Social Economic Debates
Volume 14, Issue 1, 2025
ISSN 2360-1973; ISSN-L 2248-3837

The Measurement of the Company's Performances with Support of the Statistic Techniques for Data Analysis

Mihaela GRUIESCU, Phd
Romanian American University Bucharest
E-mail: mihaela.gruiescu@rau.ro

Cristina Teodora BĂLĂCEANU, PhD
University of Bucharest
E-mail: cristina-teodora.balaceanu@faa.unibuc.ro

Cristina COCULESCU, PhD
Romanian American University Bucharest
E-mail: cristina.coculescu@rau.ro

Oana Mihaela BOIȚĂ, PhD
E-mail: boita.oana@gmail.com

Abstract: In this paper I have developed an analysis methodology for the company's performance, based on micro-data (a sample of 2565 companies in the furniture industry) from a cross sectima data. For each enterprise I have recorded the level of some indicators for the same time unit. The research I have carried out has provided significant information, which has established that the analysis has to be made on types of enterprises defined by their size (waist). The company's size has been identified as the main factor which determines the amount of information contained in the available database.

Keywords: gross profit, gross loss, inventory, labor productivity, the correlation matrix, the factorial correlations, cluster analysis.
JEL Classification: E24, G31, C38.

1. Introduction

The micro-data is recorded data which show the characteristics and individual behavior of firms; many important questions about the behavior of individuals, firms, households, etc.. can have an objective response only if it's based on micro-data.

A first advantage of using micro-data is that we can avoid the distortions induced by the use of aggregated data. Secondly, by using micro-data we can highlight the properties and the stability links for all the activity in the furniture industry, and statistical laws can be formulated..

The data used in the analysis come from a recording of the main indicators of the firms in the furniture industry for 2023. According to the statistical classification of industrial activity, in the furniture industry is included the following activities: chairs production, office and shop furniture production, kitchen furniture production, other types of furniture production, mattresses production.

The primary indicators for which data were collected are: sold production, raw material costs, staff costs, wages costs, gross profit; gross loss, average number of employees, inventory, assets, and turnover.

For the analysis, we used a number of derived indicators, calculated on the primary indicators basis from the initial database.

First we will detect the information we have from the available data, for the sample of the firms in the furniture industry. For this we will use data analysis techniques and the cluster analysis. We need to use these techniques because the sample's size is relatively large. In such cases the statistical analysis should use methods and techniques that are normally used for large databases.

2. The Main Components Analysis

The main components analysis seeks to extract the smallest number of components, that can provide as much information as possible from the total information contained in the original data.

The identified components express new attributes of the population studied, and each one of the identified variables is a linear combination of original variables. As a descriptive method of data analysis, the main components analysis is used only for quantitative variables and large tables.

Indicators and concepts associated with the main components analysis:

- *proper values and proper vectors* are associated with the original variables correlation matrix. A proper value greater than 1, for a component, indicates that, the component has a greater contribution than an original variable, so it is best to be extracted. The proper vectors, together with the proper values are the weights in the calculation of those linear combinations;
- *the main components scores* are the coordinates of individuals in the new axes, and are given by the selected proper vectors. The average of a scores column is zero;
- *the decrease chart* provides information on the proper values, and their decrease rates;
- *the Mahalanobis distance* is used to measure the distance between an individual and the gravity center of the cloud of points;
- *the correlations circle*'s coordinates consist of the correlation coefficients between the original variables and the main components of the retained factors;
- *the Kaiser-Guttman rule* is used for establishing the number of main components: we have as many components as proper values greater than 1. (However the final number of components will be determined according to the concrete interpretation that they receive);
- *the inertia criteria* is used for obtaining the main component and has the advantage of the geometrical approach and is much more complex than the correlation criteria and the dispersion criteria proposed by Hotteling;
- *the overall quality of the first main component*. For this we use the decomposition of the total inertia into the inertia explained by the first axis and the residual inertia of the cloud around the first factorial axis;
- *the main components* are abstract vectorial variables defined as some linear combinations of original variables and have two fundamental properties: are not correlated in pairs of two and the first component is a normalized linear combination which has a mximum variance;
- the loading coefficients are exactly the correlation coefficients between the original variables and the scores. They express the importance of each original variable in explaining each new component.

Table 1: ***Processed data***

ProdV	Sold production	*Stocuri*	Inventory
ChMP	Raw material costs	*ActiveCirc*	Assets
ChPers	Staff costs	*CA*	Turnover
ChSal	Wages costs	*ProdImob*	Fixed Production
ProfitB	Gross profit	*Productiv*	Labor productivity
Pierd	Gross loss	*RataProfit*	Profit rate
NrMedSal	Average number of employees		

The analysis involves the following steps:

1. **Calculating the linear correlation coefficients** (the correlation matrix).

Table 2 presents a correlation matrix. The intense correlations are marked in red (absolute values greater than 0,7). From the table we can extract all the intense correlated variables pairs. Two intense correlated variables, contribute, together, to form a main component (an abstract factor). They contain, mostly, common information, the information migrates in the formation process of the main components to one of the components. The main components consist of pure information, they are absolutely not correlated in pairs of two and form a new orthogonal axes system, based on the informational separation principle.

Table 2: The correlation matrix

	ProdVAc	ChMPAc	ChPersAc	ChSalAc	ProfitBAc	PierdBAc	NrMedSal	StocuriAc	ActiveCircAc	CAAc	ProdImob Ac	ProductivAc	RataProfit
ProdVAc	1.000	0.979	0.770	0.766	0.645	0.159	0.665	0.175	0.275	0.986	0.057	0.177	-0.003
ChMPAc	0.979	1.000	0.745	0.739	0.604	0.160	0.662	0.165	0.251	0.975	0.048	0.170	-0.008
ChPersAc	0.770	0.745	1.000	1.000	0.423	0.257	0.967	0.218	0.268	0.811	0.091	0.066	-0.004
ChSalAc	0.766	0.739	1.000	1.000	0.420	0.252	0.965	0.212	0.263	0.808	0.093	0.065	-0.004
ProfitBAc	0.645	0.604	0.423	0.420	1.000	-0.012	0.349	0.091	0.187	0.648	0.033	0.149	0.343
PierdBAc	0.159	0.160	0.257	0.252	-0.012	1.000	0.252	0.022	0.026	0.154	0.052	0.011	-0.018
NrMedSal	0.665	0.662	0.967	0.965	0.349	0.252	1.000	0.245	0.286	0.721	0.092	0.031	0.015
StocuriAc	0.175	0.165	0.218	0.212	0.091	0.022	0.245	1.000	0.871	0.209	0.056	0.079	-0.018
ActiveCircAc	0.275	0.251	0.268	0.263	0.187	0.026	0.286	0.871	1.000	0.300	0.075	0.202	-0.008
CAAc	0.986	0.975	0.811	0.808	0.648	0.154	0.721	0.209	0.300	1.000	0.057	0.169	-0.004

ProdImobAc	0.057	0.048	0.091	0.093	0.033	0.052	0.092	0.056	0.075	0.057	1.000	-0.007	0.028
ProductivAc	0.177	0.170	0.066	0.065	0.149	0.011	0.031	0.079	0.202	0.169	-0.007	1.000	0.015
RataProfit	-0.003	-0.008	-0.004	-0.004	0.343	-0.018	0.015	-0.018	-0.008	-0.004	0.028	0.015	1.000

The sold production is strongly correlated with the indicators that express the raw material and wages costs and with the average number of employees; so is strongly correlated with quantitative/extensive indicators. Surprisingly, the sold production is not significantly correlated with the labor productivity. Most correlations are positive.

The variables that express costs: raw material costs, staff costs, wages costs have the same type of correlations as the variables included in the analysis and the connection’s intensity level is similar in each case. We must observe the weak correlation of the labor productivity indicators.

The gross profit is positively and strongly correlated with indicators, which in some form, express the volume of the activity (sold production, turnover) or are in a strong positive dependent relationship with the volume of the activity.

The gross loss is not significantly correlated with all the indicators included in the analysis. However we observe a stronger correlation with the indicators related to the workforce (average number of employees, wages costs).

The average number of employees has relatively high intensity links with the indicators that express directly or indirectly the volume of the activity.

The inventory is generally poorly correlated with the indicators included in the analysis. A higher level of the coefficients is recorded for the variables that express the workforce, but the dependence remains very weak.

The assets have the same type of correlation as the inventory, both in terms of links and in its intensity. We observe the stronger dependence between the assets and the profit rate compared to the correlation between the inventory and the profit rate.

The turnover has the same type of correlation as the sold production, with only slight differences in intensity.

The fixed production is not correlated with the indicators included in the analysis.

The labor productivity is not correlated with the indicators included in the analysis, the level of the correlation coefficient is very small in all the cases.

The profit rate is significantly correlated only with the indicators that are underlying its calculation.

At this point of the analysis we can observe, only the variables that introduce redundancy in the data. These are the variables that appear in the table.

2. Calculating the main components’ variance, the distribution of information. The main components’ variance, the amount of pure information reflected by the components is given by the proper values of the correlation matrix (table 3).

Table 3: The proper values

Nr. value	Proper value	Percentage of the covered variance	Cumulative proper values	Cumulative covered variance
1	5.812686	44.71297	5.81269	44.7130
2	1.720338	13.23337	7.53302	57.9463
3	1.342814	10.32934	8.87584	68.2757
4	1.070446	8.23420	9.94628	76.5099
5	0.956763	7.35972	10.90305	83.8696
6	0.924162	7.10894	11.82721	90.9785
7	0.686502	5.28079	12.51371	96.2593
8	0.305945	2.35342	12.81966	98.6127
9	0.116429	0.89561	12.93609	99.5083
10	0.042537	0.32721	12.97862	99.8356
11	0.011548	0.08883	12.99017	99.9244
12	0.009495	0.07304	12.99967	99.9974
13	0.000334	0.00257	13.00000	100.0000

The component 1 covers over 44% of the existing variance in the data; the component 2 covers 13% etc. The first 5 components cover over 83% of the variance. Therefore the dominant type of pure information covers 44% of the global information provided by the data. The purpose of the analysis is to identify in some way the actual meaning of this pure information.

3. Calculating the factorial correlations. The factorial correlations are the correlations between the main components and the variables (Table 4). These correlations shows which variable has provided more information to the main component.

Table 4: The factorial correlations

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Factor 11	Factor 12	Factor 13
ProdV	**0.93**	0.13	-0.14	0.15	-0.04	-0.09	0.20	-0.15	-0.01	0.06	-0.02	0.07	0.00
ChMP	**0.91**	0.14	-0.12	0.16	-0.04	-0.09	0.21	-0.21	0.01	-0.10	-0.05	-0.04	0.00

ChPers	**0.92**	0.10	0.24	-0.07	0.07	0.02	-0.24	0.06	0.00	0.07	-0.02	-0.02	-0.01
ChSal	**0.92**	0.11	0.24	-0.07	0.07	0.02	-0.25	0.06	0.00	0.08	-0.02	-0.02	0.01
ProfitB	0.62	0.12	-0.60	-0.10	0.05	-0.01	0.23	0.41	0.02	-0.01	-0.01	0.00	0.00
PierdB	0.24	0.10	0.45	-0.23	-0.30	0.67	0.37	0.04	0.00	0.00	0.00	0.00	0.00
NrMed Sal	**0.87**	0.05	0.29	-0.13	0.11	0.04	-0.32	0.08	-0.01	-0.14	0.02	0.03	0.00
Stocuri	0.32	**-0.90**	0.05	-0.07	0.15	0.03	0.07	-0.02	0.23	0.00	0.00	0.00	0.00
Active Circ	0.41	**-0.88**	-0.04	-0.02	0.05	0.03	0.06	0.01	-0.24	0.00	0.00	0.00	0.00
CA	**0.95**	0.10	-0.11	0.13	-0.01	-0.09	0.15	-0.11	0.00	0.02	0.09	-0.03	0.00
Prod Imob	0.10	-0.08	0.11	-0.60	-0.59	-0.51	0.04	0.00	0.01	0.00	0.00	0.00	0.00
Pro-ductiv	0.17	-0.20	-0.36	0.41	-0.67	0.26	-0.34	0.02	0.03	0.00	0.00	0.00	0.00
Rata Profit	0.04	0.08	-0.62	-0.62	0.15	0.35	-0.20	-0.21	-0.01	0.01	0.00	0.00	0.00

In Table 4 are marked in red the correlations greater than 0,7 (absolute value) and with a gray background the correlation with the absolute value between 0,5 and 0,69 (absolute value). It may be noted that the first main component is strongly correlated with the variables that reflect, generally, costs, but also the average number of employees and the sold production. Behind all these information seems to be firm's size. For the large companies we record higher values for these variables. The component 1 could be considered as a firm's "size". The component 2 refers to the inventory and the assets. This factor seems to collect information related to inventory. Assets are related to inventory. The remaining 4 components that also have significant correlations with the variables are harder to interpret. Some possible interpretations of the components may be:

Component 3 - profit;

Component 4 - fixed production + its implications on the profit;

Component 5 - productivity + the impact of the fixed production on the productivity;

Component 6 - gross loss.

The factorial correlations chart after the first two main components (axes 1 and 2) is:

Social Economic Debates
Volume 14, Issue 1, 2025
ISSN 2360-1973; ISSN-L 2248-3837

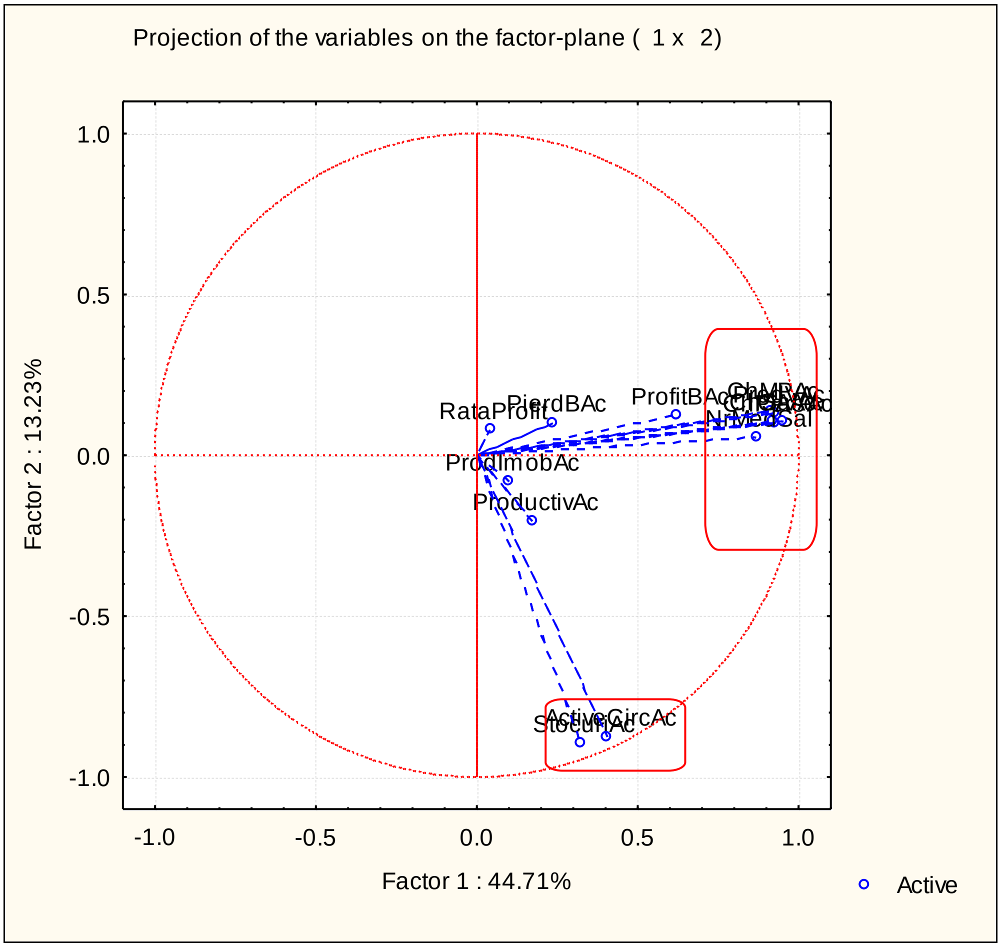

Figure 1. ***The factorial correlations in the first two factors plan***

We can observe the significant groups of variables for the components 1 (right of the chart) and 2 (bottom of the chart).

4. Calculating the main components. The projection of individuals on the new axes. In order to highlight the place of the individuals, when the main components are formed, we need to calculate:

- the projection of individuals on axes;
- the relative contribution of an individual in the making of a component comared with its contribution in the making of the other components;
- the relative contribution of an individual in the making of a component comared with the contributions of the other individuals in the making of the same component.

Because there are too many individuals, the three tables are attached in the Excel file, ACP results. xls. A chart of the individuals is shown in Figure 2. We can observe the firms hat have a particular placement from the Component 1 (the size): 2554, 2558, 2565, 2564, or the firm 1099 with a particular placement from the Component 2. If we study the average number of employees indicator, the 4 companies for factor 1, record values from the upper limit of the indicator's variation range for the studied sample: 1206 employees, 1256 employees, 3350 employees, and 2053 employees.

If we study the firm 1099, which is important for axis 2, we notice that it has 4 employees, if the inventory and the assets levels are 30% higher than the turnover.

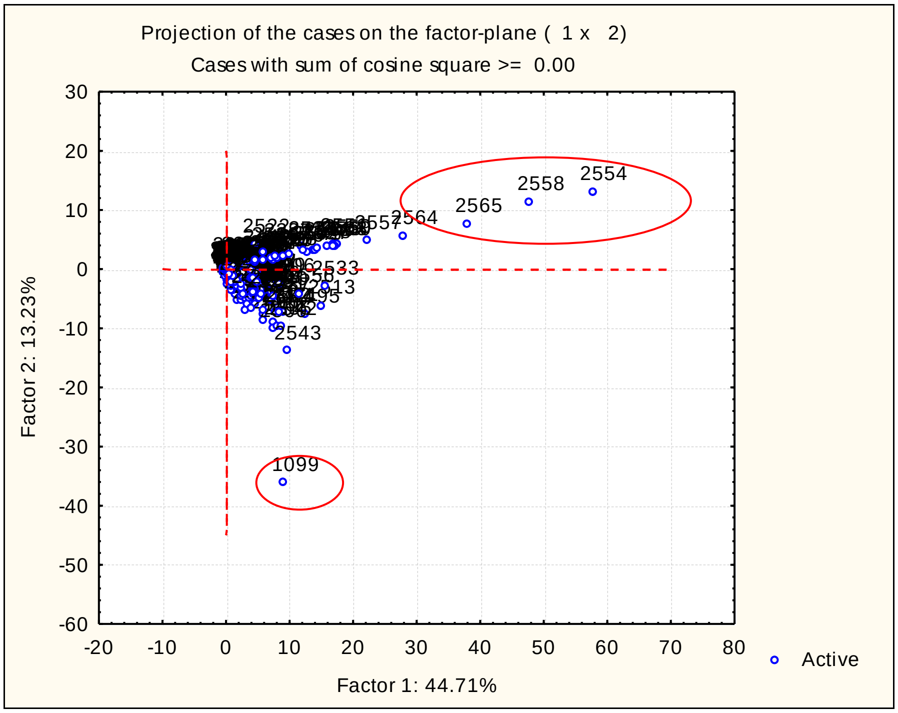

Figure 2. ***The projection of individuals on the plan "size"-"invetory". Factors 1 and 2***

We can observe those charts that are relevant for the analysis. For example, in Figure 3. the individuals are projected on the 'profit' - "productivity" plan, and in Figure 4 on the „profit"–„loss" plan. In order to have a correct interpretation, the data must take into account the correlation between the main component and the variables that define it. For example, in Figure 3, the low values on the profit axis for the firms 2522, 2287, 2558, 2554, show they have a good "profit", because the Component 3 is inversely correlated with the variables gross profit and profit rate.

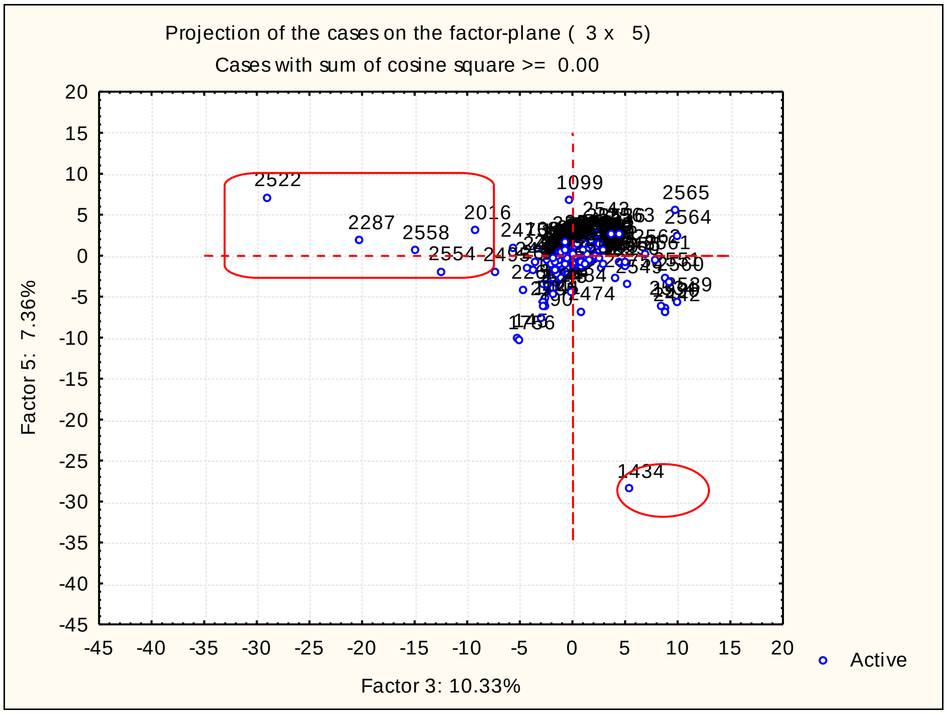

Figure 3. ***The "profit"-"productivity" plan***

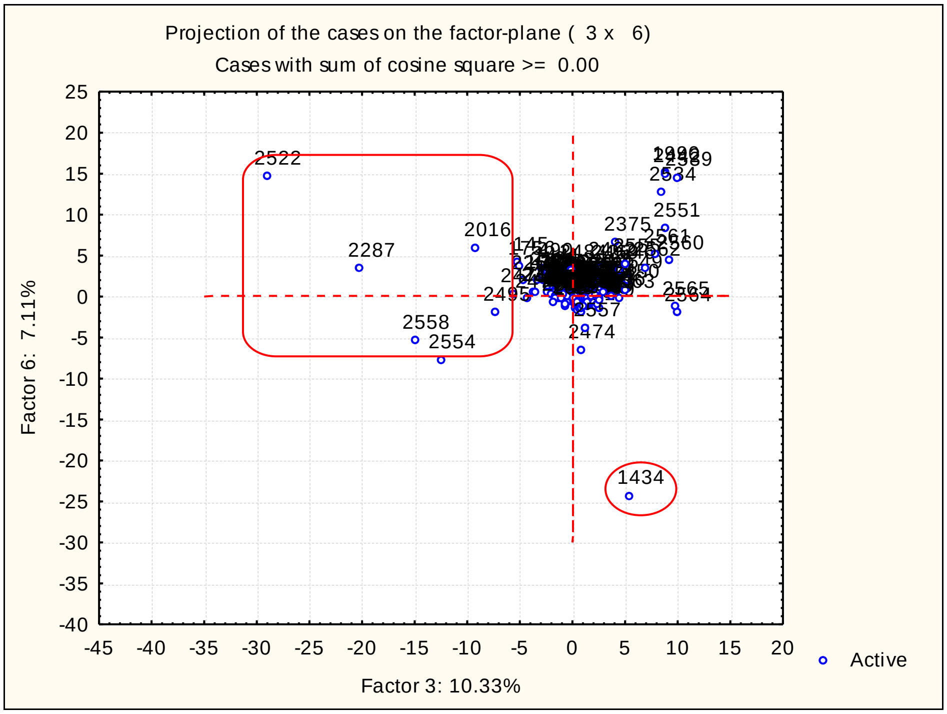

Figure 4. ***The "profit"-"loss" plan***

5. Calculating the individuals' contributions in the making of the new axes. An individual's contribution in the making of an axis is calculated reporting the squared projection of the individual on that axis to the sum of the squared projections of all the individuals on this axis. The calculation highlights the relative importance of the individual for that axis.

6. The quality of the projection of an individual on an axis. We calculate by reporting the square projection of the individual on that axis to the sum of the squared projections on the other axis. The higher this value is the more important the individual is for this axis.

We can observe the high values in the two tables for the individuals with a particular behavior towards the axes, towards they have this behavior (Axis 1: 2554, 2558, 2565, 2564, Axis 2: 1099, Axis 3: 2522, 2287, Axis 5 and 6:1434).

3. The Cluster Analysis

The cluster analysis is used to determine the natural classifications of the data, in order to provide a convenient division of the data into groups. The cluster analysis combines two types of unsupervised classification methods: hierarchical methods and unhierarchical methods. The hierarchical methods offer an easier and a more natural interpretation of the results; that is why they are the most used in practice and implemented by the software products, specialized in data analysis.

The basic concept in the hierarchical classification methods is **the hierarchy**. The hierarchical methods must identify a hierarchy that helps the representation of the points $w_1, w_2, ..., w_n$, which are related to the individuals or to the variables in the analysis. A hierarchy, denoted by H, is an ordered set of sets, denoted by h, consisting of the elements of the set Ω, of the individuals or of the variables, aggregated at a certain level, and with the following properties:

$\Omega \in H$, ie the subset aggregated at the top level contains all the individuals;
for each w_i, $i=1,n$, (n is the number of individuals), there is $\{w_i\} \in H$, that forms base terminal subsets;
for each $h, h' \in H$, there is $h \cap h' \neq \emptyset \Rightarrow h \subset h'$ or $h' \subset h$;

The data that is directly processed by the hierarchical algorithms is represented by the values of the relationships between the individuals and the variables, that are studied in pairs of two. The recording of these values is represented by the distance between individuals matrices or by the correlation between variables matrices. Therefore, the analyzed values are, already, the results of some calculations: the correlation coefficients, the differences between normal or standard values etc..

Generally, the cluster analysis is used, more, to classify the individuals and less to classify the variables. For the study of the variables, we can use the main components analysis and the canonical analysis.

A cluster, in a hierarchy, is a subset *h* (a group of individuals) existing at some point. ***A partition***, in the hierarchy, is a group of clusters, in which individuals are spread, at some point, in their development from the basic subsets to the top level subset. ***The maximum stability partition*** corresponds to a maximum distance range.

The result of the classification, the hierarchy, is represented in the dendogram chart. This chart has a forked structure (a binary tree) and divide graphically the groups from the lowest level, that contain groups consisting of one individual, to the top level, ie the subset consisting of all the individuals. In the dendogram tree, a cluster is a dot and a partition is a level.

The aggregation rule is the rule that unites two clusters at some point and form a top level cluster. This rule determines the hierarchy's content and the dendogram's form. The aggregation rule is chosen in order to correspond to the purpose of the analysis. The mixing rules more often used are:

- the simple connection – unites two clusters with a minimal distance between them. The distance between two clusters is given by the minimum distance between two individuals;
- the complete connection - ditto. The distance between two clusters is given by the maximum distance between two individuals;
- the average connection - ditto. The distance between two clusters is given by the distance between the centers of the two clusters;
- the Ward method – unites those two clusters that cause a minimum inertia variation in the community.

In this study we applied the Ward method and the average connection and for the 2565 firms. The dendogram charts are in figures 5 and 6. In the first dendogram the values on the Ox axis reveal inertia variations from a junction to another, and in the second dendogram are revealed the absolute distances between the group centers. The Ward method is better suited to perform a data informational structure classification. The average connection method builds a hierarchy in which the firms are classified according to their size. Therefore the dendogram tree is so unbalanced. We will further comment the hierarchy constructed through the Ward method. As we can observe from the chart, the maximum stability partition consists of two clusters. Usually the big picture of the dendogram highlights the partitions that have a higher stability. In this study, the dendogram recommends ***the creation of partitions with a small amount of groups*** because they are more stable. The clusters will have a higher number of individuals.

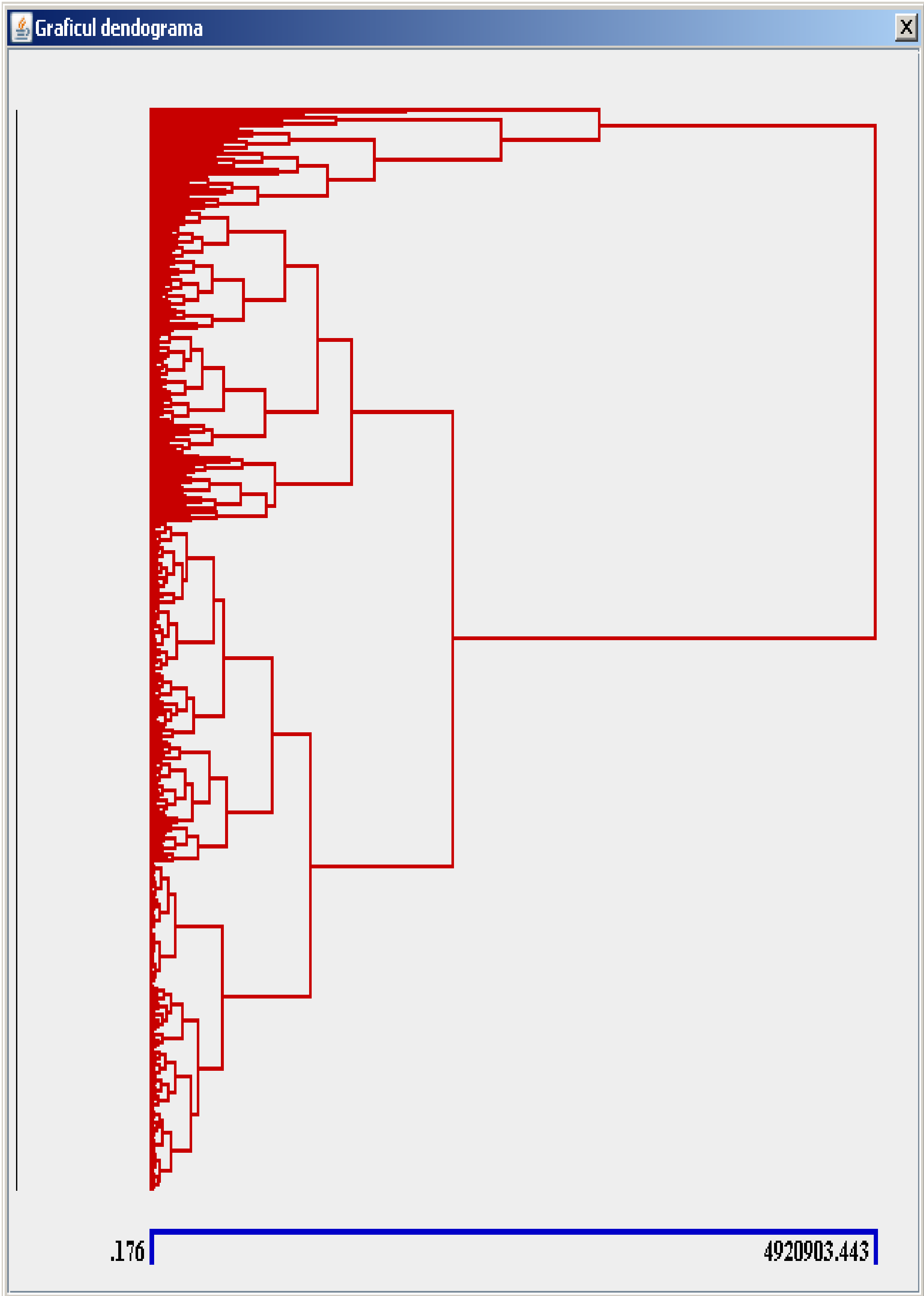

Figure 5. The dendogram chart constructed through the *Ward method*

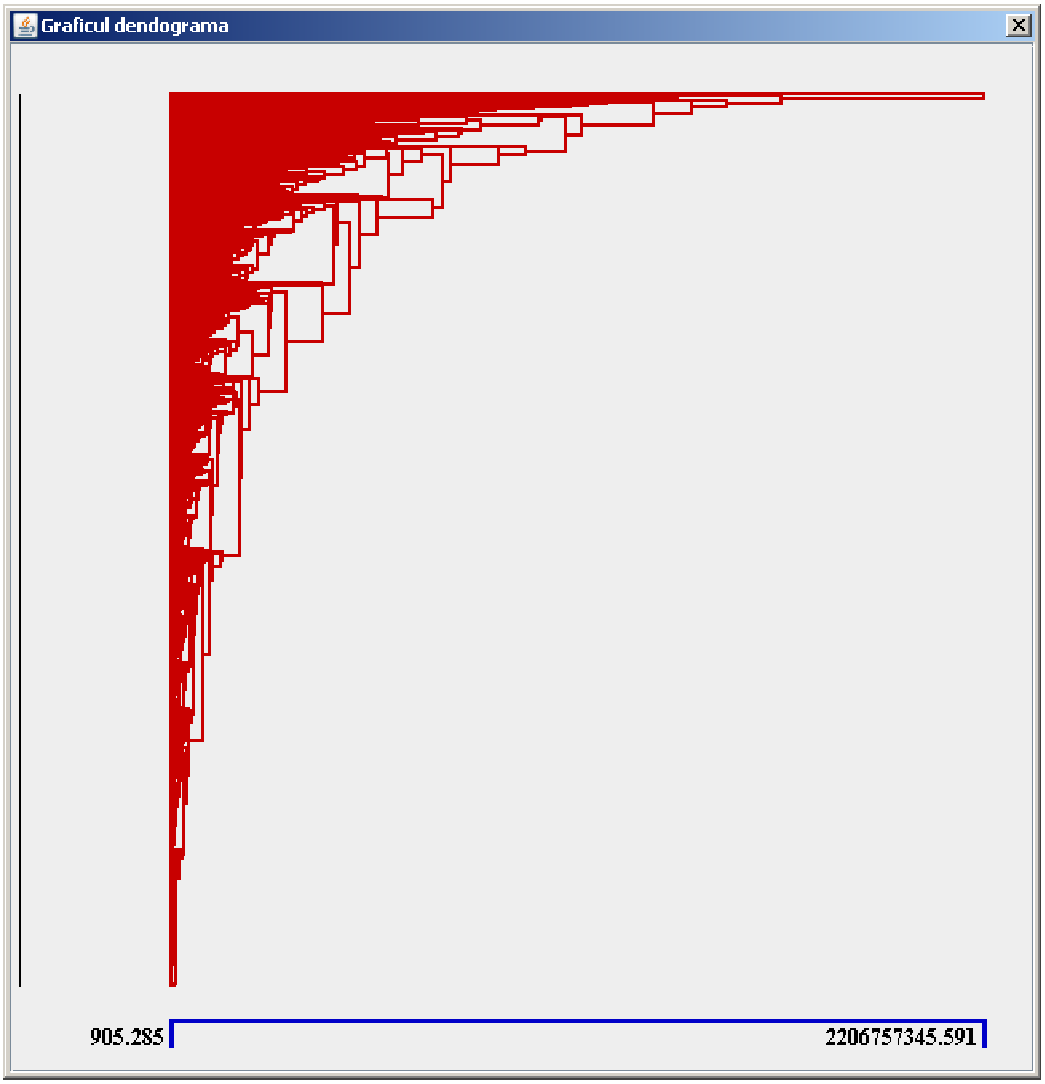

Figure 6. The dendogram chart constructed through the average connection method

If the number of individuals is very large, the dendogram chart must sectioned and analyzed in detail. In Figure 7, we can observe the first section of the 40 sections, in which the dendogram chart has been detailed.

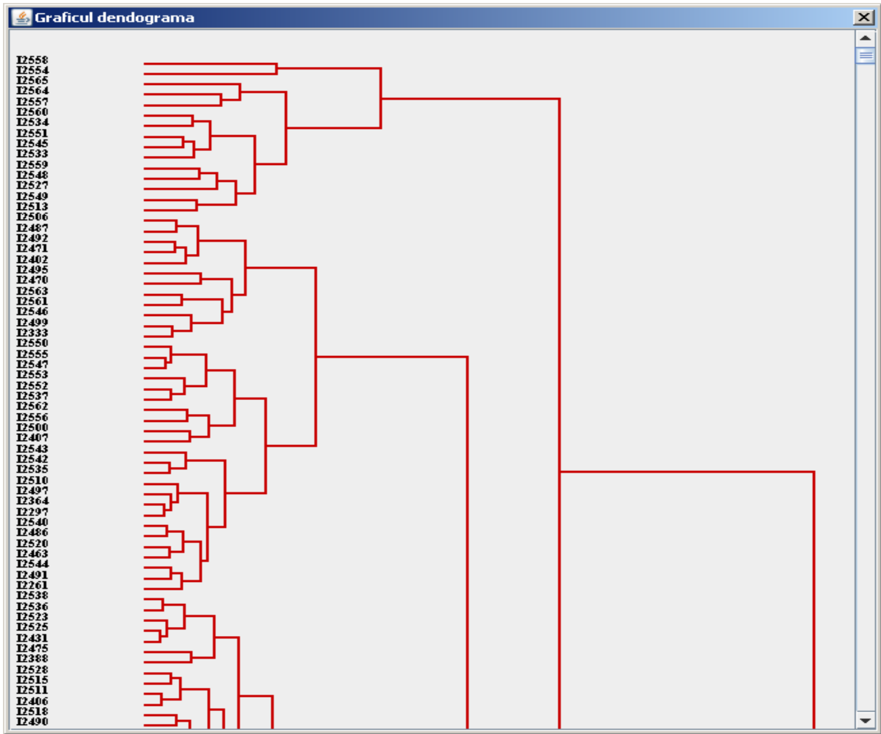

Figure 7. A *section in the dendogram chart*

A partition can be observed on each one of the clusters that form it. We notice that one of these clusters has a smaller number of individuals. Figure 8 shows the dendogram chart for this cluster only.

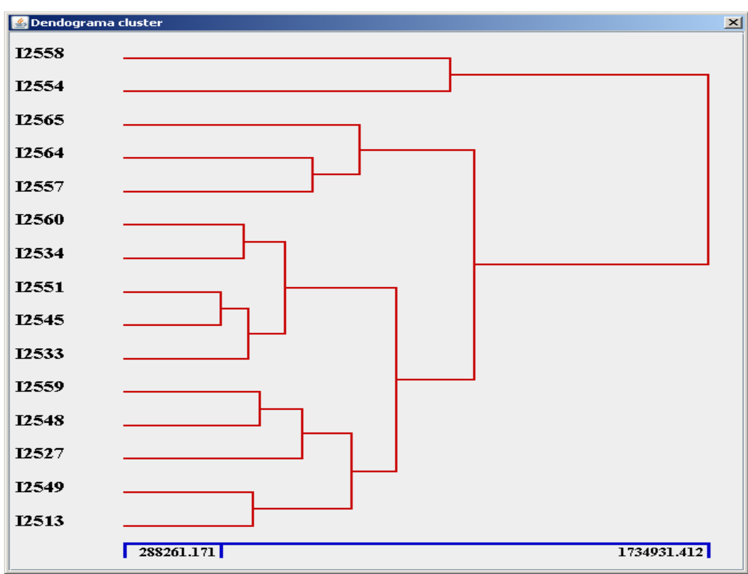

Figure 8. The *cluster 1 from the partition with 3 classes*

Social Economic Debates
Volume 14, Issue 1, 2025
ISSN 2360-1973; ISSN-L 2248-3837

The Cluster in Figure 8 has the composition:

I2558, I2554, I2565, I2564, I2557, I2560, I2534, I2551, I2545, I2533,

I2559, I2548, I2527, I2549, I2513,

Thie cluster groups the companies that have a good representation on the „size" axis from the main components analysis (Figure 2. Main components analysis). In figure 2, in the components analysis, the highlighted firms are bolded.

The first cluster from the hierarchy is formed between firms 216 and 174, at a 0,176 distance and is followed at a small distance by the formation of a large number of clusters from two groups.

CONCLUSIONS

The cluster analysis confirms the conclusions of the main components analysis. The dendogram charts show a hierarchy of the firms after their size. Some companies which have a significant representation on the 2,3,4,5 axis in the main components analysis, have a special behavior, which is highlighted, as well, in the cluster analysis. For example, the firm 1099 with a good contribution to the axis 2, forms the cluster I2522, I2287, I1099, only in the last stages of the hierarchy at a 715223.9784358442 distance. The firm 1434 has the same special behavior, and formes the first cluster at a 371133.1444514692 distance. This company has a good contribution in the formation of axis 5 and 6.

The main components analysis and the cluster analysis have provided important information which led to the fact that the analysis must not be conducted for the entire sample of firms, but on types of firms that are according to their size. A company's size has been identified as the main factor which determines the amount of information contained in the database available. Also, the clusters, have the tendency to form themselves according to the firms' size. Also, the main components analysis shows the importance of a fundamental indicator of the financial performance: the profit (the factor 3) and the productivity (the factor 5).

REFERENCES

1. Abdi, H. and Williams, L.J., 2010. Correspondence analysis. In: N.J. Salkind, ed. *Encyclopedia of Research Design*. Thousand Oaks, CA, USA: Sage.

2. Asheim, B., Cooke, P. and Martin, R. eds., 2006. *Clusters and Regional Development*. London: Routledge..
3. Baltagi, B.H., 2008. *Econometric Analysis of Panel Data* (4th ed.). Wiley.
4. Barca, Ș., 2011. Energy, property, and the industrial revolution narrative. *Ecological Economics*, 70(7).
5. Barthelemy, M., 2016. *The Structure and Dynamics of Cities: Urban Data Analysis and Theoretical Modeling*. Cambridge University Press.
6. Bartlett, M.S., 1950. Tests of significance in factor analysis. *British Journal of Statistical Psychology*, 3(2), pp.77-85.
7. Baum, C.F., 2006. *An Introduction to Modern Econometrics Using Stata*. Stata Press.
8. Danciu, A.R., Niculescu Aron, I.G., Gruiescu, M., 2007. *Statistică şi econometrie*. Bucureşti: Editura Enciclopedică. ISBN: 978-973-45-0562-3.
9. Gelman, A., & Hill, J., 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge; New York: Cambridge University Press.
10. Greene, W.H., 2008. *Econometric analysis* (6th ed.). Upper Saddle River, N.J.: Prentice Hall.
11. Griffiths, A., & Wall, S., 1999. *Applied Economics: An Introductory Course*. Financial Times/ Prentice Hall.
12. **Gruiescu, M**., (2010), *Statistical methods to analyse economic performance*, The Universitaria Publishing House, Craiova;
13. Jaba, E., Pintilescu, C., 2003. Analiza statistică a dezvoltării regionale, folosind analiza factorială a corespondenţelor şi analiza componentelor principale. *Revista Română de Statistică*, Anul III, Nr. 2, Seria nouă, Bucureşti, pp. 1-13.
14. Torres-Reyna, O., n.d. Panel data analysis: Fixed and random effects using Stata (v. 6.0). Retrieved from Princeton University website.
15. Wooldridge, J. M. (2010) *Econometric analysis of cross section and panel data*. 2nd ed. Cambridge, MA: MIT Press.
16. Institutul Național de Statistică. (2022). *Anuarul Statistic al României 2023*. București: INS.