Examination of European Union Economic Cohesion: A Cluster Analysis Approach

Jiří Mazurek

School of business administration in Karviná

8. January 2014

Examination of European Union Economic Cohesion: A Cluster Analysis Approach

Jiří Mazurek
School of Business Administration in Karviná,
The Czech Republic,
mazurek@opf.slu.cz

Abstract
In the past years majority of EU members experienced the highest economic decline in their modern history, but impacts of the global financial crisis were not distributed homogeneously across the continent. The aim of the paper is to examine a cohesion of European Union (plus Norway and Iceland) in terms of an economic development of its members from the 1st of January 2008 to the 31st of December 2012. For the study five economic indicators were selected: GDP growth, unemployment, inflation, labour productivity and government debt. Annual data from Eurostat databases were averaged over the whole period and then used as an input for a cluster analysis. It was found that EU countries were divided into six different clusters. The most populated cluster with 14 countries covered Central and West Europe and reflected relative homogeneity of this part of Europe. Countries of Southern Europe (Greece, Portugal and Spain) shared their own cluster of the most affected countries by the recent crisis as well as the Baltics and the Balkans states in another cluster. On the other hand Slovakia and Poland, only two countries that escaped a recession, were classified in their own cluster of the most successful countries.

Keywords*: cluster analysis, cohesion, European Union, economics.*

JEL classification: C38, O11, O52, O57.

1. Introduction

From the start of the global financial crisis in 2007 in the USA European countries experienced a difficult period of time both from economic and social point of view as almost all national economies fell into recessions during 2008 and 2009; in 2010 they returned to the growth, but later another surge of recessions took place which lasts until present days. However, this development was not homogenous across Europe as some countries were able to overcome negative global economic trends (such as Slovakia or Poland), while other countries experienced the deepest economic downturns in their modern history (Baltics states, Greece, Ireland, Portugal, etc.), especially in the Southern wing of EU and on European periphery, see e.g. Aiginger (2011, 2013), Mian and Sufi (2010), Verick and Islam (2010), Beblavý et al. (2011), Mazurek and Mielcová (2013) or Mazurek (2013).
Though crisis' impacts differed across the continent, some regions or group of countries might experience the similar economic development. The aim of this paper is to examine European Union economic cohesion in the recent five years (from the 1st of January 2008 to the 31st of December 2012), where a term 'cohesion' is used thereinafter as a synonym of 'similarity', 'compactness'or 'homogeneity' among countries. By this scrutiny claims about southern Europe and periphery problems can be empirically tested, and also other regions with a similar economic development can be found. Because countries are to be grouped in accord with their economic performance a cluster analysis was chosen as an appropriate method for this task. For the clustering the following five economic indicators were selected: real GDP PPP growth rate, inflation rate, unemployment rate, labour productivity per person employed and the difference between government net borrowing and net lending.
These indicators belong among the most important in economics and also they are easily available for all EU (27) countries. The data for the examined period (the 1st of January 2008

to the 31st of December 2012) were obtained from Eurostat databases. Because the data were available also for Iceland and Norway, these two countries were included in the study as well. The paper is organized as follows: in section 2 the data and the method of study is described, section 3 provides results with a brief discussion, and conclusions close the article.

2. The data and the method

2. 1. The data

The data for presented study were retrieved from Eurostat economic databases, see Eurostat (2013). The data for each of 29 countries (EU27+Norway and Iceland) include:

- Real GDP growth rates (volumes) – percentage change on previous year, abbreviated as GDP thereinafter, source: Eurostat (2013a),
- Unemployment rate – annual average, (in %), not seasonally adjusted, (UNEMP), Eurostat (2013b),
- HICP Inflation rate– annual average rate (in %), (INFL), Eurostat (2013c),
- Labour productivity per person employed (EU27 = 100%) , (PROD), Eurostat (2013d),
- Net lending (+)/Net borrowing (-) under the EDP (Excessive Deficit Procedure) as a percentage of GDP (DEBT), Eurostat (2013e).

For all countries annual data (five values) from the 1st January 2008 to the 31st December 2012 were obtained and averaged for the whole period, see Table 1.

Table 1. Average values of selected economic indicators of EU countries, Iceland and Norway from the 1st of January 2008 to the 31st of December 2012.

Country	**Real GDP PPP** (%)	**UNEMP** (%)	**INFL** (%)	**PROD** (EU=100%)	**DEBT** (%)
Belgium	0.44	7.6	2.56	127.46	-3.6
Bulgaria	0.74	9.26	4.66	41.4	-1.7
Czech Republic	0.38	6.42	2.74	73.88	-4.06
Denmark	-0.88	6.42	2.4	109.74	-1.58
Germany	0.8	6.76	1.76	106.24	-1.62
Estonia	-0.44	11.78	4.56	67.56	-0.76
Ireland	-1.46	12.34	0.58	136.52	-14.6
Greece	-4.34	14.36	2.86	93.78	-10.94
Spain	-0.92	19.22	2.28	107.22	-9.08
France	0.1	9.36	1.9	116.04	-5.6
Italy	-1.4	8.4	2.42	110.26	-3.9
Cyprus	0.2	7.04	2.76	91.26	-4.64
Latvia	-2.26	15.18	4.78	56.86	-5.4
Lithuania	-0.12	13.14	4.76	66.1	-5.72
Luxembourg	-0.3	4.9	2.7	163.9	0.24
Hungary	-0.98	10.16	4.86	70.7	-2.06
Malta	1.4	6.54	2.84	95.58	-3.58
Netherlands	-0.14	4.2	1.88	111.88	-3.72
Austria	0.62	4.3	2.3	115.82	-2.9
Poland	3.4	8.94	3.7	67.22	-5.6
Portugal	-1.1	11.98	1.92	75.38	-6.86
Romania	0.5	6.88	5.76	49.26	-6.02
Slovenia	-1	6.94	2.68	80.36	-4.84

Slovakia	2.02	12.78	2.66	80.26	-5.48
Finland	-0.58	7.7	2.74	109.58	-0.62
Sweden	0.96	7.78	1.88	114.28	0.36
United Kingdom	-0.62	7.38	3.28	106.66	-8.06
Iceland	-1.08	6.18	9.36	97	-8.58
Norway	0.64	3.16	1.92	154.76	13.44

Source: own calculations from Eurostat (2013) data.

2.2. The method

For the clustering analysis k-means clustering method was selected, see e.g. Hartigan (1975), Meloun and Militký (2006) or Chiang and Mirkin (2009). This method enables to divide objects into clusters (groups) where the division is based on objects' similarity or proximity, which is appropriate for the presented study, as k-means clustering will result in a set of (small) groups of countries with similar economic development in examined period.
The method was performed via statistical software NCSS2000, see NCSS (2013). The method divides a set of n observations into K clusters so that observations in each cluster are similar (close) to each other. 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}.

The method proceeds as follows, see e.g. Hartigan and Wong (1979) or Lloyd (1982):

1. The number of clusters K is specified by a user.
2. Random K clusters are generated and cluster centres (*centroids*) are computed 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 the 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, and it finds only a local minimum. To eliminate both problems repeated clustering is necessary (typically 25 runs are used). The most suitable number of clusters can be determined by various criteria such as elbow (bend)

rule, Hartigan index, Gap statistics, average silhouette, Aikake information criterion, etc., see Meloun and Militký (2006) or Chiang and Mirkin (2009).

As an addition to k-means clustering analysis also hierarchical clustering (a dendrogram) was performed.

3. Results

From the data shown in Table 1 correlation matrix of all indicators were computed, see Table 3, to ascertain how much are the indicators independent. The highest absolute value of Pearson's correlation coefficient $r = –0.539$ was found for the pair UNEMP-DEBT, other pairs of indicators were less correlated, hence all indicators can be considered rather low linearly dependent and suitable for further cluster analysis.

For k-means clustering the statistical software NCSS2000 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.

In the presented study the most suitable number of clusters was selected by the elbow rule, see Table 2 and Figure 1. According to this simple rule the optimal number of clusters is derived from the graph of percent variation PV_k given by (2) as the number for which a steep downward curve bends markedly more horizontally (in other words adding another cluster does not lower PV_k significantly). In this case the optimal value $k = 6$, so examined countries should be divided into six clusters. From the NCSS2000 programme menu 6 clusters, 5 to 15 random starts and 10 to 20 iterations were chosen for each run. The programme was run 25 times. Table 4 shows the final division of countries into clusters for the lowest percent of variation $PV_6 = 27.42$ of all runs.

Table 2. Percent of variation (2) as a function of a number of clusters. Source: own

clusters	1	2	3	4	5	6	7	8	9	10
PV	100	72.77	53.72	42.82	34.67	27.42	23.74	20.12	17.14	15.66

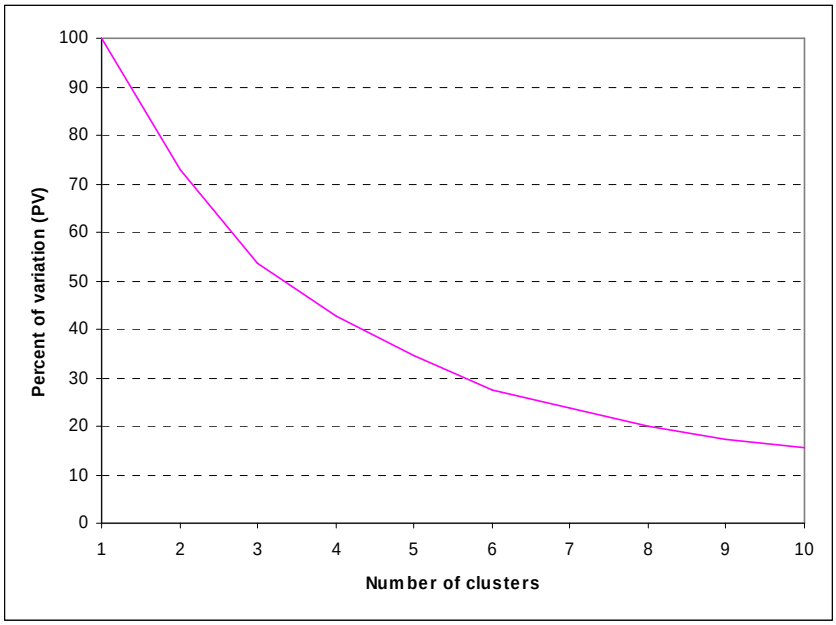


Figure 1. Percent of variation (2) as a function of the number of clusters. Source: own

Numbers in braquets in Table 4 following country codes correspond to countries' numbers in Figures 3a)-h). Geographical distribution of clusters is provided in Figure 2.

- Cluster 1 containing Luxembourg and Norway is characterized by a slight GDP growth over examined period, low unemployment, medium inflation, high productivity and high debt.
- Cluster 2 contains only one item, namely Iceland. This country experienced negative economic growth, rather low unemployment, but high inflation, under average productivity and above average debt.
- Cluster 3 includes three Baltics states, two Balkans states and Hungary. Countries associated in this cluster are characterized by by an economic decline, above average unemployment and inflation, very low productivity and under average debt.
- Cluster 4 groups together Greece, Spain, Portugal and Ireland. These countries suffered the highest economic decline, the highest unemployment and also the highest deficit in the examined period.
- Cluster 5 includes Poland and Slovakia, two countries that escaped a recession and were able to grow despite global crisis. From Table 4 it can be seen they also share some problems, namely relatively high unemployment and the second lowest productivity among all clusters.
- Cluster 6 is the most populated cluster with 14 countries mainly from Western and Central Europe. Average values of all five economic indicators in this cluster are close to global average; hence this cluster can be considered a compact 'core' of European Union countries with rather homogenous economic development.

Generally clusters 2, 3 and 4 contain countries the most affected by the crisis, while in clusters 1 and 5 the most successful countries were grouped together.
In cluster 3 transition economies of the former Soviet-bloc from the Baltics and the Balkans were placed together. Problems of these countries were already examined thoroughly in the literature, see e.g. Kattel and Raudla (2013) and Karasavvoglou and Polychronidou (2014). In general, these countries are open and vulnerable economies strongly dependent on foreign investment and demand from their more developed counterparts on the West. In times of economic boom they grow more swiftly than the rest of Europe (what can be observed until 2008), but in times of economic crisis they decline more rapidly as well.
Cluster 4 is formed by three South European (along with Ireland) countries with long-term economic problems caused mainly by large budget deficits and low productivity and competitiveness leading to deep declines in real GDP and record high unemployment rates in times of the global crisis. Also, adoption of euro worsened problems as above mentioned countries could not soften the crisis impacts by devaluation of their currencies.
Furthermore, the number of clusters and the value of percent of variation PV_k defined by relation (2) reflect homogeneity of the data as for the uniform data the number of clusters would be 1 and $PV_k = 0$, and the higher are both values the higher are differences in the data. The lower is a value of PV_k the closer are objects together in all clusters.
For the whole examined period from 2008 to 2012 the value of PV_6 was 27.42, but in 2008 PV_k was 29.02 while in 2012 PV_6 was 23.99. Hence, in 2012 examined countries in clusters were closer to each other than in 2008, which means a cohesion of European union (plus Norway and Iceland) increased in the last 5 years.
Figure 4 provides a dendrogram of countries based on the data from Table 1. According to this diagram Norway, Iceland and Greece were among the most dissimilar countries of the whole group.

Until 1989 examined countries were separated by the Iron curtain in two blocs: the West and the East under Soviet influence. Hence, it would be also interesting to find out whether this division pertains (in the sense of this study) to present days. To answer this question cluster analysis with only two clusters was performed by NCSS2000, and the result is shown in a self explanatory Figure 5.

Table 3. Correlation matrix of all indicators.

Indicator	GDP	UNEMP	INFL	PROD	DEBT
GDP	1	–0.379	–0.068	–0.062	0.356
UNEMP		1	0.034	–0.370	–0.539
INFL			1	–0.518	–0.113
PROD				1	0.271
DEBT					1

Source: own.

Table 4. Partition of all countries into six clusters with average cluster values of all indicators. (Numbers in brackets correspond to countries' numbers in Figures 3a)-h).

Cluster No.	**Countries**	**GDP (%)**	**UNEMP (%)**	**INFL (%)**	**PROD (%)**	**DEBT (%)**
1	LUX (15), NOR (29)	0.17	4.03	2.31	159.33	6.84
2	ISL (28)	-1.08	6.18	9.36	97.00	-8.58
3	BUL (2), EST (6), LAT (13), LTU (14), HUN (16), ROU (22)	-0.43	11.07	4.90	58.65	-3.61
4	IRL (7), GRE (8), ESP (9), POR (21)	-1.96	14.48	1.91	103.23	-10.37
5	POL (20), SVK (24)	2.71	10.86	3.18	73.74	-5.54
6	BEL (1), CZE (3), DEN (4), GER (5), FRA (10), ITA (11), CYP (12), MLT (17), NED (18), AUT (19), SLO (23), FIN (25), SWE (26), GBR (27)	0.02	6.92	2.44	104.93	-3.45

Source: own.

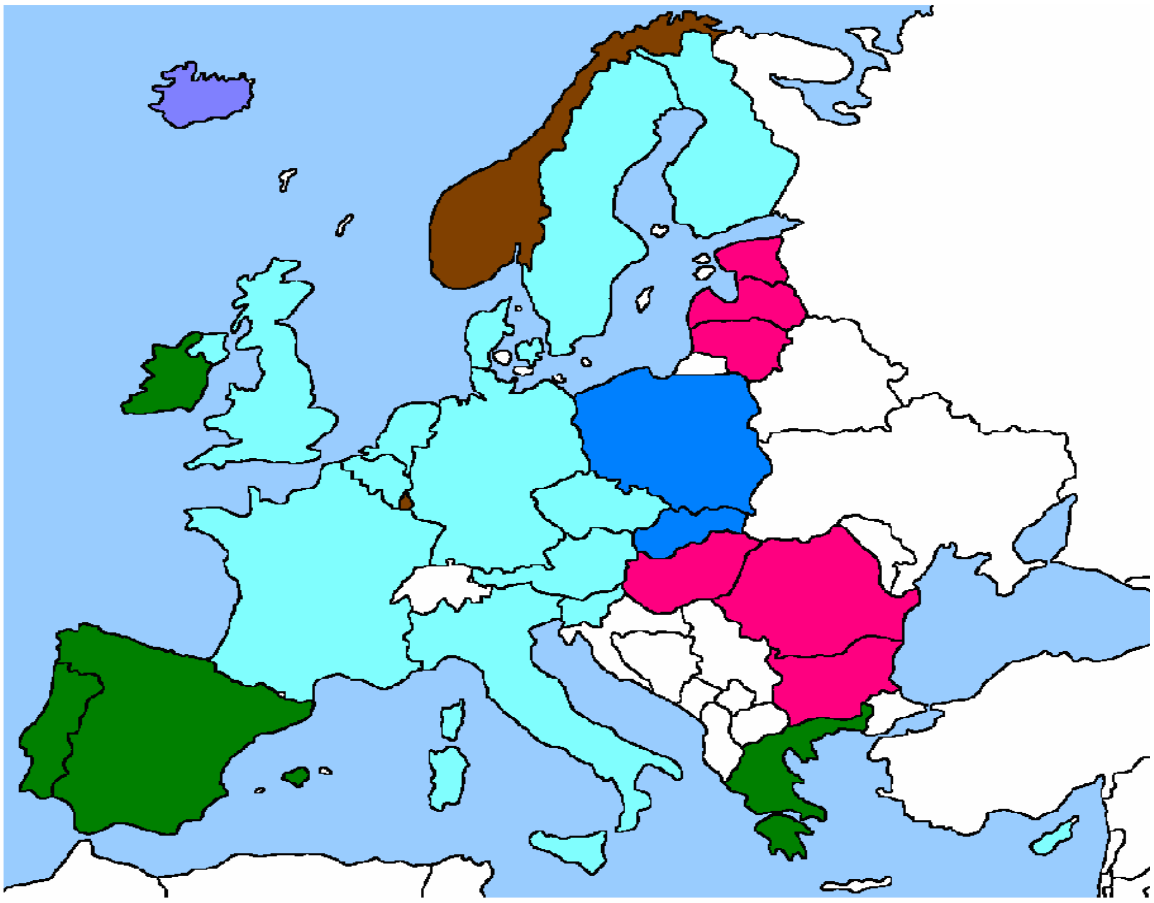
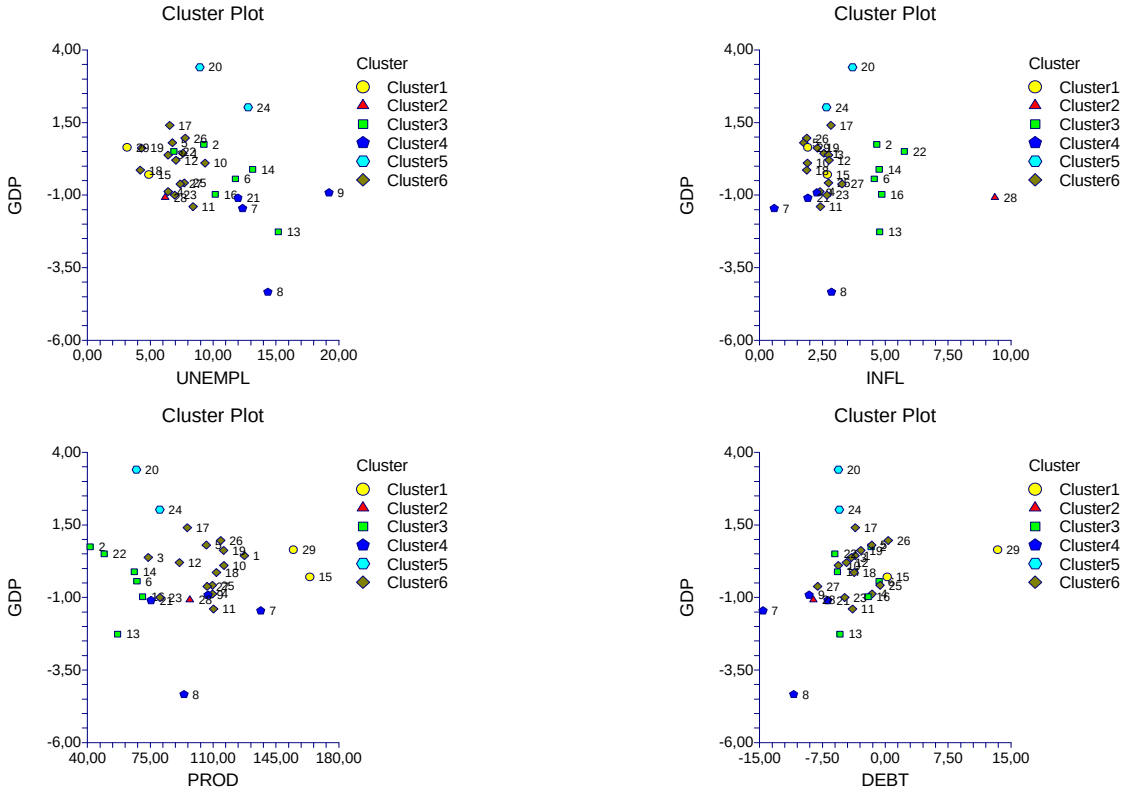
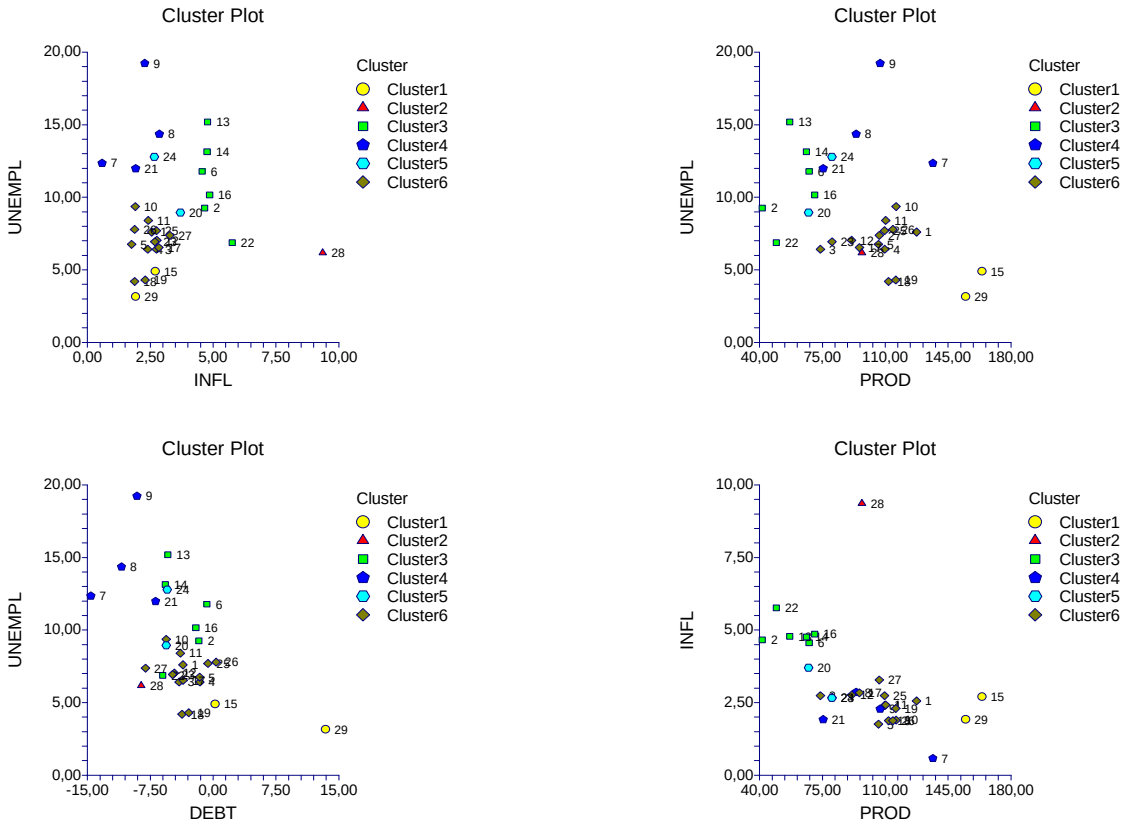


Figure 2: Geographical distribution of clusters. Source: own.





Figures 3a)-h). Cluster plots for selected pairs of variables. Source. Own.

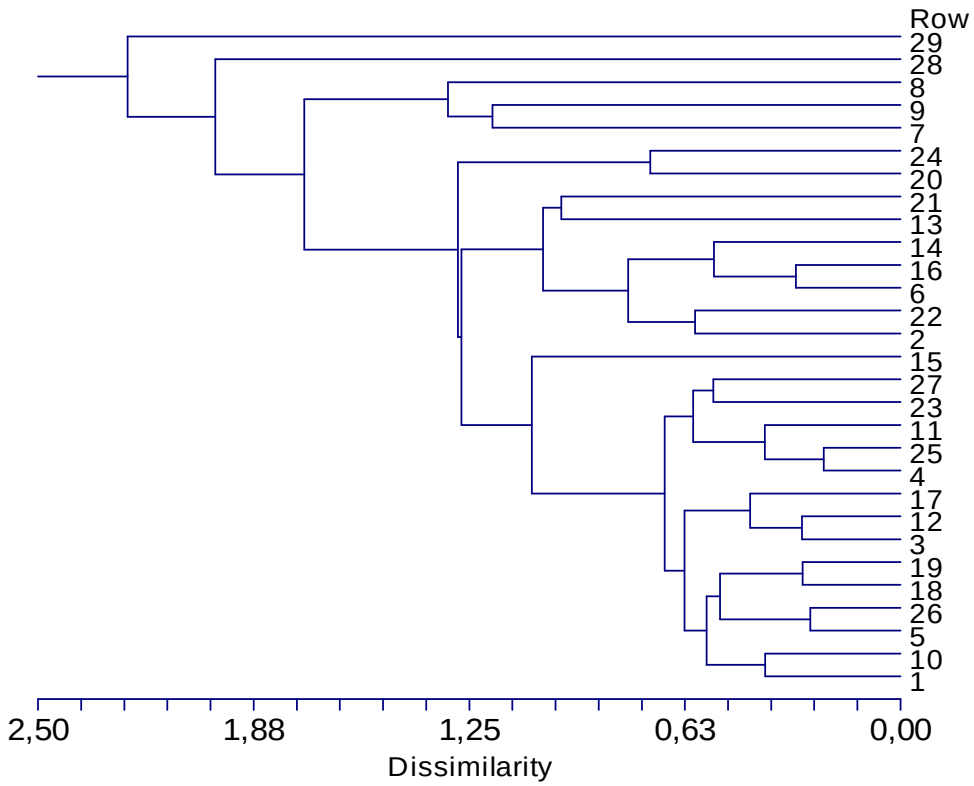


Figure 4. Dendrogram of countries. Source: own.

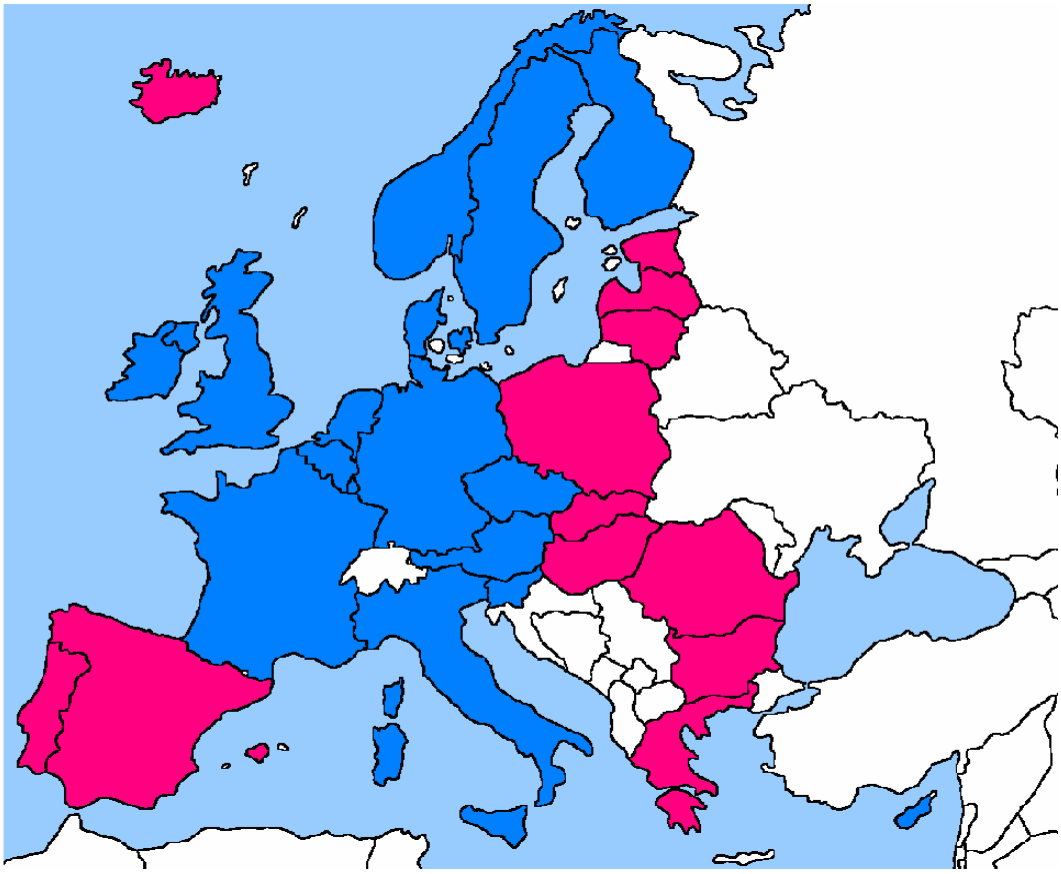


Figure 5. A division of all countries in two clusters. Source: own.

4. Conclusions

In this study a cohesion of European Union countries (plus Norway and Iceland) during 2008-2012 was examined by a cluster analysis. Fourteen countries with rather average values of all economic indicators (almost a half of 29 countries altogether) occupied one cluster (no. 6), hence this cluster can be considered an economic core of EU. A composition of this cluster ranged from Western to Central and Northern Europe. Furthermore, European 'Southern wing' countries associated with the largest budget deficits in EU (along with Ireland) formed they own cluster so their similarity (in terms of this study) was confirmed. In cluster 3 transition countries of Eastern Europe were paired together, while cluster 5 contained the only two EU countries that escaped recession: Slovakia and Poland.
Further research might focus for example on examination of a dynamics of a cohesion to find out year-to-year changes. Also, other economic indicators or different time span can be used for the evaluation, as well it is possible to use other clustering methods such as medoids or fuzzy clustering as economic data are rather imprecise due to their nature.

References

Aiginger, K. (2011). Why Performance Differed Across Countries in the Recent Crisis How Country Performance in the Recent Crisis Depended on Pre-crisis Conditions. *WIFO Working Papers*, No. 387, 29 pages.

Aiginger, K. (2013). A new strategy for the European periphery. *WWWforEurope Policy Paper*, No. 1. pages: 86

Beblavý, M., Cobham, D. , Ódor, L. (Eds.). (2011). *EURO area and the financial crisis*. Cambridge University Press.

Chiang, M. M., Mirkin, B. (2009). Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads. *Journal of Classification* 27 (2009), pages: 38

Eurostat. (2013). National accounts. On-line, cit. [2013-12-10]. Available from WWW: <http://epp.eurostat.ec.europa.eu/portal/page/portal/national_accounts/data/main_tables >.

Eurostat. (2013a). On-line, cit. [2013-12-10]. Available from WWW: http://epp.eurostat.ec. europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tec00115

Eurostat.(2013b). On-line, cit. [2013-12-10]. Available from WWW: http://epp.eurostat.ec. europa.eu/portal/page/portal/product_details/dataset?p_product_code=une_rt_a

Eurostat. (2013c). On-line, cit. [2013-12-10]. Available from WWW: http://epp.eurostat.ec. europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tec00118

Eurostat. (2013d). On-line, cit. [2013-12-10]. Available from WWW: http://epp.eurostat.ec. europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tec00116

Eurostat. (2013e). On-line, cit. [2013-12-10]. Available from WWW: http://appsso.eurostat.ec. europa.eu/nui/show.do?dataset=gov_dd_edpt1&lang=en

Hartigan, J. A. (1975). *Clustering algorithms*. John Wiley & Sons, Inc.

Hartigan, J. A., Wong, M. A. (1979). A K-means Clustering Algorithm: Algorithm AS136. *Appl. Statistics*, 28, pp. 100-108.

Karasavvoglou, A., Polychronidou, P. (Eds.). *Economic Crisis in Europe and the Balkans: Problems and Prospects*, 2014, Springer.

Kattel, R., Raudla, R. The Baltic Republics and the Crisis of 2008 - 2011, *Europe-Asia Studies*, 65(3), 2013, pp. 426-449.

Lloyd, S. P. (1982). Least Squares Quantization in PCM. In *IEEE Transactions on Information Theory* 28 (2), pp. 129-137.

Mazurek, J. (2013). A Comparison of GDP growth of European countries during 2008-2012 period from regional and other perspectives, *MPRA Paper* 51178, University Library of Munich, Germany.

Mazurek, J., Mielcová, E. (2013). The Evaluation of Economic Recession Magnitude: Introduction and Application. *Prague Economic Papers*, vol. 2013(2), p. 182-205.

Meloun, M., Militký, J. (2006). *Kompendium statistického zpracování dat*. 2nd edition. Prague: Academia. ISBN: 80-200-1396-2.

Mian, A., Sufi, A. (2010). The Great Recession: Lessons from Microeconomical Data, *American Economic Review: Papers & Proceedings*, 100 (2), 1–10.

NCSS. (2013). Statistical and Power Analysis Software Home Page. On-line, cit. [2010-09-30]. Available from WWW: <http://www.ncss.com/ncss.html>.

Verick, S., Islam, I. (2010). The great recession of 2008 - 2009: Causes, consequences and policy responses, Bonn: IZA.