ISSN 1392-1258. EKONOMIKA 2011 Vol. 90(2)

# RATING OF ENTERPRISES' ACTIVITIES BY THE MODIFIED CLUSTER METHOD

**Tetyana Kravets*, Glib Kuznetsov**
*Kyiv National Taras Shevchenko University, Ukraine*

***Abstract.*** *The problem of improving the methods of enterprise rating estimation is considered in the article. It is proposed to apply an integrated approach to the analysis of the aggregate financial performance of enterprises, based on the modified clustering method which takes into account the possibilities of Kohonen's self-organizing maps. The effectiveness of the largest non-financial sector enterprises of Ukraine over a particular period has been estimated by constructing self-organizing Kohonen maps of the aggregate financial performance and the selection of clusters of related enterprises in characteristic features.*

***Key words:*** *rating of an enterprise, Kohonen's selforganizing maps, cluster analysis*

## Introduction

The determination of the rating positions of the enterprises is one of the options of the enterprises` financial analysis that permits to get a comprehensive estimation of the financial status of the enterprises and to assess its significance for the economy as a whole. The improvement of estimation methods of economic objects' activities is an actual problem for many reasons. Let us note that the development of the rating technology for determining the largest companies is complicated by the problem of weak reporting consolidation within the limits of one informal business group. This problem exists in the European Union and in the USA; however, it is not solved absolutely at the legislation level in Ukraine. The weak reporting consolidation complicates the process of constructing adequate rating models. Statistical information on enterprises' activities should be used correctly and in the greatest possible volumes in order to improve the results of the modelling. The convenient visualisation of the obtained results is of special significance in simplifying the investment decisions. Thus, an important problem is the estimation of the rating models' quality and their comparison.

The purpose of the paper was to develop the technique of rating enterprises' activities. We offer to use a certain rating model and the modified clustering method constructed on the basis of the Kohonen neural networks. A combination of the multidimensional analysis

* *Corresponding author*:
Faculty of Economics, Kyiv National Taras Shevchenko University, 90-a, Vasylkivska St., Kyiv, 03022, Ukraine;
e-mail: tankravets@univ.kiev.ua

of a set of financial indicators and the aggregated indicator with a convenient visualisation of the obtained results was the feature of such approach. The characteristic features of the clusters were defined under the segmentation process. The obtained information was used to analyse the results of the aggregated indicator work. The offered approach has been applied to analyse of the activity results of large Ukrainian enterprises during the period of the greatest fall of production due to the global financial crisis.

The remaining part of the paper is organized as follows. The next section contains the literature review. The third section describes the modified clustering method of Kohonen's self-organizing maps methodology and provides its interpretation in the context of this study. The empirical results are presented in the fourth section. The final section concludes with a brief summary and directions for future research.

## Literature review

Building the rating indicators is traditionally based on the chosen financial performance indicators and the method of averaging some parameter ranks (Zimin, Trishin, 2006), or using a weighted sum of previously normalized parameters (Baranov, Skufyina, 2008; Shapran et al., 2008). The method of constructing the rating performance by the methods of multivariate statistical analysis (principal components method, factor analysis, cluster analysis) is offered in the paper of Kravets (Kravets, 2009).

The usage of statistical methods for data analysis requires an easy visualization of research results, which can be achieved by using the Kohonen neural networks (Kohonen's self-organizing maps – SOM) (Debok, Kohonen, 2001; Kohonen, 2008). Methods of self-organizing maps are widely used in economic research, in particular, for the analysis of consumer portfolio (Holmbom et al., 2008), of the marketing technology (Nikishina, 2003). The cluster approach was used to delimite the regions by the human capital requirements (Bezrukov, Kolosova, 2008) while studying the financial and economic situation of construction enterprises (Kovalenko et al., 2010), etc.

## Methods

The clustering of the database was based on the Kohonen SOM using Viscovery SOMine 5.2 (Viscovery Software GmbH, Austria, 1998–2009). The Viscovery SOMine implements a modified algorithm of SOM-Ward clusterisation, which combines the technique of mapping data by self-organizing maps with the classical hierarchical Ward clustering algorithm (Debok, Kohonen, 2001). SOM operates in two modes as well as in most artificial neural networks – training (without teacher) and mapping (Kohonen, 2008). The process of training contains three basic steps, the initialization is not included: sub-sampling, finding the maximum correspondance and adjustment. As a result of training, the sorting of incoming information in the form of one- or two-dimensional maps has being organized. Each multidimensional vector has its own coordinate on this map, and the closer coordinates of two vectors on a map, the closer they are in the input space. This topographic map gives a visual representation of data structure in the multidimensional input space.

Note that the topographical maps keep proximity relation only locally: close regions on the map are close in the input space, but not vice versa. In general, there is no display to reduce the dimension and maintains the proximity relation globally. The convenient tool for data visualization is a topographical map coloration similarly as in ordinary geographical maps. Each feature datum generates its own coloring of maps' cells by the average value of this feature in the data entered in this cell. If all the features are gathered in a whole map, we obtain a topographical atlas which gives a generalized idea about the structure of multidimensional data.

The Ward clustering method begins with the initialization of each node as a cluster. Two clusters with the minimum distance are merged at every step of the algorithm. The measure (distance) used by the Ward method is based on the dispersive criteria which provide a low variance inside the cluster and a large variance between the clusters. The two clusters whose merger leads to the smallest proportion of variance are combined. The distance between the clusters is defined as follows:

$$d_{rs} = \frac{n_r \cdot n_s}{n_r + n_s} \cdot \left\| \overline{x}_r - \overline{x}_s \right\|^2 ,$$

where $r$ and $s$ define two specific clusters, $n_r$ and $n_s$ are the number of data points in clusters, $\overline{x}_r$ and $\overline{x}_s$ are cluster centroids, $\| \ \|$ is the Euclidean norm. The following formulas are used to determine the centroid and the number of elements of the new cluster:

$$\overline{x}_r^{(new)} = \frac{1}{n_r + n_s} \cdot \left( n_r \cdot \overline{x}_r + n_s \cdot \overline{x}_s \right), \quad n_r^{(new)} = n_r + n_s .$$

The Ward method was modified for SOM-Ward clustering as the measure uses a nodal character of the map and a topological location of clusters. The feature is the initialization of the distances matrix which takes into account the number of data records that match the node map. Let $r$ and $s$ be two nodes for which the distance is calculated, $n_r$, $n_s$ being the number of data records that match the nodes $r$ and $s$ (Ward classical method puts $n_r = 1$), $\overline{x}_r$ and $\overline{x}_s$ – their nodal vectors. Then the distance $d'_{rs}$ is defined as:

$$d'_{rs} = \begin{cases} d_{rs} & \text{if clusters } r \text{ and } s \text{ are adjacent}, \\ \infty & \text{differently;} \end{cases}$$

this means that any two nonadjacent clusters on the map are never united.

## Results

The research was carried out on the base of the financial performance of the largest enterprises of the non-financial sector of Ukraine in the first half of 2009 (Shapran, Duhnenko, 2009). In connection with the absence of some data on the enterprises, the number of elements of the database was reduced to 193.

In the paper of Shapran (Shapran et al., 2008), the authors build a rating system based on the outcome of the integrated indicator: $((1 + 6NI) + (1 + 4S) + (1+2T))\times 1000$, which connects such indicators as the share of net income in the total sample (NI), the share of payroll and deductions for a social activities ($S$), the share of tax payments ($T$) via multipliers. The multipliers are introduced for a greater influence of the enterprise's scale, its social significance and the fiscal component in the integral indicator.

The analysis was based on the following indicators: the integrated indicator, the revenue, the percentage change of the revenue, the net income from product sales, the Value Added Tax (the VAT), taxes paid without the tax credit, costs of labour and social activities, the assets and the equity. The enterprises were ranked by a decrease of the integrated indicator values in the database. The enterprise with the rating number 1 had the best rating position.

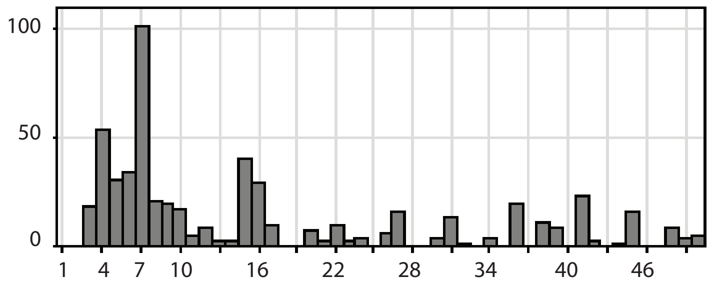


FIG. 1. **The cluster indicator chart**

*Source*: authors' calculations.

The clustering process started with building the Kohonen SOM by means of Viscovery SOMine 5.2. The map size (number of neurons) was chosen at the level of 2000 nodes in order to have a qualitative clustering and a clear data visualization. The clustering of the input data was performed by the SOM-Ward method. The Viscovery SOMine program calculated the cluster indicator for each possible number of clusters (Fig. 1) to be sure that the chosen number of clusters was optimal. The indicator is displayed on a chart in which the number of clusters is indicated on the horizontal axis. The vertical axis shows the value of the indicator for each system of clusters, which can be interpreted as follows: if the indicator value is high for a particular system of clusters, then the clustering can be considered as "natural" for the constructed map. Accordingly, when the indicator is low for a system of clusters, the clustering is "artificial". Thus, the peaks of the cluster indicator on the figure show the true clustering. Figure 1 shows that the indicator reached the highest value when the system of 7 clusters was chosen. Other important technical indicators of the constructed map had a normalized distortion (0.05) and a quantization error (0.00013) (a measure of how well the data vectors of the initial data set had been defined by the selected map node which shows the quality of map training).

Figure 2 presents a clustering map of non-financial sector enterprises of Ukraine. It consists of seven clusters marked from C1 to C7. The numbers on the map show the distribution of the enterprises among the segments. These numbers are the rating numbers of the enterprises. Let us analyze the map of clustering and maps of separate indicators (Fig. 3). Maps of separate indicators are convenient to be analyzed in order to determine the characteristics of clusters. These maps show the distribution of data on the map that matches a certain characteristic. Thus, a certain gradation of colour is used: light colours (closer to white) show a low and dark colours a high rate. Note that the map of enterprise clusters (Fig. 2) contains the darkened areas whose presence is consistent with the maps of separate indicators (Fig. 3). At the top of the map, on the verge of the first and the third clusters, there are companies that have the largest tax payment debts. At the bottom of the second and the fourth clusters, there are businesses with high taxes. The right bottom corner of the map shows the companies with the highest percentage change of the revenue, i.e. these areas indicate the leaders or outsiders according to certain characteristics.

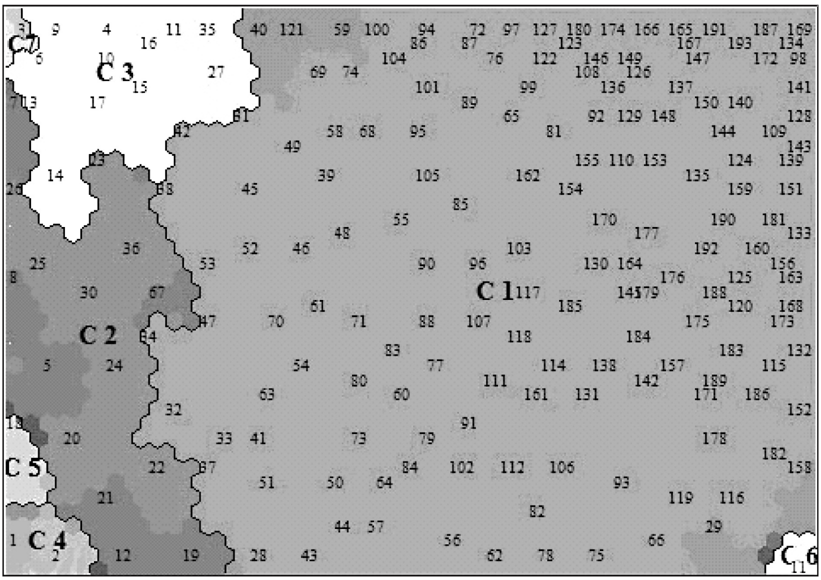


FIG. 2. **A clustering map of non-financial sector enterprises of Ukraine**

*Source*: authors' calculations.

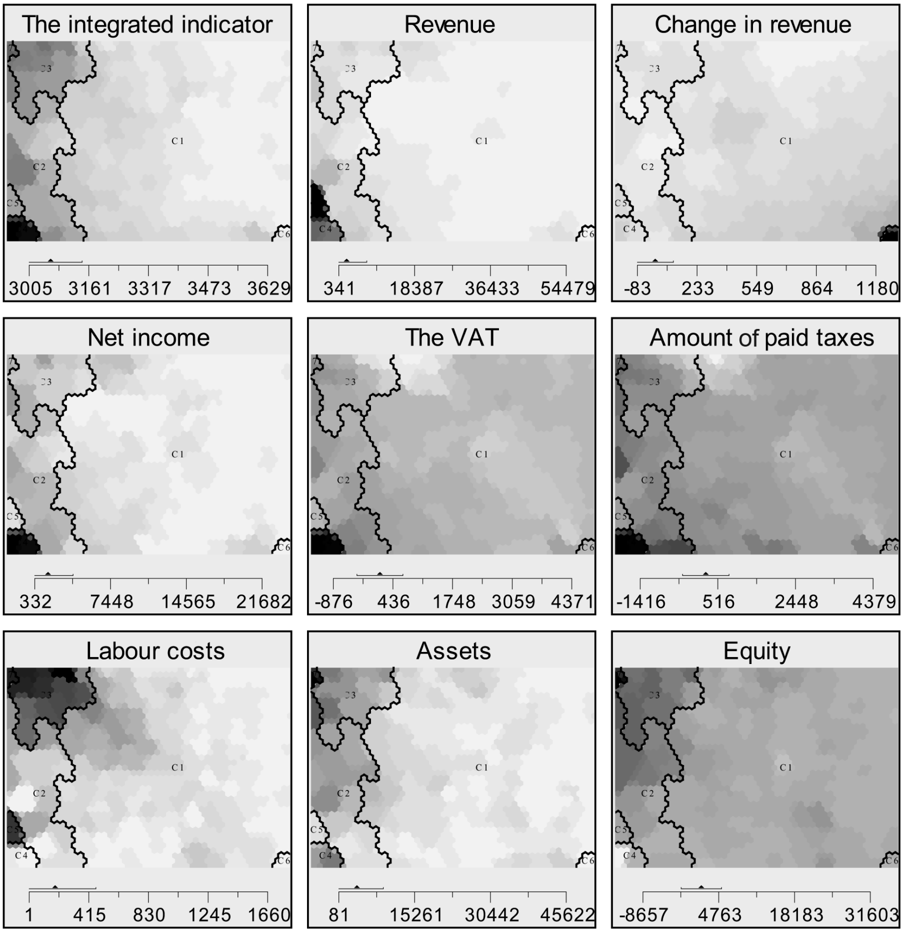


FIG. 3. **The atlas of maps of enterprises' separate indicators**

*Source*: authors' calculations.

More convenient information for interpreting the clusters may be obtained from the charts of aggregated indicators (Fig. 4). The value of these indicators is defined as a difference between the average rate of the indicator of a cluster and the mean value of a sample adjusted by the standard deviation.

Cluster 1 is the biggest cluster without prominent characteristics. All indicators of the group of companies are below the average level. The centroid of the cluster is the CJSC "Ukrainian International Airlines" (the rating number 107). The other companies have the most significant features in the cluster. The JSC "Dniprooblenergo" (59) has the highest revenues and net income in the cluster and at the same time the biggest tax payment debts. The JSC "Khartsyzsk Pipe Plant" (29) has the highest integrated indicator and VAT.

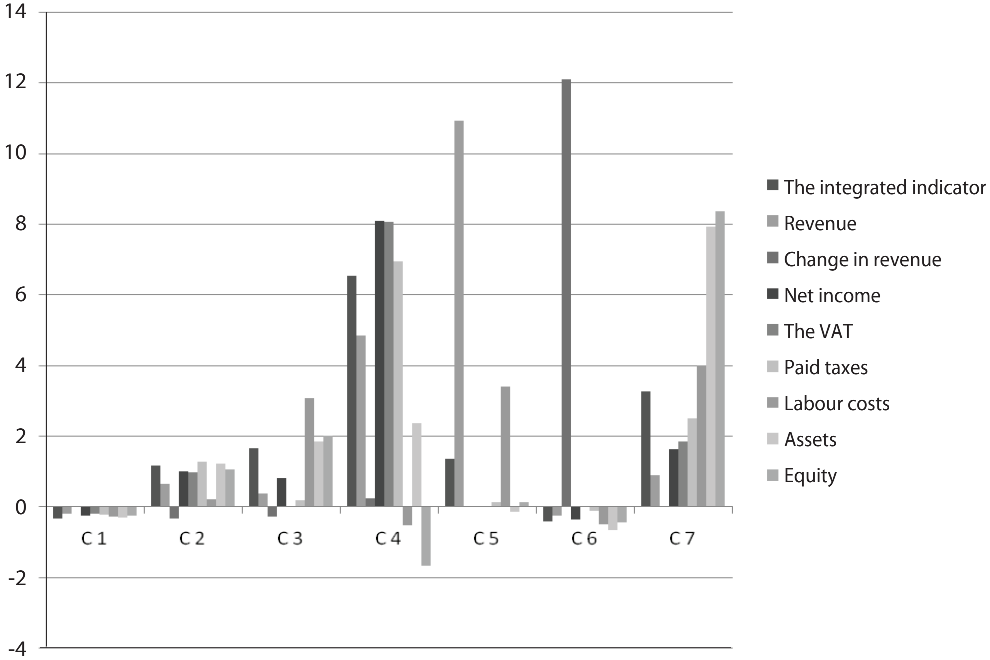


FIG. 4. **The profile of indicators contribution in the clusters forming**

*Source*: authors' calculations.

Cluster 2 shows the companies that have the lowest level of negative revenue changes, the third largest net income, taxes and equity. The centroid of the cluster is the CJSC "Donetskstal-Metallurgical Plant" (24). The JSC "Azovstal Iron and Steel Works" (7) and the Corporation "The Industrial Union of Donbass" (5) are representatives of the cluster with special features. They have the highest level of cluster integrated indicator, income and assets along with a fairly low percentage of the revenue changes.

Cluster 3 has one of the lowest average values of revenue changes, a high equity and assets. All other indicators are average. The centroid of the cluster is the NC "Odessa Railway" (15). The JSC "Arcelor Mittal Kryviy Rih" (4) and the JSC "Ilyich Iron and Steel Works of Mariupol"(14) have a special place in the cluster. These companies have the highest revenues, net income, equity and the lowest negative percentage change of the revenues.

Cluster 4 consists of only two companies: the NC "Gas of Ukraine" NJSC "Naftogaz Ukraine" (1) and the NC "Energorynok" (2) which have the highest average levels of the integrated indicator, net income, paid taxes and the highest percentage change of the revenues. Thus, they have the lowest equity and payroll.

Cluster 5 contains one enterprise – the NC "Ukrposhta" (18). The characteristic feature of this enterprise is the highest revenues per sample (54 479 mill. grn.) and the average large payroll.

Cluster 6 contains the JSC “Pervomaysk Milk Canning Plant” (113). This company has the largest percentage change of revenues in the sample (1180.3%) and rather low all other indicators.

Cluster 7: the NC National Nuclear Energy Generating Company “Energoatom” (3). A characteristic feature of this company is the highest amount of assets (45 622 mill. grn.), equity and payroll.

Several clusters, consisting of one or two companies, were allocated during the clustering. They are definitely special. Therefore, in order to identify the distinctive features of other companies, we will make clustering in the absence of the above mentioned companies (Clusters 4–7). Figure 5 presents a chart of the cluster indicator values of reclustering. The largest value of the indicator corresponds to a system that consists of five clusters. The other characteristics of the map, such as the normalized quantization error and the distortion level, were equaled to 0.117 and 0 (approximately), respectively.

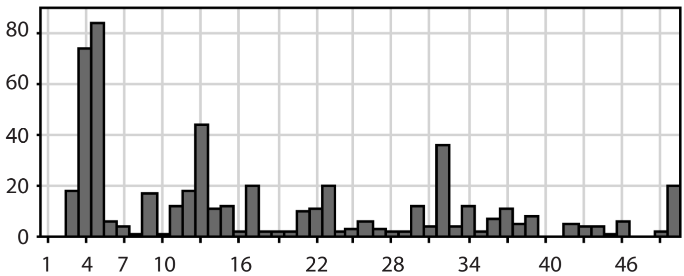


FIG. 5. **A chart of the cluster indicators**

*Source:* authors' calculations.

The results of reclustering are presented on a map (Fig. 6). Let us analyze the map of reclustering, maps of individual indicators (Fig. 7) and a chart of aggregated indicators of enterprises (Fig. 8) for a detailed information on the clusters' characteristics.

Cluster 1: all indicators of the group of companies are at the level that is lower than the average or close to it. It is easy to note that the left part of the cluster on the map of reclustering is formed by large enterprises which have been most affected by the crisis. The cluster includes engineering, chemical, steel companies and automakers. These companies have great percentages of revenue fall, but the overall figures show the average stable situation because of their own reserves and state support (no problem with VAT return). The centroid of the cluster is the JSC “Azot” (70). Typical representatives are the CJSC “Novokramatorsky Engineering Plant” (52), the JSC “Alchevsk Coking Plant” (63), the JSC “Electrometallurgical A.N. Kuzmin Works Dneprospetsstal” (89), the CJSC “Makiyivka Metallurgical Plant” (76), the SC “Ukrainian Motor Corporation” (97).

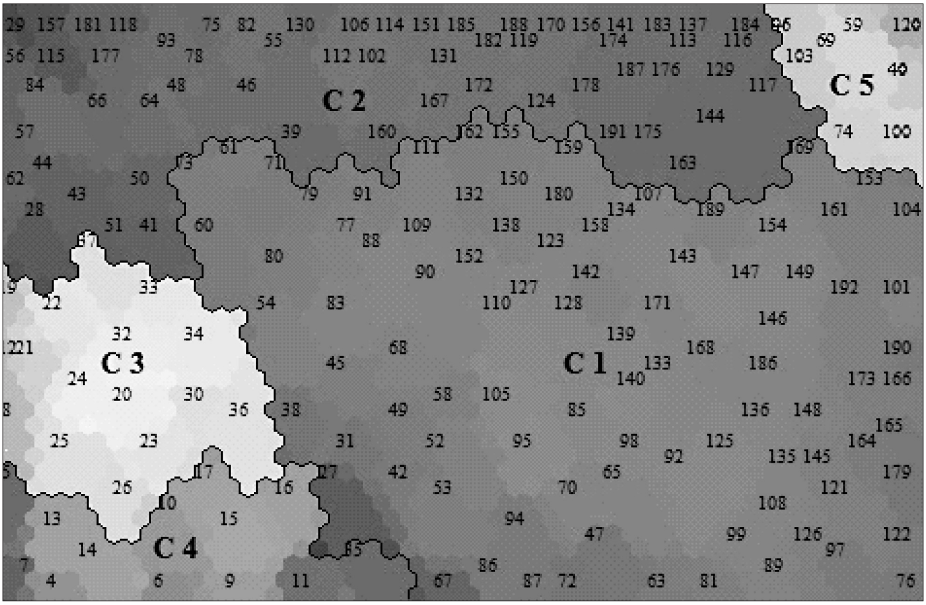


FIG. 6. **The map of reclustering**

*Source*: authors' calculations.

Cluster 2: the characteristic of the enterprises is the highest positive level of the percentage change of the revenue; the other indicators are at a low or a middle levels. The centroid of the cluster is the Firm "Soyuz Victan Limited" (102). Representatives of the cluster are the "Nibulon" Company (66), the CJSC "Cargill AT" (64), the CJSC "V.A.T. – Priluki" Tobacco Company (62), the "Fozzy-Food" Limited (84).

Cluster 3 has the largest volume of paid taxes (including VAT), the second largest revenue, net income and other indicators. The centroid of the cluster is the CJSC "Ukrainian Mobile Communications (MTS-Ukraine)" (25). The representatives are "Kyivstar GSM" (8), CJSC "Donetskstal-Metallurgical Plant" (24).

Cluster 4 includes companies with the highest levels of the integrated indicator, revenue, net income, payroll, assets and equity over the sample. However, this group has negative percentage changes of the revenue. The centroid of the cluster is the NC "Donetsk Railroad" (6). The representatives are the JSC "Arcelor Mittal Kryviy Rih" (4), the NC "Lviv Railway" (16), JSC "Azovstal Iron and Steel Works" (7), the Coropration "Industrial Union of Donbass" (5).

Cluster 5: its characteristic is the negative values of paid taxes and the VAT, indicating the debt of these companies to the state and the debt of the state to the companies through the VAT refund. The centroid of the cluster is the JSC Concern "Galnaftogaz" (100). The representatives of this cluster are the electricity distribution companies JSC "Kyivenerho" (40), JSC "Dniprooblenergo" (59) and CJSC "Ukrtatnafta" (120).

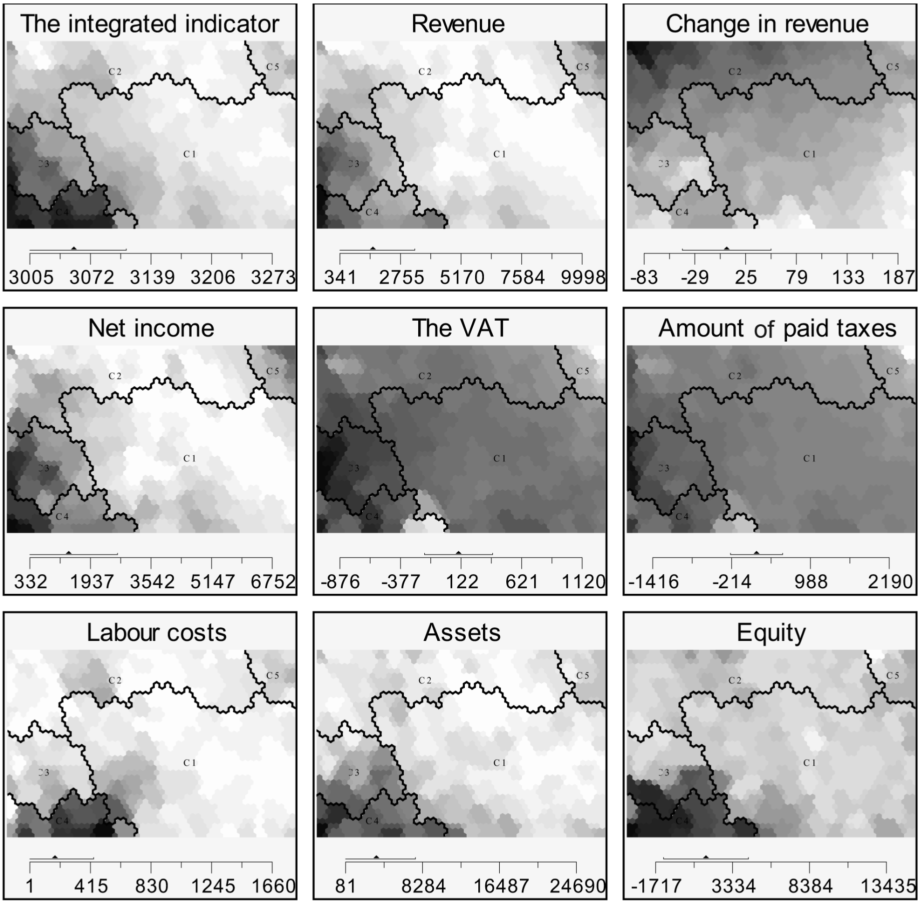


FIG. 7. **The atlas of maps of individual indicators of the enterprises**

*Source*: authors' calculations.

It should be noted that a separate indicator is dominating under the enterprises' segmentation in the multidimensional space of indicators, if the borders of all clusters almost coincide with the borders of the map of this indicator. The rating model is applicable if the integrated indicator is dominating. In the current research, the atlas of maps of separate enterprises` indicators (Fig. 7) shows that there are no definitely dominating indicators. Thus, the pairs of indicators that have almost identical maps are easily defined: revenue and net income; values of paid taxes and VAT; assets and equity. The coincidence of the maps means a strong correlation of the indicators. Only one indicator from each pair should be included in the aggregated indicator.

The integrated indicator used in the current research considers the net income and the paid taxes, but does not consider assets and equity. Besides, the integrated indicator does not consider the percentage change of the revenue. The map of the integrated indicator

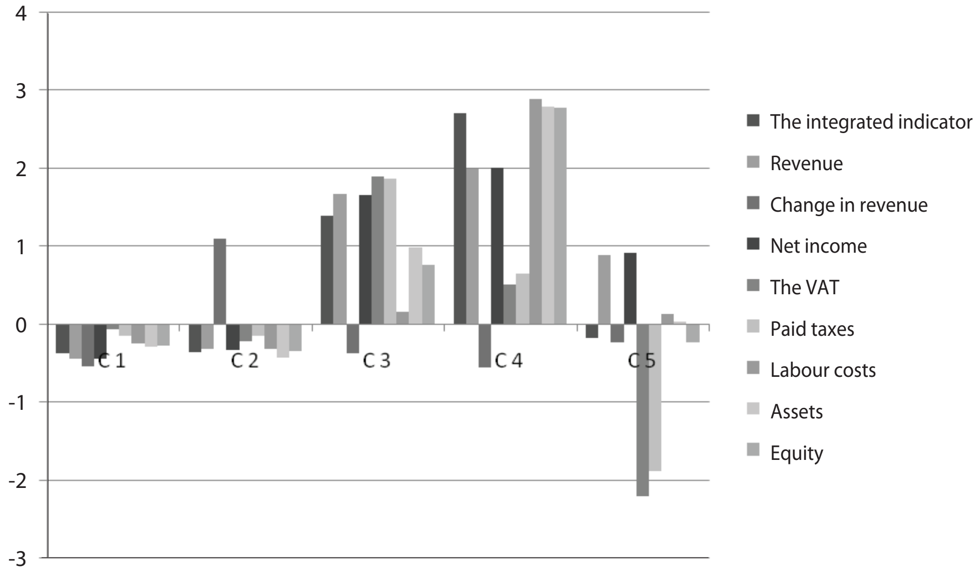


FIG. 8. **The profile of indicators' contribution to the clusters**

*Source*: authors' calculations.

shows that it allocates clusters 3–5 better than it is done by its components. However, the integrated indicator is not a dominating indicator. Its map is deformed in the field of clusters 1 and 2. First of all, it is so because of ignoring changes in revenue percentage. This indicator is especially important during the crisis period. Enterprises with a high positive percentage change of revenue have shown an effective management and stability in the crisis conditions. The rating positions of these enterprises should be reconsidered and improved. The maps of such indicators as assets and equity have blackouts in clusters 3 and 4; this means that they influence on their formation. Thus, these indicators are poorly connected with the formation of the others clusters. Therefore, assets and equity can be not considered while correcting the rating list.

A detailed analysis of the enterprises' reclustering map has revealed some anomalies in the distribution of the companies on the map. The rating of the company does not correspond to the place of the company on the map from the point of view of the ratings of its nearest neighbours in the cluster. For example, the JSC "Odessa Portside Plant" (Cluster 1) has the rating number 101, but its neighbours on the map have significantly worse ratings. Therefore, it is necessary to review the formally calculated rating and make appropriate adjustments or add explanatory comments to the rating list. Similar considerations require to improve the rating positions of the companies "Astarta-Kyiv" Limited (Cluster 1, rating number 180), NC "Milkiland-Ukraine" (Cluster 2, rating number 182) and others.

The presence of anomalies on the enterprises' clustering map is connected with the errors of the rating model. The quantity of such errors characterises the quality of the used model to a certain extent. In the end, it is necessary to note that the integrated indicator employed in this study is acceptable during the period of production growth, and it is conditionally acceptable and demands correction in the period of crisis.

## Conclusions

The rating of enterprises is based on the analysis of a great volume of statistical information. Thus, it is necessary to perform a selection of informative indicators and corresponding multipliers for detecting the aggregated rating indicators. The process of modelling and the verification of the adequacy of the rating model can be facilitated and improved by using the SOM-Ward clusterisation method. The application of this method allows projecting the multidimensional data in a two-dimensional space and gives the possibility to analyse visually the obtained system of clusters.

In this case, the technique of rating estimation consists of the following actions. In the first stage, the segmentation of the whole database by the SOM-Ward method is performed. The atlas of financial indicator maps is analyzed, and the informative and not correlated indicators are selected. An aggregated integrated indicator with certain initial multipliers is constructed based on the chosen indicators. Then the SOM-Ward method is applied to the integrated indicator and a set of selected indicators. Maps of separate indicators are compared with the map of the integrated indicator, and the dependence of the integrated indicator on the contributions of its components is estimated. The multipliers of the integrated indicator are corrected so that this indicator becomes dominating. In the last stage, anomalies on the enterprises' clustering map are analyzed. The corresponding updating of the rating list of the enterprises is carried out, and a conclusion about the quality of the rating model is made.

The further research should include investigations concerning the assessment of the rating models' adequacy and the construction of formal quality criteria for these models by the SOM-Ward method.

### REFERENCES

Baranov, S.V.; Skufyina, T.P. (2008). New techniques and results of research of inter-regional differentiation on the basis of a method of the main componens. Vestnic MGTU, Vol. 11, No. 2, pp. 201–210.

Bezrukov, S.V.; Kolosova, E.V. (2008). Ways of regional clustering on parameters of the human capital on the basis of self-trained neural networks. Upravleniye v sotsialno-ekonomicheskikh sistemakh, No. 1(15), pp. 96–102.

Debok, G.; Kohonen, T. (2001). Analysis of Financial Data using Selforganizing Maps. Moscow: ALPINA.

Holmbom, A.H.; Eklund, T.; Back, B. (2008). Customer portfolio analysis using the SOM. Proceedings of 19th Australasian Conference on Information Systems, Christchurch, pp. 412–422.

Kohonen, T. (2008). Selforganizing maps. Moscow: BINOM.

Kovalenko, A.V.; Urtenov, M.H.; Zaikina, L.N. (2010). The cluster analysis of a financial and economic condition of the enterprises of building branch. KubGAU, No. 60 (06), pp. 1–11.

Kravets, T. (2009). Rating estimations of the enterprises: problems and modeling. Ekonomika, No. 88, pp. 83–89.

Nikishina, Y.V. (2003). Market segmentation using selforganizing maps. Entsiklopediya marketinga. <http://www.marketing.spb.ru>

Shapran, V.; Duhnenko, V.; Kornilyuk, R. (2008). Atlantas on clay feet. Expert. No. 45, pp. 48–59.

Shapran, V.; Duhnenko, V. (2009). Champions of anti-recessionary struggle. Expert, No. 49, pp. 20–27.

Zimin, V.S.; Trishin, V.N. (2006). The analysis of cost of the largest raw companies of Russia on the basis of the journal “Forbes” 2006 rating and their estimated activity. Ekonomicheskiye strategiyi, No. 5–6, pp. 50–61.