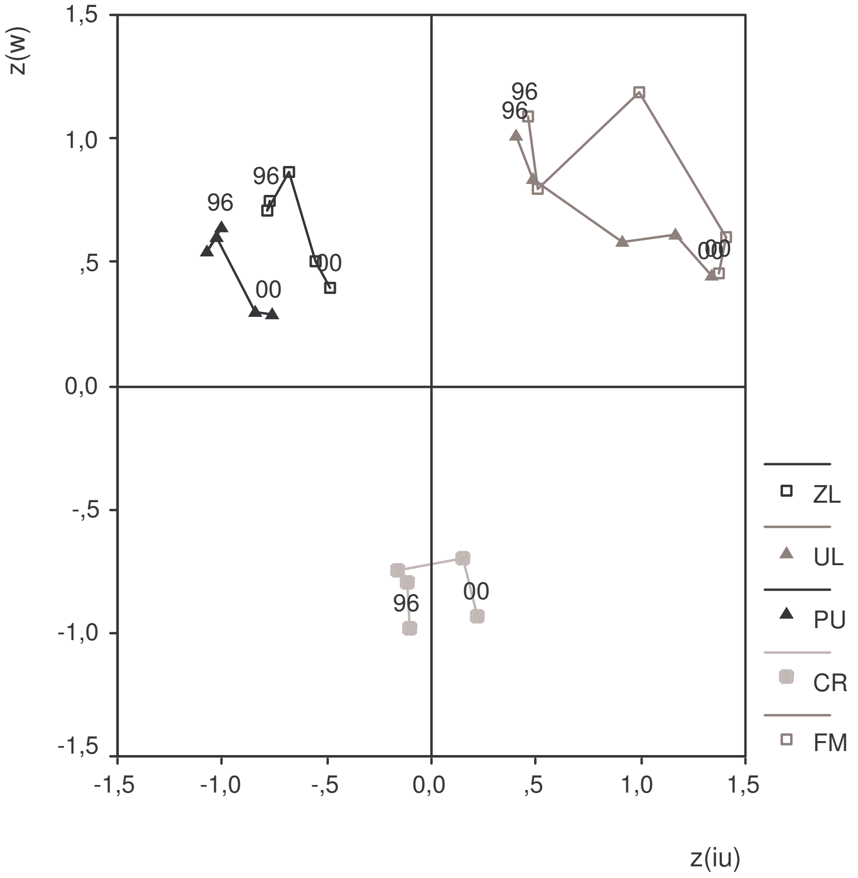
In case of this type of classification, the following analysis of regional transition between the above-defined types of regions is necessary. The majority of the districts belong to a single type of regions during the entire analyzed period. There are, however, some districts, where a change occurs. These transitions (changes) can be evaluated positively or negatively, too, depending on the relative position of a district toward the other districts of the Czech Republic. The transitions between the classification types of regions are summarized in Table 4.9.

Table 4.9: Transitions between the classification types of regions

Change	Transition			District
positive	Lower *w* and lower *iu*	→	Higher *w* and low *iu*	BE
positive	Lower *w* and higher *iu*	→	Lower *w* and lower *iu*	PT; TC; JI; LI; PB; ZR
positive	Higher *w* and high *iu*	→	Lower *w* and higher *iu*	DC, KD
neutral	Lower *w* and lower *iu*	↔	Lower *w* and higher *iu*	BK, ST, CK,
neutral	Lower *w* and higher *iu*	↔	Higher *w* and high *iu*	PR
negative	Higher *w* and low *iu*	→	Lower *w* and lower *iu*	ZL, PU
negative	Lower *w* and lower *iu*	→	Lower *w* and higher *iu*	CR
negative	Lower *w* and higher *iu*	→	Higher *w* and high *iu*	FM, UL

Further, we will focus on the five districts the development of which has been labeled "negative". The following graph shows that the districts can be found in different quadrants. That points out the different reasons for the development.

Figure 4.9: Graph of districts development with a negative transition



The districts of Frýdek-Místek and Ústí nad Labem have the same end-positions on one hand, but on the other hand, the development in these two districts ran in a different way. On contrary, the districts of Zlín and Pardubice experienced similar development in the period observed. In the above-mentioned districts the development of wages did not keep up with the rate of wage increase in the prosperous territories. Furthermore, in case of Frýdek-Místek and Ústí nad Labem these changes are accompanied by the significant increase in *iu*. Taking the district of Chrudim, the increase in *iu* was caused by the inadequate employment structure due to which the accrued unemployed people could not be absorbed during the period of recession.

What is to be pointed out here is that the changes stated in this section ought to be perceived with relativity i.e. in the context of the other Czech regions development. It is because this classification is being applied on the standardized data.

4.2.2 Dynamic Classification System

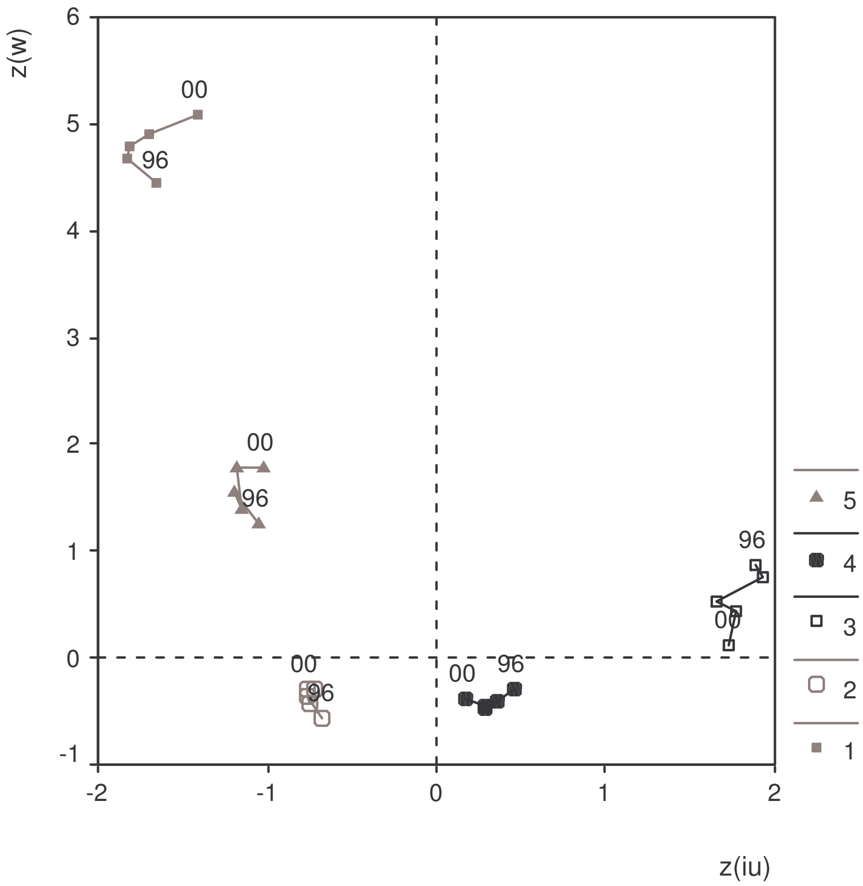
Dynamic classifications postulate that classification systems depend on time and it is possible to study the development of the classification types. The following classification system is based on this assumption. The data are the same as in the previous section. The coordinates of centroids in the individual years are summarized in Table 4.10.

Table 4.10: Centroids of regions classification types

	YEAR	**Very high *w* and low *iu***	**Lower *w* and low *iu***	**Higher *w* and high *iu***	**Lower *w* and high *iu***	**Higher *w* and low *iu***
Zscore(*w*)	1996	4,441	-0,568	0,863	-0,286	1,258
	1997	4,683	-0,355	0,755	-0,403	1,543
	1998	4,789	-0,438	0,535	-0,483	1,391
	1999	4,912	-0,291	0,447	-0,459	1,774
	2000	5,098	-0,285	0,114	-0,386	1,787
Zscore(*iu*)	1996	-1,654	-0,674	1,879	0,460	-1,057
	1997	-1,822	-0,763	1,934	0,363	-1,192
	1998	-1,816	-0,748	1,652	0,281	-1,157
	1999	-1,693	-0,724	1,775	0,289	-1,182
	2000	-1,412	-0,768	1,731	0,171	-1,028

For better understanding the development of centroids is shown in the following graph.

Figure 4.10: Development of classification types centroids



Classification types

1. *Very high w and low iu* – in this type of a region only capital Prague is included. That is why its development is identical with the development of Prague. The position of Prague with respect to the other districts is characteristic by relatively strong increase in wage in the entire period observed and since 1998 by increase in the rate of unemployment, too.

2. Lower w and lower iu – the position of the centroid of this classification type changes slightly. It concerns mostly the agricultural districts manifesting strong resistance toward the changes.

3. Higher w and high iu – the districts with a high proportion of employment in the field industry that has undergone restructuralization. The centroid position development suggests that the wages in these districts are relatively decreasing in proportion to the ones in other districts.

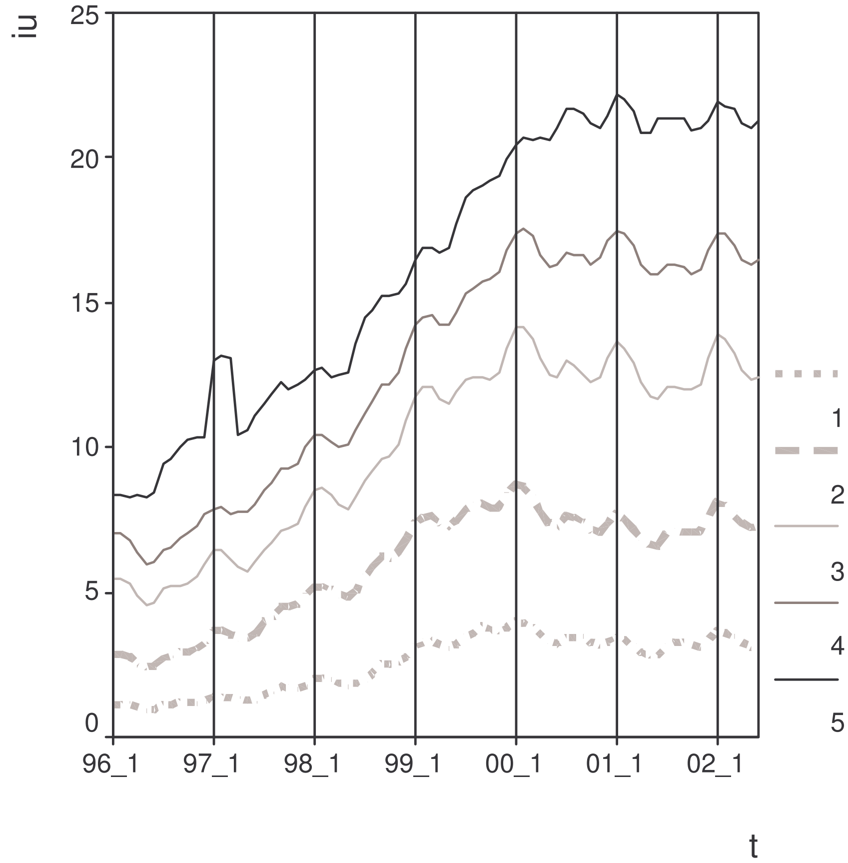
4. Lower w and higher iu – a slight relative decrease in unemployment occurs in these districts.

5. Higher w and low iu – districts with some prosperity where the growth of wages in the year 2000 did not keep up with the wage growth in Prague.

4.2.2.1 Time series clustering – absolute differences

The classification system was created on the basis of the unemployment rate monthly data in the time period from January 1996 through June 2002 using the aggregation method of average distance between groups. The differences between the regions have been measured via the Euclid distance. The typology suggested helps us to identify the problematic districts with regard to the unemployment rate. The typology consists of five regions, whose centroids of the development are shown in the following graph.

Figure 4.11: Centroids of *iu* development classification – absolute differences



Type of regions with regard to *iu* development:

Type 1 – totally non-problematic districts: Prague and satellite districts Prague-east, Prague-west, Benešov, Mladá Boleslav, Pelhřimov.

Type 2 – non-problematic districts: the largest group formed by 52 districts.

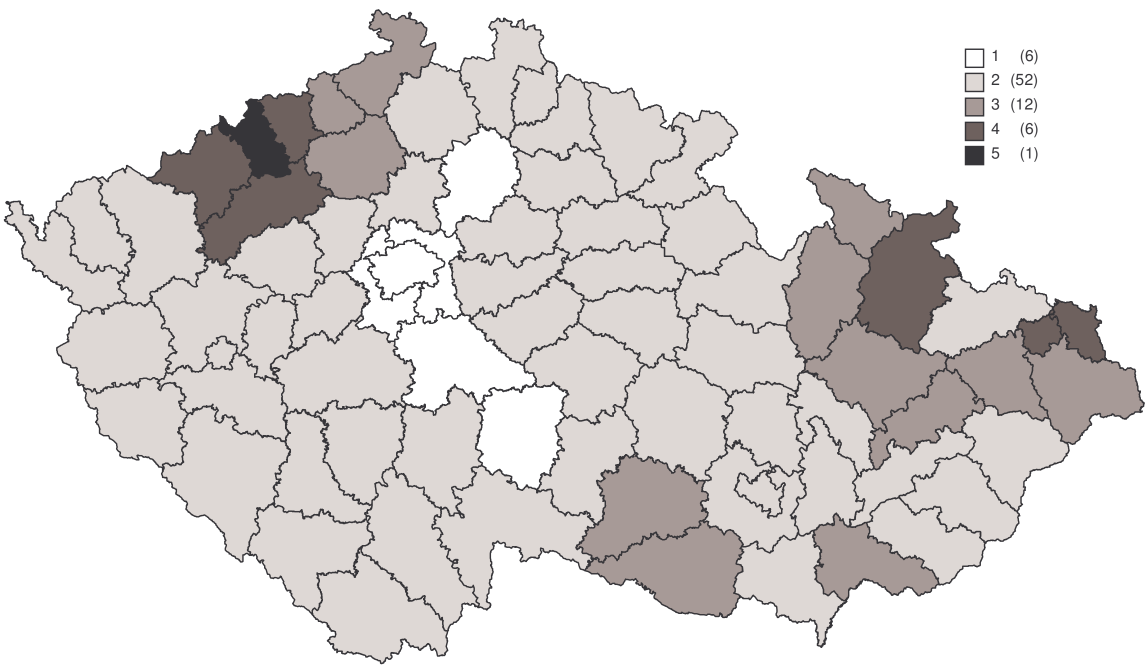
Type 3 - problematic districts: Děčín, Litoměřice, Ústí nad Labem, Třebíč, Hodonín, Znojmo, Jeseník, Olomouc, Přerov, Šumperk, Frýdek - Místek, Nový Jičín.

Type 4 – highly problematic districts: Chomutov, Louny, Teplice, Bruntál, Karviná, Ostrava-city.

Type 5 – the most problematic district: Most.

As it can be seen, there is an increase in the absolute regional differences concerning the rate of unemployment among the individual region types during the studied period. The most serious problem occurs in case of the type 4 and 5 where after 1999 the rate of unemployment had not been cut down and it has been still fluctuating around the level of 15%. For better understanding this typology is shown via the cartogram (see Figure 4.12).

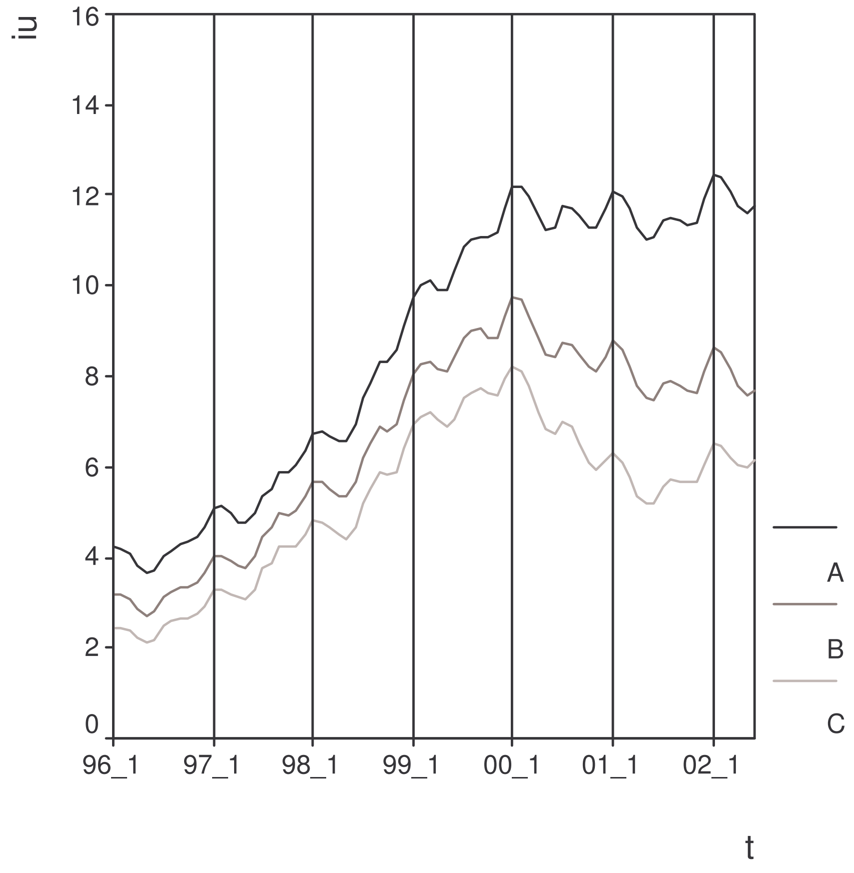
Figure 4.12: Cartogram of suggested typology with regard to the unemployment rate development (absolute differences)



4.1.6 Time series clustering – relative differences

The following classification system has been created using the same method and being based on the same data as the previous one but this time the Pearson's correlation coefficient has been used for the proximity measurement. In geographical analyses these proximities are usually called spatial autocorrelations. It is convenient to work with this typology at the lower level i.e. with the larger amount of clusters. The suggested classification system is designed for 14 classes but because of the small number of districts in case of many clusters, the centroids of only three largest clusters are shown in the following graph (see Figure 4.13).

Figure 4.13: Centroids of *iu* development classification – relative form



Types of regions:

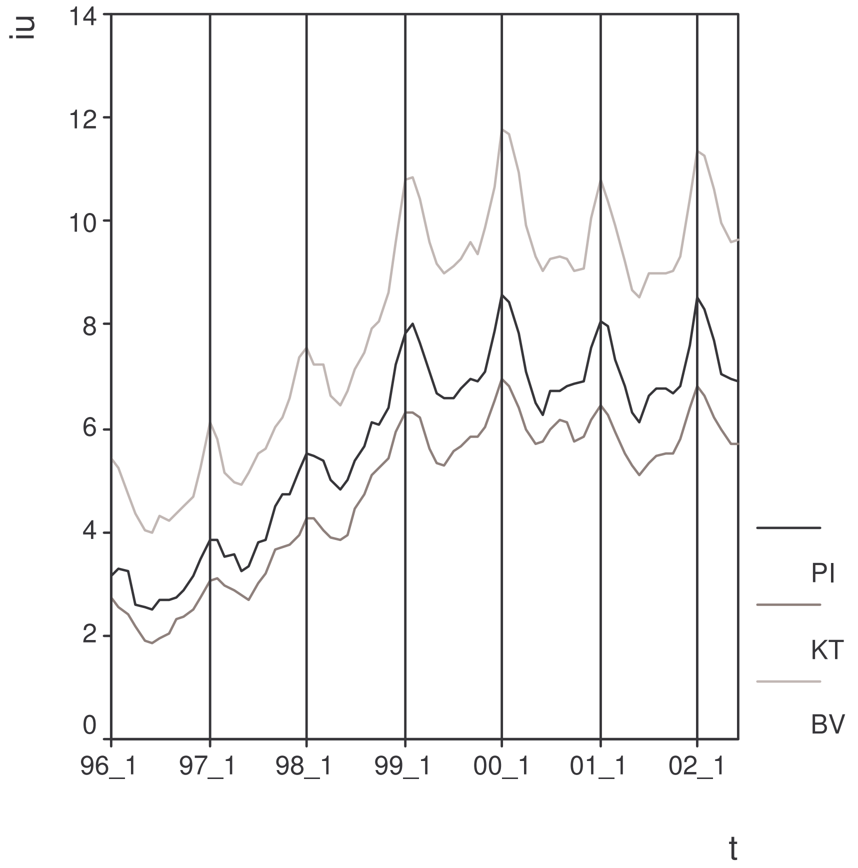
Type A –23 districts: Praha, Kutná Hora, Praha - east, Karlovy Vary, Sokolov, Louny, Most, Teplice, Ústí nad Labem, Svitavy, Brno - city, Brno – country, Hodonín, Šumperk, Kroměříž, Uherské Hradiště, Zlín, Bruntál, Frýdek - Místek, Karviná, Nový Jičín, Opava, Ostrava – city

Type B – 23 districts: Beroun, Kolín, Mělník, Nymburk, Praha - west, Rakovník, České Budějovice, Tábor, Plzeň - city, Plzeň - south, Plzeň – north, Rokycany, Chomutov, Litoměřice, Semily, Chrudim, Pardubice, Ústí nad Orlicí, Třebíč, Vyškov, Olomouc, Prostějov, Přerov

Type C – 9 districts: Domažlice, Cheb, Hradec Králové, Jičín, Havlíčkův Brod, Žďár nad Sázavou, Blansko

When interpreting the results shown in Figure 4.13, it is necessary to realize that we are not interested in the magnitude of unemployment but just in its course. The similarity of unemployment rate development in the individual types of regions was much higher during the recession (i.e. untill 1999) than in the subsequent years. Thus the sharp rise in *iu* i.e. the development of the difference between the regions has been apparent since the year 2000. In case of some of the districts a decline of *iu* occurs (type B), but in case of many other districts the augmented rate of unemployment is persistent (type A). When its level is relatively low, then there is no need to be worried (see Prague and its satellite districts). However, in case of a high level of the unemployment rate (i.e. more than 15%), this development is a serious problem that should be handled without any delay. In case of these regions the degradation of labor force occurs by the process of hystersis.

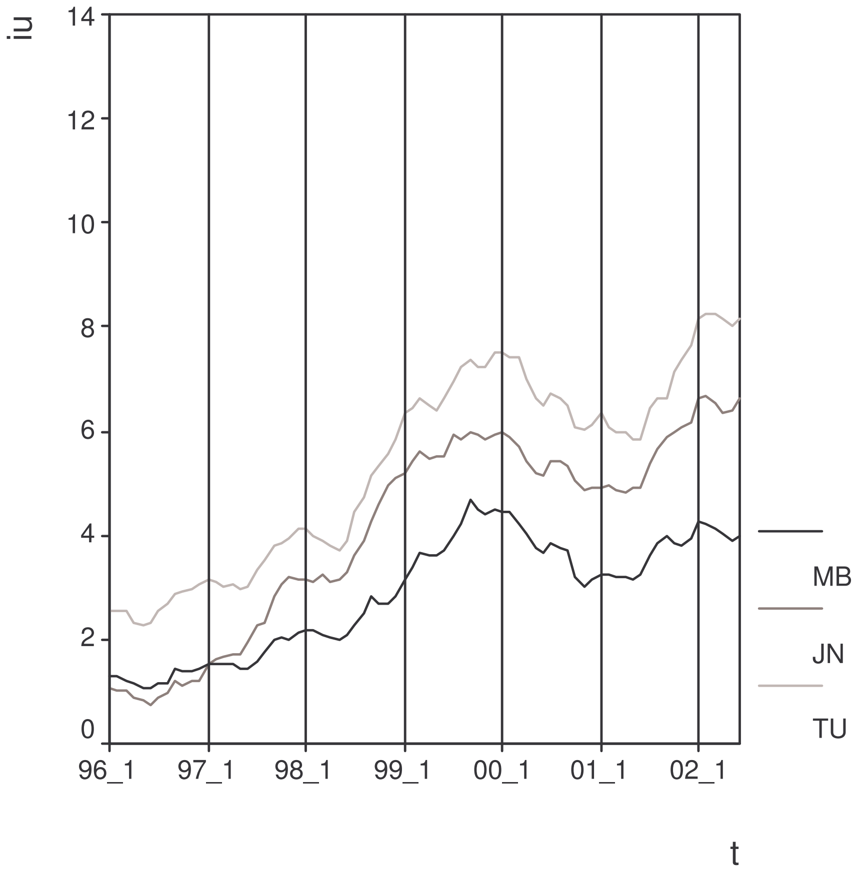
Figure 4.14: *iu* development – higher seasonal variability



Note: PI-Písek, KT-Klatovy, BV-Břeclav.

Let us return to the classification. It has its use even at the lowest level where the following question emerges, for example: "Why does the unemployment rate have almost the same character in case districts of Brno-country and Zlín?" In the graphs presented in Figure 4.14 and 4.15 there are the districts with higher and lower seasonal variability selected. The similarity can in most cases be clarified by the employment structure in a region but these explanations often fail. In such cases the factors of institutional nature in its broadest sense are usually mentioned. However, these factors are difficult to measure.

Figure 4.15: *iu* development – lower seasonal variability



Note: MB-Mladá Boleslav, JN – Jablonec nad Nisou,TU-Trutnov.

4.2.3 Conclusion

The use of region classification methods gives us some information about disparities among regions and they enable us to identify the problematic regions. However, when suggesting these methods, it is necessary to consider carefully the choice of descriptors and carry out their analysis. By making the classification systems dynamic, we are able to determine even a group of "potentially problematic regions" in the context of development of the other regions.

The practical application of these classifications in the individual regions at the municipal level (NUTS5) is also very fertile. For this level it is possible to use the data from SDLB, the Regional Information Systems (RIS) or the demographic data from the registers that are compiled by the Czech Statistical Office, see for example (Šotkovský – Tvrdý, 2002).

The next field in which development of the unemployment rate classifications with the relative differences find their use is predictions for the individual classes. Some reduction of the number of models occurs here and also the analysis of these time series could bring better results.

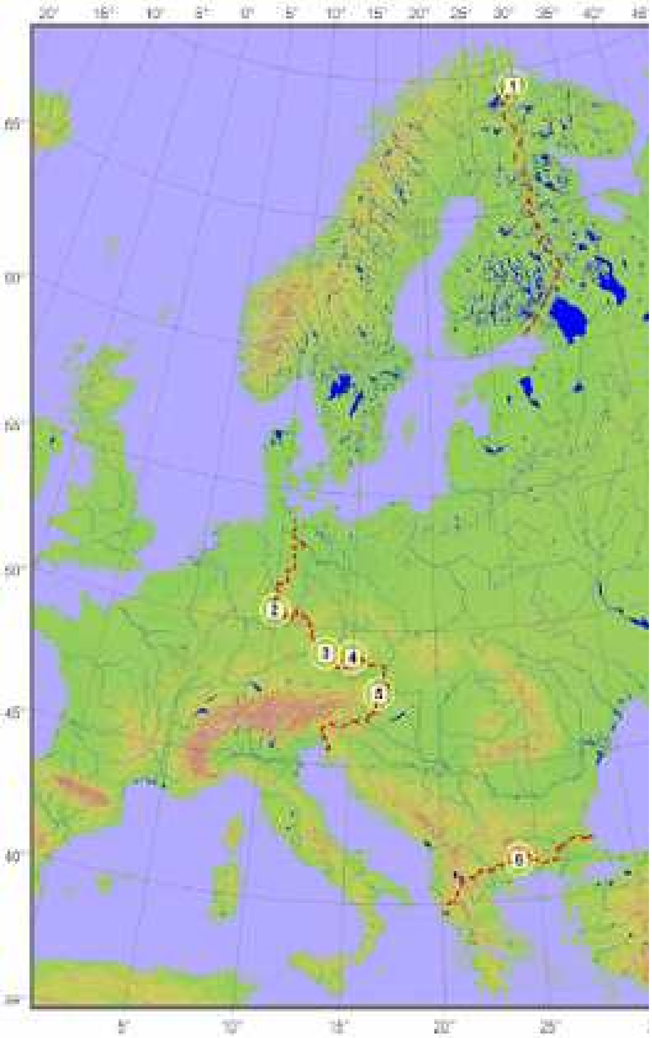
4.3 Classification of Regions through Expert Estimates[1]

The data used for practical applications of the new approach originate in a 5th FP RTD project “Iron Curtain” QLRT-2000-01401 (“Innovative models of critical key indicators as planning and decision support for sustainable rural development and integrated cross border regional management in former Iron Curtain areas based on north to south European reference studies.”) The project is focused on 6 reference areas along the former “Iron Curtain”.

---

[1] This chapter was written in a cooperation with Petr Koran (GEO, a.s. - Czech Republic), and George Lev (INTERCONSULT, ASA - Norway)

Figure 4.16: Reference areas
(Source: Iron Curtain report: Deliverable 4.1)



Reference areas:

RA1
NORWAY/RUSSIA
Pasvik Valley

RA2
GERMANY/GERMANY
Biosphere Reserve
Rhön

RA3
GERMANY/CZECH
REPUBLIC
Bavarian Forest/
Šumava

RA4
CZECH
REPUBLIC/AUSTRIA
Waldviertel/Trebon
Basin

RA5
AUSTRIA/HUNGARY
Kekfrankos Area

RA6
GREECE/BULGARY
Nestos/Mesta River

One of the goals of the project is to compare a regional competitiveness level and territorial capital of selected pilot areas. The twelve principal dimensions of territorial competitiveness as defined by the LEADER methodology concept are mentioned in the end of the section 2.

4.3.1 REDIF method

The REDIF (REgional DIFferential) method has been developed within the IRON CURTAIN project as a tool to deal with the high variability of reference area descriptors in terms of actual problem definition and data availability. It provides a unifying concept allowing the analyst to bring the assessment of territorial competitiveness of such different areas on the comparable level using only limited set of widely available indicators and quantified expert estimates. The quantification of expert estimate is inspired by semantic differential introduced by Osgood and his colleagues (Osgood et al, 1957 and Neuman, 2000).

REDIF is a combined qualitative – quantitative method to assess the status of an area according to the LEADER concept of Territorial competitiveness (LEADER, 1999). Its main strength is perceived in the ability to combine the qualitative and quantitative data assessment.

REDIF is based on a combination of the following approaches:

Quantitative assessment of territorial competitiveness components using basic set of indicators with defined scales from 0 (worst value) to 10 (best value) for each reference area.

Qualitative assessment by expert estimates based on the same scale with the limit values of 0 and 10 referring to overall verbal description of the worst and the best status for the component according to the LEADER concept.

The assessment has been done using the table covering both qualitative and quantitative approaches (see the example of social welfare in Table 4.11).

Table 4.11: Example of assessment of area capital component (Norway side RA1 – social welfare)

The worst	orign	Scale											The best
bad housing conditions, (no housing – homeless / displaced persons, or housing not affordable), insecure area of high criminality, high poverty, low life expectancy,		0	1	2	3	4	5	6	7	8	9	10	good housing conditions, accessible family houses, secure area with very low criminality, low number of poor people, long life expectancy, health population
			Rescaled value										
Unemployment rate (%)	*4*	***30***	***8,7***									***0***	
Number of unemployed on 1 vacancy	*na.*	***100***										***1***	
Life expectancy men	*75*	***55***	***8,7***									***78***	
Life expectancy women	*81*	***57***	***8,6***									***85***	

REDIF method is very suitable for comparison of different reference areas. Current status of twelve regions within the six IRON CURTAIN reference areas was assessed by region experts. The result of the REDIF assessment is depicted in Table 4.12. The data in the table represent an input matrix for statistical methods used for comparison of the areas.

Tab. 4.12: The results of REDIF assessment in IRON CURTAIN areas

Region	Physical resources	Environment conservation	Human resources	Cultural Identity	Know-How and Skills	Governance	Institutional capacity	Activities and business firms	Markets and external relations	Economic structure	Image and perception	Social welfare
RA1_NO	6	8	8	9	8	5	8	6	5	7	7	9
RA1_RU	3	5	2	3	4	2	4	4	3	2	2	2
RA2_GE_W	9	8	6	10	9	7	5	6	4	8	10	9
RA2_GE_E	9	8	3	8	9	8	5	6	5	6	9	9
RA3_GE	7	7	8	9	6	7	8	9	9	9	8	9
RA3_CZ	9	9	4	8	5	5	5	3	4	6	9	7
RA4_AU	8	7	6	9	6	8	8	7	6	6	8	8
RA4_CZ	6	7	8	10	5	6	4	6	6	4	9	7
RA5_AU	8	8	8	9	7	8	8	8	7	6	8	8
RA5_HU	8	7	6	9	6	7	8	6	7	6	9	6
RA6_GR	7	6	7	7	6	7	6	6	5	6	7	7
RA6_BG	6	6	8	7	4	6	4	4	4	3	7	4

The method of factor analysis was used to decrease the dimensionality of the input matrix and compare the reference areas. Three factors explaining of about 84 % of the matrix variability were calculated by factor analysis. The representation of particular TC components within the elicited factors is summarized in Table 4.13.

Tab. 4.13: Representation of TC components in three factors according to the factor analysis

Rotated Component Matrix

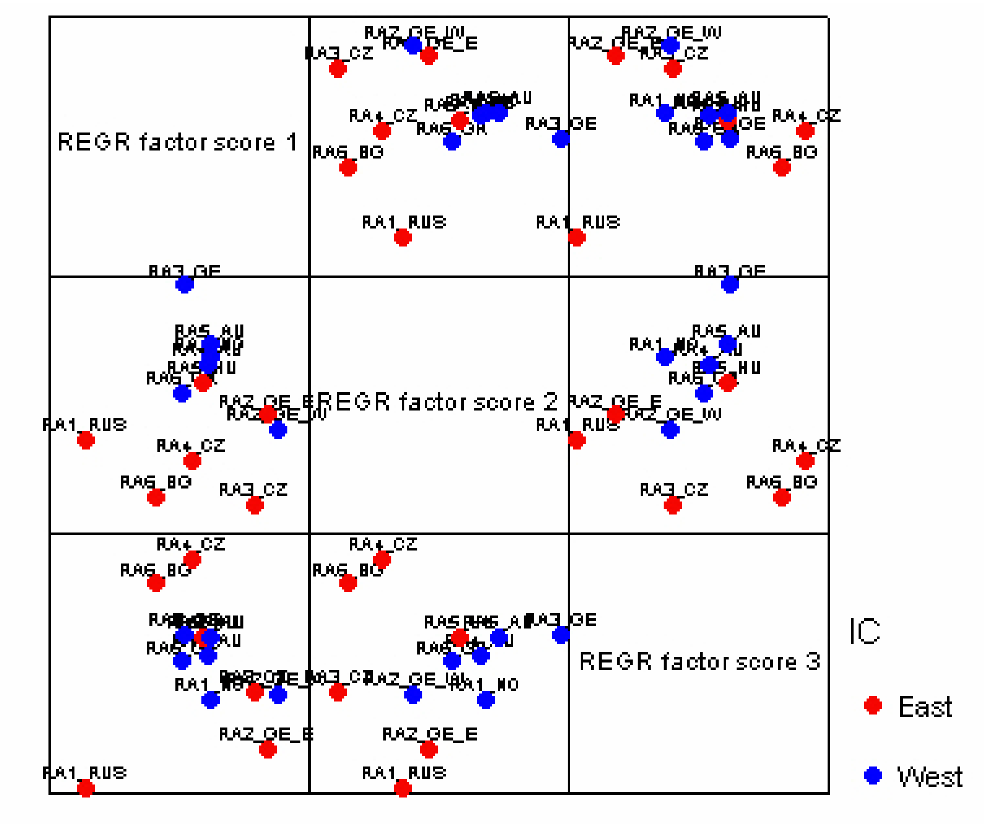
	Component		
	1	2	3
Physical resources	,921		
Environmental conservation	,887		
Image and perception of region	,877		,445
Social welfare	,799	,511	
Know-How and Skills	,745	,381	-,313
Cultural Identity of the region	,730	,288	,554
Economic structure	,659	,634	
Governance	,616	,416	,406
Activities and business firms	,149	,901	,271
Institutional capacity	,161	,854	,107
Markets and external relations	,104	,775	,486
Human resources		,361	,816

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Factor 1 is the most aggregated, prevailing dimension of the area competitiveness is related to the physical conditions, social dimension and economical dimension. Factor 2 describes the economical dimension both in terms of private and public sector (Institutional capacity). Factor 3 is influenced mainly by human resources.

For calculation of factor scores the method of regression was selected. The results of factor analysis can be visualized in a three-dimensional picture using the calculated factor scores (see Figure 4.17).

Figure 4.17: Visualization of the regions using three factors



Classification of the regions has been carried out by the Hierarchical Cluster Analysis, which is suitable only for a smaller amount of objects. This procedure identifies relatively homogeneous groups of regions.

The classification procedure was the following:

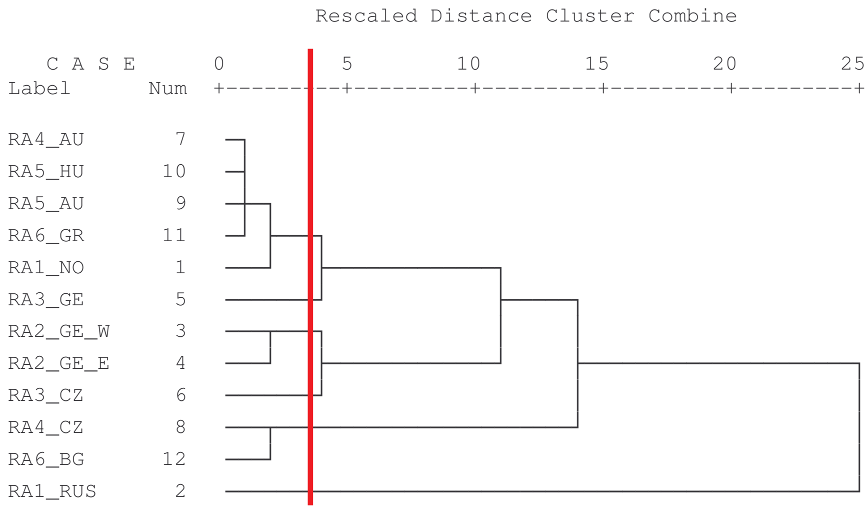
Matrix of distance was calculated from the factor scores.

Squared Euclidean Distance was used as the measure of dissimilarity.

The aggregation method called Average Linkage between Groups has been chosen, where the distance between groups is equal to average distance of every pair of points between groups.

The corresponding dendrogram can be found in Figure 4.18.

Figure 4.18: Dendrogram using Average Linkage (Between Groups)

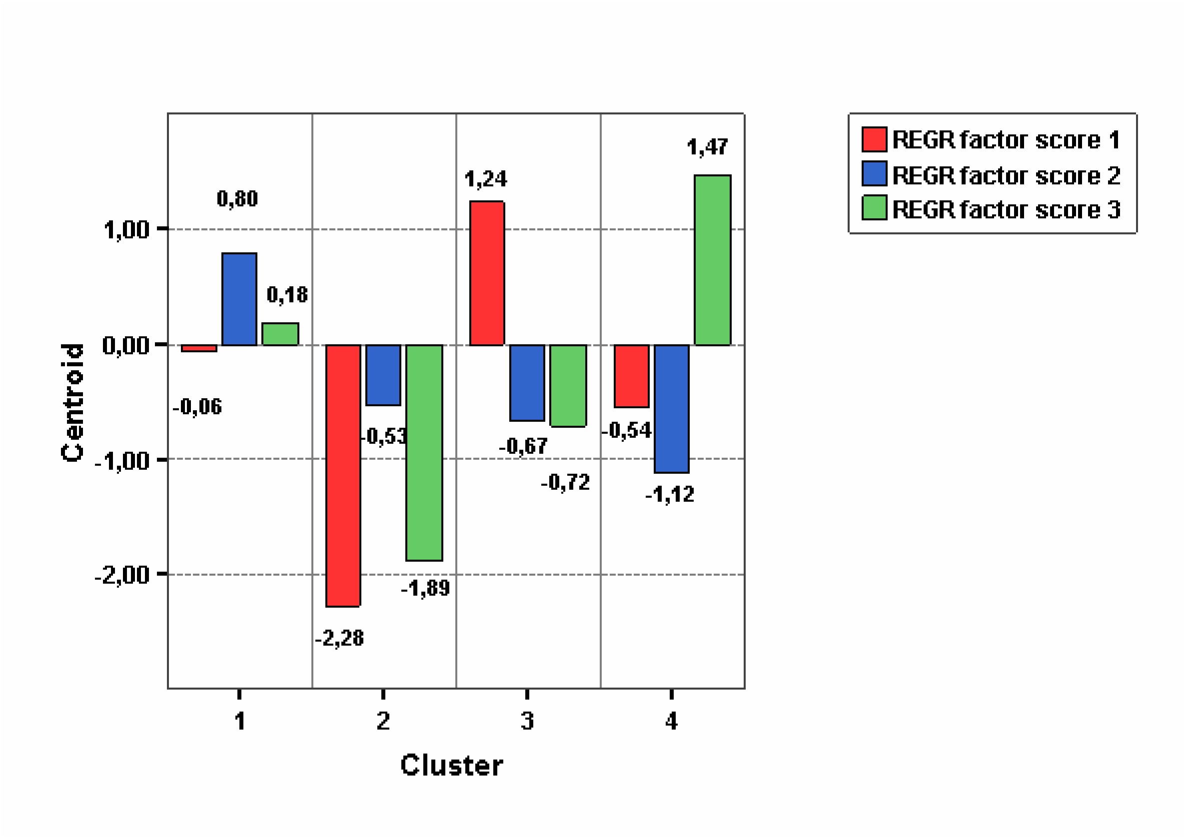


The dendogram shows continuous connection of particular countries into the final group and the rising distance between connected groups. Final clusters can be projected according to Figure 4.18 in the following way:

Cluster 1 - RA1_NO, RA3_GE, RA4_AU, RA5_AU, RA5_HU, RA6_GR
Cluster 2 - RA1_RUS
Cluster 3 - RA2_GE_W, RA2_GE_E, RA3_CZ
Cluster 4 - RA4_CZ, RA6_BG

For a support of each type of clusters the centroids are used. The centroid is defined as an arithmetic average of all objects in group for the particular variable. Four types of previously mentioned centroids are depicted in Figure 4.19.

Figure 4.19: Centroids of Clusters



Cluster 1 is closest to the average level since it contains majority of the objects. The positive values are mostly connected with factor 2. (Activities, firms and institutional capacities). In this cluster the high amount of regions from so called „West Europe“ are comprised. It means the regions with higher competitiveness comparing to other analyzed regions.

On the contrary, the values of cluster 2 are all under average, mostly in factor 1. (Physical resources, Environmental conservation) and factor 3 (Human resources). This region has the lowest competitiveness.

A characteristic for cluster 3 is higher value of factor 1. In these regions there are more nature reserves. The regions are preserved and there is no room for development of new entrepreneurial activities.

Higher value of factor 3 and negative values of factor 2 are typical for cluster 4. It follows, that these regions have some potential in human resources, but there is a low activity of firms. These potentials should be activated with some tools of regional policies and used for rising up competitiveness of these regions.

4.3.2 Conclusion of the section

The conclusions of the described classification have been verified by cluster analysis, which was applied directly to the input matrix. The assumption of stability of the proposed classification system has been thus confirmed. The application of regional descriptors based on quantified expert estimates is identical with the application of quantitative data. Nevertheless, evaluation of these estimates requires careful consideration - linear dependences between estimates from lower amount of regions may occur and the resulting matrix may not be a positive definite type. The method of factor analysis cannot be applied to any other matrix type.

5 Conclusion of the chapter

The research results presented in this chapter show a successful and meaningful use of multivariate techniques in the regional classification.

The expert estimates incorporated into the classification of regions enables, on one hand, to substitute a lack of data sources by expert's expressions, but, on the other hand, also to combine qualitative and quantitative appraisal of regions, that contributes to increasing duality of results and validity of problem solving.

The dynamic approach applied to regional classification process enables to map not only regional allocation, but also to follow the paths of their development either separately or in the frame of clusters. This part could be widened by examining the prediction of successfulness or exploiting consequences of simulations of regional policies. There are, however, problems connected with the extracted factors stability in time.

In the field of time-series clustering we have applied two approaches for measuring proximity – absolute and relative. In the given case study we discussed the advantages and disadvantages of those approaches in a more detail. The absolute form is adequate for measuring the position of the region in the relation to the other regions and by the relative form especially internal structure of the character of regions with no respect to the absolute level that have been compared.

The methods of multi-variate statistics are suitable for revealing the problem areas and setting the adequate instruments of regional policy.

REFERENCES

BERG, B. L. Qualitative Research Methods for the Social Science. Needham Heights: Allyn & Bacon, 2001. ISBN: 0-205-31847-9.

BERKA, K. Měření. Pojmy, teorie, problémy. Praha: Academia, 1977.

BOLTON, R. (1980). Multiregional Models: Introduction to a Symposium. Journal of Regional Science, Vol. 20, pp. 131-142.

CZSO „Opatření Českého statistického úřadu ze dne 27. dubna 1999 k 1.lednu 2000“ Sbírka zákonů 1999, Částka 33

GLICKMAN, N.J. (1982). Using Empirical Models for Regional Policy Analysis. In: M.Albegov et al. (eds.) Regional Development Modelling: Theory and Practice. Amsterdam: North-Holland.

HAIR. J. F. a kol. (1998). Multivariate data analysis. New Jersey: Prentice Hall, Inc.. ISBN 0-13-894858-5

HAMPL, M. a kol. (1983). Sociálněgeografická regionalizace ČSR. Praha: Univerzita Karlova.

HANČLOVÁ, J. ET al. (2002). Modelling and classification of regional labour markets. Ostrava: Technical University of Ostrava, 147 p.,ISBN 80-248-0220-1.

ISSAEV B., NIJKAMP P., RIETVELD, P., SNICKARS F. (1982). Multiregional Economic Modelling: Practice and Prospect. Amsterdam: North-Holland. ISBN 0 444 86485 7.

KLINE, P. (1994). An Easy Guide to Factor Analysis. Routledge, London.

IRON CURTAIN.(2004) EU 5th FP RTD Project, QLK5-CT-2001-01401) – Final report. Writing. (Innovative models of critical key indicators as planning and decision support for sustainable rural development and integrated cross border regional management in former Iron Curtain areas based on north to south European reference studies).www.ironcurtainproject.com

LEADER (1999). Territorial competitiveness. Creating a territorial development strategy in light of the LEADER experience. Part 1. [on-line] Brusel: European LEADER Observatory. [cit. 2004-02-14]. http://europa.eu.int/comm/archives/leader2/rural-en/biblio/compet/competitivite.pdf.

LUKASOVÁ, A. - ŠARMANOVÁ, J. (1985). Metody shlukové analýzy. SNTL, Praha. s. 212.

NEUMAN, W. L. (2000) Social Research Methods. Qualitative and Quantitative Approaches. Needham Heights: Allyn & Bacon 2000. ISBN:0-205-29771-4.

NORUSIS,M.J. (1990). SPSS/Pc+ Statistics 4.0, SPSS Inc.

OPENSHAW, S. (1995). Census Users' Handbook. Geoinformation International and John Wiley & Sons, Cambridge.

OSGOOD, E.C., et al (1957). The measurement of meaning. Urbana, University of Illinois Press..

OZIMEK, J. (1993). Targeting For Success: A Guide to New Techniques for Measurement and Analysis in Databáze and Direkt Response Market. McGraw-Hill, Berkshire.

PAELINCK, H. J. - NIJKAMP, P. (1975). Operational Theory and Method in Regional Economics. 1st ed. Farnbrough: Saxon House Studies. 473 p. ISBN 0-347-01117-9

RAMÍK, J. (1999). Vícekriteriální rozhodování - analytický hierarchický proces (AHP). Karviná: Slezská univerzita v Opavě, Obchodně podnikatelská fakulta, s. 211. ISBN: 80 7248 047 2

SDA - Sociologický datový archív [online].Praha : Sociologický ústav AV ČR, c1998-1999 [cit 1.11.2001] Dostupný z WWW: <http://archiv.soc.cas.cz/czech/archivtop-f.html>

SIROVÁTKA, T. – ŽIŽLAVSKÝ, M. (2002). Český trh práce ve druhé polovině 90. let. Politická ekonomie. č. 3 s. 419-435

SOARES, J.O., MARQUES, M. M., MONTEIRO C. M. (2002). A multivariate methodology to uncover regional disparities. A contribution to improve European Union and governmental decision. European Journal of Operational Research. http:// www. sciencedirect.com/

SPSS (1997). SPSS 7.5 Statistical Algorithms. Chicago: SPSS Inc. ISBN: 1-56827-185-9

ŠOTKOVSKÝ, I. - TVRDÝ, L. (2002). Změny reprodukčního a prostorového chování obyvatel Moravskoslezského kraje v letech 1992 až 2000. In: Mezinárodní vědecká konference „Ekonomické a adaptační procesy 2002“, sekce č. 4: Regionální politika ČR před vstupem do EU. Sborník referátů. Ostrava: VŠB-Technická univerzita Ostrava, Ekonomická fakulta, ISBN: 80-248-0186-8.

TOLSTOVOVÁ, J. N. et al. Typologie a klasifikace v sociologických výzkumech. Praha: Svoboda. 1989.

CONTACTS

Faculty of Economics,
VŠB – Technical University, Ostrava, Sokolská 33, 701 21 Ostrava 1

Hančlová, Jana, Doc. Ing. CSc. jana.hanclova@vsb.cz
Tvrdý, Lubor, Ing. lubor.tvrdy@vsb.cz

Annex 1: Methodology Redif for RA1_RUS Pechenga District

	The worst	orign	Scale			The best
Physical resources	Unregulated exploitation of natural resources, low natural resources potential, high land fragmentation, weak forestry and agriculture, low biodiversity, no use of renewable energy sources, no tourism, poor water and soil conditions		0	1 2 3 4 5 6 7 8 9	10	sustainable land use (functional planning legislation), low fragmentation, rich in natural resources, strong ecological agriculture, forestry, high biodiversity, excellent water and soil conditions, high usage of RES, attractive touristic area
Environmental conservation	Disregard for environment, preference of monetary gains over preserving nature, valuable ecosystems not protected resulting in badly polluted water and air, endangered fauna and flora, noise pollution		0	1 2 3 4 5 6 7 8 9	10	Effective conservation, high percentual share of protected areas/natural reserves resulting in clean air and water, high biodiversity,
Human resources	Unfavourable age structure of population, high gender inequality, high emigration/ brain drain, high share of poor social groups, high percentage of old population, high migration / population exchange		0	1 2 3 4 5 6 7 8 9	10	high rate of inhabitants of productive age, gender equality, people are moving into the area, strong middle class, young population, low migration
				rescaled value		
	Change population 2001 - 1991 (%)	*-26,4*	-35	1,7	15	
	Net migrantion rate (‰)	*-15,2*	-37	4,2	15	
	Natural increas rate (‰)	*-1,96*	-10	3,2	15	
Cultural Identity of the region	Low and damaged architecture and landscape heritage, key regional actors not interested in the area development, actors not sharing a common vision of future development		0	1 2 3 4 5 6 7 8 9	10	rich in architectural and landscape sightseeings, inhabitants have strong relationship with the region, it's history and landscape,
Know-How and Skills	Few university educated (illiteracy), non-competitive skills (outdated), unstable labour force (itinerant workers), no research capacity, low usage of advanced technologies (not accessible)		0	1 2 3 4 5 6 7 8 9	10	high rate of university educated, wide usage of modern technologies, stable and skilled labour force, high research capabilities
				rescaled value		
	Share of University graduate (%)	*na.*	*3*		30	
	Share of Elementary school (%)	*na.*	*55*		10	
Governance	Corruption, favouritism, nepotism, missing or insufficient democratic mechanisms, local administration goes into debt (incompetent administration), very low financial resources both in public and private sectors		0	1 2 3 4 5 6 7 8 9	10	democratic mechanisms working well, effective administration and justice, enough financial resources, balanced public budgets, effective funding programmes, high savings, rich public and private subjects

	The worst	orign		Scale		The best
Institutional capacity	No or limited access to public services and facilities, poor infrastructure, no horizontal administration work, weak scientific capacities, no collaboration between research and industry, no free time possibilities		0	1 2 3 **4** 5 6 7 8 9	10	public services and facilities of high level, very good infrastructure, presence of scientific subjects collaborating with industry, rich in free time possibilities
Activities and business firms	None or very limited market diversity, high share of big companies (monopolism..), one-side oriented, high centralisation of economic activites, low number of SMEs, no innovative firms,		0	1 2 3 **4** 5 6 7 8 9	10	high diversity of economical activities and company sizes, suitable spatial distribution of businesses, high number of SME's, innovation on company level
Markets and external relations	High rate of grey / black economy, low local production, no export, badly connected regional markets to external markets, low /non existent business credibility, non-fuctional enterprise networks, no investments into the region, low regional income, high		0	1 2 **3** 4 5 6 7 8 9	10	strong regional market, regional products exported, working networks of enterprises, good connection with external markets, high regional production and income, attractive for investments, low inflation
Economic structure	Low GDP (low income) - low/no economical reserves of large share of population, low buying power of average income, high unemployment, economical remoteness of the region in the frame of country		0	1 **2** 3 4 5 6 7 8 9	10	high local GDP, high buying power, low unemployment, important share of region on the country GNP
				rescaled value		
	GDP per capita (EURO)	*9700*	1000	2,2	40000	
	Share of Service *(%)*	*41,3*	25	2,5	90	
Image and perception of region	Bad internal image, bad emotional relation of inhabitants to the region, region percepted as not good for living externally and internally, bad image in press and TV		0	1 **2** 3 4 5 6 7 8 9	10	Positive perception of the area (internal and external), historical and ethical values highly appreciated, percepted as attractive area for living or investing because of environment and socio-economical conditions, regional positively presented in TV and
Social welfare	Bad housing conditions, (no housing - homeless/displaced persons, or housing not affordable), insecure area of high criminality, high poverty, low life expectancy,		0	1 **2** 3 4 5 6 7 8 9	10	good housing conditions, accessible family houses, secure area with very low criminality, low number of poor people, long life expectancy, health population
				rescaled value		
	Unemployment rate (%)	*15,27*	*30*	4,9	*0*	
	Number of unemployed on 1 vacancy	*na.*	*100*		*1*	
	Life expectancy men	*60,1*	*55*	2,2	*78*	
	Life expectancy women	*71,5*	*57*	5,2	*85*	