Cluster analysis of selected world development indicators in the fields of agriculture and the food industry in European Union countries

MARIAN REIFF*, ZLATICA IVANICOVA, KVETOSLAVA SURMANOVA

University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Bratislava, Slovak Republic

*Corresponding author: marian.reiff@euba.sk

Reiff M., Ivanicova Z., Surmanova K. (2018): Cluster analysis of selected world development indicators in the fields of agriculture and the food industry in European Union countries. Agric. Econ. – Czech, 64: 197–205.

Abstract: The paper analyses the disparity in the performance of the agriculture and food industry sectors in EU countries during the period from 2002 to 2013, and identifies significant differences between countries as well as the dynamics of change. The individual countries are clustered according to the long-term average of the World Bank collection of development indicators such as agricultural raw materials exports, agricultural raw materials imports, crop production index, food production index, livestock production index, cereal yield, agriculture value added and agriculture value added per worker. The analysis reveals convergence in the given period and identifies significant differences in the indicator of agriculture value added per worker at the end of the analysed period.

Keywords: agriculture value added per worker, multivariate analysis, Ward's method

In recent decades, successive reforms of the Common Agricultural Policy (CAP) and the enlargement of the European Union (EU) have stimulated research interest in studying the distinct differences in the performance of the agricultural and food industry sectors. In addition to standard market self-regulation, the regulation of the industry by means of the CAP has played an important role in this sector. The CAP was created to regulate and support European agriculture. Inter alia, the aim of the CAP is to assist the development of agriculture in EU member states and to eliminate differences in performance between different countries. Its aims are to combine strong economic performance with sustainable use of natural resources in the field of agriculture (European Council 2001). Significant reforms in the CAP have been made in recent years, notably in 2003 and during the CAP Health Check in 2008, to modernize the sector and render it more market-oriented. In the Europe 2020 strategy, the role of the CAP is to contribute more to the development of intelligent, sustainable,

and inclusive growth, through its response to new economic, social, environmental, climate-related and technological challenges in society (European Commission 2016). Predictions concerning the impacts of changes in the CAP after 2013 has been dealt with by Ciaian et al. (2014). Besides reforms to the CAP, a further important factor has been the various enlargements of the EU by the additions of new member states which have had to adapt to the CAP.

The impacts caused by the enlargement of the EU and analyses of the impacts of the CAP measured by various agricultural performance measures have been the topic of many studies. For example, Giannaskis and Bruggeman (2015) investigated the factors that lie behind the differential performance of agriculture across the twenty-seven EU countries, based on gross-value-added farms and land and labour productivity indicators. Significant differences were revealed between the Northern-Central counties and the continental peripheries (Mediterranean, Eastern, and Northern Scandinavian). The authors identified

Supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences Vega (Project No. 1/0245/15 and 1/0248/17).

human capital characteristics, environmental conditions and technical efficiency of crop production as factors underlying this differential performance. Agricultural sectors characterized by a young and better-trained farm population are more likely to attain high economic performance. On the other hand, the wheat and tomato yield variables highlight the importance of both environmental conditions and technical efficiency on farm economic performance.

Szabo and Grznár (2015a) ranked individual EU countries according to the long-term average of their agricultural produce per unit of area into seven segments, and the relationship between the input of fixed assets, intermediate consumption, labour force, numbers of animals, and other aspects were identified in connection with agricultural production. The conducted analysis showed close relationship between production, fixed and variable assets, livestock counts, and the provided support.

Svoboda et al. (2014) compared agricultural subsidies in the member states of the EU during the period of 2004–2012, based on the Farm Accountancy Data Network database. The authors concluded that there has been a slight increase in operational subsidies due to the fact that variability shows a decreasing trend. With the help of cluster analysis, the member states were divided into groups according to their operational subsidies, total production and costs.

Dos Santos (2013) characterized and segmented the farms of the twenty-seven member states of the European Union. For this purpose, she adopted the technique of cluster analysis, and cases were clustered according to different farm segments, based on a sample of farms of the Farm Accountancy Data network. Variables used for segmenting farms were of four types: structural variables, use of inputs, technology (intensification) and financial variables. The results showed the existence of four types of farms in the EU that are distinguishable by their structural characteristics, financial characteristics, guidance and the importance of subsidies.

Spicka (2013) investigated the differences in farm income and its determinants between the old (EU-15) and new EU member states (EU-12) before and after EU enlargement during the period of 2001–2011. Using cluster analysis, the specific structural and economic features within the EU were identified. The author concluded that the rankings of the EU-27 countries changed after EU enlargement. However, the European countries with highly intensive agriculture still occupied the top positions. For example, the average labour input in the EU-12 is substantially higher than in the EU-15. This fact, together with the lower fixed capital consumption, points to a lower level of technical equipment and farming technologies in the EU-12.

The goal of this paper was to characterize the disparity in the performance of the twenty-eight member states of the European Union in the agriculture and food industry sectors, and also to identify the significant differences, so as to show the dynamics of selected world development indicators of agriculture and the food industry of EU member states from 2002 to 2013. To analyse disparities in performance, we have selected a cluster analysis approach. We have grouped countries in homogeneous clusters based on the loss of homogeneity minimization criterion. Our aim was to identify different groups containing mutually similar countries based on selected agricultural and rural development indicators. Our study is driven by the following question: is the performance of agriculture and the food industry, if measured using selected world development indicators, different across the countries of the European Union?

MATERIAL AND METHODS

Cluster analysis is a multivariate statistical technique that entails division of a large group of observations into smaller and more homogeneous groups. In our analysis of the European Union agriculture and food industries, we employed cluster analysis to classify EU member states according to the long-term average of the World Bank collection of agriculture & rural development indicators such as, agricultural raw materials exports, agricultural raw materials imports, crop production index, food production index, livestock production index, cereal yield, agriculture value added measured as % GDP, annual % growth and agriculture value added per worker. The source of the data is the World Bank database of world development indicators and a detailed description of the indicator can be found on the website listed in the references (World Bank 2016). We also provide a short description of the analysed indicators as follows: Agricultural raw materials comprise crude materials. The crop production index shows agricultural production for each year relative to the base period of 2004-2006. The food production index covers food crops that are considered edible and that contain nutrients. The livestock produc-

tion index includes meat and milk from all sources, dairy products such as cheese, and eggs, honey, raw silk, wool, and hides and skins. Cereal yield, measured as kilograms per hectare of harvested land, includes wheat, rice, maize, barley, oats, rye, millet, sorghum, buckwheat, and mixed grains. Value added in agriculture measures the output of the agricultural sector less the value of intermediate inputs. Agriculture comprises value added from forestry, hunting, and fishing as well as cultivation of crops and livestock production. Agriculture value added per worker is a measure of agricultural productivity. Data are presented in constant 2005 U.S. dollars. The data were averaged across three reference periods: 2002-2005, 2006-2009 and 2010-2013 in order to mitigate specific effects in particular years, caused by fluctuations either in production due to, for example, bad weather conditions or because of input or output prices in world markets.

Our aim was to identify groups of countries that are similar to each other but different from other groups of countries based on the studied characteristics. We have selected a minimum-variance criterion to determine which clusters are merged at successive steps. Minimizing variance inside each cluster emphasizes inner homogeneity, which implies preference of Ward's method (Ward 1963; Lance and Williams 1967). The method uses the approach of analysis of variance and evaluation is based on squared Euclidean distances between clusters. Cluster membership is assigned based on calculation of the total sum of squared deviations from the mean of a cluster. Two clusters are merged if this results in the smallest increase in the overall sum of squared within-cluster distances. In order to determine the number of clusters, relatively large merging distances are considered. Ward's method is an iterative process that is repeated until a desired number of clusters are achieved, or each cluster is consolidated into a single massive cluster. Cluster analysis was performed in MATLAB, by applying Ward's method. The same clustering procedure was performed for three different periods: 2002–2005, 2006–2009 and 2010–2013.

RESULTS AND DISCUSSION

Results obtained from the clustering procedure for the periods of 2002–2005, 2006–2009 and 2010–2013 are depicted on the dendrograms, tables and maps that follow. Results are discussed below.

To verify differences between evidence clusters, it is appropriate to use methods that reveal these differences. To identify indicators that are significantly different in one cluster compared to another, the Kruskal-Wallis rank test procedure was used. The Kruskal-Wallis rank test was performed on clusters one, two and five. Cluster numbers are assigned in Table 1 and Figure 1. This analysis indicates that statistically significant differences between clusters one, two and five at the 0.05 level of significance are seen in the following variables: food production index, livestock production index, agriculture value added (annual % growth), agriculture value added (% of GDP) and agriculture value added per worker. The conclusions are based on *p*-values which were

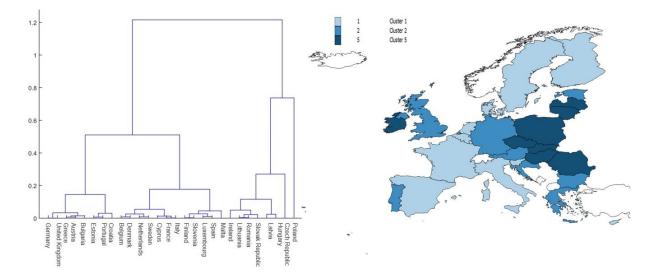


Figure 1. Dendrogram for 2002–2005 and its depiction on the map of Europe

Analysis of clu	stering in the period 2002–2005 – Ward's method		
Clusters	Creation of clusters from European Union countries	Distance form clusters	Number of countries
Z1	{(Germany, United Kingdom)+[(Greece, Austria)+Bulgaria]}	0.04	5
Z2	[(Estonia, Portugal)+Croatia]	0.03	3
Cluster2	Z1 + Z2	0.15	8
Z3	{[(Belgium, Denmark)+(Netherland, Sweden)]+[(France, Italy)+Cyprus]}	0.05	7
Z4	{[(Finland, Slovenia)+(Luxembourg, Spain)]+Malta}	0.045	5
Cluster 1	Z3 + Z4	0.18	12
Cluster 3	Z1 + Z2 + Z3 + Z4	0.52	20
Z5	[:{[(Lithuania, Romania)+Slovak Republic]+Ireland}+(Latvia, Hungary):]	0.052	6
Z6	Czech Republic		1
Cluster 4	Z5 + Z6	0.28	7
Z7	Poland		1
Cluster 5	Z5 + Z6 + Z7	0.74	8
Final Cluster	Cluster 3 + Cluster 5	1.205	28

Table 1. Description of dendrogram for 2002–2005

Table 2. Mean values for significant variables

Cluster means of variables:	Food production index (2004–2006 = 100)	Livestock production index (2004–2006 = 100)	Agriculture value added (annual % growth)	Agriculture value added (% of GDP)	Agriculture value added per worker (constant 2005 US\$)
Cluster 1	100.988	102.349	0.07732	2.22031	41 823.6
Cluster 2	98.7556	98.2722	0.99408	3.84902	16 021
Cluster 5	97.1434	99.745	2.68937	5.00662	7 893.17

compared with the level of significance ($\alpha = 0.05$). Table 2 shows a statistical summary; cluster means were calculated for each of the significant variables.

Clusters one, two and five are characterized by higher values of the significantly different variables that are highlighted in bold in each column of Table 2. For example, cluster one, is characterized by higher values for the food production index, livestock production index and agriculture value added per worker, and cluster five is characterized by larger values for agriculture value added (annual % growth) and agriculture value added (% of GDP).

Results obtained by applying Ward's method for 2006-2009 are depicted in Figure 2. The Kruskal-Wallis rank test was again conducted on clusters one, two and four. Cluster numbers are assigned

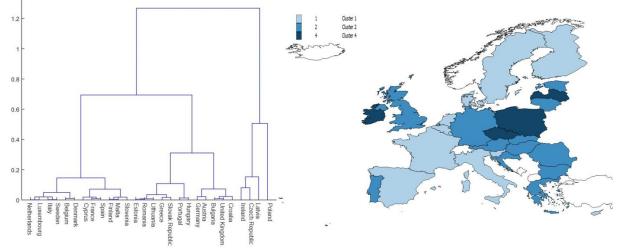


Figure 2. Dendrogram for 2006–2009 and its depiction on the map of Europe

Table 3. Description of dendrogram for 2006–2009

Analysis	of clustering in the period 2006–2009 – Ward's method		
Clusters	Creation of clusters from European Union countries	Distance form clusters	Number of countries
Z1	{[(Netherlands, Luxembourg)+(Italy, Sweden)]+(Belgium, Denmark)}	0.05	6
Z2	{[(France, Spain)+Cyprus]+[(Finland, Malta)+Slovenia]}	0.07	6
Cluster1	Z1 + Z2	0.15	12
Z3	[:{[(Estonia, Romania)+Lithuania]+(Greece, Slovak Republic)}+ +(Portugal, Hungary):]	0.12	7
Z4	{[(Bulgaria, United Kingdom)+Croatia] +(Germany, Austria)}	0.08	5
Cluster 2	Z3 + Z4	0.3	12
Cluster 3	Z1 + Z2 + Z3 + Z4	0.7	24
Z5	[(Ireland, Czech Republic)+Latvia]	0.17	3
Z6	Poland		1
Cluster 4	Z5 + Z6	0.5	4
Final Clu	ster Cluster 3 + Cluster 4	1.24	28

Table 4. Summary statistics for the variable "agriculture value added per worker (constant 2005 US\$)"

	Mean	Median	Standard Deviation	Minimum	Maximum
Cluster 1	47 024.739	43 496.181	20 047.565	11 378.811	91 809.388
Cluster 2	15 660.621	12 228.723	8 996.1742	7 493.8176	32 940.6
Cluster 4	6 831.3933	6 249.8854	3 868.9664	2 801.7511	12 024.051

in Table 3. A statistically significant difference between clusters one, two and four at the 0.05 level of significance was only observed for the agriculture value added per worker variable.

The Kruskal-Wallis rank test was again conducted on clusters one, two and four. Cluster numbers are assigned in Table 5. A statistically significant difference between clusters one, two and four at the 0.05 level of significance was again only observed for the agriculture, value added per worker variable. Based on the results of Ward's method, we divided European Union countries into three homogeneous groups. The first group of countries is represented by cluster one. This group contains the same twelve countries in all three analysed periods: 2002–2005, 2006–2009 and 2010–2013; namely, Belgium, Denmark, Cyprus, Finland, France, Italy, Luxembourg, Malta, the Netherlands, Slovenia, Spain and Sweden. These countries are the most homogenous, and the variability in the analysed indicators

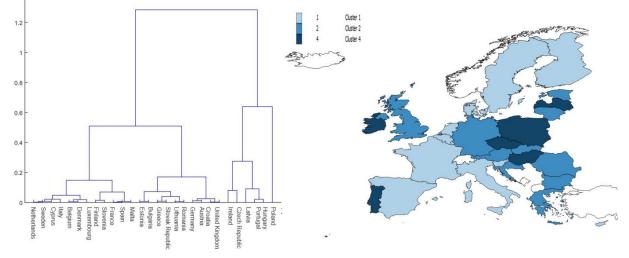


Figure 3. Dendrogram for 2010–2013 and its depiction on the map of Europe

Clusters	Creation of clusters from European Union countries	Distance form clusters	Number of countries
Z1	[:{[(Netherland, Sweden)+Cyprus]+Italy}+[(Belgium, Denmark)+Luxembourg]:]	0.05	7
Z2	{(Finland, Slovenia)+[(France, Spain)+Malta]}	0.08	5
Cluster1	Z1 + Z2	0.15	12
Z3	{(Estonia, Bulgaria)+[(Greece, Slovak Republic) +(Lithuania, Romania)]}	0.08	6
Z4	[(Germany, Austria)+(Croatia, United Kingdom)]	0.02	4
Cluster 2	Z3 + Z4	0.19	10
Cluster3	Z1 + Z2 + Z3 + Z4	0.52	22
Z5	{[(Portugal, Hungary)+Latvia]+(Ireland, Czech Republic)}	0.28	5
Z6	Poland		1
Cluster4	Z5 + Z6	0.67	6
Final Cluste	r Cluster 3 + Cluster 4	1.28	28

Table 5. Description of dendrogram 2010–2013

Table 6. Summary statistics for variable "agriculture value added per worker (constant 2005 US\$)"

	Mean	Median	Standard Deviation	Minimum	Maximum
Cluster 1	55 075.04	51 944.74	30 003.686	13 489.403	135 039.16
Cluster 2	19 156.195	15 143.258	8 837.2316	10 259.49	31 698.315
Cluster 4	8 340.2009	8 235.9708	3 495.445	3 158.2919	12 976.06

exhibits a decreasing trend. From Tables 1, 3 and 5 we can see that variance drops from a value of 0.18 in the period of 2002-2005 to a value of 0.15 in the period of 2010–2013. The countries in cluster one are significantly different from other groups of countries in the period of 2002–2005 in terms of the food production index, livestock production index and agriculture value added per worker variables, for which they reached the highest values, and in terms of the value added (annual % growth) and agriculture value added (% of GDP) variables, for which they had the smallest values. In the periods of 2006-2009 and 2010-2013, countries grouped in cluster one are significantly different from other groups of countries only in terms of the agriculture value added per worker variable. Figure 4 visualizes the mean values of the agriculture value added per worker variable for different clusters and periods. Figure 5 visualizes the dynamics of the analysed variable over the time periods of 2002–2005, 2006–2009 and 2010-2013.

We will skip the second group, and now discuss the third group of countries that is represented by cluster five in the period of 2002–2005 and cluster four in the periods of 2006–2009 and 2010–2013. In the period of 2002–2005, this group contains the 2004 EU accession countries, namely, the Czech Republic, Hungary, Latvia, Lithuania, Poland, Romania and the Slovak Republic and, from the older members of the European Union, Ireland. These counties attain the highest values in the following indicators: agriculture value added (annual % growth) and agriculture value added (% of GDP), and the lowest in food production index and agriculture, value added per worker. In the period of 2006–2009 Hungary, Lithuania, the Slovak Republic and Romania leave this group. In the period of 2010–2013, Hungary returns to this

□2002-2005 □2006-2009 □2010-2013

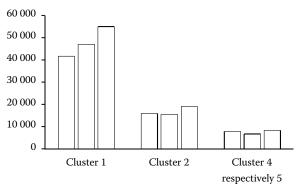


Figure 4. The mean values of the "agriculture value added per worker" variable for clusters one, two and five, as well as four, in the periods of 2002–2005, 2006–2009 and 2010–2013

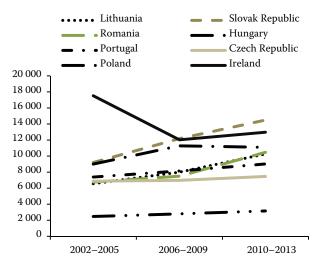


Figure 5. The trend over time of the mean values of the variable "agriculture value added per worker" for selected countries in the periods of 2002–2005, 2006–2009 and 2010–2013

group and Portugal joins this group. In the periods of 2006–2009 and 2010–2013, this group is significantly different from other groups of countries in terms of the agriculture value added per worker indicator (see Figure 4). This group is the most heterogeneous because of the highest values of variance, but there is also a decreasing trend. Variability has dropped from 0.7 in the period of 2002–2005 to 0.67 in the period of 2010–2013. From the dendogram, it can be seen that Poland significantly increases heterogeneity in all three periods studied.

The second group of countries is represented by cluster two. This group contains different numbers of countries in all the three periods analysed: 2002–2005, 2006–2009 and 2010–2013. In the period of 2002 to

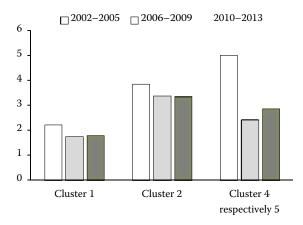


Figure 6. The mean values of the "agriculture value added (% of GDP)" variable for clusters one, two and five, as well as four, in the periods of 2002–2005, 2006–2009 and 2010–2013

2005, this group contains Austria, Bulgaria, Croatia, Estonia, Germany, Greece, Portugal and the United Kingdom. In the period of 2006–2009, this group is joined by Hungary, Lithuania, the Slovak Republic and Romania; in the period of 2010–2013, Hungary and Portugal leave this group. Again, this group of countries is significantly different from other groups in terms of the agriculture value added per worker indicator (see Figure 4). The homogeneity of this group as measured by variance has increased from an initial value of 0.15 in the period of 2002–2005 to 0.19 in period of 2010–2013.

Since the agriculture value added per worker indicator is statistically significant in all the three periods analysed, Figure 5 plots the time trend of the means of this variable for the selected countries, in particular the countries which moved between the second and third groups of countries in the periods 2006-2009 and 2010-2013. Lithuania, the Slovak Republic and Romania exhibit a relatively pronounced increasing trend in agriculture value added per worker compared to the Czech Republic, Poland, Portugal, Ireland and Hungary, which all exhibit decreasing trends in the period studied. The relationship between labour and performance in agriculture is discussed in more detail in, for example, Szabo and Grznár (2015b). Our results are in agreement with the results of Spicka (2013) that show that the average labour input per 100 hectares in the new EU member states is substantially higher than in the older EU member states and generates relatively low income per hectare. Spicka also relates this fact to the lower level of technological equipment and farming technologies in new EU member states. It is also important to mention that agriculture, besides its primary function of supplying food and fibre, is the main contributor to rural vitality as it generates rural employment, impedes rural depopulation, and keeps traditions alive (OECD 2001).

For our discussion, we also select an indicator that was significant only in the first period analysed, namely, agriculture value added (% of GDP) in the period of 2002–2005. In Figure 6, we can see a decreasing trend of cluster means in the period analysed. Despite the decline in the relative economic weight of the primary sector as an inevitable consequence of economic progress (Byerlee et al. 2009), the economic role of agriculture remains significant in many rural areas. Indeed, the economic importance of agriculture is generally much greater in Eastern and Southern Europe than in the west and north (European Commission 2013).

CONCLUSION

This article describes a multivariate cluster analysis of agriculture and food industry sector performance in EU countries in the period of 2002 to 2013, and identifies significant differences as well as the dynamics of change. The individual countries are clustered based on selected World Bank agriculture and rural development indicators. Based on the results obtained by applying Ward's method of clustering, we were able to group EU countries into three relatively homogeneous groups. The first group consisted of the same twelve countries in all three analysed periods: 2002–2005, 2006–2009 and 2010–2013; namely, Belgium, Denmark, Cyprus, Finland, France, Italy, Luxembourg, Malta, Netherlands, Slovenia, Spain and Sweden. The second group was represented by different numbers of countries in all three periods analysed: 2002-2005, 2006-2009 and 2010-2013. In the period of 2002-2005, the second group contained Austria, Bulgaria, Croatia, Estonia, Germany, Greece, Portugal and the United Kingdom. In the period of 2006–2009, this group was joined by Hungary, Lithuania, the Slovak Republic and Romania coming from the third group. Hungary and Portugal left this group in the period of 2010–2013 and joined the third group. In the first period of 2002–2005, the third group of countries comprised the Czech Republic, Hungary, Latvia, Lithuania, Poland, Romania, the Slovak Republic and Ireland.

In the next step, the differences between groups were verified. To identify indicators that were of a significantly different level in one class compared to another, the Kruskal-Wallis rank test procedure was used. The Kruskal-Wallis rank test indicated the presence of statistically significant differences between groups at a level of significance of 0.05 for a number of analysed variables, including food production index, livestock production index, agriculture value added (annual % growth), agriculture value added (% of GDP) and agriculture value added per worker only in the first analysed period of 2002-2005. In the subsequent periods of 2006–2009 and 2010–2013 the only significant difference was in the agriculture value added per worker.

In conclusion, although the countries analysed in our research did not historically have the same starting position, their economic trajectory after European Union enlargement has identical directions. The results of the analysis indicate convergence of analysed indicators in the given period and reveal one remaining significant difference in the agriculture value added per worker indicator at the beginning and end of the analysed period. For the EU countries in our study, the presence of more than one cluster strongly suggests that achieving particular future outcomes is likely to be a function of common strategies of national development within each cluster, thus confirming the importance of the design of the future CAP.

REFERENCES

- Byerlee D., De Janvry A., Sadoulet E. (2009): Agriculture for development: toward a new paradigm. Annual Review of Resource Economics, 1: 15–35.
- Ciaian P., Kancs D., Swinnen, J. F. M. (2014): The impact of the 2013 reform of the common agricultural policy on land capitalization in the European Union. Applied Economic Perspectives and Policy, 36: 643–673.
- Dos Santos M.L. (2013): Segmenting farms in the European Union. Agricultural Economics Czech, 59: 49–57.
- European Commission (2013): Structure and Dynamics of EU Farms: Changes, Trends and Policy Relevance. EU Agricultural Economics Briefs No 9. Brussels.
- European Commission (2016): Agriculture and Rural Development: the Common Agricultural Policy after 2013, Available at http://ec.europa.eu/agriculture/ cap-post-2013/index_en.htm (accessed Jan 31, 2016).
- European Council (2001): Presidency Conclusions. Goteborg, June 15–16 2001, SN 200/1/01 REV 1.
- Giannakis A., Bruggeman E. (2015): The highly variable economic performance of European agriculture. Land Use Policy, 45: 26–35.
- Lance G.N., Williams W.T. (1967): A general theory of classificatory sorting strategies. I. Hierarchical systems. The Computer Journal, 9: 373–380.
- MATLAB 8.0 and Statistics Toolbox 8.1. The MathWorks, Inc., Natick, Massachusetts.
- OECD (2001): Multifunctionality Towards an Analytical Framework. Organisation for Economic Cooperation and Development, Paris.
- Spicka J. (2013): The economic disparity in European agriculture in the context of the recent EU enlargements. Journal of Economics and Sustainable Development, 4: 125–133.
- Svoboda J., Lososová J., Zdeněk R. (2014): Analysis of operating costs of subsidies in the field of agriculture of EU countries. Agris on-line Papers in Economics and Informatics, 7: 161–173.
- Szabo Ľ., Grznár M. (2015a): Agriculture in the EU and position of the Slovak Republic. Agricultural Econonomics – Czech, 61: 493–501.

- Szabo Ľ., Grznár M. (2015b): Pracovné sily a výkonnosť poľnohospodárstva v SR. Ekonomika poľnohospodárstva, 15: 4–13.
- Ward J.H. (1963): Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58: 236–244.
- World Bank (2016): World Development Indicators. Available at http://data.worldbank.org/indicator/all (accessed June 6, 2016).

Received June 27, 2016 Accepted November 22, 2016 Published online July 21, 2017