Comparison of impacts of the financial crisis on the Czech Republic regions by cluster analysis

Jiří Mazurek
Silesian University in Opava
School of Business Administration in Karviná, Department of Mathematical Methods in Economics
Univerzitní náměstí 1934/3
Karviná, 733 40
The Czech Republic
mazurek@opf.slu.cz

***Abstract**:*
In 2008 the Czech Republic´s economy fell into recession following the global financial crisis. Given unique socio-economic situation each Czech region suffered in a different way during the crisis. The aim of this article is to compare impacts of the financial crisis by partitioning 14 regions of the Czech Republic into groups (clusters) with similar changes in selected macroeconomic indicators before and after the crisis. To solve the problem, cluster analysis via k-means clustering method was employed. Results showed regions formed following clusters: A = {Ústecký, Moravskoslezský}, B = {Prague}, C = {Středočeský, Plzeňský, Karlovarský}, D = {Královéhradecký, Pardubický, Vysočina, Olomoucký, Zlínský} and E = {Jihočeský, Jihomoravský, Liberecký}. Globally, regions in clusters A, B and E were less affected and regions in clusters C and D more affected by the crisis. Presented results can be used to stimulate collaboration between regions, to adopt similar policies and to coordinate measures needed for overcoming crisis consequences.

Keywords: Cluster analysis, Czech regions, Czech Republic, financial crisis impacts.
JEL codes: C38, O18, R11.

1. Introduction

The recent global financial crisis started in the U.S.A. in the summer 2007 by a liquidity shortfall in the U.S. banking system. The trigger of the crisis was a collapse of a US housing bubble that led to fall of large financial institutions or its bailout by national governments, and to large share drops around the world.

In the Czech Republic recession began in the fourth quarter of 2008 and lasted to the third quarter of 2009. According to the Eurostat (2010) during this period the gross domestic product (GDP) fell by 0.7, 3.8 and 0.5 % respectively in quarter-to-quarter comparison (see Figure 1). The budget deficit in 2009 reached monumental 192 milliard crowns (CSO, 2010), which is 5.93 % of GDP, the number unprecedented in the Czech history. This development led to tax reforms and drastic cuts in government spending. Further, the data for the Czech Republic (CSO, 2010) indicate that between 2007 and 2009 unemployment grew by alarming 52 %, sales of own goods and services incidental to industry fell by 12.8 %, elementary construction production dropped-off by 22 %, number of guests decreased by 6.6 %, only number of registered economic entities recorded slight increase by 3.6 %.

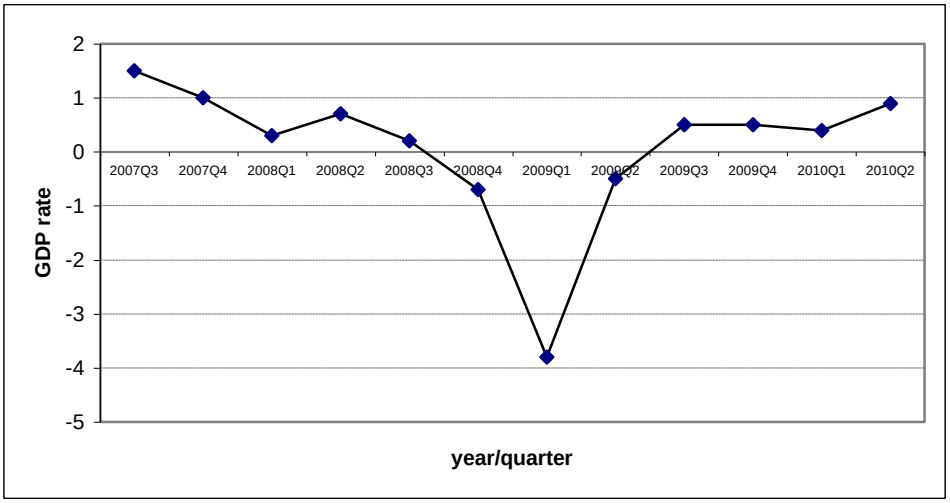
However, impacts of the crisis varied across regions. Given unique social and economic situation each region suffered in a different way during the crisis, and therefore have to adopt its own policy to successfully bridge the crisis consequences. Regions with similar crisis impacts may take advantage of similar methods to enhance economics and to coordinate its efforts.

Understanding the region‘s evolution is crucial because importance of the role of regions on the national economic development is growing, mainly due to region’s innovations (Urbančíková and Burger, 2010). So far, there are not many regional studies concerning the last financial crisis, with the exception of Mielcová (2010), who studies unemployment ties between Moravskoslezský region and Prague.

The aim of this article is to compare impacts of the recent economic crisis by partitioning regions of the Czech Republic into groups (clusters) with similar changes in selected macroeconomic indicators before and after the crisis. To solve the problem, cluster analysis via k-means clustering

method was employed. For the crisis impact comparison the data from the end of 2007 and the end of 2009 were chosen. The data from 2007 describe the state of economics before crisis and the data from 2009 the state of economics after the crisis.

Figure 1: Inter-quartal GDP rates (in %), the Czech Republic, Q3/2007-Q2/2010.



Source: Eurostat (2010).

2. The data

The data for regions comparison were taken from the *Statistical Yearbook of Regions* of the Czech Statistical Office (CSO, 2010) from the end of 2007 and 2009 respectively. The data in the form of percent change from 2007 to 2009 are provided in Table 1.

CSO's *Region Statistical Yearbooks* contain elementary geographic, demographic and macroeconomic indicators for each region. From macroeconomic section the following indicators were chosen for comparison:

- unemployment,
- sales of own goods and services incidental to industry
- measure of economic activity,
- registered economic entities,
- elementary construction production,
- number of guests in collective accommodation establishments.

One of the most important macroeconomic indicator, gross domestic product, was not included in the study because the data were not available for each region. Also the data from agriculture couldn't be used, for the data format changed (e.g. total cereals sales were changed to total harvest) between 2007 and 2009.

In the first step possible dependences among indicators were examined. Correlation among indicators is presented in Table 2. The correlation coefficient exceeded value 0.5 (bold figures) only for three pairs of indicators, other pairs of indicators were rather uncorrelated.

Table 1: Percentual changes of all indicators from the end of 2007 to the end of 2009.

Region	*unemployment*	*industry*	*ec. activity*	*entities*	*construction*	*guests*
Jihočeský	67	–7	–2	2	–24	–6
Jihomoravský	49	–21	–1	2	–12	–11
Karlovarský	49	–15	0	4	–37	–3
Královéhradecký	61	–24	–1	1	–39	–8
Liberecký	80	–17	–1	2	–16	–5
Moravskoslezský	22	–14	2	2	–35	–7
Olomoucký	75	–17	0	1	–43	–10
Pardubický	74	–14	0	2	–26	–13
Plzeňský	85	–13	1	6	–50	–7
Prague	72	5	1	8	–11	–3
Středočeský	68	–21	0	5	–33	–10
Ústecký	20	–7	0	2	–21	–11
Vysočina	76	–29	–3	2	–31	–13
Zlínský	68	–11	–3	0	–34	–14

Source: CSO (2010).

Table 2: Correlation matrix for indicators.

Indicators	*unemployment*	*industry*	*ec.activity*	*entities*	*construction*	*guests*
unemployment	1	–0.14	–0.32	0.21	–0.16	–0.01
industry	–0.14	1	0.36	**0.56**	0.40	0.42
activity	–0.32	0.36	1	**0.54**	–0.12	0.47
entities	0.21	**0.56**	**0.54**	1	0.1	**0.57**
construction	–0.16	0.40	–0.12	0.1	1	0.1
guests	–0.01	0.42	0.47	**0.57**	0.1	1

Source: author.

3. The method

For the study k-means clustering method was used. Method's name was coined by MacQueen (1967) and standard algorithm comes from Lloyd (1982). The method divides a set of *n* observations into *K* clusters so that observations in each cluster are similar (close) to each other. It is widely used in economics, data mining, pattern recognition, image processing, bioinformatics, biology, etc. The method requires variables that are (preferably) continuous with no outliers, as discrete data may cause problems.

Each observation *j* is an *m*-dimensional vector x_{ij}, $i = 1$ to m. Let's assume *k*-th cluster contains n_k objects. Then the clustering aims to partition *n* observations into *K* sets ($K< n$) so as to minimize the within-cluster sum of squares for *K* clusters (WSS_K), see Meloun and Militký (2006):

$$WSS_K = \frac{mn}{mn-m} \cdot \sum_{k=1}^{K} \sum_{i=1}^{m} \sum_{j=1}^{n_k} \left(1-\delta_{ijk}\right)\left(z_{ij}-c_{ik}\right)^2, \qquad (1)$$

where c_{ik} is an average of *i*-th variable in *k*-th cluster, δ_{ijk} denotes (eventual) missing value of *i*-th variable in *j*-th object for *k*-th cluster, and z_{ij} is a standardized value of x_{ij}.

In general Lloyd´s algorithm consists of the following four steps:

1. The number of clusters (*K*) is chosen.
2. Random *K* clusters are generated and cluster centres (*centroids*) are computed e.g. as an average of coordinates of all observations (represented as points).
3. All observations are assigned to the nearest cluster centre.

4. New cluster centres are computed as an average of observation‘s coordinates and Step 3 is repeated until algorithm converges (no cluster is changed by repeating procedure).

Goodness-of-fit is given by percent of variation PV_K see Meloun and Militký (2006):

$$PV_K = \frac{WSS_K}{WSS_1} \cdot 100\,, \qquad (2)$$

where index K is a number of clusters. PV_K gives the within-sum of squares for K clusters (WSS_K) as a percentage of within sum of squares without clustering (WSS_1).

Advantage of the method is its simplicity, speed and possibility of running algorithm on large databases. However, algorithm solution depends on the initial random assignment of cluster centres, number of clusters and number of iteration, thus giving different results in different runs. The second disadvantage is that algorithm minimizes within-cluster variance, but it finds only local minimum, not global minimum. To eliminate both problems repeated clustering is necessary (typically 25 runs are used).

Number of clusters is usually selected by a researcher, however the most proper number of clusters for given data can be found via *average silhouette*, criterion that expresses the tightness of cluster arrangement of the data on the interval from –1 to +1, the latter value indicating absolutely strong and tight sructure, for details see Meloun and Militký (2006). In the presented study four to five clusters were in consideration, average silhouette was larger for five clusters (0.43 comapred to 0.41 for four clusters), so five clusters were chosen for the analysis.

For computation the statistical software NCSS 2000 was employed (see NCSS home page). NCSS uses AS136 algorithm by Hartigan and Wong (1979) which is a slight modification of Lloyd’s algorithm in that it allows to by-bass a local optimal solution by swapping points between clusters. NCSS 2000 facilitates calculation, performs goodness-of-fit and significancy tests, evaluates distances of all objects to all cluster centres (Table 4) and enables to visualize clusters in bivariate scatter plots, see Figures 3: a)-h). From programme menu 5 clusters, 5 to 15 random starts and 10 to 20 iterations were chosen for each run. The programme was run 25 times. Tables 3 and 4 show the best result with the lowest percent of variation $PV_K = 31.76$ of all runs.

4. Results

Regions partition into clusters is shown in Tab. 3 along with its labelling from cluster A (cluster nr. 1) to cluster E (nr. 5). The capital city Prague is a solitaire; other 13 regions are divided into 4 clusters. Tables 4: a)-e) show average values of all indicators for each cluster. Table 5 gives regions distances to all cluster centres. Figure 2 shows the geographic distribution of clusters. Bivariate scatter plots of selected indicators, Figures 3: a)-h), are presented at the end of the chapter along with Figure 4, which displays clusters differences in the graphic form. Region numbers and cluster numbers correspond to the same numbers in scatter plots.

Table 3: Partition of regions into clusters.

A = cluster 1	*B = cluster 2*	*C = cluster 3*	*D = cluster 4*	*E = cluster 5*
Ústecký	Prague	Středočeský	Královéhradecký	Jihomoravský
Moravskoslezský		Plzeňský	Pardubický	Jihočeský
		Karlovarský	Vysočina	Liberecký
			Olomoucký	
			Zlínský	

Source: author.

4.1 Cluster A

Cluster A consists of two regions: Ústecký and Moravskoslezský. Both regions were similar in rather small unemployment growth compared to other regions (20 % and 22 % respectively), but still remain among regions with the highest unemployment. The heaviest downturn for both regions was registered in elementary construction production; in other indicators did the regions rather well. Both

regions have alike economic structure. In Ústecký region the manufacturing industry and an opencast coal-mining are leading economic branches, in Moravskoslezský region leads heavy industry with steel production and coal mining. Hence its assignment to the same cluster is not surprising.

Table 4a): Average changes (in %) of indicators for the cluster A.

unemployment	*industry*	*ec. activity*	*entities*	*construction*	*guests*
21	–11	1	2	–28	–9

Source: author.

4.2 Cluster B

Cluster B includes only the capital city of Prague, ensuring its privileged economic position (see the recent study Martinčík (2008) that confirms its exceptional place among Czech regions). During the crisis unemployment in Prague grew by 72 %, but in absolute numbers there is still the lowest unemployment in the country. In the remaining indicators Prague succeeded to stay in the "black numbers" with exception of slight losses in the elementary construction production and number of tourists. Along with regions in clusters A and E Prague coped with the recession better than others.

Table 4b): Average changes (in %) of indicators for the cluster B.

unemployment	*industry*	*ec. activity*	*entities*	*construction*	*guests*
72	5	1	8	–11	–3

Source: author.

4.3 Cluster C

Cluster C contains Středočeský, Plzeňský and Karlovarský regions. Středočeský region belongs to the most developed Czech regions with excellent engineering, chemistry, grocery and agriculture. Economics of Plzeňský region is based on grocery, electrotechnics and engineering and Karlovarský region with manufacturing industry and services. Plzeňský region recorded the highest unemployment growth (80 %) and also the highest decline in elementary construction production (–50 %), which makes it one of the most crisis affected regions in the country. In spite of this fact, the number of registered economic entities in the region grew by 6 % over the same period. Globally, the cluster C experienced losses mainly in the elementary construction production which declined by 40 % during two years period, but on the other hand economic activity didn´t changed and number of registered entities even grew by 5 %.

Table 4c): Average changes (in %) of indicators for the cluster C.

unemployment	*industry*	*ec. activity*	*entities*	*construction*	*guests*
67	–14	0	5	–40	–7

Source: author.

4.4 Cluster D

Cluster D is the most populous cluster including regions of eastern Bohemia and central Moravia, which belong to less developed regions in the country with agriculture and manufacturing industry as leading branches. The cluster is characterized by above average unemployment growth (71%), but the largest losses occurred in the sales of own goods and services incidental to industry (–20 %), where Vysočina, Olomoucký and Královéhradecký regions placed the 1st, 2nd and 3rd. In the elementary construction production decline Olomoucký and Královéhradecký regions placed the 2nd and 3rd. Zlínský, Pardubický and Vysočina regions suffered the largest decrease of tourists, placing the 1st, 2nd and 3rd respectively in this category. Globally, the cluster experienced the worst economic

results in four out of six indicators. Along with Plzeňský region from the cluster C regions from the cluster D rank among the most severely affected regions at all.

Table 4d): Average changes (in %) of indicators for the cluster D.

unemployment	*industry*	*ec. activity*	*entities*	*construction*	*guests*
71	–20	–1	1	–35	–12

Source: author.

4.5 Cluster E

Cluster E includes Jihomoravský, Jihočeský and Liberecký region. Jihočeský region is agricultural with long fishpond management and cereals production, in Jihomoravský region leads manufacturing industry and agriculture, and Liberecký region is industrial with production of glass, jewellery and plastics, and engineering. The cluster is characterized by an average unemployment growth and average downturn of sales of own goods and services incidental to industry. Measure of economic activity slightly decreased, number of registered entities moderately grew. Losses in elementary construction production and efflux of guests were again close to average. Globally, regions in clusters A, B and E rank among regions with lower recession´s impacts compared to regions in clusters C and D. Interestingly, if the Czech Republic was treated as a region then it would be assigned to this cluster.

Table 4e): Average changes (in %) of indicators for the cluster E.

unemployment	*industry*	*ec. activity*	*entities*	*construction*	*guests*
65	–15	–1	2	–17	–7

Source: author.

Table 5: Region’s distances to all cluster centres.

Region	*Cluster*	*Dist1*	*Dist2*	*Dist3*	*Dist4*	*Dist5*
1 Prague	2	4.701	0.000	3.785	5.563	4.171
2 Středočeský	3	2.870	3.826	1.122	2.175	2.259
3 Jihočeský	5	3.233	3.966	2.655	2.453	1.249
4 Plzeňský	3	4.174	4.260	1.397	3.484	3.849
5 Karlovarský	3	2.618	3.950	1.481	3.130	2.572
6 Ústecký	1	1.144	4.715	3.497	3.361	2.814
7 Liberecký	5	3.654	4.085	2.776	2.538	1.032
8 Královéhradecký	4	3.086	5.625	2.406	1.275	2.218
9 Pardubický	4	2.971	4.711	2.556	1.531	2.008
10 Vysočina	4	4.584	6.444	3.748	1.634	2.847
11 Jihomoravský	5	2.745	4.985	3.382	2.278	1.561
12 Olomoucký	4	3.482	5.791	2.412	1.368	2.793
13 Zlínský	4	4.003	6.128	3.865	1.784	2.795
14 Moravskoslezský	1	1.144	4.958	2.909	3.695	3.496

Source: author.

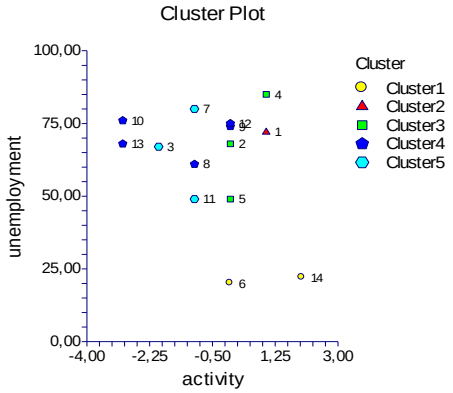
Figure 2: The geographic distribution of regions into clusters.

Source: author.

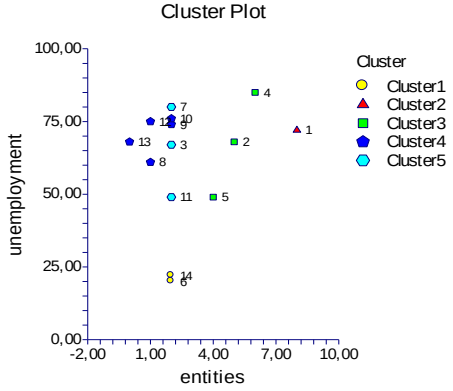
Figure 3: a)-h): Scatter plots for selected pairs of indicators.

a)

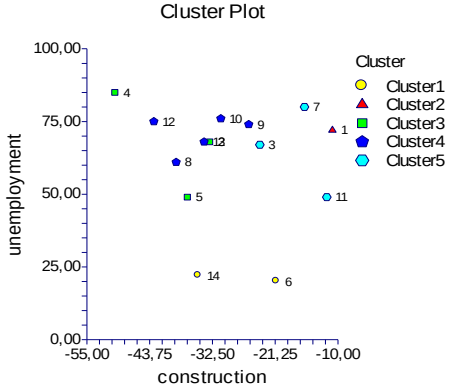
b)

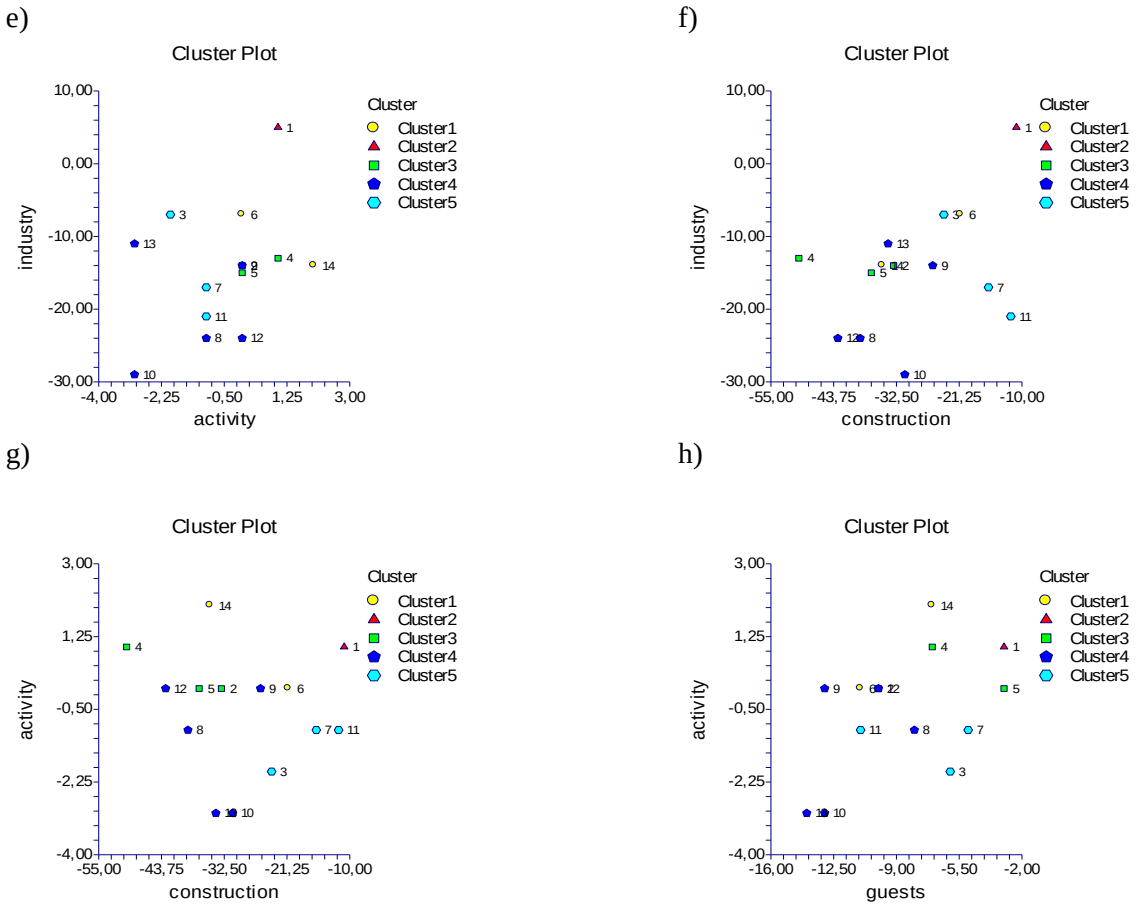


c)



d)





Source: author.

Figure 4. Graphical comparison of average indicator's changes, all clusters.

Source: author.

5. Conclusions

In the study fourteen regions of the Czech Republic were partitioned into five clusters on the basis of similarity in changes of six different macroeconomic indicators between the end of 2007 (before the economic crisis) and the end of 2009 (after the crisis).

Results showed regions formed following clusters: A = {Ústecký, Moravskoslezský}, B = {Prague}, C = {Středočeský, Plzeňský, Karlovarský}, D = {Královéhradecký, Pardubický, Vysočina, Olomoucký, Zlínský} and E = {Jihočeský, Jihomoravský, Liberecký}.

Globally, regions in clusters A, B and E were less affected and regions in clusters C and D more affected by the crisis.

Cluster A, traditionally associated with the highest unemployment, inhibition of heavy industry and coal mining in last decades, and also sharp social problems surprisingly recorded the lowest unemployment growth and was affected mainly in elementary construction production. The capital city Prague formed a solitaire cluster B, demonstrating its exceptional place in the Czech economics. It suffered heavy unemployment growth, but sales of own goods and services incidental to industry, measure of economic activity and elementary construction production expanded in spite of the recession. Cluster C experienced the largest downturn in elementary construction production, with Plzeňský region hit by the largest unemployment growth and construction downturn among all regions. Cluster D suffered by the second highest unemployment growth, the largest decline of sales of own goods and services incidental to industry and the second highest fall of elementary construction production. These findings put regions of cluster D (along with Plzeňský region from cluster C) among the most severely affected regions by the crisis. Cluster E recorded the second largest decline in sales of own goods and services incidental to industry, in other indicators it was rather under average.

Presented results can be used as guidance for policymakers and stimulation to collaborate between regions borders, to adopt similar policies and to coordinate measures necessary to overcome crisis consequences, as well as to get inspired by regions which policy was more successful. In the future, the scope of the study is going to be broadened to include regions of the Slovak Republic and Poland.

References

Czech Statistical Office (*CSO*). (2010). Regional Yearbooks. On-line, cit. [2010-09-30]. Available from WWW: <http://www.czso.cz/csu/redakce.nsf/i/rocenky_souhrn>.

Eurostat. (2010). National accounts. On-line, cit. [2010-09-30]. Available from WWW: <http://epp.eurostat.ec.europa.eu/portal/page/portal/national_accounts/data/main_tables>.

HARTIGAN, J. A., WONG, M. A. (1979). A k-means clustering algorithm: Algorithm AS136. *Appl. Statistics*, 28, p. 100-108.

LLOYD, S. P. (1982). Least squares quantization in PCM. In: *IEEE Transactions on Information Theory* 28 (2), p. 129-137.

MACQUEEN, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. In: *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, p. 281-297.

MARTINČÍK, D. (2008). Ekononomicko-sociální úroveň krajů – komplexní srovnávací analýza. *E+M Ekonomie a Management*, roč. 11, č. 1, p. 14-24.

MELOUN, M., MILITKÝ, J. (2006). *Kompendium statistického zpracování dat*. 2. vydání. Prague: Academia. ISBN: 80-200-1396-2.

MIELCOVÁ, E. (2010). Computer Estimations of Expected Unemployment Rate Decrease at the End of Recession - the Case of Moravian-Silesian Region. In: *6th International Scientific Symposium on Business administration: Global Economic Crisis and Changes*. ISBN 978-80-7248-594-9.

NCSS. (2010). Statistical and power analysis software home page. On-line, cit. [2010-09-30]. Available from WWW: <http://www.ncss.com/ncss.html>.

URBANČÍKOVÁ, N., BURGER, P. (2010). The Level of Regionalization of Innovation Policies and Their Impact on Innovation Performance of Regions. *E+M Ekonomie a Management*, roč. 13, č. 1, p. 23-36.