Zeszyty Naukowe Uniwersytet Ekonomiczny w Krakowie

11 (947)

ISSN 1898-6447
Zesz. Nauk. UEK, 2015; 11 (947): 37–52
DOI: 10.15678/ZNUEK.2015.0947.1103

Mária Vojtková
Eva Sodomová
Department of Statistics
University of Economics in Bratislava, Slovakia

Classification of EU Countries according to Selected Indicators from the Field of Business Demography Using Self-organising Maps*

Abstract

Business Demography monitors the file of active enterprises – those actually doing business regardless of when they obtained the authorisation to operate, closed enterprises – those out of business regardless of when they were actually legally terminated, and enterprises surviving for a given time since their establishment. The result of this monitoring is a group of basic indicators used to characterise the number of newly established, closed and surviving businesses in the EU as well as indicators related to the number of people these enterprises employ.

The main aim of this article is to classify EU countries by selected derived data focused on business demography using a neural network model – Kohonen self-organising maps for the last monitored period. This model involves creating homogeneous groups of countries to be characterised by demographic indicators associated with the birth, survival, and deaths of enterprises and the related employment, all of which are inter-

* The article came into being within the project VEGA no. 1/0501/14 entitled "Business in terms of Demography – an Instrument for Assessing Changes in Growth and Employment Enterprise".

connected. This means the classic method of classification cannot be employed. Kohonen maps, however, can be, and may be enabled by various software. We decided to use the tools offered by the statistical analytical system SAS Enterprise Miner (SAS®EM).

An additional objective was to introduce business demography and a description of selected indicators derived in the EU from 2008 to the last reporting period, with a focus on economic development in Slovakia and in the Visegrad Group (V4).

Keywords: Kohonen self-organising maps, clustering, business demography, classification of EU countries.

1. Introduction

The current period is characterised mainly in the European Union by an emphasis on the corporate sector, which forms the basis for innovation and economic development. In 2000 in Lisbon the Union laid down one of its fundamental objectives: to build a "dynamic, knowledge-based economy, with continuous economic growth, with more and more jobs and greater social cohesion". The Lisbon strategy was refined in 2005 under the official name "Growth and employment strategy". In this context, the need arose to rapidly develop comparable data on business demography in order to monitor the implementation of policy measures in the social and economic spheres.

The main aim of this article will be to classify EU countries by selected derived data focused on business demography using a neural network model – Kohonen self-organising maps for the last monitored period. This is the creation of homogeneous groups of countries, which are characterised by demographic indicators associated with the birth, enterprise deaths and the related employment, which are interconnected, so the classic method of classification is unavailable. Use of Kohonen maps is possible through various software. We decided to use the tools of statistical analytical system SAS Enterprise Miner (SAS®EM).

The additional objective is to introduce business demography and description of selected indicators derived in the EU from 2008 to the last reporting period, with a focus on economic development.

2. Business Demography

The commercial register contains a list of legal entities (companies, organisations, institutions, businesses registered in the commercial register, etc.) and entrepreneurs (entrepreneurs, freelancers, self-employed farmers, etc.), who are entrepreneurs or are engaged in any gainful activity under special regulations. The register is a public list in the extent provided by law. The register itself does

not provide sufficient information on economic activities of the organisations and people registered.

There was therefore a need in the EU and OECD countries to develop a statistical or administrative data register that would register not only the organisations themselves, but also their activities. In the Slovak Republic, registration data on economic entities are collected in the Statistical Register of Organisations (REGIS). These data are largely descriptive in nature and are intended for purposes of state statistics. The register is harmonised with international registries, thereby ensuring international comparability of statistical data.

According to the definitions of the types of statistical units of Council Regulation (EEC) No. 696/93 on the statistical units, for the purposes of statistical surveys and analysis of economy in the Community are all legal and natural persons registered in the register of organisations of "legal entity". Legal units that produce products or services are also statistical units of the "enterprise"[1].

National Statistical Offices and hence the Statistical Office of the Slovak Republic create a file of so-called statistical businesses, namely economic entities that are economically active and also subject to statistical reporting and the basis for the drafting and presentation of statistical data on economic development. There is a set of statistical data, and the activity that deals with the collection of such data is called business demography. The Statistical Office of the European Union, Eurostat, maps and gathers comparable data on business demography in EU member states.

Business Demography monitors the file of active enterprises actually incurred, i.e. those actually doing business regardless of when they obtained the authorisation to operate; closed enterprises, i.e. that that are out of business regardless of the actual termination under the legal framework; and enterprises surviving for some time since its establishment. The result of this monitoring a group of basic indicators is to characterise the number of newly established, closed and surviving businesses in the EU as well as indicators related to the number of employed in these enterprises. From this point of view it is very important to properly define two basic concepts: newly established, and closed businesses.

According to the methodology of business demography, newly established enterprises are considered subjects, legal and natural persons-entrepreneurs, that had revenues or employees in a given year, and had no revenues or employees in the two preceding years. Newly established enterprises are also those formed as legal entities or natural persons-entrepreneurs by combining production factors, but without the participation of another enterprise. Closed businesses, on the other hand, are those that ceased to exist as legal entities or natural

[1] http://portal.statistics.sk/files/Sekcie/sek_200/Metodicke_pokyny/vseob_2010.rtf, accessed: 15.10.2014.

persons-entrepreneurs by termination of production factors with the restriction without the participation of another enterprise (Sodomová & Coss 2011).

Another part of the database on business demography is kept on derived indicators calculated from basic indicators. In our analysis we focused on these derived (relative) indicators (Eurostat 2007):

– *birth rate* (BR) – the number of enterprise births in the reference period (*t*) divided by the number of enterprises active in *t*;

– *death rate* (DR) – the number of enterprise deaths in the reference period (*t*) divided by the number of enterprises active in *t*;

– *employment share of enterprise births* (EB) – the number of persons employed in the reference period (*t*) among enterprises newly born in *t*, divided by the number of persons employed in *t*, among the stock of enterprises active in *t*;

– *survival rate 2* (SR) – the number of enterprises in the reference period (*t*) newly born in $t - 2$ that survived to *t* divided by the number of enterprise births in $t - 2$.

The source of this data is the Eurostat database, the data are updated with an 18-month shift, so the most recent data relate to the year 2012, but some countries have so far not indicated values for certain indicators, so we replaced them with the most recent data available. Greece does not provide such data and Malta started to provide data only in the last reporting period.

3. Development of Selected Business Demography Indicators

The analytical section begins by looking at the development of selected indicators of business demography derived for the period from 2008 to the last reporting period in the countries of the Visegrad Group and briefly also in the European Union.

Figures 1–4 show the development of the first three indicators in each V4 country. The development of the last reporting indicator will be shown separately because it is represented on a different scale (Fig. 5).

All V4 countries reported a decline in enterprise births in 2012 over the previous year, and also, with the exception of the Czech Republic, over the reference year 2008. The rate developed in each country, however, at significantly different speeds. Poland and Hungary have more balanced development, suggesting the existence of a more stable business environment in these countries. Slovakia had the highest number of births in the years and in the years 2008, 2009 and 2011, while Poland enjoyed that distinction in 2010 and 2012.

The enterprise death rate is monitored with the annual shift. In 2011 there was an increase in the death rate in all V4 countries over the previous year. Moreover,

the enterprise death rate was higher than the birth rate in the V4, which is negative. In Hungary, the death rate outstripped the birth rate for the entire period; in 2011 the difference between the death and birth rates reached 5.32%.

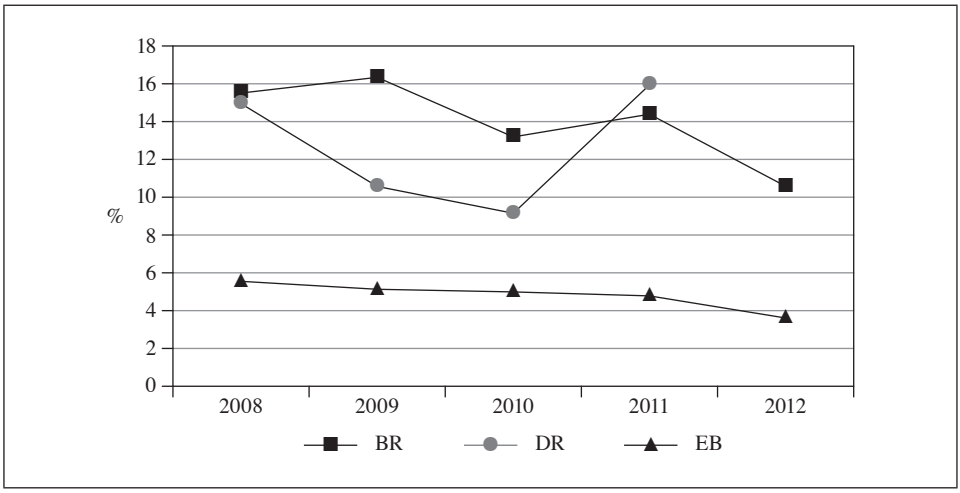

Fig. 1. Development of the Birth Rate (BR), Death Rate (DR) and Employment Share of Enterprise Births (EB) in Slovakia in the Years 2008–2012

Source: Eurostat and the authors' own elaboration.

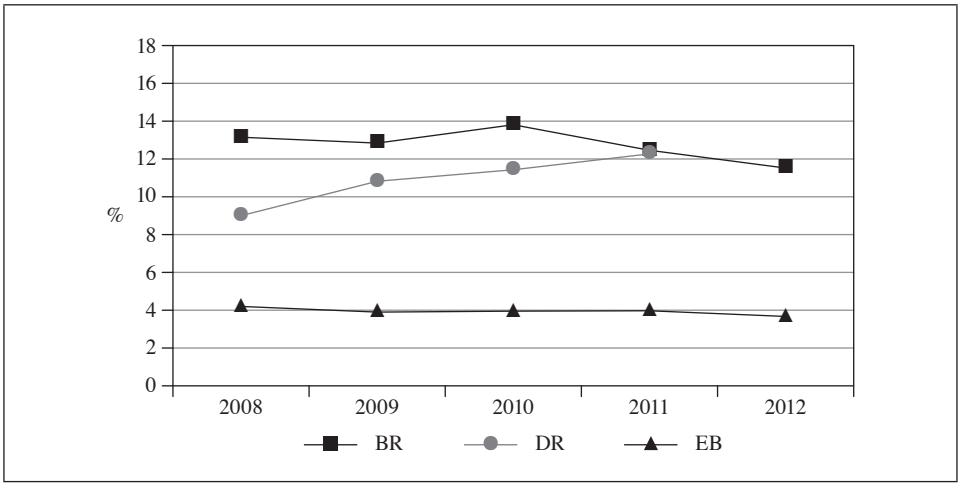

Fig. 2. Development of the Birth Rate (BR), Death Rate (DR) and Employment Share of Enterprise Births (EB) in Poland in the Years 2008–2012

Source: Eurostat and the authors' own elaboration.

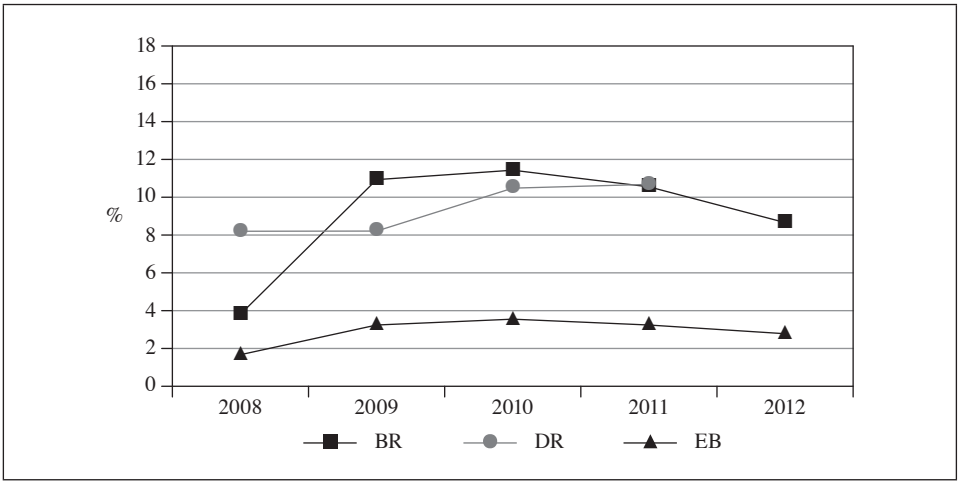

Fig. 3. Development of the Birth Rate (BR), Death Rate (DR) and Employment Share of Enterprise Births (EB) in Czech Republic in the Years 2008–2012

Source: Eurostat and the authors' own elaboration.

When the indicator of share of employment in newly born enterprises rises, development may be considered stable. In the last reporting year, there was a slight decline, though in Slovakia the decline was only 1.17%, while elsewhere in the V4 it did not breach even the 1% mark.

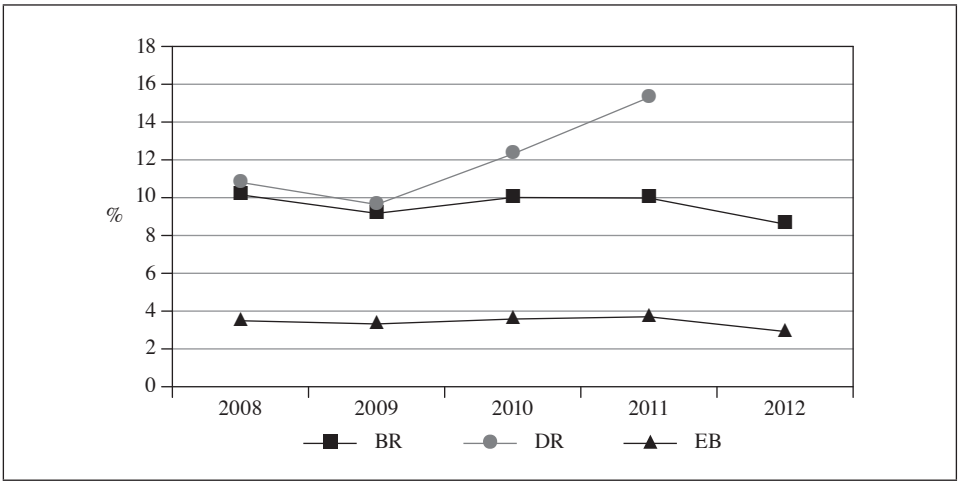

Fig. 4. Development of the Birth Rate (BR), Death Rate (DR) and Employment Share of Enterprise Births (EB) in Hungary in the Years 2008–2012

Source: Eurostat and the authors' own elaboration.

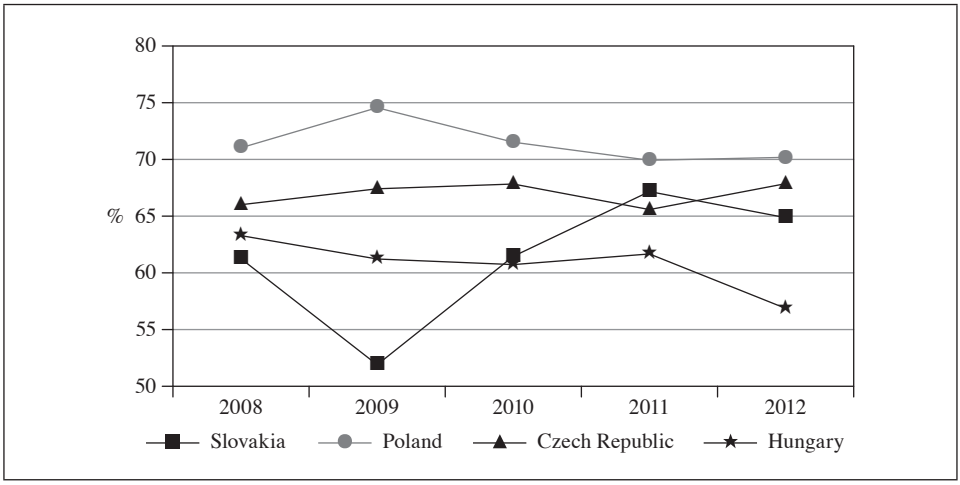

Fig. 5. Development of the Survival Rate in Slovakia, Poland, Czech Republic and Hungary in the Years 2008–2012

Source: Eurostat and the authors' own elaboration.

The highest enterprise survival rate values in all reporting years were achieved by Poland, followed by the Czech Republic (except in 2011). The lowest values were observed for Hungary. Slovakia reported a significant increase in the enterprise survival rate between 2010 and 2011 but it decreased slightly in 2012.

Overall, analysis of the graphs reveals that the most stable development of all monitored parameters was achieved by Poland, while similar, but slightly lower values were recorded in the Czech Republic. In 2009 and 2010, Slovakia had a high enterprise birth rate and low enterprise death rate, but in 2011 the situation took a turn for the worse and the country had negative values. The country's enterprise survival rate was positive in 2011 – it scored the highest level in the reporting period. This says a lot about the high survival rate of enterprises established in 2009, which grew to 67.07%. Hungary showed the least favourable development across all of the monitored parameters. The country has a low enterprise birth rate, enterprise survival rate, a high enterprise death rate and the second lowest share of employment in newly established enterprises (after the Czech Republic).

Development of selected indicators of business demography can be observed even in individual EU countries by years. As the next section will be devoted to the classification of EU countries in the last reporting period, the Fig. 6 simplifies it. It is a star plot, in which each side of the quadrilateral illustrates the size of the business demography indicator according to the key glyph (Fig. 7). The shape, size and length of the individual sides can be compared to see which countries are similar to each other.

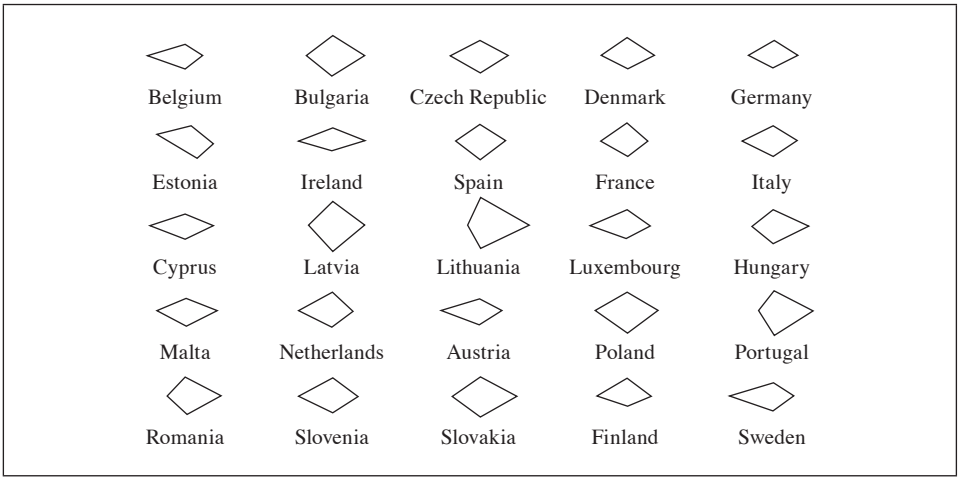

Fig. 6. Star Plots EU Countries in 2012 by Selected Business Demography Indicators
Source: Eurostat and the authors' own elaboration.

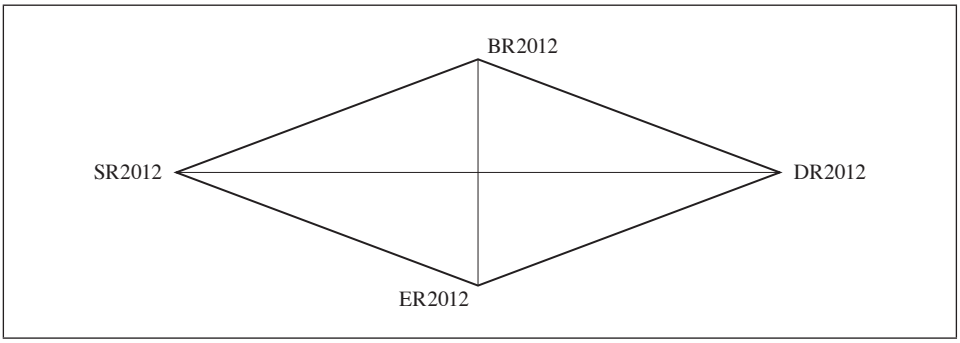

Fig. 7. Key Glyph of Selected Indicators of Business Demography
Source: Eurostat and the authors' own elaboration.

To take one example, the quadrilaterals representing Bulgaria and Slovakia are similar. Such a visual view of the country, however, is not always sufficient, so it will be built upon in the next section.

4. Classification Methodology Using the Algorithm Self-organising Maps

Kohonen self-organising map (SOM) is actually a method intended for the characterisation (mapping) of multidimensional space into a spaceless dimen-

sional, usually a two-dimensional one. Statistical non-linear relationships between multidimensional data are transformed into simple geometric relationships shown in less dimensional space. Specific data (patterns) that are close in the input space cause response to neurons in the grid that are close in the output space. Each neuron has only external inputs – the neurons are not connected to each other in any way. The output neurons compete among themselves as to who will have a vector of weights closest to the input vector. For each input can be calculated for neuron output of similarity (match closeness) of their weights. The neuron which is closest to the input is deemed the winner. The winning neuron can then, together with neighbouring neurons, adjust the values of its weights depending on the parameter learning setting.

The Kohonen algorithm for SOMs (SAS Institute 2011) is very similar to the Kohonen algorithm for adaptive vector quantisation (AVQ). The Kohonen SOM algorithm requires a kernel function $K^s(j, \mathrm{n})$, where $K^s(j, j) = 1$ and $K^s(j, n)$ is usually a non-increasing function of the distance (in any metric) between seeds j and n in the grid space (not the input space). Usually, $K^s(j, n) = 0$ for seeds that are far apart in the grid space. As each training case is processed, all the seeds are updated as:

$$C_n^{s+1} = C_j^s\left(1 - K^s\left(j, n\right)L^s\right) + X_i K^s\left(j, n\right)L^s,$$

where C_j^s is the seed for the j-th cluster on the s^{th} step, X_i is the input vector for the i^{th} training case, and L^s is the learning rate for the s^{th} step[2].

An SOM works by smoothing the seeds in a manner similar to kernel estimation methods, but the smoothing is done in neighbourhoods in the grid space rather than in the input space (Mulier & Cherkassky 1995). The smoothing process takes place in the following steps:

1. Initialise the seeds:

$$n = \arg\min_j \left\| C_j^s - X_i \right\|.$$

2. Repeat the following two steps until convergence or boredom:

a) read the data, assigning each case to the nearest (using Euclidean distance) seed. While you are doing this, keep track of the mean and the number of cases for each cluster,

b) do a non-parametric regression using $K^s(j, n)$ as a kernel function, with the grid points as inputs, the cluster means as target values, and the number of cases in each cluster as a case weight. Replace each seed with the output of the non-parametric regression function evaluated at its grid point.

---

[2] Numerous similar algorithms have been developed in the neural net and machine learning literature; see Hecht-Nielsen (1990) for a brief historical overview and Kosko (1992) for a more technical overview of competitive learning.

Work done in the module SAS®EM is practically based on the formation of nodal graphs whose nodes are formed by individual tasks (Terek, Horníková & Labudová 2010). Node SOM/Kohonen performs uncontrolled learning, which attempts learn the data structure. Maps are produced in the form of grids and individual characteristics can be viewed graphically using a browser that shows the results. Node actually provides an analysis of the results in the form of an interactive map that illustrates the characteristics of the clusters. In the settings panel of the node, the target number of clusters through the target dimensions 2-dimensional Kohonen maps are set (row, column). In general, larger dimension Kohonen maps provide a higher chance that the map will contain "dead" clusters (clusters which were not activated after the learning any input value of a data set).

5. The Results of Segmentation of EU Countries according to Selected Indicators of Business Demography

The main aim of this chapter was to classify, typology or map a certain similarity of EU countries characterised by the means of selective derived indicators of business demography. Since the selected indicators are interconnected, we decided to use the SOM/Kohonen method for the classification. The number of input countries led us to create 6 clusters (2 rows and 3 columns). Fig. 8 shows the result of the clustering focused on the distribution of frequencies via Kohonen map. The colouring of the fields shows that there is no dead cluster, i.e. each cluster contains at least one EU country. Each cell in the network represents one cluster, while the parameters of the dashed cell are shown in the Figure. In addition to the notification of dimensions, there is also information about the number of objects in each cluster.

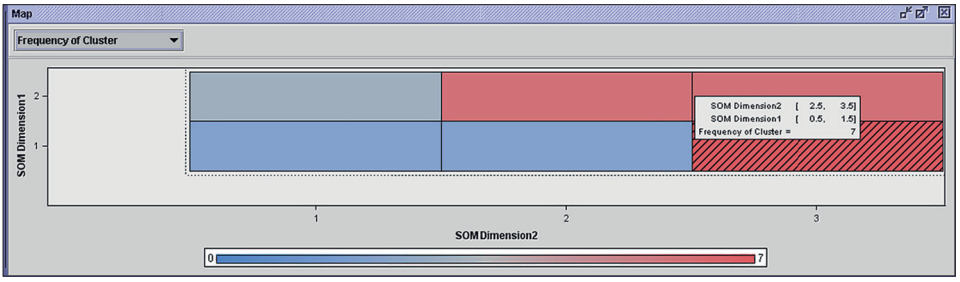

Fig. 8. Kohonen Self-organising Map of Frequency Distribution of EU Countries according to Selected Business Demography Indicators in 2012

Source: Eurostat and the authors' own elaboration.

Another output (Table 1, Fig. 9) identifies the absolute and relative size of selected clusters, which we chose in the classification of EU countries according to selected business demography indicators. The allocation of the countries can be considered as a distribution into well-isolated, mutually non-overlapping, multi-component clusters.

Table 1. The Distribution of Frequencies EU Countries according to Clusters

Segment	Count	Percent
1	2	7.7
2	2	7.7
3	7	26.9
4	3	11.5
5	6	23.1
6	6	23.1

Source: Eurostat and the authors' own elaboration.

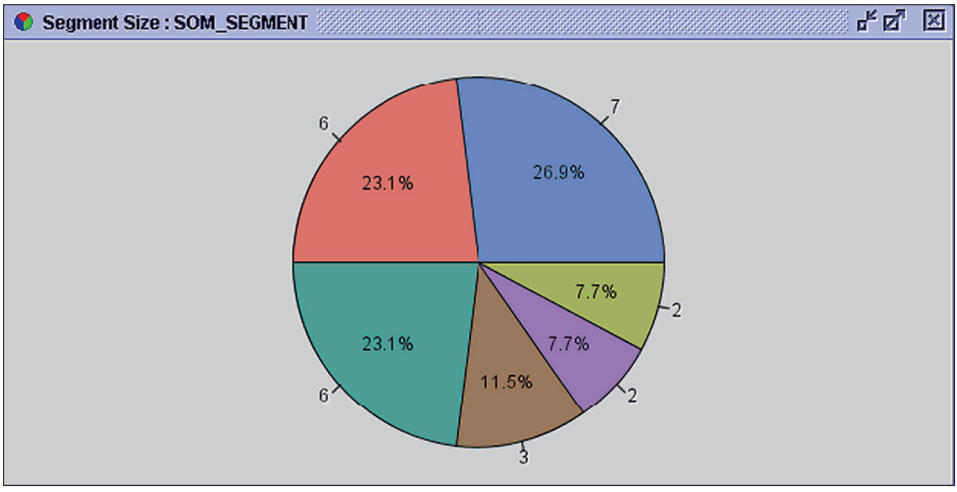

Fig. 9. Pie Chart of the Absolute Size of Individual Clusters
Source: Eurostat and the authors' own elaboration.

A similar Kohonen map can be constructed for the standard deviation of variables forming a cluster, which ilustrates the quality of clustering (Fig. 10). This characteristic reflects the intragroup variability of the cluster. The lower value means lower variability, and thus higher cluster homogeneity. The disadvantage of this characteristic is that it is not arranged in a fixed interval. This character-

istic obtains low values in all clusters, witch the highest value (1.027449) obtains cluster 6 (2nd row, 3rd column – top right) as evidenced by its paler color.

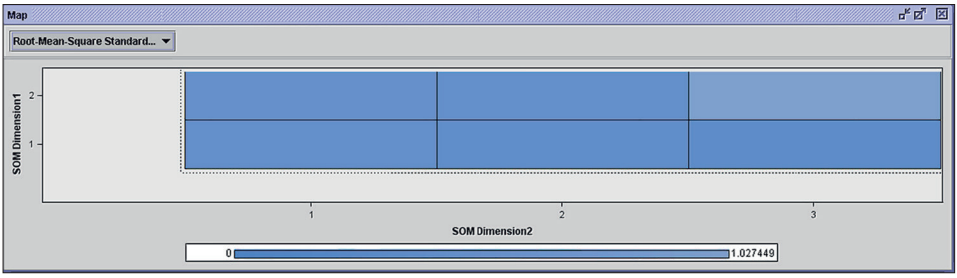

Fig. 10. Kohonen Self-organising Map of Standard Deviations Forming a Cluster
Source: Eurostat and the authors' own elaboration.

Additional maps can be created for a maximum distance from the cluster representative, the distance from the nearest cluster and for the input variables as well. The scale of shading of individual fields is always displayed at the bottom of the map. Fig. 11 shows a graphical evaluation of the indicator *birth rate*. According to this map, we can see that the lowest average value of this ratio is in the third cluster (1st row, 3rd column – bottom right), while the highest value can be found in the first cluster (1st row, 1st column – bottom left).

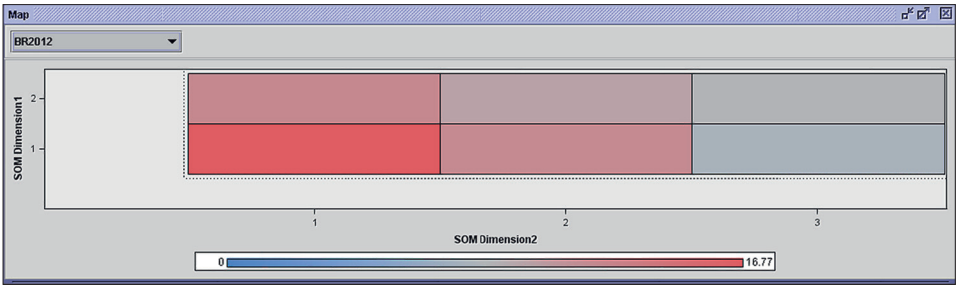

Fig. 11. Kohonen Self-organising Map of Clustering Centroids of the Birth Rate
Source: Eurostat and the authors' own elaboration.

Kohonen self-organising maps can be constructed in similar fashion for other input of variables under different clusters. In respect of the range, the characteristics of the clusters are reported in detail in the tables below (Table 2 and Table 3).

Table 2. Cluster Means of Selected Indicators of Business Demography

Cluster Means				
Cluster	BR2012	DR2011	EB2012	SR2012
1	16.770	21.480	4.365	46.945
2	11.800	14.095	3.595	64.545
3	7.743	6.520	1.970	78.263
4	12.090	13.910	3.770	58.047
5	9.798	8.943	2.025	63.815
6	8.483	10.810	2.410	70.542

Source: Eurostat and the authors' own elaboration.

Table 3. Cluster Standard Deviations of Selected Indicators of Business Demography

Cluster Standard Deviations				
Cluster	BR2012	DR2011	EB2012	SR2012
1	6.081	4.398	0.120	0.445
2	1.725	2.609	0.021	0.346
3	2.435	2.440	0.368	3.874
4	4.255	1.549	1.190	2.790
5	1.509	1.443	0.961	2.186
6	2.654	1.637	0.957	2.280

Source: Eurostat and the authors' own elaboration.

Consequently we proceed with an interpretation of individual clusters according to selected indicators of business demography.

The first group consists of two EU member states: Lithuania and Portugal. This is a dual element cluster, and has the highest average value of the variable of birth rates and death rates of enterprises. The fact that the average death rate is higher than the average birth rates by up to 4.71% can be considered a negative. Both indicators have reached the highest variability as well. The highest average level of employment share of enterprise births with the second lowest variability is another interesting occurrence. However, the average survival rate of enterprises in the cluster is the lowest (46.94%), indicating a negative development of the business environment in this cluster.

The second cluster consists of two relatively new EU countries: Slovakia joined the EU in 2004 while Bulgaria followed in 2007. Both showed the second highest average death rate value, which was 2.3% higher than the average birth rates.

Other business demography indicator values turned in moderately higher mean values with relatively low variability.

The third group includes seven EU countries: the three foundation countries – Belgium, Luxembourg and Netherlands, two countries that joined the EU in 1995 – Austria and Sweden), and two that joined in 2004 – Slovenia and Cyprus. The survival rate of enterprises in this group was evaluated as the most important and entered the highest average value. Other indicator values, on the other hand, reached the lowest average values with moderate to moderately lower variability. This group also shows relatively stable business demography as evidenced by the enterprise birth rate being 1.22% higher than the death rate.

Eastern bloc countries Latvia, Hungary and Romania make up the fourth group. At 58.05%, the average scale of survival rate of enterprises in this group is the second lowest. However, the average birth rate and the proportion of employed people in newly born enterprises earned the second highest value. In this group, however, the death rate slightly exceeded the birth rate (by 1.82%).

The fifth cluster consists of six members: Denmark, Germany, Estonia, Spain, France and Finland. Average business demography indicator values were relatively low, with birth rates exceeding the death rate by 0.85%.

The last group consists of the remaining EU countries: Czech Republic, Ireland, Italy, Malta, Poland and United Kingdom. This group contains countries that entered the EU at different times. The most important indicator for this group, again, is the average enterprise survival rate, which, at 70.54%, was the second highest value for this category. Other average values were moderately lower and had a lower level with relatively low variability.

6. Conclusions

Due to the strong connectivity of the input variables of demographic businesses we used for the classification, we use an alternative methods of non-hierarchical *k*-averages, which is known as Kohonen self-organising maps for the compartmentalisation of EU countries. The one advantage of these maps is that they are less likely to get stuck in bad configurations for highly non-linear data. In this paper, a node SOM/Kohonenen was used for its application, which is part of the statistical tools software system SAS Enterprise Miner (SAS®EM).

Our main aim was to create a typology of EU countries according to selected business demography indicators. Our results did not confirm significant spatial or chronological arrangement of EU countries into clusters. The enterprise survival rate was the most important indicator in the input parameter aggregation we used. The second cluster turned in the highest average enterprise survival value, while

the third cluster showed the lowest average values. Rating enterprise survival rate indicator provides some idea of the stability of the business environment in these clusters. To interpret them, the average values of other indicators of business demography also needed to be accounted for.

That the average enterprise birth rates reached higher values than the average death rates is positive for all EU countries in the years we looked at. However, if we look at the classification of the last current period, not all groups of countries did as well. Clusters 3 and 5, which contain almost all foundation EU countries as well as mostly countries that joined the EU prior to 2000, both showed a decline. In the other clusters the average enterprise death rate exceeded the average birth rate. The first cluster had by far the highest birth and death rates and employment share of enterprise births. However, in the context of the lowest average indication of business survival, the situation must be evaluated in a negative light.

Bibliography

Eurostat (2007), *Eurostat – OECD Manual on Business Demography Statistics*, Office for Official Publications of the European Communities, Luxembourg.

Hecht-Nielsen R. (1990), *Neurocomputing*, Reading, Addison-Wesley, Massachusetts.

Kosko B. (1992), *Neural Networks and Fuzzy Systems*, Englewood Cliffs, Prentice-Hall, New Jersey.

Mulier F., Cherkassky V. (1995), *Self-organization as an Iterative Kernel Smoothing Process*, "Neural Computation", nr 7, http://dx.doi.org/10.1162/neco.1995.7.6.1165.

SAS Institute (2011), *SAS/STAT User's Guide*, SAS Institute, Cary, North Carolina.

Sodomová E., Coss S. (2011), *Business Demography – An Instrument for Assessing Changes in Growth and Employment* (in:) *Contemporary Problems of Transformation Process in the Central and East European Countries*, Lviv Academy of Commerce Publishing House, Lviv.

Terek M., Horníková A., Labudová V. (2010), *Hĺbková analýza údajov*, Iura Edition, Bratislava.

Data source

http://epp.eurostat.ec.europa.eu/portal/page/portal/european_business/data/main_tables.

Zastosowanie sieci Kohonena do klasyfikacji państw Unii Europejskiej na podstawie wybranych wskaźników z zakresu demografii przedsiębiorstw

(Streszczenie)

Demografia przedsiębiorstw zajmuje się opisem populacji przedsiębiorstw aktywnych ekonomicznie, a więc realnie prowadzących działalność, niezależnie od momentu jej formalnego rozpoczęcia, przedsiębiorstw zlikwidowanych, czyli nieaktywnych – bez względu na termin rzeczywistego zakończenia działalności w świetle obowiązujących

regulacji prawnych, oraz przedsiębiorstw, które funkcjonowały przez pewien okres od chwili ich założenia. Rezultatem takiego opisu jest grupa podstawowych wskaźników, charakteryzujących zarówno liczbę przedsiębiorstw nowo powstałych, zlikwidowanych oraz tych, które przetrwały, jak i poziom zatrudnienia w tych przedsiębiorstwach w poszczególnych państwach Unii Europejskiej.

Głównym celem artykułu jest przeprowadzenie klasyfikacji państw UE na podstawie wybranych wskaźników z zakresu demografii przedsiębiorstw i z wykorzystaniem specyficznych modeli sieci neuronowych – samoorganizujących się map (sieci Kohonena) dla danych z ostatniego obserwowanego okresu. Dokonano tego poprzez wyłonienie jednorodnych grup państw, charakteryzowanych za pomocą wskaźników demograficznych odnoszących się do powstania, przetrwania oraz zaprzestania działalności przez przedsiębiorstwa z tych krajów, a także do poziomu zatrudnienia w poszczególnych kategoriach przedsiębiorstw. Ponieważ wskaźniki te są ze sobą powiązane, nie można zastosować w tym wypadku klasycznych metod klasyfikacji. Budowanie sieci Kohonena możliwe jest za pośrednictwem różnego rodzaju oprogramowania. Na potrzeby klasyfikacji prezentowanej w artykule wykorzystano narzędzia statystycznego systemu analitycznego – SAS Enterprise Miner (SAS ®EM).

Dodatkowym celem artykułu jest prezentacja demografii przedsiębiorstw oraz opis wybranych wskaźników dla Unii Europejskiej od 2008 r. do ostatniego raportowanego okresu, ze szczególnym uwzględnieniem rozwoju ekonomicznego Słowacji oraz państw Grupy Wyszehradzkiej (V4).

Słowa kluczowe: sieci Kohonena, sieci samoorganizujące, grupowanie, demografia przedsiębiorstw, klasyfikacja państw Unii Europejskiej.