Investment Attractiveness Modeling Using Multidimensional Statistical Analysis

Volodymyr Shinkarenko, Maksym Matskul

Odessa National Economic University, 8, Preobrazhenska Str., Odesa, 65000, Ukraine
shinkar@te.net.ua, maksym.matskul@gmail.com

Dean Linok

Fujitsu Technology Solutions, Textorial Park, 17, Fabryczna Str., 90-344, Łódź, Poland
dean@linok.net

Abstract. The article examines the investment attractiveness of the main branches of the food industry of Ukraine as a latent variable. For the first time in this area, a combination of various methods of multivariate statistical analysis is used for research (cluster analysis and factor analysis – the principal component method). These methods made it possible to use a large number of various indicators of the activities of industries to characterize investment attractiveness. As a result, the set of the branches was divided into three groups-clusters: "leaders" are the most attractive sectors for investment, "middle peasants" are attractive branches for investment, and "outsiders" are the least attractive branches for investment. The generalizing factors (principal components), which influence the resulting factor – investment attractiveness, were found. The interrelation of the generalizing factors and initial indicators is established. As a result of the research, it was possible to make an objective assessment of the investment attractiveness (as a latent indicator) of the main branches of the food industry in Ukraine, using instead of a multitude of indicators only three latent factors.

Keywords: food industry, investment attractiveness, latent variables, cluster analysis, Principal Components Analysis.

1 Introduction

One of the global problems of the world is to provide the population with food. The agro-industrial complex and the food industry as the final link of this complex are engaged in solving this problem at the regional, state and world levels. In recent years, the food industry of Ukraine has come to the fore among the branches of the national economy. It provides the highest rates of industrial growth (with a contribution of more than 31%), more than 10% of the cost of products sold, is one of the leaders among the sectors of Ukraine in filling the state budget. It should be noted that the food industry (as part of the agro-industrial complex) is the export leader and the only sector of the national economy with a positive balance of foreign trade. The Institute of Food Resources of the National Academy of Agrarian Sciences of Ukraine, which is a member of the Ukrainian Research and Training Consortium, deals with the economic problems of the food industry. It is necessary to note the merits of the Institute in the development of national food quality standards, their harmonization with international ones, which allows enterprises to manufacture products at the level of the best world samples. The Institute helps the food industry to master innovative technologies and promote their products in international markets. Many studies have been devoted to the study of the state and trends in the development of the food industry and its industries, including the monographs [1, 2] and the article [3]. The food industry is considered an investment-attractive industry due to the relatively short payback period of the investments and is the leader among the processing industries. In recent years, foreign direct investment in food enterprises has averaged about $ 3 billion per year. To ensure stable growth, the food industry (especially some of its branches) requires constant technical and technological renewal and increased innovation. To solve these problems it is necessary to attract investments. The study of the investment attractiveness of enterprises and branches of the food industry of Ukraine was carried out in [4-7]. Note that among them only in article [8] an attempt was made to apply the method of hierarchical cluster analysis in the study. In recent years, in the study of various economic objects and processes, methods of multivariate statistical analysis have been widely distributed (see, for example, the monograph [9]). And in the work of one of the authors [10] by similar methods (and, additionally, regression on the latent structure) the competitiveness of food enterprises was investigated. So serious research (based on mathematical modeling) of investment attractiveness is unknown to the authors. The purpose of this article is to study the investment attractiveness (as a latent indicator) by the methods of multivariate statistical analysis.

2 Materials and Methods

The data on the performance indicators of the main branches of the food industry for 2017 are taken on the website of the State Statistics Service of Ukraine [11]. Preliminary data processing was carried out in MS Excel spreadsheets. When modeling and computing was used DELL STATISTICA software, version 12.

2.1 Cluster Analysis

Cluster analysis is one of the methods of multivariate statistical analysis. This method allows you to divide a set of objects into groups-clusters according to some latent (obviously unobservable) indicator, the values of which are manifested through a combination of signs-symptoms. The complete procedure consists of three steps:

- Step 1: Tree Clustering (Joining). At this step, the set of objects is ranked using one of the methods. As a measure of the proximity of objects, various metrics of the multidimensional feature space are used.

- Step 2: K-Means Clustering. The method allows to divide all the set of objects into clusters (more than one). The number of clusters is determined by the researcher.
- Step 3: Two-Way Joining Clustering. This step gives us the opportunity to find out which of the attributes have affected the inclusion of objects in the cluster.

Note that the methods of cluster analysis do not allow to identify generalizing factors affecting the latent index under study. Therefore, it is necessary for more comprehensive studies to apply other methods of multivariate statistical analysis.

2.2 Principal Components Analysis (PCA)

The state of most objects (especially economic) is characterized by a very large number of indicators, which are often interrelated (correlated). Therefore, there is a problem of identifying the main factors (Principal Components) that have the most significant impact on the studied result. This problem is solved by one of the methods of factor analysis – the Principal Components Analysis (PCA). This method based on the correlation matrix (matrix of paired correlation coefficients between source variables). The factorization (special representation) of the correlation matrix allows instead of the original feature space of large dimension to consider the space of the Principal Components, the dimension of which is much less than the original one. Since the Principal Components are orthogonal, the problem of multicollinearity is simultaneously solved. Note that in economic research it is necessary to solve an additional problem – the correct (from an economic point of view) interpretation of the Principal Components.

3 Results and Discussion

3.1 Cluster Analysis

The investment attractiveness of 11 main branches of the food industry of Ukraine is investigated as a latent indicator: C1 – the production of meat and meat products; C2 – processing and preservation of fish, crustaceans and mollusks; C3 – processing and preserving fruits and vegetables; C4 – the production of vegetable oils and animal fats; C5 – dairy products; C6 – production of the flour-and-cereals industry, starches and starch products; C7 – production of bread, bakery and flour products; C8 – production of other food products; C9 – production of finished animal feed; C10 – beverage industry; C11 – production of tobacco products. The variable (latent indicator) “investment attractiveness” (as the ability to effectively absorb investments) manifests itself as a result of the effect of explicit variables (indicators-symptoms) x_j (j=1..18): x_1 – volume (billion UAH) of the industry’s annual output; x_2 – volume (million USD) of the industry’s annual export; x_3 – current ratio (= current assets/current liabilities); x_4 – quick ratio (= (current assets-reserves)/current liabilities); x_5 – absolute liquidity ratio (= cash/current liabilities); x_6 – ratio between current receivables and payables (= receivables/current liabilities); x_7 – the ratio of current assets with own funds (= (current assets-current liabilities)/current assets); x_8 – the coefficient of ensuring own working capital stocks (= (current assets-current liabilities) / stocks); x_9 – autonomy or financial independence ratio (= equity/liabilities); x_{10} – working capital ratio (= (current assets-current liabilities)/equity); x_{11} – concentration ratio of borrowed capital (= borrowed capital/liabilities); x_{12} – financial stability ratio (= equity / borrowed funds); x_{13} – financial leverage ratio (= long-term liabilities/equity); x_{14} – financial stability ratio (= (equity + long-term liabilities) / liabilities); x_{15} – return on assets (= net profit/assets) – the amount of net profit per unit of funds invested in assets; x_{16} – return on equity (= net income / equity); x_{17} – operating profitability; x_{18} – profitability of all activities. The source data for multivariate statistical analysis is a matrix (see Table 1).

In this table x_{ij}; i=1..11, j=1..18 are the values of the j-th attribute for the i-th object (branch of the food industry).

Table 1. Indicators of investment attractiveness of the main branches of the food industry of Ukraine for 2017.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
C1	62.92	531.24	0.14	0.12	-6.28	0.09	-6.28	-38.94	0.25	-16.43	5.01	0.05	0.91	0.48	0.02	0.07	6.7	1.7
C2	3.78	26.377	1.17	0.78	0.14	0.69	0.14	0.43	0.12	1.06	0.88	0.13	1.15	0.25	0.02	0.18	2.8	1
C3	13.14	176.5	1.03	0.79	0.03	0.61	0.03	0.1	-0.01	-2.23	1.01	-0.01	-26.09	0.26	-0.06	-5.38	1.2	-4.9
C4	240.16	4605	1.02	0.66	0.02	0.51	0.02	0.05	-0.07	-0.17	1.07	-0.07	-4.36	0.25	-0.11	-1.51	-0.4	-7.1
C5	51.56	494.21	1.08	0.9	0.08	0.82	0.08	0.47	0.11	0.51	0.89	0.13	1.8	0.32	-0.05	-0.49	1.5	-4.1
C6	28.11	181.89	1.91	1.19	0.48	1.06	0.48	1.27	0.4	0.98	0.6	1.53	0.45	0.57	0	0	4.3	0.1
C7	30.7	296.41	0.97	0.7	-0.03	0.6	-0.03	-0.12	0.19	-0.1	0.81	0.23	1.37	0.45	-0.07	-0.37	4.1	-4.5
C8	9.01	1210.5	1.34	1.08	0.26	0.92	0.26	1.29	0.37	0.5	0.63	0.59	0.24	0.46	-0.03	-0.09	7.5	1.4
C9	15.84	15.55	1.08	0.61	0.08	0.57	0.08	0.18	0.09	0.7	0.91	0.1	1.85	0.25	-0.01	-0.11	0.9	-0.6
C10	46.9	209.24	1.08	0.61	0.08	0.57	0.08	0.18	0.09	0.7	0.91	0.1	1.85	0.25	-0.01	-0.11	0.9	-0.6
C11	23.52	355.73	1.89	1.1	0.13	0.03	0.1	0.1	-0.01	-0.05	1.05	-0.05	-3.37	0.25	-0.1	-1.3	-0.1	1.3

Step 1. Note that all variables (signs-symptoms) are stimulators (when more their value, then better), except x_{11} and x_{13}, which are de-stimulators (when more their value, then worse). Before conducting the multivariate statistical analysis, we will make a replacement $x_{11}^{stimulator}=1-x_{11}$, $x_{13}^{stimulator}=1-x_{13}$, which translates all signs into stimulators. For the correct ranking of object-branches, we add 2 more objects to the considered set: the "etalon" C12, for which the values of all signs are maximum, and the "anti-etalon" C13, for which the values of all signs are minimal. Note that ignoring the procedure for creating "etalon" and "anti-etalon" objects often leads researchers to inaccurate conclusions (see, for example, the article [7]). In addition, we will perform data standardization (a mandatory requirement of all multivariate statistical analysis methods) according to the formulas: $z_{ij} = \dfrac{x_{ij} - \overline{x_j}}{\sigma_j}$, $j = \overline{1,18}$, where are $\overline{x_j}$ the mean values, σ_j are the standard deviations for all objects for the j-th attribute. This transformation leads to the fact that all new variables have average values equal to 0 and standard deviations (as well as variances) equal to 1. Thus, the matrix $Z_{13\times18} = \left(z_{ij}\right); i = \overline{1,13}; j = \overline{1,18}$ will be analyzed. At the first step, using the "nearest neighbor" method and choosing the Euclidean distance (distance d_{ps} between p-th and

s-th objects: $d_{ps}=\sqrt{\sum_{j=1}^{18}\left(z_{pj}-z_{sj}\right)^{2}}$) as a measure of the proximity of objects, we get the "Tree Clustering" in the form of a diagram (Fig. 1).

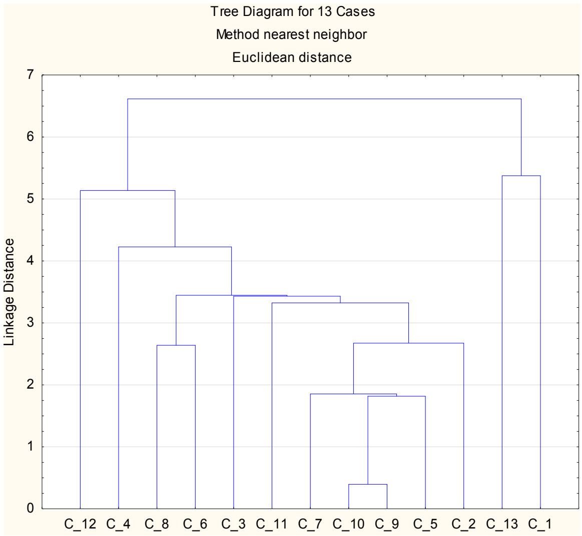

Fig. 1. Diagram of distribution the aggregate of objects (branches).

To determine the number of cluster groups into which we will break our set of industry objects, we will construct a graph of the union in steps (Fig. 2).

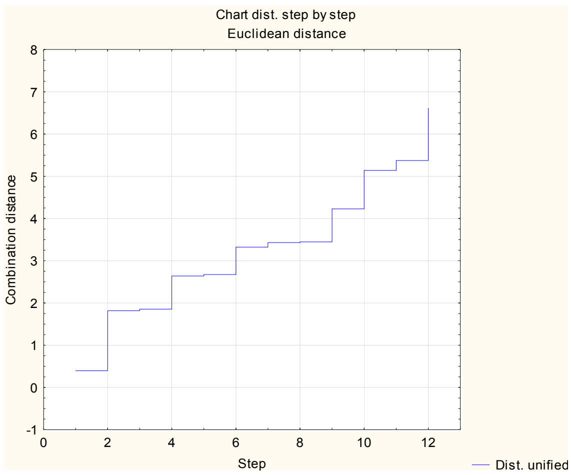

Fig. 2. Diagram of the aggregate of objects step be step.

Analyzing the above graphs, we conclude about the possibility of splitting the set of objects into 3 clusters.

Step 2. Considering the results obtained in the first step, in the second stage, using the K-means method. Set the required number of clusters, equal to three. We get:

Cluster 1 – 8 objects:

Observ.	unified
C_2	0,713673
C_3	0,685849
C_4	0,928692
C_5	0,303296
C_7	0,485469
C_9	0,347048
C_10	0,322583
C_11	0,660865

Cluster 2 – 3 objects:

Observ.	unified
C_6	0,483053
C_8	0,529164
C_12	0,799856

Cluster 3 – 2 objects:

Obser.	unified
C_1	0,633323
C_13	0,633323

Thus, we obtained a stable (robust) partition of the set of objects into 3 clusters (groups): "Leaders" – branches C6, C8, C12; "Middle peasants" – branches C2, C3, C4, C5, C7, C9, C10, C11 ("Best" of which are the branches C3, C4, C11); "Outsiders" – branches C1, C13 (see Table 2).

Table 2. Splitting the set of objects into 3 clusters.

Groups – clusters	Branch – objects
Group 1 – "Leaders"	C6, C8, C12
Group 2 – "Middle peasants"	C2, C3, C4, C5, C7, C9, C10, C11
Group 3 – "Outsiders"	C1, C13

Note that the robustness of clustering is easy to verify using discriminant analysis methods. The same methods determine the ownership of a new object to a particular cluster. This is especially important when investing in the newly built enterprises of the food industry.

Step 3. (Two-Way Joining Clustering). We set the threshold level value in such a way that our set of objects is divided into 3 blocks-clusters. As a result of the third step of the Cluster Analysis procedure, we obtain the reordered matrix of objects-attributes. The graphic image of this matrix is presented in the diagram (Fig. 3), which shows the rearrangement of variables-objects.

This matrix shows which groups of attributes and to what extent influenced the formation of clusters. At the end of this item of research we conclude that almost all branches of the food industry in Ukraine (except for sector C1) are investment attractive.

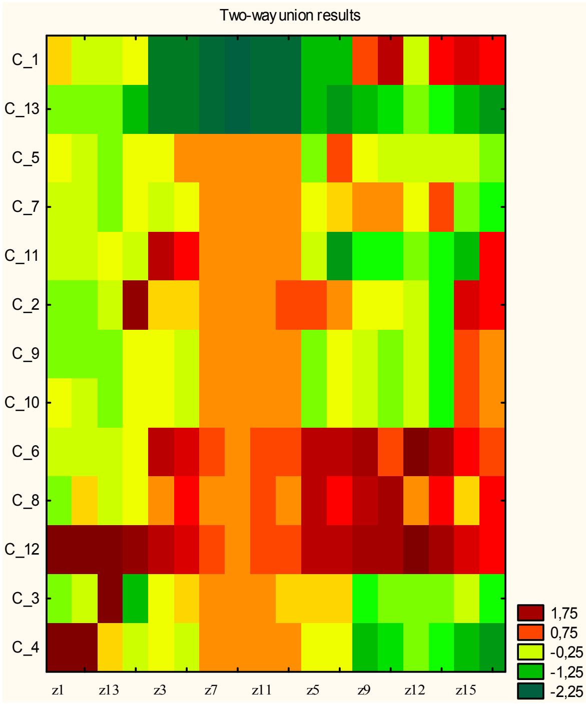

Fig. 3. The graphic image of the reordered matrix of objects-attributes.

3.2 Principal Components Analysis

In this part of the research, a correlation matrix is used, the elements of which are the pair correlation coefficients between all the variables-attributes (Table 3).

Analyzing the eigenvalues of the correlation matrix. For clarity, we use the following so-called “scree chart” (Fig. 4).

This diagram is used to highlight the Principal Components. The Kaiser method is commonly used. According to this method, components are selected that correspond to eigenvalues exceeding 1. We conclude that there are 3 main factors (Principal Components), the action of which causes more than 81% of the total variation (see Table 4).

As a result of applying the PCA, we obtain a factor solution (Table 5).

Next we find the so-called “factor solution”. At this stage, we obtain the decomposition of the Main Components (factors) through the initial variables-simptoms (see Table 6).

Independent (orthogonal) latent factors: the factor F_1 (financial condition) is appreciably loaded under influence the indicators-symptoms z_3–z_{12}, i.e.:

1) $F_1 \approx -0,86z_3 - 0,89z_4 - 0,89z_5 - 0,87z_6 -$
$-0,78z_7 - 0,76z_8 - 0,71z_9 - 0,78z_{10} - 0,79z_{11} - 0,79z_{12}$;

2) the factor F_2 (profitability of production) is appreciably loaded under influence the indicators-symptoms z_{14}–z_{18}, i.e.:

$$F_2 \approx -0,69z_{14} - 0,6z_{15} - 0,3z_{16} - 0,76z_{17} - 0,47z_{18};$$

3) the factor F_3 (production potential) is appreciably loaded under influence the indicators-symptoms z_1, z_2, i.e.:

$$F_3 \approx 0,87z_1 + 0,87z_2 .$$

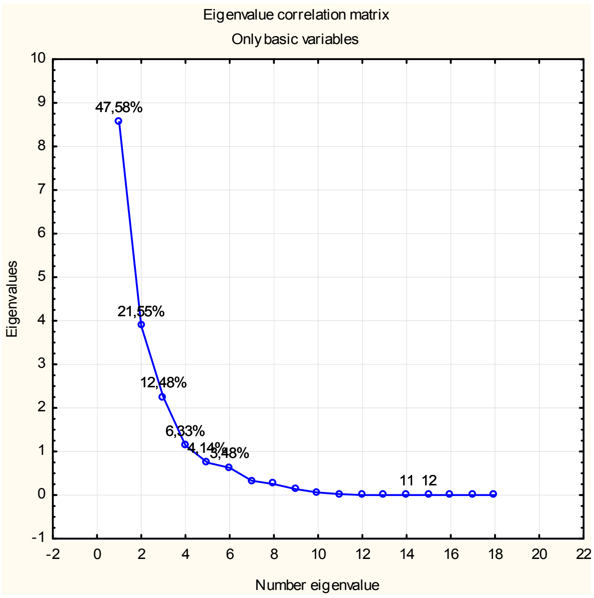

Fig. 4. Diagram of distribution the eigenvalues of the correlation matrix.

Table 3. Correlation matrix.

Correlation (Table.sta)
Main of level p <,05000
N=13

Variable	z1	z2	z3	z4	z5	z6	z7	z8	z9	z10	z11	z12	z13	z14	z15	z16	z17	z18
z1	1,000	0,964	0,215	0,165	0,223	0,232	0,154	0,144	0,098	0,153	0,141	0,310	0,453	0,244	-0,009	0,342	0,150	-0,100
z2	0,964	1,000	0,272	0,261	0,367	0,292	0,209	0,199	0,144	0,201	0,200	0,341	0,481	0,261	-0,057	0,342	0,221	-0,075
z3	0,215	0,272	1,000	0,956	0,743	0,593	0,811	0,792	0,413	0,802	0,806	0,634	0,324	0,306	0,175	0,428	0,174	0,454
z4	0,165	0,261	0,956	1,000	0,814	0,700	0,834	0,819	0,468	0,816	0,837	0,628	0,343	0,356	0,150	0,387	0,267	0,387
z5	0,223	0,367	0,743	0,814	1,000	0,807	0,626	0,606	0,668	0,601	0,644	0,782	0,404	0,572	0,338	0,480	0,556	0,375
z6	0,232	0,292	0,593	0,700	0,807	1,000	0,688	0,674	0,663	0,688	0,715	0,736	0,272	0,514	0,483	0,482	0,477	0,218
z7	0,154	0,209	0,811	0,834	0,626	0,688	1,000	0,999	0,189	0,993	0,998	0,322	0,240	0,013	0,100	0,337	-0,011	0,170
z8	0,144	0,199	0,792	0,819	0,606	0,674	0,999	1,000	0,166	0,992	0,997	0,289	0,229	-0,011	0,082	0,320	-0,028	0,152
z9	0,098	0,144	0,413	0,468	0,668	0,663	0,189	0,166	1,000	0,206	0,230	0,839	0,078	0,915	0,710	0,540	0,912	0,658
z10	0,153	0,201	0,802	0,816	0,601	0,688	0,993	0,992	0,206	1,000	0,993	0,321	0,156	0,010	0,134	0,394	-0,007	0,202
z11	0,141	0,200	0,806	0,837	0,644	0,715	0,998	0,997	0,230	0,993	1,000	0,346	0,227	0,050	0,121	0,348	0,029	0,179
z12	0,310	0,341	0,634	0,628	0,782	0,736	0,322	0,289	0,839	0,321	0,346	1,000	0,318	0,829	0,514	0,492	0,643	0,441
z13	0,453	0,481	0,324	0,343	0,404	0,272	0,240	0,229	0,078	0,156	0,227	0,318	1,000	0,172	0,059	0,113	0,181	-0,008
z14	0,244	0,261	0,306	0,356	0,572	0,514	0,013	-0,011	0,915	0,010	0,050	0,829	0,172	1,000	0,491	0,361	0,870	0,433
z15	-0,009	-0,057	0,175	0,150	0,338	0,483	0,100	0,082	0,710	0,134	0,121	0,514	0,059	0,491	1,000	0,696	0,655	0,745
z16	0,342	0,342	0,428	0,387	0,480	0,482	0,337	0,320	0,540	0,394	0,348	0,492	0,113	0,361	0,696	1,000	0,515	0,627
z17	0,150	0,221	0,174	0,267	0,556	0,477	-0,011	-0,028	0,912	-0,007	0,029	0,643	0,181	0,870	0,655	0,515	1,000	0,611
z18	-0,100	-0,075	0,454	0,387	0,375	0,218	0,170	0,152	0,658	0,202	0,179	0,441	-0,008	0,433	0,745	0,627	0,611	1,000

Independent latent factors by according the significance influence on the level of investment attractive (resulting latent factor F) are put as following order: F_3, F_2 and F_1. For clarity, let us show on the plane of the first two Principal Components how the original features are scattered (grouped) along these components.

Table 4. Factors (Principal Components) and their contribution to the total variation.

Components	PCA -Eigenvalue Eigenvalue	% total var.
1	8,564522	47,58068
2	3,879773	21,55429
3	2,245572	12,47540
4	1,140186	6,33437
5	0,744677	4,13709
6	0,625664	3,47591
7	0,328563	1,82535
8	0,251792	1,39884
9	0,136768	0,75982
10	0,059331	0,32962
11	0,022119	0,12288
12	0,001034	0,00574

Table 5. Results of PCA.

Component	PCA-Results Number components = 3 82,2001% SS R2X	R2X(Cum.)	Eigenvalue	Q2	Limit	Q2(Cumm.)
1	0,462193	0,462193	9,706049	0,158949	0,126984	0,158949
2	0,250791	0,712984	5,266608	0,343864	0,136364	0,448156
3	0,109018	0,822001	2,289370	0,001094	0,147368	0,448759

Table 6. The importance of indicators in regard to the allocated principal components.

Variable	Factor scoures (on correlations) Factor 1	Factor 2	Factor 3
z1	-0,322763	-0,012103	0,871116
z2	-0,391765	0,007772	0,865764
z3	-0,857168	0,292811	-0,060598
z4	-0,887320	0,286191	-0,068549
z5	-0,892380	-0,050333	0,062015
z6	-0,867913	0,002402	-0,010514
z7	-0,777495	0,599929	-0,128712
z8	-0,758020	0,618527	-0,134304
z9	-0,714043	-0,647521	-0,140689
z10	-0,776692	0,578658	-0,164824
z11	-0,793686	0,570032	-0,143279
z12	-0,794195	-0,404238	0,120584
z13	-0,367898	0,060851	0,564043
z14	-0,570762	-0,689629	0,119186
z15	-0,493197	-0,598249	-0,315778
z16	-0,643265	-0,297523	-0,018541
z17	-0,544338	-0,759026	0,006649
z18	-0,517530	-0,471789	-0,419246

4 Conclusion

We developed and mathematically proved a new method for evaluating the investment attractiveness of the main branches of the food industry of Ukraine, which does not contain the subjective estimations and it takes into account many different indicators of activity of branches as possible. A mathematical model is proposed, which is based on a combination of methods of multivariate statistical analysis (Cluster Analysis and Principal Components Analysis). Economic and mathematical modeling allowed us to obtain the following results: the set of the main branches of the food industry of Ukraine divided into clusters-groups according to the latent sign “investment attractiveness” (with ranking of branches); the use of Principal Components Analysis allowed to identify and evaluate the main factors that most significantly affect the investment attractiveness. From the conducted research it follows that when deciding on investing in food industry enterprises, it is necessary (mostly) to assess its financial condition (factor F_1) and profitability of production (factor F_2).

References

1. Gadzal, J.M.: Food industry of Ukraine: strategic aspects of development. IFEAF NASU, Kyiv (2016)
2. Krysanov, D.F.: Integration of agriculture-food sector of Ukraine is in single normative space of European Union. IFEAF NASU, Kyiv (2016)
3. Kaletnyk, H.M., Kovalenko, O.V., Broiaka, A.A.: Modern progress of food industry trends and its place are in the economy of the state. Economics, Finance and Management: pressing questions of science and practice. 8, 7–26 (2017)
4. Goncharuk, A.G., Yatsyk, A.A.: On the factors of investment attractiveness of food enterprises. Ekonomika kharchovoi promyslovosti. 1(13), 34–36 (2012)
5. Koshelok, G.V.: The system of factors that affect the assessment of the investment attractiveness of the food industry. http://dspace.oneu.edu.ua/jspui/bitstream/96.pdf (2014). Accessed 25 Feb 2019
6. Kozak, Y., Matskul, V., Chepurna, O., Cherevko, Y., Chernyshev, V., Kyrylova, L., Podvalna, N., Ropotan, I., Shinkarenko, V., Voropai, N.: Mathematical methods and models for master of economics. http://dspace.oneu.edu.ua/jspui/handle/123456789/6000 (2014). Accessed 24 Feb 2019
7. Shamanska, O. I.: The mechanism of increasing the investment attractiveness of food industry enterprises. Investytsii: praktyka ta dosvid. 22. http://www.investplan.com.ua/pdf/22_2012/8.pdf (2012). Accessed 26 Feb 2019
8. Yashchenko, I.O.: Investment Attractiveness of Food Industries in Ukraine. Scientific Bulletin of the National Academy of Statistics, Accounting and Audit. 1-2, 68–76 (2017)
9. Yankovy’j, O.G.: Latent features in economics. Odesa (2015)
10. Matskul, M.: A multivariate factor analysis: application for evaluating the competitiveness of enterprises. European Journal of Economics and Management. **3**(6), 99–105 (2017)
11. The State Statistics Service of Ukraine. http://www.ukrstat.gov.ua (2019). Accessed 20 Feb 2019
12. Shinkarenko, V., Matskul, M., Linok, D.: Investment attractiveness modeling using multidimensional statistical analysis methods. SHS Web of Conferences. **65**, 04007 (2019). doi:10.1051/shsconf/20196504007