

## SEGMENTATION OF THE EU USING LEADING INDICATORS

Nikolay Kulbakov

---

### Abstract

This paper is a continuation of the series of articles about the segmentation of EU countries using cluster analysis of economic indices. The paper contains the output from the cluster analysis based on the time series of macroeconomic indices for the EU. This paper compared to previous articles, focuses on the analysis based on leading indicators.

**Key words:** cluster analysis, EU, macroeconomic segmentation, leading indicators, economic expectations, MATLAB.

**JEL Code:** E01, C38

---

### Introduction

This article continues the list of works on segmentation of the EU countries by means of cluster analysis and macroeconomic indicators. Also see the articles by Kulbakov (MSED 2013) and Kulbakov (VS 2013). This work aims at investigating a group of indicators presented by Eurostat in the Business and Consumer surveys by the cluster analysis methods described in the papers by Rezanková, H., & Snásel, V. (2009), Löster, T. (2011) and others. This investigation makes it possible to divide the EU countries by groups relative to one another: economic optimists and pessimists. Also, this investigation helps to monitor the dynamics of changes of economic sentiments over time. The author's secondary goal is development of socially accessible tools for making a similar analysis in the MATLAB medium.

## 1 Parameters of the research

### 1.1 Goal Setting

The goal of this investigation is dividing the EU countries into economic pessimists and optimists and making relevant conclusions.

## 1.2 What kind of data was used for the analysis

**Sampled population by countries.** I used a set of data for countries from the database of Eurostat (Eurostat 2014), therefore all countries of the EU were included in the sample.

**Sampled population by time.** I decided to use annual data for several years. Eurostat provides statistics for the period 1985 – 2014 years.

**Sampled population by data type.** I compare the countries by leading indices. The leading indicators are derived from a periodic investigation carried out under the auspices of Eurostat Business and consumer surveys.

By Eurostat definition, the leading indicators are: *A leading indicator is an economic statistical indicator that changes before general economic conditions and therefore can be used to predict turning points in the business cycle.*

**Sampled population by macroeconomic indicators.** It was decided to split the countries into groups based on the following criteria and using the following indicators<sup>1</sup>:

**Tab. 1: Codes of indices**

BS-CCI-BAL	BS-FS-LY	BS-IEME	BS-PE3M
BS-CEME-BAL	BS-FS-NY	BS-IEOB	BS-SABC
BS-COB-BAL	BS-GES-LY	BS-IOB	BS-SAEM
BS-CPE-BAL	BS-GES-NY	BS-IPE	BS-SARM
BS-CTA-BAL	BS-MP-NY	BS-IPT	BS-SCI
BS-FLBA1-PC	BS-MP-PR	BS-ISFP	BS-SEEM
BS-FLBA2-PC	BS-PT-LY	BS-ISPE	BS-SERM
BS-FLBA3-PC	BS-PT-NY	BS-RAS	BS-CSMCI-BAL
BS-FLBA4-PC	BS-SFSH	BS-RCI	BS-ESI-I
BS-FLBA5-PC	BS-SV-NY	BS-REBS	BS-ICI-BAL
BS-FLBA6-PC	BS-SV-PR	BS-REM	BS-RCI-BAL
BS-FLBA7-PC	BS-UE-NY	BS-ROP	BS-SCI-BAL
BS-CSMCI	BS-ICI	BS-RPBS	Totla count: 51

Source: own

**Difficulties.** The principal problem of the chosen package of data is partial absence of information on some countries of the European Union relative to some temporal intervals and indicators of interest. Therefore it is impossible to do the clustering for each year or for all the indicators, which would result in distribution of all the EU countries by the clusters. It means that some countries drop out of the investigation.

All the start-up data from the Eurostat database are indicator values over months. My task is to translate them into yearly temporal rows.

<sup>1</sup> Full names of indices: [http://epp.eurostat.ec.europa.eu/NavTree\\_prod/everybody/BulkDownloadListing?file=dic%2Fen%2Findic.dic](http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing?file=dic%2Fen%2Findic.dic)

### 1.3 Methods

In the research described in this article, I use the hierarchical clustering. This method, the description of algorithms, and other details are presented in Kulbakov (MSED 2013).

For carrying out this analysis, in addition to the previous developments I have written new modules. All these tools are realized in the MATLAB medium and perform the following functions:

1. Loading of the said sets of indicator temporal rows straight from Eurostat. I used datasets with the help of Eurostat's bulk download facility.
2. Creation of analytic cubes, i.e. data transfer of *.tsv* files from Eurostat to the data structure for MATLAB.
3. Aggregation of analytic cubes data from temporal rows by months to rows by years. The aggregation is only possible by the arithmetic mean, geometrical mean, sum total, minimum, maximum, and the first month data.

The start-up code is accessible for loading at the URL given below.

## 2 Results of the research

### 2.1 First outputs

At the output of the research I received better distribution of matrix groups. For convenience, I picked up the names of the groups according to the ratings. Rating has been calculated based on the average normalized values of the indicators for each group given the nature of the indicators.

For a visual analysis of the nearest neighbours, will be given only one dendrogram based on 2013 year. Dendrogram for the remaining years can be obtained by running the application MATLAB which was mentioned earlier, or request from the author.

**Tab. 2: Distribution of countries by groups based on 51 indices**

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Austria							-1	0	-1	0	0	-1	0	0	1	1	1
Belgium	-1	0	-1	-1	1	1	-1	0	-1	0	0	-1	0	0	1	1	1
Bulgaria						0	-1	0	0	0	0	0	0	0	0	1	0
Cyprus												0	0	0	0	0	-1
Czech Republic						1	1	0	0	0	1	1	0	0	0	1	0
Germany					-1	-1	-1	-1	-1	0	0	-1	0	0	1	1	1
Denmark														0	1	1	1
Estonia						1	1	0	0	1	1	-1	-1	0	1	1	1
Greece	-1	-1	0	0	0	1	-1	0	-1	0	0	0	0	-1	-1	-1	-1
Spain	0	0	-1	-1	1	1	-1	0	-1	0	0	-1	-1	0	0	0	0
Finland	1	1	1	1	1	1	-1	0	1	1	1	1	1	0	1	1	1
France							-1	0	-1	0	0	-1	0	0	1	1	0
Croatia												0	0	0	0	1	0
Hungary						1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0
Ireland																	
Italy							-1	0	-1	0	0	-1	0	0	0	0	0
Lithuania									0	1	1	-1	-1	0	0	1	1
Luxembourg																	
Latvia						1	1	0	0	1	1	-1	-1	0	0	1	1
Malta															0	1	1
Netherlands							-1	0	-1	0	0	1	0	0	1	1	0
Poland							-1	-1	-1	0	0	1	0	0	0	1	0
Portugal												-1	0	0	0	0	-1
Romania									0	0	0	1	0	0	0	1	0
Sweden		0	-1	1	1	1	-1	0	1	1	1	1	1	1	1	1	1
Slovenia						1	1	0	0	0	1	0	0	0	0	1	0
Slovakia						1	0	1	0	0	1	1	0	0	0	1	0
UK	0	0	-1	-1	1	1	-1	0	-1	0	0	-1	-1	0	0	1	1

Source: own calculations

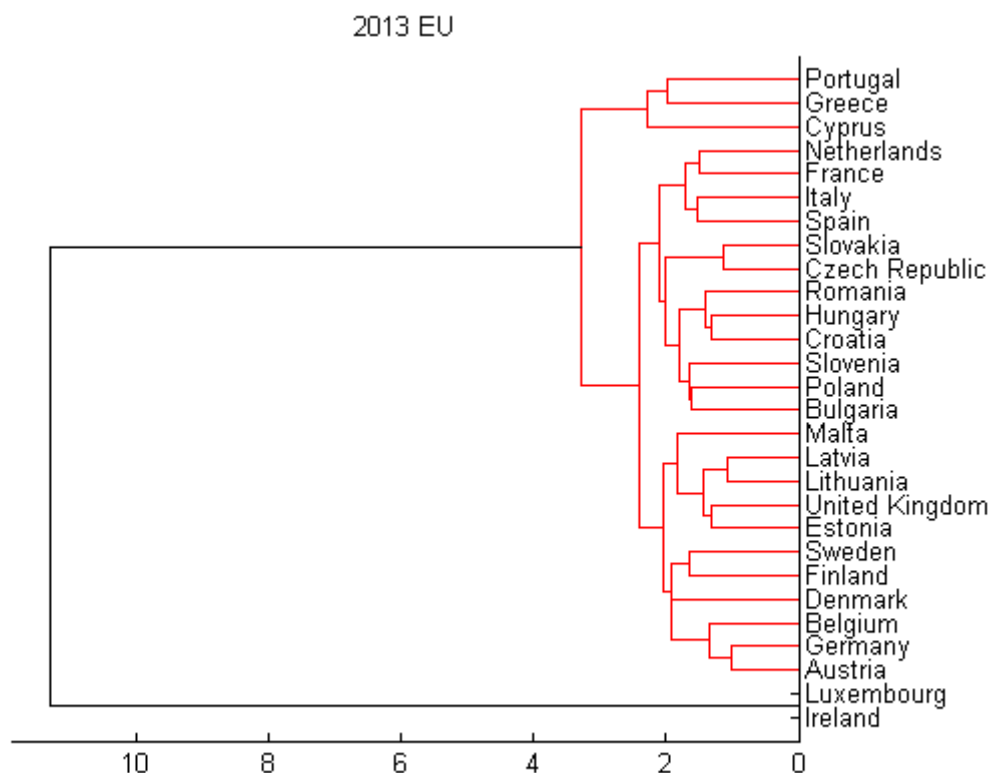
Table 2 contains the distribution of the EU countries by three clusters: pessimists, optimists, and in-between operators. The boxes with value -1 mean that the country is in the cluster of the most pessimistic countries, i.e. its citizens believe that their economic condition will be worse than thought about their economic condition by citizens of other countries.

+1 Boxes mean that the country is in the cluster of the most optimistic countries, i.e. their citizens believe that their economic condition will be better than thought about their economic condition by population of other countries not included in that cluster.

0 Boxes mean that the country is in the cluster between pessimist and optimist countries.

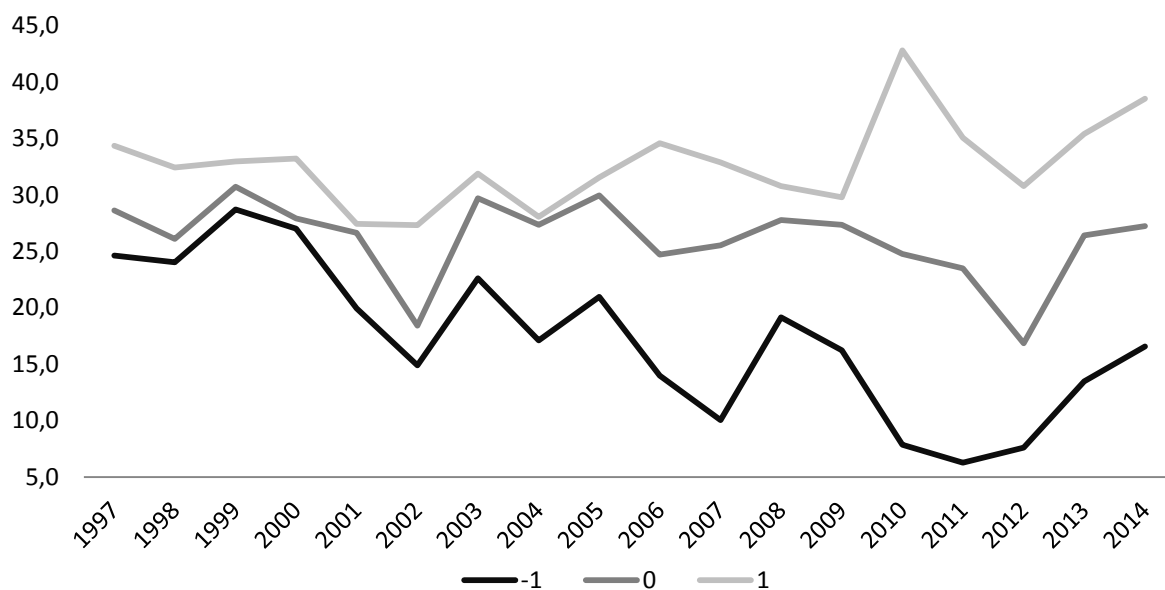
If the value is missing, it means that the country under consideration does not participate in the cluster distribution for the given year; its position is unknown and is not in any way considered for estimating the positions of other countries. As a rule, the deeper we look into the history, the less data we discover and, hence, the more blanks we spot. The investigation results contain no data at all on Luxemburg and Ireland for lack of the required amount of initial data. That is why further cluster analysis was carried out by the minimum set of indicators in order to distribute all the EU countries and to simplify the set of indicators for subsequent investigations.

**Fig. 1: Output: dendrogram 2013 EU28 by 51 indices**



Source: own calculations

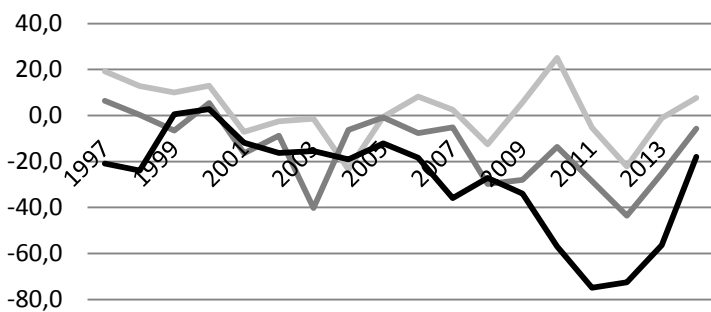
**Fig. 2: Rating of clusters EU**



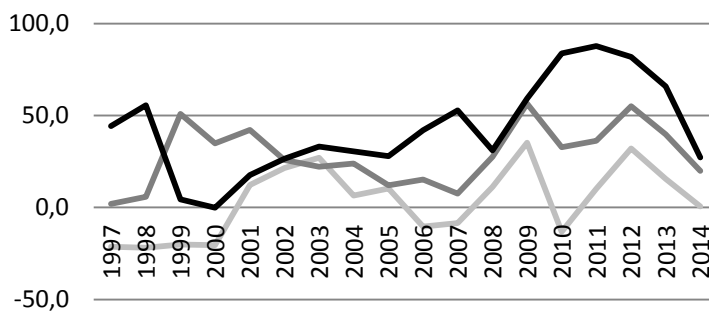
Source: own calculations

**Fig. 3-6: Aggregate indices by clusters**

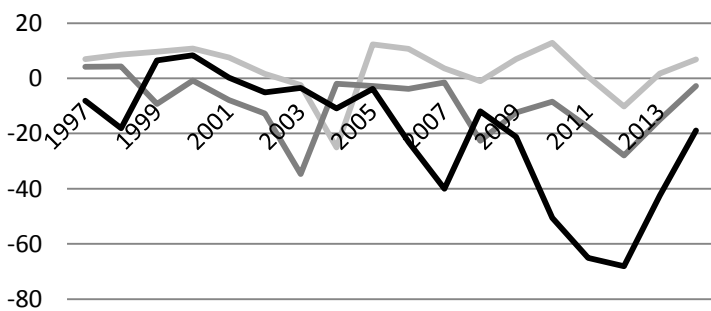
**General economic situation**  
over the next 12 months



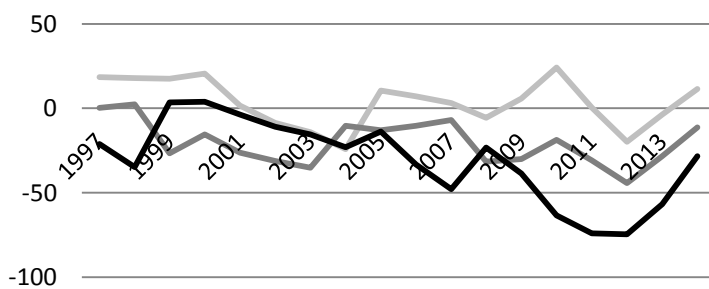
**Unemployment expectations**  
over the next 12 months



**Financial situation**  
over the next 12 months



**Consumer confidence**  
indicator



Source: own calculations based on Eurostat data

## 2.2 Commentary on the results

The table shows, for instance, that in 2008 through 2009 more countries got into the cluster showing the relatively pessimistic anticipations of the pending economic development in the wake of we know which crisis. Also, we can see the impacts of the 2013 banking crisis in Cyprus.

The immediate neighbours' diagram shows that in 2013 Greece, Cyprus, and Portugal were in this aspect closest than others. Those countries encountered economic difficulties. The Czech Republic's closest neighbour is Slovakia. The optimists are represented mostly by the northern countries of the European Union.

Observation of the rating pattern or normalized indicators' mean values for clusters led me to an idea of isolating a minimum from the 51st indicator which would have similar characteristics, i.e. would outpace the business cycle by several years. As a result, I chose 4 indicators:

- Consumer confidence indicator
- General economic situation over the next 12 months
- Financial situation over the next 12 months
- Unemployment expectations over the next 12 months

The diagrams of those indicators (Fig. 3-6) show, for instance, that the mean values of the chosen indicators for clusters of relative optimists/pessimists start to diminish 1 to 3 years prior to the 2009 crisis. It is fair to say that they forecast a crisis. Consequently, the unemployment anticipation indicator increases.

Further on, I decided to repeat the cluster analysis only for those 4 indicators .

### 2.3 Second output

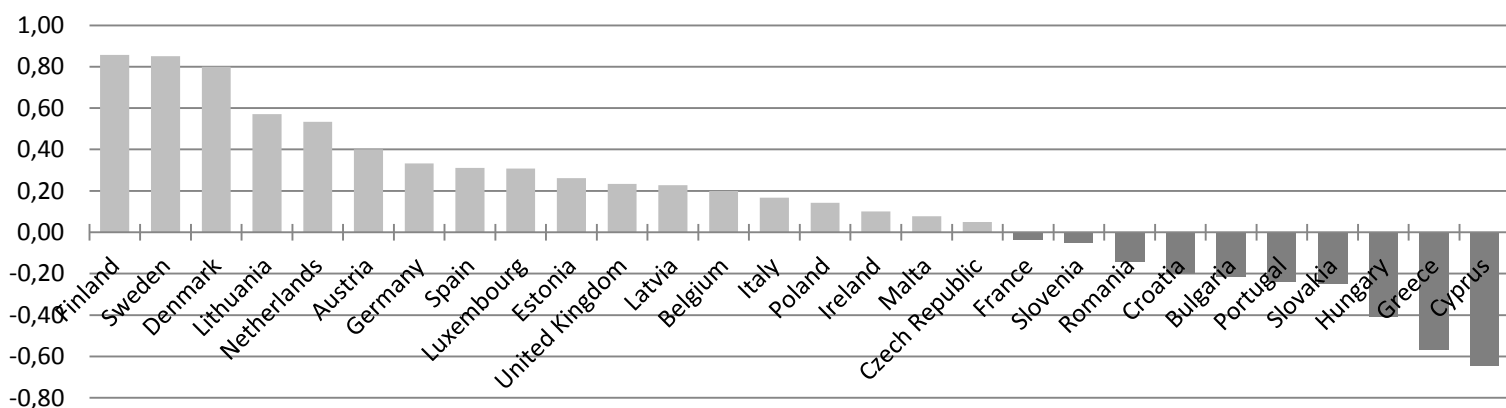
Tab. 3: Distribution of countries by groups based on 4 indices

	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014		
Austria											0	0	0	0	1	0	1	0	1	1	0	0	1	0	0	0	1	0	1	1		
Belgium	-1	0	0	0	0	1	0	0	0	-1	0	0	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	1	0	0	1	
Bulgaria																	0	-1	0	0	0	-1	0	0	-1	0	0	0	0	0	0	
Cyprus																	0	-1	0	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1	0	
Czech Republic												0	1	0	-1	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	
Germany	1	1	1	0	0	1	0	0	0	-1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1	1	1	1	
Denmark	1	1	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	
Estonia								-1	-1	-1	0	1	0	0	0	-1	0	0	1	1	1	1	1	1	0	0	0	1	1	1	1	
Greece	1	-1	-1	-1	0	0	-1	-1	1	-1	-1	0	0	-1	0	0	0	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
Spain		0	0	1	0	1	0	0	0	-1	0	1	1	1	1	0	1	0	1	1	1	0	0	0	-1	0	0	1	0	0	1	
Finland			1	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	
France	-1	0	-1	0	0	1	-1	0	0	-1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Croatia																						0	0	0	0	-1	-1	0	0	0	0	
Hungary									-1	-1	-1	-1	-1	0	0	-1	0	0	0	0	0	0	-1	-1	-1	-1	0	0	0	0	1	
Ireland	-1	-1	-1	-1	0	1	-1	0	1	-1	0	1	1	1	1	1	1	0	0	1	0	0	0	-1	-1	0	0	0	1	1	1	
Italy	0	0	0	0	0	1	0	0	0	-1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Lithuania																	0	0	1	1	1	1	1	0	-1	0	1	1	1	1	1	
Luxembourg																			0	1	1	0	0	0	0	0	0	1	0	0	1	
Latvia									-1	-1	-1	0	0	0	0	0	1	0	1	1	0	1	1	0	-1	0	1	1	1	1	1	
Malta																			0	1	0	-1	-1	0	0	0	0	0	0	1	1	
Netherlands	1	1	1	1	1	1	0	0	0	-1	0	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1	0	0	1	1	
Poland																	-1	-1	0	0	0	0	0	1	1	0	0	1	0	0	1	
Portugal		0	1	1	0	1	1	0	0	-1	0	0	0	0	1	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	
Romania																	0	-1	0	1	0	0	0	0	-1	-1	0	0	0	0	0	
Sweden											0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	
Slovenia												1	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	
Slovakia																	-1	-1	-1	-1	-1	0	0	0	1	0	-1	0	0	0	0	1
UK	0	0	1	1	-1	-1	0	0	1	-1	0	1	1	0	1	0	1	0	1	1	0	0	0	-1	0	0	0	0	0	1	1	

Source: own calculations

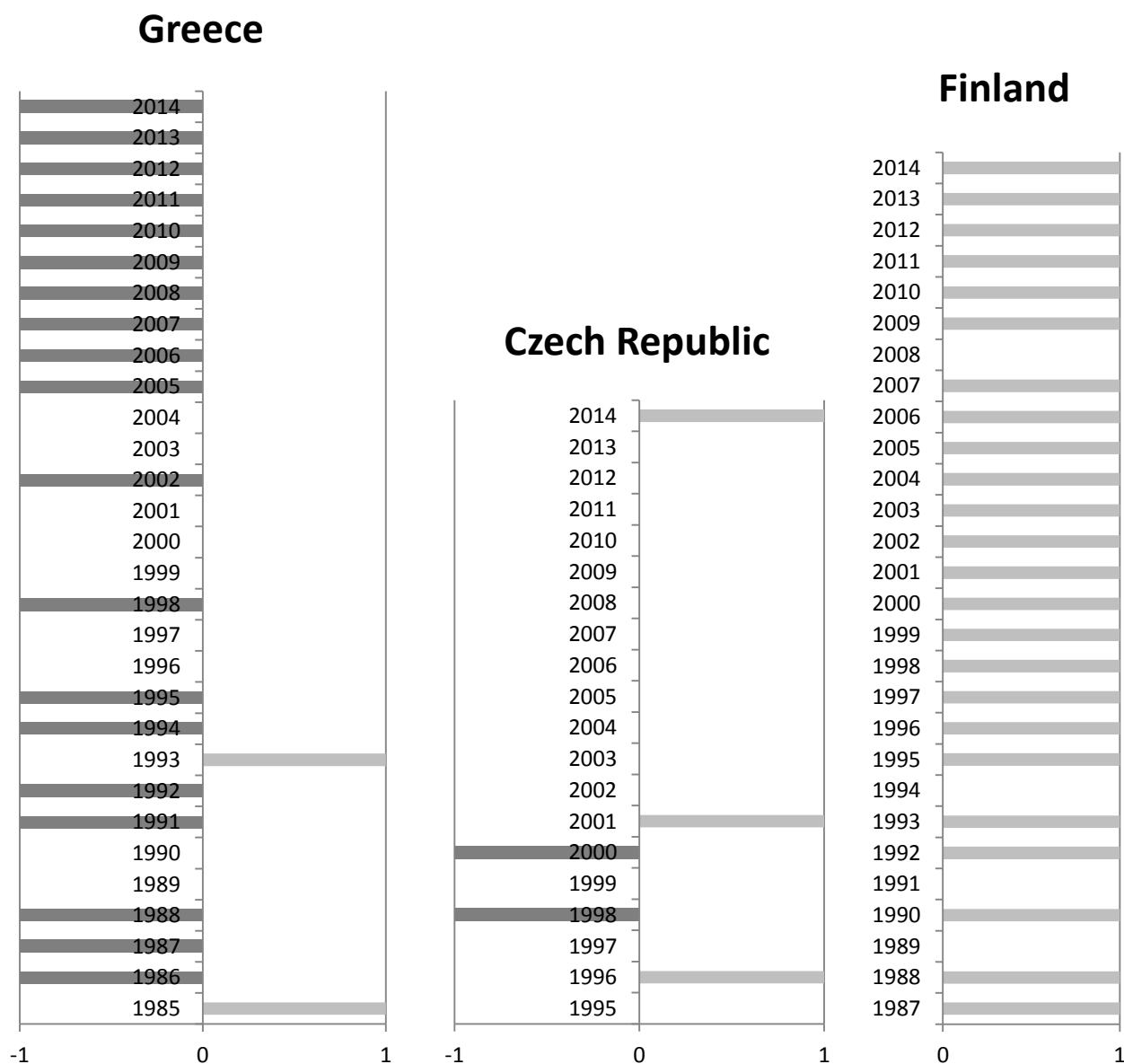


**Fig. 7: Optimism's / pessimism's rating of EU countries based on 4 indices**



Source: own calculations

**Fig. 8-10: Bad, normal or good years by economic expectations**



Source: own calculations

## 2.4 Commentary on the results

Based on table 3, I sorted out the countries by the optimism criterion, which is shown in diagram 7. The countries that had been more often in the cluster with relatively better anticipations are optimists. Those countries are more often at the top of the rating. The rating is calculated based on the arithmetic mean of the -1/0/+1 values.

It is possible to consider in detail the poles, i.e. the most optimistic country, Finland; a middling like the Czech Republic; and the pessimistic Greece.

Diagrams 8 through 10 show that in Greece the sentiments over the economic situation have been falling as compared to Finland, where from year to year the citizens feel increasingly optimistic about their economy. The Czech Republic is mainly in the middle, while the number of years when the Czechs deviate from the mean attitude to pessimists or optimists is demonstratively close to fifty-fifty.

## Conclusion

The investigation showed that the cluster analysis by leading indicators can be used to piece together a picture showing the division of the EU countries into those whose population is more or less optimistic over their countries' prospective economic development. This is related both to the domestic economic situation and the state of the global business cycle.

As a result of the research was created the tools with which it is possible to carry out the cluster analysis with similar parameters quickly and efficiently. The experience that was received can be used in the future for different sets of data and for different sets of countries.

The results of this research could be interesting to the public, and could be used as material for journalists and teachers.

## 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

- Rezankova, H., & Loster, T. (2013). Shlukova analyza domacnosti charakterizovanych kategorialnimi ukazateli. *E+M. Ekonomie a Management*, 16(3), 139-147. ISSN: 1212-3609
- Baker, F., & Hubert, L. (1975). Measuring Power of Hierarchical Cluster-Analysis. *Journal of the American Statistical Association*, 70(349), 31–38. doi:10.2307/2285371
- Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques. In J. Kogan, C. Nicholas, & M. Teboulle (Eds.), *Grouping Multidimensional Data* (pp. 25–71). Springer Berlin Heidelberg. Retrieved from
- Cattinelli, I., Valentini, G., Paulesu, E., & Borghese, N. A. (2013). A Novel Approach to the Problem of Non-uniqueness of the Solution in Hierarchical Clustering. *Ieee Transactions on Neural Networks and Learning Systems*, 24(7), 1166–1173. doi:10.1109/TNNLS.2013.2247058
- Cernakova, V., & Hudec, O. (2012). Quality of Life: Typology of European Cities Based on Cluster Analysis. *E & M Ekonomie a Management*, 15(4), 34–48.
- Eurostat. (2014). *Eurostat research database*. Retrieved from [http://epp.eurostat.ec.europa.eu/NavTree\\_prod/everybody/BulkDownloadListing](http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing)
- Faber, V. (1994). Clustering and the continuous k-means algorithm. *Los Alamos Science*, 22, 138–144.
- Hallet, M. (2002). Regional Specialisation and Concentration in the EU. In P. J. R. Cuadrado-Roura & P. M. Parellada (Eds.), *Regional Convergence in the European Union* (pp. 53–76). Springer Berlin Heidelberg. Retrieved from
- Löster, T. (2011). *Hodnocení výsledků metod shlukové analýzy*. FIS VŠE. Retrieved from <https://isis.vse.cz/auth/lide/clovek.pl?id=8340;zalozka=7;studium=92375>
- MathWorks. (2013). Cluster Analysis. Retrieved from <http://www.mathworks.com/help/stats/cluster-analysis.html>
- Papageorgiou, T., Michaelides, P. G., & Milios, J. G. (2010). Business cycles synchronization and clustering in Europe (1960-2009). *Journal of Economics and Business*, 62(5)
- Rezanková, H., & Snásel, V. (2009). *Shluková analýza dat*. Praha: Professional Publishing.
- S&P. (2013, May). Sovereigns Rating. Retrieved from <http://www.standardandpoors.com/ratings/sovereigns/ratings-list/en/us?sectorName=null&subSectorCode=39>
- Kulbakov, N.(MSED 2013) *Segmentation of EU/EA Countries via Cluster Analysis of Macroeconomics Indicators*. Retrieved from <http://msed.vse.cz/files/2013/196-Kulbakov-Nikolay-paper.pdf> or <http://www.ilovecz.ru/research/matlab01.zip>
- Kulbakov, N.(VS 2013) *Segmentation of EU27 into four groups and their dynamics*. Retrieved from <http://www.ilovecz.ru/research/matlab02.zip>
- Kulbakov, N. (VS 2013) *Source for MATLAB cluster analysis, input and output data*. Retrieved from <http://www.ilovecz.ru/research/matlab03.zip>

**Contact**

Nikolay Kulbakov

University of Economics, Prague

W. Churchill Sq. 4

130 67 Prague 3

Czech Republic

kulbakov@gmail.com