OECONOMIA
COPERNICANA

2012 NR 2
ISSN 2083-1277

Marta Kuc[*]
Nicolaus Copernicus University

THE USE OF TAXONOMY METHODS FOR CLUSTERING EUROPEAN UNION COUNTRIES DUE TO THE STANDARD OF LIVING

JEL Classification: *C38*

Keywords: *standard of living, taxonomy methods, comparative analysis, Ward's method, k-means clustering*

Abstract: *This paper proposes the application of taxonomic tools to study the differentiation of standard of living in the European Union countries. The aggregate distance between given countries is the basis for grouping member states in terms of their internal structure of the studied characteristics. The analysis is based on two chosen methods–the Ward's and k-means method. The study included 24 member states of the European Union in 1995-2010. Depending on the distance between the object, the countries were divided into two or four clusters. Similar configuration of each group obtained using both methods has led to the conclusion that these methods can be used both complementarily and separately.*

Date of submission: May 24, 2011; date of acceptance: April 21, 2012.

[*] Contact: martakuc@doktorant.umk.pl, Nicolaus Copernicus University, Faculty of Economic Sciences and Management, ul. Gagarina 13a, 87-100 Toruń.

INTRODUCTION

Taxonomy as a science begins to play an increasingly important role in the studies of various socio-economic processes.

The research goal of this article is to define the diversity of the European Union countries in terms of standard of living and designation the group of countries with similar standard of living. The standard of living is understood as the degree of satisfying the population's needs of consumption of material goods and services, as well as natural and social environment benefits (Bywalec, Wydmus 1992). The analysis used a hierarchical procedure constructed by J.H. Ward and the method of k-means.

The taxonomic measure has been used to describe changes in the standard of living which were taking place in EU countries during 1995-2010. 35 diagnostic variables have been used. All those variables according to formal and essential conditions are crucial to describe the examined phenomenon. The empirical material was taken from databases published by Eurostat, Euromonitor and the World Health Organization.

THE STANDARD OF LIVING

Nowadays, the standard of living has an increasing role in the European Union integration process. Without doubts there is a need to analyze the standard of living issue because it is a source for defining goals and measuring the effectiveness of social policy. The choice of specific definition of standard of living has a significant impact on the selection of diagnostic variables. As mentioned before definition proposed by Bywalec and Wydmus (1992) is used in this article. Preparation of reliable and comparable data for the standard of living is problematic due to the lack of data. Also the dynamic analysis can be difficult because of the lack of data from the past to describe phenomena relevant in the present time. In this paper, the author focused on variables which can be gathered from statistical databases. The set of diagnostic variables was divided into the following categories: population, labor market and job security, health and social care, education, recreation living conditions, transport and communications, social security, population incomes and expenditures and the degradation and protection environment (see table 1).

The analysis of such demographic variables as total fertility rate and demographic dependency ratio of elderly people gives some relevant information about the standard of living. Tapering replacement rate and longer length of human life contribute the transition from stationary to the regressive society age structure. The demographic situation directly deter-

mines the countries' opportunity of development, characterizing the potential labor force in a given country.

The basic indicator of the situation at the labour market is unemployment rate. It is obvious that high unemployment rate negatively affects the society's moral and their perception of the standard of living. The number of deaths due to accident at work is a diagnostic variable because we want to have decent working condition corresponding with safety and health at work.

The disproportion in the health care can be described in terms of access to various types of medical services measured by the number of doctors, nurses and hospital beds per 100 thousand people. However, not only the number of medical stuff determines the standard of living but mainly good health care organization. The indicators describing the number of deaths from lifestyle diseases (for example diabetes), illustrate the effects of health care activity, promoting healthy lifestyles and disease prevention and society's health awareness.

The standard of living is also affected by the ability and quality of education. In this study, society's level of education is characterized by the indicator specifying the number of students per 1,000 people. The intellectual potential of society can be developed through public education at the university level. The quality of education is measured by the number of teachers per one student. Increased activity in academic environment stimulates entrepreneurship and promotes improving skills.

In this paper, the differences in the availability and use of leisure time and cultural objects are described by the average annual number of cinema trips per capita and the number of hotels per 1,000 people. It is worth mentioning that the access to higher-order goods, including the culture, is an indicator of increased standard of living.

Table 1. The final set of variables

Variable's symbol	Variables
1. Population	
$X_{1,1}$	Total fertility rate
$X_{1,2}$	Demographic dependency ratio of elderly people (in %)
2. Labour market and job security	
$X_{2,1}$	Unemployment rate (in %)
$X_{2,2}$	Number of deaths due to accident at work per 100000 inhabitants

Table 1 Continued

Variable's symbol	Variables
	3. Health and social care
$X_{3,1}$	Number of deaths due to cancer per 100000 inhabitants
$X_{3,2}$	Number of deaths due to diabetes per 100000 inhabitants
$X_{3,3}$	Number of new AIDS cases per 100000 inhabitants
$X_{3,4}$	Number of doctors per 100000 inhabitants
$X_{3,5}$	Number of nurses per 100000 inhabitants
$X_{3,6}$	Number of hospital beds per 100000 inhabitants
$X_{3,7}$	Obese population (BMI 30kg/sq m or more) as a percentage of population aged 15+
	4. Education
$X_{4,1}$	Number of university students per 1000 inhabitants
$X_{4,2}$	Number of academic teachers per 1 student
	5. Recreation, culture and leisure time
$X_{5,1}$	Annual cinema trips per capita
$X_{5,2}$	Number of hotels per 1000 inhabitants
	6. Living conditions
$X_{6,1}$	Number of newly built dwellings per 1000 households
	7. Transport and communication
$X_{7,1}$	Number of newly registered cars per 1000 inhabitants
$X_{7,2}$	Length of expressways in km per 1 sq km of land
$X_{7,3}$	Proportion of paved roads as a percentage of total road network
$X_{7,4}$	Density of road network in km per 1 sq km of land
$X_{7,5}$	Length of public railway network operated per 1000 sq km of land
$X_{7,6}$	Number of mobile phones subscribers per 100 inhabitants
$X_{7,7}$	Railway passenger traffic in million passenger-km per 1000 inhabitants
$X_{7,8}$	Airline passenger traffic in millions of passenger-km per 1000 inhabitants
	8. Social security
$X_{8,1}$	Number of suicides and self-harm per 100 thousand inhabitants
$X_{8,2}$	Number of divorces per 1000 inhabitants
$X_{8,3}$	Number of crimes per 100 thousand inhabitants
	9. Population incomes and expenditures
$X_{9,1}$	Annual average rate of inflation (in %)
$X_{9,2}$	Gross domestic product per capita in USD
$X_{9,3}$	Household saving as % of disposable income

Table 1 Continued

Variable's symbol	Variables
10. Degradation and protection of the environment	
$X_{10,1}$	Sulfur oxides emissions in kg per capita
$X_{10,2}$	Nitrogen oxide emissions in kg per capita
$X_{10,3}$	Carbon monoxide emissions in kg per capita
$X_{10,4}$	Nationally protected areas as a percentage of land
$X_{10,5}$	Forest land as a percentage of land

Source: Author's own study.

The housing situation significantly determines the standard of living. In this study, due to lack of data, only one variable is considered – i.e. number of newly built dwellings per 1000 households. Without doubt, improving housing conditions is one of the most important socio-economic problems. For less developed countries is extremely difficult to overcome the "housing gap" in relation to more developed countries. The cause of this condition may be a low level of housing investment, rising housing prices and declining housing assistance.

Undoubtedly, the ability to communicate and society's mobility are affected by such factors as: a well-developed network of highways and railways, the number of mobile subscribers, Internet access and so on. Well-developed transport infrastructure stimulate the entrepreneurship and conductive business environment.

Social security seems to be another important factor influencing the standard of living. In recent years, the increase of threats taking the form of car theft, breaking into homes, as well as robberies, assaults and even killings are noticeable. Number of divorces is also included in this group, affecting the psychological sense of security and stability.

The condition of the environment significantly affects the standard of living, indirectly affecting the health and length of life. It is characterized by natural conditions (the degree of forest cover and protected area of the country) and the environment pollution (air pollution caused by emissions of harmful gases).

DESCRIPTION OF CHOSEN CLUSTERING METHODS

Isolating groups of similar objects can be carried out by various methods. One can use methods based on the synthetic measures of development, as well as methods based on taxonomic similarity of objects. In economic studies, hierarchical clustering methods based on taxonomic similarity is the one commonly used. Among the hierarchical clustering methods we can distinguish agglomerative and divisive methods. Agglomerative methods are based on the assumption that every object is a separate class, and pairs of clusters are merged as one moves up the hierarchy, so called "bottom up" approach. The representative of agglomerative clustering methods is discussed below – Ward's method. This method was proposed by J.H. Ward in 1963. The Ward approach can be described in five steps:

1. assumption that every object $Q_i(i = 1, \dots, m)$ is a separate class,
2. based on the lowest value in the distance matrix a pair of the most similar object p and q is pointed out,
3. p and q objects are formed into one cluster, reducing the number of clusters to $m - 1$,
4. the distance between newly formed cluster and other objects is calculated,
5. steps 2-4 are repeated until sample units are combined into a single large cluster of size m.

In Ward's approach the distance between newly created group and other objects is the difference between the error sum of squares within the individual units from the centroid of group to which these objects belong to. The error sum of squares is calculated due to the formula (Rószkiewicz, 2002):

$$ESS = \sum_{i=1}^{k} x_i^2 - \frac{1}{k}\left(\sum_{i=1}^{k} x_i\right)^2, \qquad (1)$$

where:
x_i – value which is the criterion of segmentation for ith unit,
k – number of units in given group.

The results of the analysis are presented in a dendrogram, which shows the successive mergers of units, together with a minimum distance at which the merger took place. On such a graph, we establish a certain arbitrary value. The mergers which are the closest to the arbitrary value, but do not

exceed it are the optimal groups of analyzed objects (Grabiński 2003). The Ward method gives clusters with low numbers of elements. The disadvantage of this method is its sensitiveness to extreme values (Młodak 2006) and freedom of choosing the number and size of groups. (Stec, Janas 2009).

In divisive methods, the algorithm is opposite to agglomerative methods'. At the first stage, all observations are in one cluster, and splits are performed recursively as one moves down the hierarchy. It so called "top down" approach. One of the divisive methods is k-means method. The algorithm can be described as (Rószkiewicz 2002):

1. setting up a priori number of k clusters and maximum number of iterations,
2. division of units according to pre-selected representatives of each segment, the choice may be made random, arbitrary, or due to most different units,
3. revision of the obtained solution by calculating the distance between centroid of groups and each unit.
4. the procedure ends when sum of squares shows no significant changes. Sum of squares is calculated as (Rószkiewicz 2002):

$$SES = \sum_{i=1}^{k} \sum_{j=1}^{n_i} d_{jS_i}^2, \qquad (2)$$

where:
SES – sum of squares,
d_{jS_i}– distance between the jth unit and center of gravity of the ith cluster.

The essence of this approach is to create the most distinct clusters by minimizing intra-group variance while maximizing inter-cluster variance. The procedure ends when the sum of squares shows no significant changes or when the maximum change of centroids does not exceed the arbitrary value. The procedure can also end when the number of iteration is used. The advantage of this method is the result of k different as much as possible groups. However, the necessity of determining the number of k clusters beforehand can be perceived as a disadvantage of this approach (Stec, Janas 2009). Thus, I recommended to use other clustering methods to determine the initial number of groups.

The presented methods of classification differ from one another in terms of assumptions, grouping procedure and results presentation. The empirical research conducted in this paper shows the results of using Ward's and k-means methods.

EUROPEAN UNION COUNTRIES CLUSTERING USING CHOSEN GROUPING METHODS

European Union countries clustering is based on 35 socio-economical variables, divided into $p = 10$ groups, each of which consist of $k_1 = 2$, $k_2 = 2$, $k_3 = 7$, $k_4 = 2$, $k_5 = 2$, $k_6 = 1$, $k_7 = 8$, $k_8 = 3$, $k_9 = 3$, $k_{10} = 5$ representatives. Detailed reasons for selecting the indicators which determine the standard of living can be found in literature (for example Zeliaś 2000; Zeliaś 2004; Malina 2004; Młodak 2006).

Diagnostic variables tend to have different characteristics, so their direct comparisons are impossible. In this case, the normalization procedures should be applied. In this study the following transformation has been used:

$$s_{ijt} = \frac{x_{ijt}}{\max\limits_{i}\{x_{ijt}\}}, \qquad (3)$$

$$\max\limits_{i}\{x_{ijt}\} \neq 0 \;\; i = 1,2,\dots,m; j = 1,2,\dots,k; t = 1,2,\dots,n),$$

where:
s_{ijt} - normalized value of the jth variable on object i in a time t,
x_{ijt} – real value of the jth variable on object i in a time t,
$\max_i\{x_{ijt}\}$ – maximum value of the jth variable.

The value of the pattern should be regarded as a "moving target", i.e. the maximum value which can be achieved in a given year.

The set of standardized variables was used to calculate the distance between analyzed objects. This is why the Euclidean metric was used:

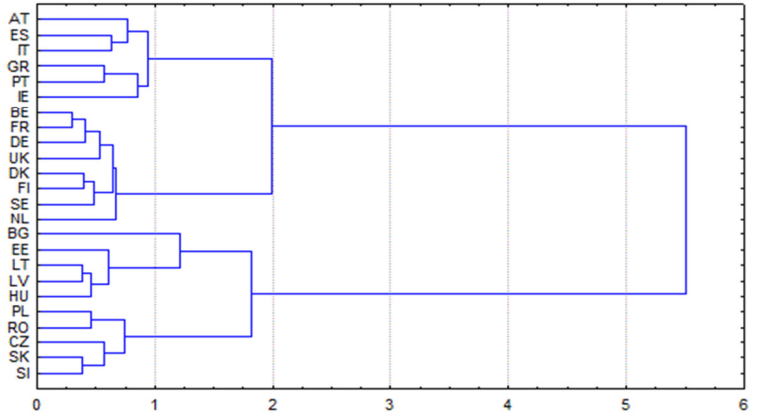
$$d_{ij} = \sqrt{\sum_{k=1}^{p}\left(x_{ik} - x_{jk}\right)^2}, \qquad (4)$$

where:
d_{ij} – the Euclidean metric between ith and jth unit,
x_{ik} – the value of k varialbe for ith country.

The basis for the determination of homogeneous groups of countries in the Ward's method is the size of the minimum joint between-within measure of distance between clusters (Rizzo 2005). So one can find groups which are tightly knit and distinct from each other (Kettenring 2006). In this article, only the results for 1995 and 2010 will be precisely described. Figure 1 and 2 present the clustering of countries using the Ward's method.

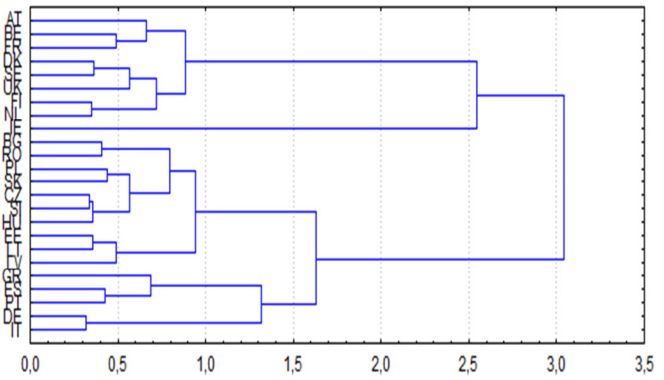
When analyzing these dendrograms one can see that there are significant differences between "old" and "new" European Union countries. However, these differences are becoming smaller, in 1995 the distance between these two groups was 5,50, in 2010 it was only 3.10. In 1995, the first cluster consists of Austria (AT), Belgium (BE), Denmark (DK), Finland (FI), France (FR), Greece (GR), Spain (ES), Netherlands (NL), Ireland (IE), Germany (DE), Portugal (PT), Sweden (SE), United Kingdom (UK) and Italy (IT). The second group includes the: Bulgaria (BG), Estonia (EE), Lithuania (LT), Latvia (LV), Poland (PL), Czech Republic (CZ), Romania (RO), Slovakia (SK), Slovenia (SI) and Hungary (HU).

Figure 1. Countries clustering using the Ward's method and Euclidean metric in 1995



Source: Author's own study.

Figure 2. Countries clustering using the Ward's method and Euclidean metric in 2010



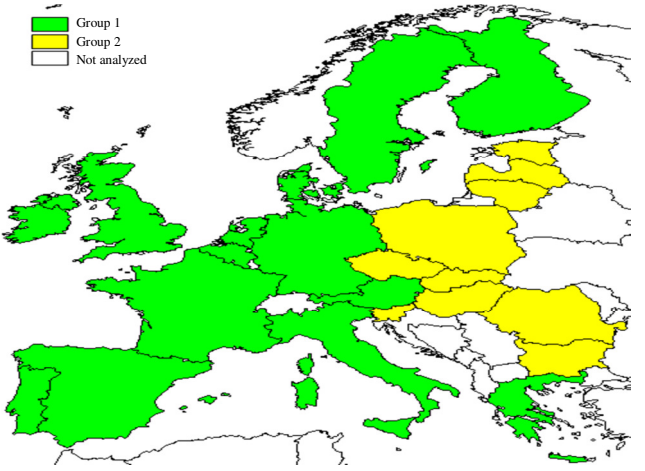
Source: Author's own study.

The set of groups was almost the same in the 1995-2010, the shift occurs only in relation to Portugal, which in 2003 and was in group 2. However, after the global crisis there was some changes in countries clustering. The countries such as: Greece, Spain, Germany, Portugal and Italy which formerly belonged to the first group in 2008-2010 were classified into second group. See figure 3 and 4.

However, the division of the European Union in two clusters of countries seems to be too trivial to extract the typologically similar objects. It was therefore decided to extract four homogeneous groups of countries. In 1995 the clusters were as follows:

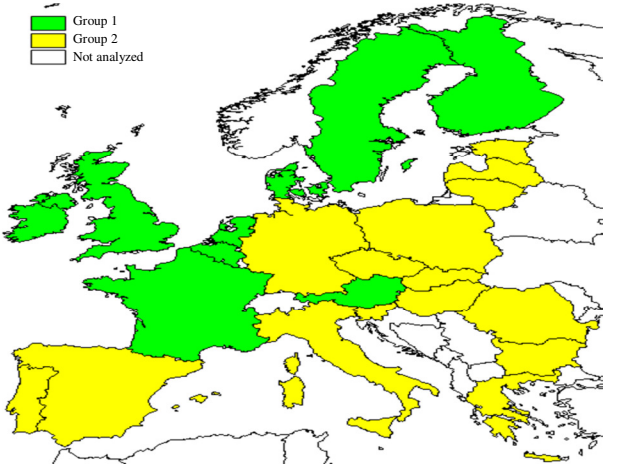
- Group 1 (high standard of living) – Austria (AT), Greece (GR), Spain (ES), Ireland (IE), Portugal (PT), Italy (IT);
- Group 2 (medium standard of living) – Belgium (BE), Denmark (DK), Finland (FI), France (FR), Netherlands (NL), Germany (DE), Sweden (SE), United Kingdom (UK);
- Group 3 (low standard of living) – Bulgaria (BG), Estonia (EE), Lithuania (LT) Latvia (LV), Hungary (HU);
- Group 4 (the lowest standard of living) – Poland (PL), Czech Republic (CZ), Romania (RO), Slovakia (SK), Slovenia (SI).

Figure 3. Countries clustering using the Ward's method and Euclidean metric in 1995 – two groups



Source: Author's own study.

Figure 4. Countries clustering using the Ward's method and Euclidean metric in 2010 – two groups

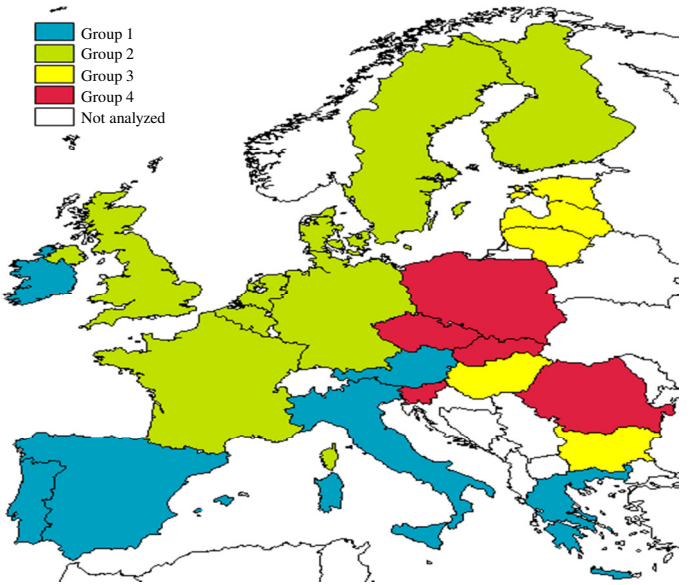


Source: Author's own study.

In 2010 the clusters were as follows:

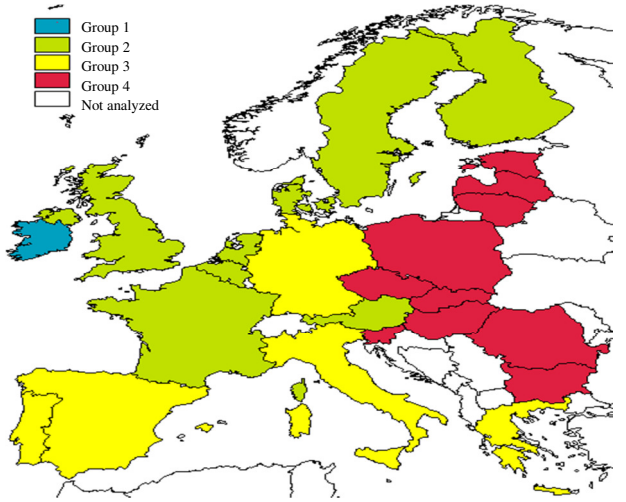
- Group 1 (high quality of life) – Ireland (IE);
- Group 2 (medium quality of life) – Austria (AT), Belgium (BE), Denmark (DK), Finland (FI), France (FR), Netherlands (NL), Sweden (SE), United Kingdom (UK);
- Group 3 (low quality of life) – Denmark (DE), Greece (GR), Portugal (PT),
  Spain (ES), Italy (IT);
- Group 4 (the lowest quality of life) – Bulgaria (BG), Estonia (EE), Lithuania (LT), Latvia (LV), Hungary (HU). Poland (PL), Portugal (PT), Czech Republic (CZ), Romania (RO), Slovakia (SK), Slovenia (SI). Figure 5 and 6 illustrates this situation.

Figure 5. Countries clustering using the Ward's method and Euclidean metric in 1995 – 4 groups



Source: Author's own study.

Figure 6. Countries clustering using the Ward's method and Euclidean metric in 2010 – 4 groups



Source: Author's own study.

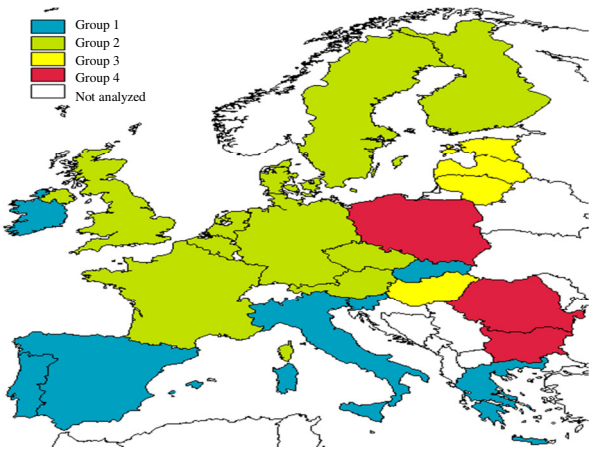
The results of Ward's classification shows that there have been significant changes in the standard of living. Already in 1996, Bulgaria significantly worsened their position, declined in relation to Estonia, Latvia, Lithuania and Hungary, and formed a separate cluster. In 1997-2000 a similar situation occurred in Romania. These countries, in 2001 formed a separate group. Significant changes are noticeable in 2002. Ireland, which had previously formed a group with Greece, Spain and Portugal, is now the separate class. Bulgaria and Romania moved closer to the level of living in Estonia, Latvia, Lithuania, Poland, Czech Republic, Slovakia and Slovenia, forming a cluster with them. In 2003 Portugal, which previously was in group together with Greece and Spain, joined the group number 3. In 2004, Greece and Spain have brought back to Ireland and formed one cluster. In the same year, the Czech Republic, Slovenia and Portugal improved their position, creating a group with Italy. However, the improvement of the situation in the Czech Republic and Portugal was only momentary, because a year later (2005) they went back to the cluster with Bulgaria, Poland, Romania, Slovenia and Slovakia. The same clusters were also observed in 2007. The results confirm the hypothesis about huge spatial differentiation of living standards in the European Union countries. The global crisis has worsened this situation. The countries such as: Greece, Spain,

Portugal and Italy, which formerly belonged to the first group, in 2008-2010 were classified into the third group.

Thus, the standard of living was significantly decreased, because of the employment crisis and property bubble. It is intrusting that despite the crisis Ireland was able to keep its strong leading position. In 2010, Ireland was the only one in the first group with the highest standard of living.

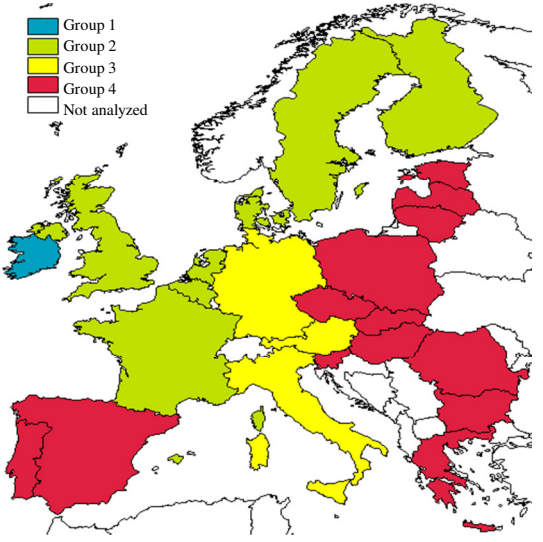
In order to verify the results obtained by Ward's approach, the countries were grouped using the k-means method. K-means is the method which minimizes an ordinary least-squares fitting function (A. Chaturvedi). Also in this case, the basis for taxonomic analysis was a set of standardized diagnostic variables. The two options were distinguished in this analysis, k = 2 and k = 4 with maximum number of iterations – 10. The classification obtained in this way is almost the same as the classification obtained by Ward's method. The differences occur in 1995, where Ward's method placed the Czech Republic and Slovakia among the "new" EU countries, while in the method of k-means, these countries belong to the group of "old" EU countries. A significant difference occurred in 1996, using the k-means method Bulgaria creates independent cluster, and the Ward algorithm classifies it with the countries of the "new" EU countries. In other cases, the classifications are consistent, with one exception (Portugal in 2004). The division into two groups, however, is not sufficient, therefore, we set up the optimal division into four homogeneous groups. Figure 7 and 8 presents the results of the classification for 1995 and 2010.

Figure 7. Countries clustering using the k-means method for k=4 in 1995



Source: Author's own study.

Figure 8. Countries clustering using the k-means method for k=4 in 2010



Source: Author's own study.

The results obtained by using both methods seems to be consistent. In 1996, 3rd and 4th groups are the same. There are only slight shifts of Portugal and Italy, which using the Ward's method are placed in the group with Ireland, Spain and Greece, and as a result of k-means clustering they belong to the first group. In 1997, the classification results do not coincide only for Portugal. As a result of Ward's algorithm in 1998, Italy belonged to the "Irish group", while the k-means method classified it in first group. The results obtained in 1999-2003 were entirely consistent. The discrepancies appeared in 2004, when the effect of the k-means method created separate cluster of Ireland, and Ward's method classified it with Greece and Spain, the same situation also took place in 2005-2006. In 2004 Ward's method created a group consisted of Portugal, the Czech Republic, Slovenia and Italy, where the k-means method classified the first two countries into the 4th group, while the remaining two countries in the 3rd group. The results for 2005 show almost complete coverage of the 3rd group obtained using the k-means method and the 2nd group obtained using Ward's method, there is only one difference in the case of Slovenia. The countries in the 4th group received using k-means approach, according to Ward's approach were divided into two separate groups, the first one – Estonia, Latvia, Lithuania, Hungary and the second one – Bulgaria, Poland, Portugal, the Czech

Republic, Romania and Slovakia. The classification in year 2006 was similar to 2005. Also, the results from 2007 shows almost complete coverage of 3rd group in the k-means method to 2nd group in the Ward's method, the shift occurs only in the case of Slovenia. The discrepancy also occurred for Greece, which was qualified with Ireland and Spain (Ward's method), but according to k-means approach it belonged to the group with Bulgaria and Romania. There is also a small difference in 2010, when Austria using k-means clustering was classified with Germany and Italy and using Ward's method was classified in second group. Using Ward's method, Germany and Italy were classified together with Greece, Spain and Portugal – the countries which are in economic chaos.

The results of the analysis confirm the assumption about the spatial differentiation of the quality of life in the European Union. The first two groups of countries have a higher standard of living. It is impossible not to note that the new member states have a lower standard of living.

TESTING RESULTS COMPATIBILITY

The earlier analysis shows that the results of clustering in both methods are very similar. However, we should conduct a detailed analysis of accuracy of the results obtained by the two methods. For this purpose, contingency tables were constructed and values of Cramer's V were calculated (Grabiński, Wydmus, Zeliaś 1983):

$$V = \sqrt{\frac{\chi^2}{m \cdot min\{(r-1),(s-1)\}}} \qquad (5)$$

where:
m – number of objects,
r and s – size of the contingency table,
χ^2– value of chi-square statistics, calculated as:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(m_{ij} - \hat{m}_{ij}\right)^2}{\hat{m}_{ij}}, \qquad (6)$$

where:
m_{ij} – values calculated inside the contingency table,
$\hat{m}_{ij}$ – theoretical values, calculated as:

$$\hat{m}_{ij} = \frac{m_{i.}m_{.j}}{m_m} \quad (i = 1, \dots, r; j = 1, \dots, s), \qquad (7)$$

where:
$m_{i.}, m_{.j}$ – boundary elements of the contingency table.

Cramer's V is defined between ⟨1,0⟩, wherein the closer a coefficient is to unity, the higher accuracy of the results. Table 2 contains the calculated values of Cramer's V for the classification results by both methods. Cramer's V calculated according to the formula (5) is compared to the threshold value V*= 0.485. If the calculated value of Cramer's V exceeds the threshold value, it means that there is agreement in the classification in the considered time period. Analyzing the results from table 2, we can see a high accuracy of the results. The only significant differences occurred between 1995 and 1996 using the k-means method. Identical results were observed in 1999-2000, using Ward's method and in the years 1998-2001 and 2004-2006 using k-means clustering.

Table 2. Cramer's V values for both clustering methods in comparable years

Comparable years	Ward's metod	k-means clustering
1995 and 1996	0,786	0,575
1996 and 1997	0,760	0,819
1997 and 1998	0,952	0,945
1998 and 1999	0,952	1,000
1999 and 2000	1,000	1,000
2000 and 2001	0,903	1,000
2001 and 2002	0,816	0,816
2002 and 2003	0,932	0,932
2003 and 2004	0,787	0,971
2004 and 2005	0,823	1,000
2005 and 2006	0,909	1,000
2006 and 2007	0,909	0,785
2007 and 2008	0,665	0,750
2008 and 2009	0,823	0,816
2009 and 2010	0,909	0,932

Sourcce: Author's own study.

The next step in the analysis was to examine the accuracy of the classification obtained from the two clustering methods. Table 3 shows the results.

Table 3. Cramer's V value between results obtained using Ward's method and k-means clustering

Year	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
V	0,711	0,911	0,945	0,911	1,000	1,000	1,000	1,000
Year	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
V	1,000	0,777	0,782	0,785	0,747	0,785	0,777	0,835

Source; Author's own study.

Analyzing the results in table 3, it can be seen that there is very strong agreement in classification obtained by both methods. Also in this case the whole set of Cramer's V is higher than V*. It should be pointed out that in the years 1999-2003 identical results were obtained. The analysis of accuracy of the results shows that the process of grouping the European Union countries was dealt with correctly.

CONCLUSIONS

The paper presents two clustering methods – Ward's method and k-means clustering, as well as the results of countries classifications based on these algorithms. The subject of empirical analysis are 24 EU member states in the years 1995-2010. Conducting empirical research on EU countries provides a basis for differentiating groups of countries with similar standard of living. Countries were divided into two or four groups depending on the linkage distance. The most explicit division is the one into two groups, which is the same as political division of "old" and "new" member countries. The specification of groups obtained using both methods are almost the same, which indicates the possibility of using these methods, both complementarily and individually.

LITERATURE

Bywalec C., Wydmus S. (1992), *Poziom życia ludności Polski w porównaniu z krajami Europejskiej Wspólnoty Gospodarczej*, „Ekonomista" nr 5.

Grabiński T. (2003), *Analiza taksonometryczna krajów Europy w ujęciu regionów,* Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.

Kettenring J. (2006), *The Practice of Cluster Analysis*, "Journal of Classification" No 23.

Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej,* Difin, Warszawa.

Rizzo M.L. (2005), *Hierarchical Clustering via Joint Between-Within Distances: Expending Ward's Minimum Variance Method*, "Journal of Classification", No 22.

Rószkiewicz M. (2002), *Narzędzia statystyczne w analizach marketingowych,* Wydawnictwo C.H. Beck, Warszawa 2002,

Stec M., Janas A. (2009), *Analiza porównawcza metod klasyfikacji województw,* „Wiadomości Statystyczne", nr 4.