EVALUATING OF THE RESULTS OF CLUSTERING IN PRACTICAL ECONOMIC TASKS

Tomas Loster – Tomas Pavelka

Abstract

Cluster analysis is a popular multivariate statistical method, whose objective is to create groups of objects. The basic objective of cluster analysis is the classification of objects into groups, in which are the objects the most similar. The key role take distances measures (similarities) between objects. There are many distance measures in the current literature and there is not strictly defined, which clustering method must be used. In some cases, for example, in current literature is stated that Ward´s method is usually used in conjunction with the squared Euclidean distance measures. Given that the different methods combined with various measure of distance may give different results, so it is appropriate to evaluate the results of clustering. Evaluate the results of clustering can be from two perspectives. Clustering methods may be compared with each other or optimal number of clusters can be determined. The main aim of this paper is to describe some methods for determining the optimal number of clusters and their application to practical economic problems - to determine the optimal number of clusters of EU countries. We used some macroeconomics indicators and we compared clusters of EU countries in two periods. One period is before the global financial crisis and the second period is during the crisis period. Data used in cluster analysis come from EUROSTAT database.

Key words: clustering, evaluating of clustering, methods, macroeconomics indicators

JEL Code: C 38, C 40, E 20

Introduction

Cluster analysis is a popular multivariate statistical method, whose objective is to create groups of objects, see Loster (2011), Gan (2007). These objects can be customers, companies or countries. Cluster analysis can be used in many areas. For example can be used for modelling of incomes see Bilkova (2011, 2012), for analysing of unemployment a wages see Loster (2011), Pavelka (2011) or for clustering of households from diferent views, see Miskolczi (2011), Rezankova (2013), Zelinsky (2012). Cluster analysis is very often used in demography, see Megyesiova (2011, 2012) or in agriculture see Simpach (2012). The basic objective of cluster analysis is the classification of objects into groups, in which are the

objects the most similar. The key role take distances measures (similarities) between objects. There are many distance measures in the current literature and there is not strictly defined, which clustering method must be used. In some cases, for example, in current literature is stated that Ward´s method is used only in conjunction with the squared Euclidean distance measures. Given that the different methods combined with various measure of distance may give different results, so it is appropriate to evaluate the results of clustering. Evaluate the results of clustering can be from two perspectives. Clustering methods may be compared with each other or optimal number of clusters can be determined, see Gan (2007). We used software SYSTAT 13, Statistica 8 and SPSS 20 for clustering and it´s evaluation.

1 Clustering methods

In current literature there are described many methods and measures, see Gan (2007). Measuring of the similarity of objects when they are characterized by quantitative variables is based on distances of objects. In the current literature there are a lot of distance measures such as Hamming, Minkowski Mahalanobis or Euclidean distance. In this paper, we used only Euclidean distance, respectively the second power.

Euclidean distance (also called geometric metric) represents the length of the hypotenuse of a right triangle, see Rezankova (2009). Calculation of this rate is based on the Pythagorean theorem, and is calculated according to formula (1).Let $\mathbf{x}_i$ and $\mathbf{x}_j$ are two objects which are represented by a column vector. Calculation of this measure distance between *i*-th and *j*-th object is based on the Pythagorean theorem and is calculated by the formula:

$$D_E(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^{t} \left(x_{il} - x_{jl}\right)^2} \tag{1}$$

where *t* is the number of quantitative variables that characterize objects.

In the current literature there are various ways of classification of methods of cluster analysis. Basic classification divides methods into hierarchical and non-hierarchical methods of clustering.

Hierarchical clustering methods are based on a hierarchical arrangement of objects into clusters. Graphical representation of process of clustering of objects is dendrogram. Hierarchical methods are often used, especially due to their simple applicability in various statistical programs. Hierarchical clustering methods may be suitable as the basis for non-

hierarchical methods. The application of different clustering methods on the same object described by identical properties can give different results. As stated in Halkidi (2001), "You can not say a priori which method is best for a given problem. Commonly, the nearest neighbor method is the least desirable method and average distance, respectively. Ward's method is in many cases the best. "

Non-hierarchical clustering, unlike hierarchical clustering methods are not centered on the creation of tree (dendrogram), but focus on the classification of objects to a previously known number of clusters. First we need to establish the initial decomposition of objects into clusters and then using iterative techniques and methods to improve the original decomposition. In this group of methods, the progressive improvement of decomposition objects may transfer an object from one cluster to another.

In the current literature there are a lot of clustering methods such as single linkage, complete linkage, the method of the average distance (average linkage, Sokal-Sneath´s method centroidní method (Gower method) Ward´s method, etc. In this paper, we used only Ward´s method.

Ward's method (Ward - Wishart method)

Ward's method is based on minimizing of the heterogeneity of clusters, ie clusters are created by maximizing intra homogenty, see Gan (2007). As a measure of homogeneity of the clusters is intragroup sum of squares of diferences of values from the average of cluster, denoted G_1, called the Ward's criterion, and is calculated using the formula:

$$G_1 = \sum_{h=1}^{k}\sum_{i=1}^{n_h}\sum_{j=1}^{t}\left(x_{hij} - \bar{x}_{hj}\right)^2 , \tag{2}$$

kde k is the number of clusters,,

n_h is the number of objects in the h-th cluster,

t is the number of variables describing objects.

The criterion for joining clusters based on the idea that in each step of clustering was minimal increase in G_1, thus minimizes the expression:

$$\Delta G_1 = \sum_{i=1}^{g}\sum_{j=1}^{t}\left(x_{gij} - \bar{x}_{gj}\right)^2 - \sum_{i=1}^{h}\sum_{j=1}^{t}\left(x_{hij} - \bar{x}_{hj}\right)^2 - \sum_{i=1}^{h'}\sum_{j=1}^{t}\left(x_{h'ij} - \bar{x}_{h'j}\right)^2 . \tag{3}$$

When using the Ward's clustering methods is usually used with square of Euclidean distance.

2 Evaluating of clustering

CHF index (also called pseudo-F index) is used to determine the optimal number of clusters. High values of this index represent a well-separated clusters, ie looking for the maximum value of this index, which indicates the optimal number of clusters k^*, ie

$$I_{\mathrm{CHF}}(k^*) = \max_{2 \le k \le n-1} I_{\mathrm{CHF}}(k). \tag{4}$$

Davies-Bouldin DB index is a measure whose values depend on the chosen method of clustering, see Rezankova (2009). The resulting Davies-Bouldinův index is calculated as the arithmetic average of the maximum rates of similarity clusters. According to Davies-Bouldinova index is the best that split objects into clusters in which the value of this index is minimal. At small values of this index are individual clusters of compact and well separated.

The optimum number of clusters is considered to be the value of k^* for which the value of the Davies-Bouldina index minimum within the predetermined maximum number of clusters, ie

$$I_{\mathrm{DB}}(k^*) = \min_{2 \le k \le n-1} I_{\mathrm{DB}}(k). \tag{5}$$

Dunn´s separation index is used to find a compact and well separated clusters, see Gan (2007). High values of the index indicate a separation Dunn compact and well-separated clusters. To determine the optimal number of clusters k * we look for the maximum of Dunn´s separation index, which determine number of clusters, ie

$$I_{\mathrm{D}}(k^*) = \max_{2 \le k \le n-1} I_{\mathrm{D}}(k). \tag{6}$$

3 Clusters of EU Countries

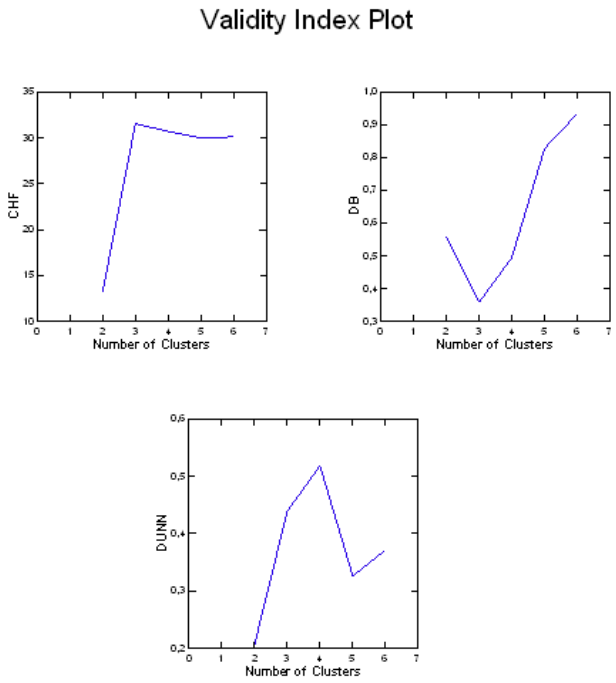
In this section we deal with the creation of clusters of EU countries. For clustering, we selected a total of 12 macroeconomic indicators, which include GDP per capita in purchasing power standards, Total Receipts from taxes and social Contributions as a Percentage of GDP, Government deficit as a Percentage of GDP, Current Account Balance as a Percentage of GDP, Total social Expenditure as a Percentage of GDP, Annual inflation rate, real GDP growth rate, trade balance in billion EUR etc. The data comes from Eurostat. It investigates two seasons. The first period is before the the global financial crisis - year 2008 and a second term year 2011. This is the period in which the still ongoing global financial crisis and are the

latest available period in the Eurostat database. Used statistical software to perform cluster analysis are: to find the optimal number of clusters is used SYSTAT version 13, we used software Statistica version 8 for dendrogram, for the analysis of the obtained clusters is used SPSS software version on the 20. For the clustering was used in both cases Ward's method in conjunction with the square of Euclid distance, because this method often gives the best results clustering, see above.

First we searched for the optimal number of clusters of EU. We used the above validation coefficients. In the literature, however, as stated above, there is no clear rule to determine the strict application of the coefficients. Generally, it is said, in the case that the results are supported by the results of several coefficients simultaneously, it is possible to regard these results as objective. However, these results are not strictly binding.

In our case, we used the CHF, DB and Dunn´s index, see picture 1. From these coefficients we found as a optimal number of clusters 3 or 4 clusters. We decided for a larger number of clusters due to obtain more detailed structure of the country and we could explore further the effects of the global financial crisis on the structure of individual clusters.

Picture 1: Validity indices



Source: Own calculations, SYSTAT

For each cluster will be set basic descriptive characteristics of the selected indicators, such as the minimum and maximum value, average value and standard deviation of the indicators.

YEAR 2008

In 2008, under the above indicators we created 4 clusters, and Luxembourg due to the significantly different values of their parameters were included in a separate cluster, thus resulting number of clusters in 2008 is 5. Assign of countries into clusters is:

Cluster 1:

Belgium, Denmark, Finland, Ireland, Germany, Netherlands, Austria, United Kingdom, Sweden

Cluster 2:

France, Italy, Cyprus, Greece, Slovenia, Spain

Cluster 3:

Bulgaria, Lithuania, Latvia, Hungary, Poland, Romania

Cluster 4:

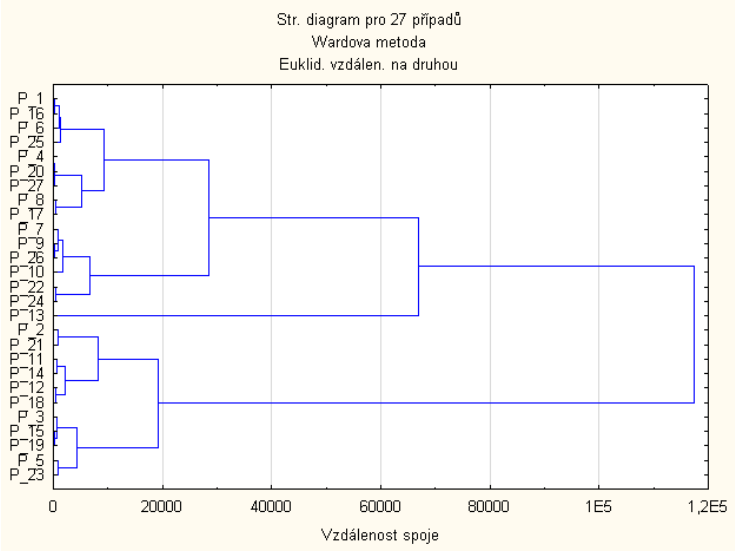
Czech Republic, Estonia, Malta, Portugal, Slovakia

Cluster 5:

Luxembourg

The process of clustering of states using Ward's method is shown in Figure 1.

Fig. 1: Dendrogram of countries, 2008



Source: Own calculations, STATISTICA

There are more detailed values of selected indicators in clusters in 2008. Table 2 contains a summary of characteristics for all 27 countries together. Table 4 contains the values of

selected aggregate characteristics in separated clusters. Table 3 contains a comparison of the values for each cluster with the EU.

Table 2: Desrciptive Statistics for the whole EU, 2008

	N	Minimum	Maximum	Mean	Std. Deviation
GDP	27	10900,00	65900,00	24766,6667	10678,95126
Taxes	27	28,50	48,40	37,1926	5,82976
Inflation	27	2,20	15,30	5,3296	3,25350
GDP_growth	27	-4,20	7,30	1,2704	2,84251
Employment	27	55,30	77,90	66,3259	5,92917
Long_unemployment	27	,40	6,60	2,2889	1,35059
Valid N (listwise)	27				

Source: Own calculations, SPSS

Table 3: Differences between cluster´s statistics and the EU statistics, 2008

Cluster	Indicator	Mean	Mean EU 27	Difference
1	GDP	30655,5556	24766,6667	5888,889
	Taxes	41,8556	37,1926	4,663
	Inflation	3,3556	5,3296	-1,974
	GDP_growth	0,1222	1,2704	-1,148
	Employment	71,5778	66,3259	5,252
	Long_unemployment	1,6444	2,2889	-0,644
2	GDP	24900,0000	24766,6667	133,3333333
	Taxes	38,4000	37,1926	1,207
	Inflation	4,1500	5,3296	-1,180
	GDP_growth	1,0667	1,2704	-,204
	Employment	64,8667	66,3259	-1,459
	Long_unemployment	2,3333	2,2889	0,044
3	GDP	13783,3333	24766,6667	-10983,333
	Taxes	32,3167	37,1926	-4,87592593
	Inflation	9,4167	5,3296	4,087
	GDP_growth	3,1833	1,2704	1,913
	Employment	61,9667	66,3259	-4,359
	Long_unemployment	2,4000	2,2889	0,111
4	GDP	18960,0000	24766,6667	-5806,66667
	Taxes	33,0000	37,1926	-4,193
	Inflation	5,6400	5,3296	0,310
	GDP_growth	1,6800	1,2704	,410
	Employment	64,4400	66,3259	-1,886
	Long_unemployment	3,4000	2,2889	1,111
5	GDP	65900,0000	24766,6667	41133,33333
	Taxes	38,2000	37,1926	1,007
	Inflation	4,1000	5,3296	-1,230
	GDP_growth	-0,7000	1,2704	-1,970
	Employment	63,4000	66,3259	-2,926
	Long_unemployment	1,6000	2,2889	-0,689

Source: Own calculations

Cluster 1

Countries in cluster 1 are characterized by the following properties: above-average value of GDP per capita in purchasing power standards, above-average value of Total Receipts from taxes and social Contributions as a Percentage of GDP, below-average value of Annual inflation rate, below-average real GDP growth rate, above-average value of Employment rate, below -average value of Long unemployment – compared with the EU average.

Cluster 2

Countries in cluster 2 are characterized by the following features: Above-average value of GDP per capita in purchasing power standards, above-average value of Total Receipts from taxes and social Contributions as a Percentage of GDP, below-average value of Annual inflation rate, below-average value of real GDP growth rate, below-average value of Employment rate, approximately the same value of Long unemployment – compared with the EU average.

Cluster 3

Countries in cluster 3 are characterized by the following features: a below-average value of GDP per capita in purchasing power standards, below-average value of Total Receipts from taxes and social Contributions as a Percentage of GDP, above-average value of Annual inflation rate, above-average value of Real GDP Growth rate, below-average value of Employment rate, approximately the same value of Long unemployment – compared with the EU average.

Cluster 4

Countries in cluster 4 are characterized by the following features: a below-average value of GDP per capita in purchasing power standards, below-average value of Total Receipts from taxes and social Contributions as a Percentage of GDP, approximately equal to the value of Annual inflation rate, below-average value of Real GDP growth rate, below-average of Employment rate, above-average value of Long unemployment – compared with the EU average.

Cluster 5

Country in cluster 5 (Luxembourg) is characterized by the following features: above-average value of GDP per capita in purchasing power standards, above-average value of Total

Receipts from taxes and social Contributions as a Percentage of GDP, below-average value of Annual inflation rate, below-average value of Real GDP growth rate, below-average value of Employment rate, below-average value of Long unemployment – compared with the EU average.

Table 4: Descriptive Statistics for clusters of countries, 2008

cluster		N	Minimum	Maximum	Mean	Std. Deviation
1	GDP	9	28200,00	33600,00	30655,5556	1850,75060
	Taxes	9	31,00	48,40	41,8556	5,28444
	Inflation	9	2,20	4,50	3,3556	,65786
	GDP_growth	9	-2,10	1,80	,1222	1,31223
	Employment	9	62,40	77,90	71,5778	4,76623
	Long_unemployment	9	,40	4,00	1,6444	1,20738
	Valid N (listwise)	9				
2	GDP	6	22700,00	26700,00	24900,0000	1658,91531
	Taxes	6	33,50	44,80	38,4000	4,56727
	Inflation	6	3,20	5,50	4,1500	,80187
	GDP_growth	6	-1,20	3,60	1,0667	1,99967
	Employment	6	58,70	70,90	64,8667	4,41573
	Long_unemployment	6	,50	3,60	2,3333	1,11116
	Valid N (listwise)	6				
3	GDP	6	10900,00	16000,00	13783,3333	2046,86753
	Taxes	6	28,50	40,10	32,3167	4,29018
	Inflation	6	4,20	15,30	9,4167	4,13058
	GDP_growth	6	-3,30	7,30	3,1833	3,92551
	Employment	6	56,70	68,60	61,9667	4,42207
	Long_unemployment	6	1,20	3,60	2,4000	,82219
	Valid N (listwise)	6				
4	GDP	5	17200,00	20200,00	18960,0000	1262,14104
	Taxes	5	29,10	35,70	33,0000	2,64197
	Inflation	5	2,70	10,60	5,6400	3,06562
	GDP_growth	5	-4,20	5,80	1,6800	3,88806
	Employment	5	55,30	69,80	64,4400	5,82349
	Long_unemployment	5	1,70	6,60	3,4000	1,98368
	Valid N (listwise)	5				
5	GDP	1	65900,00	65900,00	65900,0000	
	Taxes	1	38,20	38,20	38,2000	
	Inflation	1	4,10	4,10	4,1000	
	GDP_growth	1	-,70	-,70	-,7000	
	Employment	1	63,40	63,40	63,4000	
	Long_unemployment	1	1,60	1,60	1,6000	
	Valid N (listwise)	1				

Source: Own calculations, SPSS

YEAR 2011

In 2011, under the above indicators we created 4 clusters. Assign of countries into clusters is:

Cluster 1:

Belgium, Bulgaria, Czech Repulic, Estonia, Lithuania, Latvia, Hungary, Slovakia, Slovenia

Cluster 2:

Denmark, Finland, Cyprus, Luxembourg, Portugal, Austria, Sweden

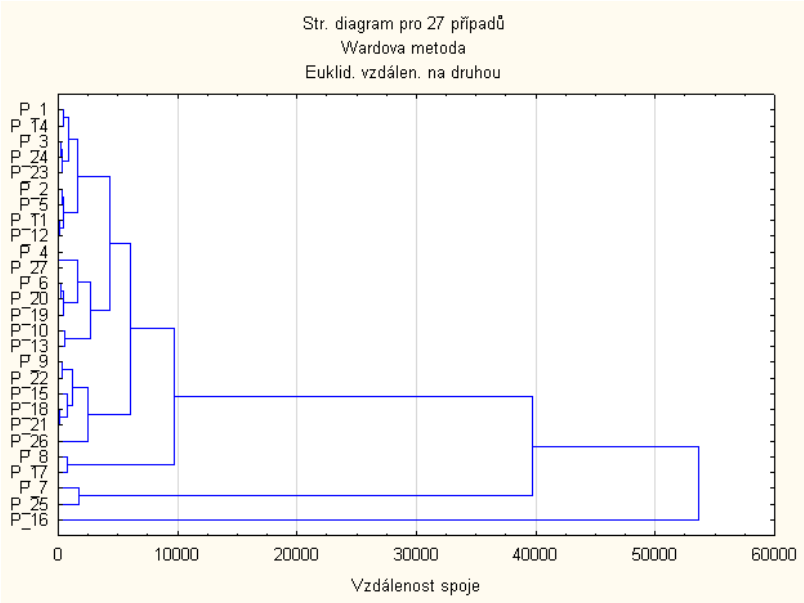
Cluster 3:

Italy, Malta, Poland, Romania, Greece, Spain

Cluster 4:

France, Ireland, Germany, Netherlands, United Kingdom

Fig. 2: Dendrogram of countries, 2011



Source: Own calculations, STATISTICA

There are more detailed values of selected indicators in clusters in 2008. Table 5 contains a summary of characteristics for all 27 countries together. Table 6 contains the values of selected aggregate characteristics in separated clusters. Table 7 contains a comparison of the values for each cluster with the EU.

Table 5: Desrciptive Statistics for the whole EU, 2011

	N	Minimum	Maximum	Mean	Std. Deviation
GDP	27	14800,00	68400,00	25915,9600	10692,016
Taxes	27	26,10	48,40	36,2852	6,414
Inflation	27	-1,60	6,10	2,0593	1,580
GDP_growth	27	-7,10	8,30	1,8926	2,686
Long_unemployment	27	1,10	9,20	4,1370	2,453
Employment	27	55,40	74,70	63,7037	5,866
Valid N (listwise)	27				

Source: Own calculations, SPSS

Table 6: Descriptive Statistics for the clusters of countries, 2011

Cluster		N	Minimum	Maximum	Mean	Std. Deviation
1	GDP	9	14800,00	29800,00	19287,5000	4751,973
	Taxes	9	26,10	45,90	32,8778	6,461
	Inflation	9	-1,20	4,70	1,8556	1,655
	GDP_growth	9	,60	8,30	3,3889	2,586
	Long_unemployment	9	3,00	9,20	5,9222	2,317
	Employment	9	55,40	66,20	60,5778	3,418
	Valid N (listwise)	9				
2	GDP	7	19400,00	68400,00	33714,1429	16039,594
	Taxes	7	35,00	48,40	41,2571	5,009
	Inflation	7	1,40	2,80	2,0429	,513
	GDP_growth	7	-1,60	3,70	1,5429	1,757
	Long_unemployment	7	1,10	6,30	2,1286	1,861
	Employment	7	65,20	73,40	69,4857	3,316
	Valid N (listwise)	7				
3	GDP	6	16200,00	25300,00	21520,0000	3700,270
	Taxes	6	27,10	42,60	33,9000	5,086
	Inflation	6	1,60	6,10	3,1833	1,808
	GDP_growth	6	-7,10	4,30	,3000	3,901
	Long_unemployment	6	2,50	7,30	4,3000	1,851
	Employment	6	56,10	59,60	58,2167	1,399
	Valid N (listwise)	5				
4	GDP	5	27000,00	32900,00	30000,0000	2739,526
	Taxes	5	30,10	45,60	38,3200	5,557
	Inflation	5	-1,60	3,30	1,1000	1,771
	GDP_growth	5	,90	3,00	1,6000	,846
	Long_unemployment	5	1,20	6,70	3,5400	2,043
	Employment	5	60,00	74,70	67,8200	5,879
	Valid N (listwise)	5				

Source: Own calculations, SPSS

Table 7: Differences between cluster´s statistics and the EU statistics, 2011

Cluster	Indicator	Mean	Mean EU 27	Difference
1	GDP	19287,5000	25915,9600	-6628,460
	Taxes	32,8778	36,2852	-3,407
	Inflation	1,8556	2,0593	-0,204
	GDP_growth	3,3889	1,8926	1,496
	Long_unemployment	5,9222	4,1370	1,785
	Employment	60,5778	63,7037	-3,126
2	GDP	33714,1429	25915,9600	7798,1829
	Taxes	41,2571	36,2852	4,972
	Inflation	2,0429	2,0593	-0,016
	GDP_growth	1,5429	1,8926	-,350
	Long_unemployment	2,1286	4,1370	-2,008
	Employment	69,4857	63,7037	5,782
3	GDP	21520,0000	25915,9600	-4395,960
	Taxes	33,9000	36,2852	-2,3851852
	Inflation	3,1833	2,0593	1,124
	GDP_growth	,3000	1,8926	-1,593
	Long_unemployment	4,3000	4,1370	0,163
	Employment	58,2167	63,7037	-5,487
4	GDP	30000,0000	25915,9600	4084,04
	Taxes	38,3200	36,2852	2,035
	Inflation	1,1000	2,0593	-0,959
	GDP_growth	1,6000	1,8926	-,293
	Long_unemployment	3,5400	4,1370	-0,597
	Employment	67,8200	63,7037	4,116

Source: Own calculations

Obtained results:

Cluster 1

Countries in cluster 1 are characterized by the following properties: below-average value of GDP per capita in purchasing power standards, below-average value of Total Receipts from taxes and social Contributions as a Percentage of GDP, below-average value of Annual inflation rate, above-average value of real GDP growth rate, above-average value of Long unemployment rate, below-average value of Employment rate – compared with the EU average.

Cluster 2

Countries in cluster 2 are characterized by the following features: Above-average value of GDP per capita in purchasing power standards, above-average value of Total Receipts from taxes and social Contributions as a Percentage of GDP, approximately the same value of

Annual inflation rate, below-average value of real GDP growth rate, above-average value of Long unemployment rate, Employment rate – compared with the EU average.

Cluster 3

Countries in cluster 3 are characterized by the following features: a below-average value of GDP per capita in purchasing power standards, below-average value of Total Receipts from taxes and social Contributions as a Percentage of GDP, above-average value of Annual inflation rate, below-average value of Real GDP Growth rate, approximately the same value of Long unemployment rate, below-average value of Employment rate – compared with the EU average.

Cluster 4

Countries in cluster 4 are characterized by the following features: above-average value of GDP per capita in purchasing power standards, above-average value of Total Receipts from taxes and social Contributions as a Percentage of GDP, below-average value of Annual inflation rate, below-average value of Real GDP growth rate, below-average value of Long unemployment rate, above-average value of Employment rate – compared with the EU average.

Conclusion

Cluster analysis is a popular multivariate statistical method that is used to classify different objects. The results of clustering depend on the using methods and on using measures of similarity or distance and therefore is suitable evaluate obtained results of clustering. In our paper we showed clustering of EU countries using Ward's method and we showed how we canto determine the optimal number of clusters of EU countries. We characterized obtained clusters in 2 period (2008 and 2011), ie the period befor of the financial crisis and during the financial crisis. In each year, we reported the average values of indicators in the clusters of countries and we compared them with the EU average.

We compared EU-27 between 2008 and 2011. Between these two years occured to: increase of value of GDP per capita in purchasing power standards, decline of value of Total Receipts from taxes and social Contributions as a Percentage of GDP, decline of value of Annual inflation rate, increase of value of Real GDP Growth rate, increase of value of Long Unemployment rate, decline of value of Employment rate. It is clear that the global financial

crisis affected adversely in some indicators, such as in long-term unemployment rate or average deceleration of Real GDP Growth Rate.

Acknowledgment

This article was created with the help of the Internal Grant Agency of University of Economics in Prague No. 6/2013 under the title „Evaluation of results of cluster analysis in Economic problems."

References

Bilkova, D. (2011). *Modelling of income and wage distribution using the method of l-moments of parameter estimation* . In Loster Tomas, Pavelka Tomas (Eds.), *International Days of Statistics and Economics* (pp. 40-50). ISBN 978-80-86175-77-5.

Bilkova, D. (2012). *Development of wage distribution of the czech republic in recent years by highest education attainment and forecasts for 2011 and 2012*. In Loster Tomas, Pavelka Tomas (Eds.), *6th International Days of Statistics and Economics* (pp. 162-182). ISBN 978-80-86175-86-7.

Gan, G., Ma Ch., Ww J. (2007): *Data Clustering Theory, Algorithms, and Applications*, ASA, Philadelphia.

Halkidi, M., Vazirgiannis (2001), M.: *Clustering validity assessment: Finding the optimal partitioning of a data set*, Proceedings of the IEEE international conference on data mining, s. 187-194.

Loster, T., & Langhamrova, J. (2011). *Analysis of long-term unemployment in the czech republic* . In Loster Tomas, Pavelka Tomas (Eds.), *International Days of Statistics and Economics* (pp. 307-316). ISBN 978-80-86175-77-5.

Megyesiova, S., & Lieskovska, V. (2011). *Recent population change in europe*. In Loster Tomas, Pavelka Tomas (Eds.), *International Days of Statistics and Economics* (pp. 381-389). ISBN 978-80-86175-77-5.

Megyesiova, S., & Lieskovska, V. (2012). *Are europeans living longer and healthier lives?*. In Loster Tomas, Pavelka Tomas (Eds.), *6th International Days of Statistics and Economics* (pp. 766-775). ISBN 978-80-86175-86-7.

Miskolczi, M., Langhamrova, J., & Fiala, T. (2011). *Unemployment and gdp*. In Loster Tomas, Pavelka Tomas (Eds.), *International Days of Statistics and Economics* (pp. 407-415). ISBN 978-80-86175-77-5.

Pavelka, T. (2012). *The minimum wage in the czech republic – the instrument for motivation to work?*. In Loster Tomas, Pavelka Tomas (Eds.), *6th International Days of Statistics and Economics* (pp. 903-911). ISBN 978-80-86175-86-7.

Řezanková, H., Húsek, D., Snášel, V.: *Shluková analýza dat,* 2. vydání, Professional Publishing, Praha, 2009.

Rezankova, H., & Loster, T. (2013). Shlukova analyza domacnosti charakterizovanych kategorialnimi ukazateli. *E+M. Ekonomie a Management*, *16*(3), 139-147. ISSN: 1212-3609.

Simpach, O. (2012). *Statistical view of the curent situation of beekeeping in the czech republic*. In Loster Tomas, Pavelka Tomas (Eds.), *6th International Days of Statistics and Economics* (pp. 1054-1062). ISBN 978-80-86175-86-7.

Zelinsky, T., & Stankovicova, I. (2012). *Spatial aspects of poverty in slovakia*. In Loster Tomas, Pavelka Tomas (Eds.), *6th International Days of Statistics and Economics* (pp. 1228-1235). ISBN 978-80-86175-86-7.

Contact

Ing. Tomas Loster, Ph.D.

University of Economics in Prague,

Department of Statistics and probability,

W. Churchill Sg. 4,

130 67 Prague 3

Tomas.Loster@vse.cz

doc. Ing. Tomas Pavelka, Ph.D.

University of Economics in Prague,

Department of Microeconomics,

W. Churchill Sg. 4,

130 67 Prague 3

PavelkaT@vse.cz