Problems of Digitalization of Collection and Analytics of Enterprises' Financial Information for Determining Industry Average Financial Indicators

SVETLANA PUPENTSOVA[1], MIKHAIL LASKIN[2], MARIA LIVINTSOVA[1]
[1]Institute of Industrial Management, Economics and Trade,
Peter the Great St. Petersburg Polytechnic University,
Polytechnicheskaya str., 29, 195251 St. Petersburg,
RUSSIA

[2]St. Petersburg Institute for Informatics and Automation,
Russian Academy of Sciences,
14th Line, 39, 199178 Saint Petersburg,
RUSSIA

Abstract: - The study aims to develop theoretical and methodological provisions for the digitalization of the collection and analysis of industry financial indicators. The relevance of the chosen topic is confirmed by the emergence of open access to the financial statements of industrial enterprises during the period of digital transformation. The process of collecting and analyzing industry financial information causes certain problems since there are no methods for determining industry average financial indicators. The study used a system and comparative analysis, methods of economic and financial analysis of the industrial enterprises` financial information, and methods of statistical assessment of the main parameters of the sample with a lognormal distribution. The main attention is paid to the application of multivariate statistical analysis, the use of the model of lognormal distributions, recommended for distributions with pronounced right-sided asymmetry. During the study, the authors proposed the stages of analysis of the determination of sectoral financial indicators and tested the hypothesis about the lognormal distribution of the marginal distributions of the main financial indicators of the sample. The paper implements a test of the hypothesis about the jointly normal distribution of the multifactorial vector of the financial indicators of the sample. Distributions with a pronounced right-sided asymmetry are constructed and an algorithm for testing the hypothesis of the joint normality of a nine-dimensional vector is presented. The median and modal values of the industrial average financial indicators are obtained. In the paper, it is demonstrated that the orientation towards the generally established recommended values of the coverage indicators, financial leverage, immobilization, profitability, and turnover seems to be incorrect due to the specific features of the enterprises' activities. It is recommended to calculate the most probable industry average financial ratios by groups depending on the size of the enterprise (revenue or capitalization).

Key-Words: - Digitalization of collection and analytics financial data, financial analysis, industry average financial ratios, financial stability of the organization, industry profitability, modal values, logarithmically normal distribution law, gas and oil industry.

Received: March 24, 2024. Revised: August 20, 2024. Accepted: September 23, 2024. Published: October 24 , 2024.

1 Introduction

In the period of overall digital transformation, the analysis of financial and economic activities is necessary for participants in the digitalization process for control and management decision-making. Sectoral financial indicators are used by analysts and appraisers in the comparative analysis of the financial ratios of the analyzed enterprise with the average industry ones, in the selection of analog objects in the business assessment, in identifying the strengths and weaknesses of the economic entity, in building the enterprise development strategy. The relevance of the chosen topic is confirmed by the emergence of open access to the financial statements of industrial enterprises during the period of digital transformation. An important advantage of digital technologies is financial data processing capability, [1], [2], [3]. Nevertheless, the process of digital collection and analysis of industry financial information causes certain problems, since there are no methods for determining industry average financial indicators based on the data obtained.

Theoretical and practical aspects are reviewed in the following directions.

1. The normative values of the financial stability ratios are considered in the works, [4], [5], [6], [7].

2. It should be noted that when analyzing financial statements, various researchers mainly consider not individual companies, but groups of companies in the same industry, selected according to some characteristic, [8]. Sectoral financial ratios were investigated [9], including research for the energy industry [10], on the example of the oil and gas sector [11], on the example of small and medium-sized businesses of the regions [12], on the example of assessing the potential of the Yamal region of Russia [13], on the example of assessing the level of industrial development based on fuzzy-multiple methods [14], [15], on the example of the analysis of the financial condition of enterprises by type of economic activity [16], [17], [18], [19], on the example of enterprises in Nigeria and Georgia, industry benchmarking reference book for US companies, [20].

3. Integral indicators of the financial condition of the enterprise by industries are considered in [21] (enterprises are divided into groups of growth, maturity, and recession according to the financial condition), [12] on the example of evaluation of the financial state of the region's metallurgical enterprises. In [22] propose to use industry ratios (multipliers) in a comparative approach. Authors evaluate in their research [23] the financial state of steel companies in Egypt using a multi-criteria decision-making model.

We consider it necessary to note that the above studies have an applied and fragmentary nature and do not disclose the general algorithm for assessing the sectoral financial indicators of enterprises.

It should also be noted that the discrepancy in the reported financial data exists due to the differences in accounting methods and the reporting standard (according to either the Russian accounting system, the system of International Financial Reporting Standards, or US GAAP). CBR Regulation No. 714-P dated March 27, 2020, "On Disclosure of Information by Securities Issuers" came into force in October 2021 for securities issues in Russia.

The aim of the study is to develop theoretical and methodological provisions for the digitalization of the collection and analysis of industry financial indicators.

The research objectives are:

– describe the stages of the analysis for determining industry average indicators;

– describe the process and test the hypothesis about the lognormal distribution of the marginal distributions of the main financial indicators of the sample;

– describe the process and test the hypothesis about the jointly normal distribution of the multifactorial random vector of the sample's financial indicators;

– obtain the most probable values of indicators of financial analysis for the oil and gas industry.

The theoretical significance of the work is the algorithm for calculating industry indicators of financial analysis proposed by the authors, which differs in obtaining modal values based on the analysis of the multidimensional lognormal distribution of the vector of the sample's financial indicators.

Practical significance. The methodological recommendations given by the authors are of a practical nature and are applicable to assessing the main financial indicators in dynamics for all sectors of the national economy.

2 Materials and Methods

2.1 Research Methodology

The study was carried out on the example of enterprises in the oil and gas industry based on the methods of systemic and comparative analysis, statistical methods, and methods of economic and financial analysis.

Preliminary selection of digital systems for disclosing the financial information of enterprises by industry is carried out according to such factors as the completeness of data, the possibility of free access, etc.

The object of the research is the data of financial (accounting) statements (balance sheet and report of financial results) of enterprises in the oil and gas industry, which are made publicly available at [24].

The subject of the research is the sectoral indicators of financial analysis.

The analysis of the selected data of the financial statements of the enterprises of the industry is carried out according to the algorithm proposed for research purposes. The main stages of the analysis of the definition of industry financial indicators are shown in Figure 1.

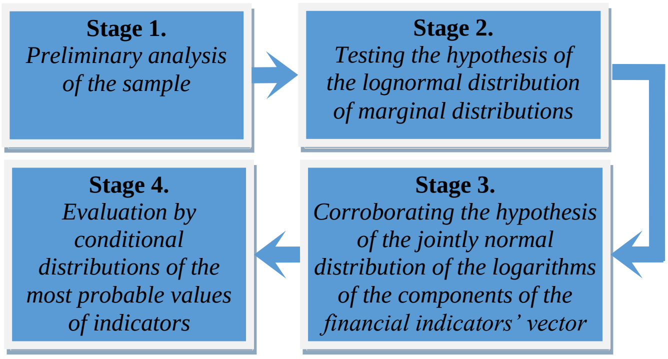

Fig. 1: The main stages of determining industry indicators of financial analysis.

2.2 Knowledge of the Problem

The algorithm for analyzing the financial and economic activities of the organization, given in [25], is reduced to the calculation of financial ratios for the analyzed company and their subsequent comparison with the values recommended in [25]. At the same time, many authors [6], [7], [18], [26], [27], [28] note that sectoral financial ratios can differ significantly from the recommended ones, therefore, in a comparative analysis, it is better to use sectoral indicators.

In the public domain on the TestFirm portal [29], created by the auditing firm "Avdeev and Co" based on data from 2.3 million enterprises of the Russian Federation, key financial indicators for 96 types of activities are given. Sectoral financial ratios are derived for the main areas of financial analysis: financial stability, solvency, profitability, and turnover. The authors of the portal recommend using median values, but it is possible to derive arithmetic mean values in dynamics since 2012. The information on the TestFirm portal [29] is up-to-date, free, and convenient for comparative analysis. But, since the initial data and the algorithm for deriving the key financial indicators of enterprises by type of activity are not presented, the confidence in the derived industry indicators is sharply reduced.

3 Results

Further, the stages of performing the analysis of financial data are considered in detail.

Stage 1. *Preliminary analysis of the sample*

For the analysis, only break-even companies with positive net profit and net total capital were selected for analysis from the one sector of the national economy. The sample size was 185 companies. For all sample enterprises, the most popular financial ratios for business valuation were calculated using the data of accounting statements (balance sheet and income statement) for 2016 from the website, [6], [24]:

– Current ratio (Current assets / Current liabilities);
– Debt-to-equity ratio, or Financial leverage (Total debt /Common equity);
– Immobilization ratio (Current assets / Fixed assets);
– Return on equity (Net income / Common equity);
– Receivables turnover (365 × Accounts receivables / Sales revenue), days;
– Accounts payable turnover (365 × Accounts payable / Cost of sales), days;
– Inventory turnover (365 × Inventory / Cost of sales), days;
– Manoeuvrability coefficient (Own working capital / Common equity).

The distribution of enterprises by revenue was constructed (Figure 2).

The obtained distributions for the main indicators have a pronounced right-sided asymmetry. Testing the hypothesis about the lognormal distribution of the marginal distributions contributes to the further estimation of the parameters by the logarithms of the indicators.

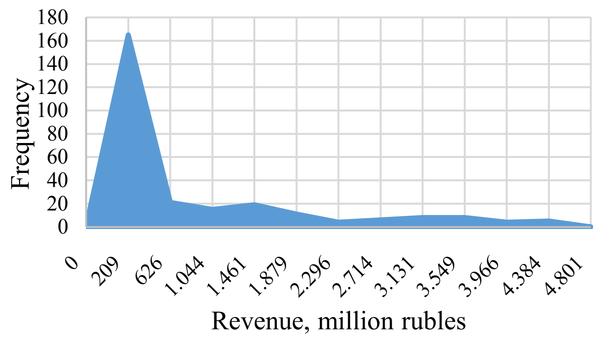

Fig. 2: Distribution of enterprises by revenue (sample size 185 enterprises)

Stage 2. *Testing the hypothesis of the lognormal distribution of marginal distributions*

We will test the hypothesis about the lognormal distribution of revenue and the main financial indicators using the Kolmogorov-Smirnov test. When testing, we will calculate the probability, *p-value*, with which the numerical series, composed of the logarithms of the initial data, satisfies the normal distribution law. The hypothesis about the agreement of the distribution of the variable with the normal law is accepted if the *p-value* is higher than the significance level of 0.05.

Table 1 below summarizes the *p-value* probabilities for similar testing of the remaining components. For the manoeuvrability coefficient, because it has negative values, we introduce an artificial variable:

$$\bar{X} = -\bar{x} + 1.01, \quad (1)$$

where $\bar{x}$ is the vector of empirical observations of the manoeuvrability coefficient.

Table 1. Results of the p-value for the logarithm of the indicators

№	Indicator	P-value
1	Sales revenue	0.711
2	Current ratio	0.204
3	Debt-to-equity ratio	0.197
4	Immobilization ratio	0.061
5	Return on equity	0.647
6	Receivables' turnover	0.633
7	Accounts payable turnover	0.455
8	Inventory turnover	0.062
9	Manoeuvrability coefficient	0.069

Thus, in the study of marginal distributions, it was shown that all of them with varying degrees of "quality", but at a significance level of 95 %, can be approximated by a model of lognormal density. Therefore, the hypothesis of a lognormal distribution of all indicators in Table 1 is not rejected and remains a working hypothesis.

Recall that the mathematical expectation (mean) of a sample with a lognormal distribution is determined by the formula:

$$e^{\mu+\sigma^2/2}, \quad (2)$$

the median of the lognormal density is calculated using the formula:

$$e^{\mu}, \quad (3)$$

and the mode is calculated using the formula:

$$e^{\mu-\sigma^2}. \quad (4)$$

where e is the base of the natural logarithm, a mathematical constant; μ is the mathematical expectation of the logarithms of the indicator; σ is the mean square (standard) deviation of the logarithms of the indicator from their mean. Details of the features of estimating the parameters of a sample with a lognormal distribution are considered in [14], [22].

In Table 2, we give a comparative analysis of the main parameters for assessing the indicators of the sample with a lognormal distribution of different volumes and give the values of the main financial indicators according to the TestFirm portal, [29].

Table 2 shows how many times the mean values and the median can differ from the mode if anomalous values are retained in the sample, leading to high standard deviations. The result of the assessment will be unstable until anomalous values are removed from the sample. It is recommended to remove brightly distinguished values from the sample by technicians known in statistics to remove outliers. Such values are always located at a considerable distance from the centre of scattering, excluding them. We significantly reduce the "spread" and make the estimates more stable.

Table 2. The results of calculating the main parameters for a sample of indicators

Indicator	Estimated values for a sample of 185			Estimated values for a sample of 138			The values of TestFirm	
	mean	median	mode	mean	median	mode	mean	median
Current ratio	2.72	1.55	0.50	2.59	1.67	0.69	2.25	1.08
Debt-to-equity ratio	25.93	1.80	0.01	2.07	0.82	0.13	0.04	1.92
Immobilization ratio	7.35	0.98	0.02	3.84	1.30	0.15	0.85	0.64
Return on equity	0.72	0.23	0.02	0.54	0.19	0.02	0.28	0.15
Receivables' turnover	318	91	8	114	68	25	371	74
Accounts payable turnover	465	140	13	153	113	62	n / a	n / a
Inventory turnover	146	28	1	44	21	5	111	14
Manoeuvrability coefficient	-4.05	-0.67	-0.02	-1.20	-0.39	-0.04	-7.19	0

After removing outliers by technical methods, 138 enterprises remained in the sample, the results of the main parameters for the sample are shown in Table 2.

Since model lognormal densities can be selected for all indicators, the next step in the study is to test the hypothesis of jointly normal distribution of the components, [30], [31].

Stage 3. *Corroborating the hypothesis of the jointly normal distribution of the logarithms of the components of the financial indicators' vector*

We combined the data on sales revenue and calculated indicators into one array by entering the following designations: X_0 – sales revenue, in billion rubles per year, X_1 – current ratio, X_2 – debt-to-equity ratio, X_3 – immobilization ratio, X_4 – return on equity, X_5 – receivables' turnover, X_6 – accounts payable turnover, X_7 – inventory turnover, and X_8 – an artificial variable $\bar{X} = -\bar{x} + 1.01$ of the manoeuvrability coefficient.

The hypothesis of the joint normality of the logarithms of the variables $X_0, \ldots, X_8$ is the hypothesis of the joint normality of a 9-dimensional random vector.

The following theorem is known: the joint distribution of the components of a random vector $(Y_1, ..., Y_n)$ is normal if and only if any linear combination of these components $c_0 Y_1$ + ... + $c_n Y_n$ has a normal distribution. Since applying library tests and removing outliers for a 9-dimensional vector presents some technical difficulties we use the following technique (implemented in the *R* statistical package):

– using the library function runif (9, 0, 1), the coefficients $c_0, ..., c_8$ are randomly generated. The generation results are normalized to $\sum_{i=0}^{8} c_i$;

– a linear combination of (c_0 $ln(Z_0)$ + ... +c_8 ln (Z_8)) / $\sum_{i=0}^{8} c_i$ checked for normality by Kolmogorov-Smirnov test. The result of the test *p-value* is recorded in the element of the array. (Here $Z_0, \dots, Z_8$ are centred and normalized logarithms of variables $X_0, \dots, X_8$ empirical observations; $Z_i = (X_i - \overline{X_i})/\sqrt{\overline{D_i}}$ where $\overline{X_i}$ is a selective average of the *i*-th variable, $\overline{D_i}$ is its sample variance, and *i* = 0, …, 8).

Figure 3 shows the result of 100,000 repetitions of the Kolmogorov-Smirnov test for linear combinations with randomly generated weights.

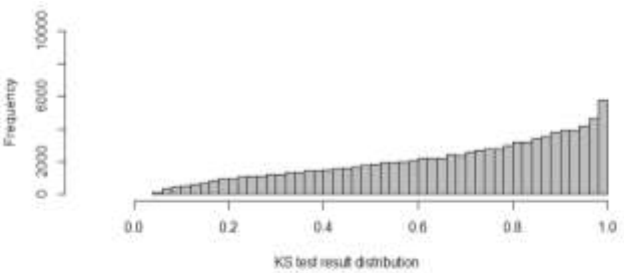

Fig. 3: 100,000 generations of Kolmogorov-Smirnov test repetitions for linear combinations

Numerical calculations showed that out of 100,000 generations in 99,685 cases the *p-value* is above the critical level of 0.05, and only in 315 cases it is below the critical level of 0.05 (which is 0.315 % of the total number of generations). The results of such testing allow us to preserve as a working hypothesis the joint lognormal distribution of economic indicators and revenues of oil and gas industry enterprises selected for research.

Stage 4. *Evaluation by conditional distributions of the most probable values of indicators*

With sufficient grounds to preserve as a working statistical hypothesis a joint lognormal distribution of the values of revenue and economic indicators (components), we construct conditional distributions of indicators, provided that the company's revenue is known and estimate the median and modal (most probable) values of indicators using conditional distributions.

Note that further research is carried out under the reasonable assumption that the random variables $X_0, \dots, X_8$ have a joint lognormal distribution, and the quantities $ln(X_0)$,...,$ln(X_8)$ are the joint normal distribution.

Let $Y_i = ln(X_i)$, *i* = 0, ..., 7, then $Y_8 = ln(X_8) = ln(-\overline{x_8} + 1.01)$, where $\overline{x_8}$ is a vector of empirical observations of the manoeuvrability coefficient.

It is necessary to construct a conditional distribution of quantities Y_1, ..., Y_8 provided that Y_0 takes a fixed value (i. e., using the logarithm of the volume of proceeds, to estimate the maximum point of the joint distribution density of the logarithms of the economic indicators of enterprises) during reverse potentiation, considering the peculiarities of the multivariate logarithmically normal distribution. This allows determining by the specified value of revenue the median or modal (most frequently repeated) values of economic indicators for the industry.

Consider the covariance matrix of a multidimensional normal random vector, $Y_0, Y_1, ..., Y_8$ in the following block form:

$$CV = \begin{pmatrix} \sigma_0^2 & cov(Y_0, \vec{Y}) \\ cov(Y_0, \vec{Y})^T & COV \end{pmatrix}, \quad (5)$$

where σ_0^2 is the variance of a random variable Y_0; COV is the covariance matrix of a random vector $\vec{Y} = (Y_1, \dots, Y_8)$; $cov(Y_0, \vec{Y})$ is a row vector of covariances Y_0 and vector components $\vec{Y}$; $cov(Y_0, \vec{Y})^T$ is a column vector of covariances Y_0 and vector components $\vec{Y}$.

The conditional mathematical expectation of the vector $\vec{Y}$ provided that $Y_0 = y$ is determined by the formula:

$$E(\vec{Y} | Y_0 = y) = \vec{\mu} + \frac{cov(Y_0, \vec{Y})^T}{\sigma_0^2}(y - \mu_0). \quad (6)$$

Conditional covariance matrix provided that $Y_0 = y$ is:

$$CV(\vec{Y} | Y_0 = y) = COV - \frac{cov(Y_0, \vec{Y})^T \times cov(Y_0, \vec{Y})}{\sigma_0^2}, \quad (7)$$

where $\vec{\mu}$ is the vector of means of the random vector $\vec{Y}$; μ_0 is the mean of the random variable Y_0.

The absolute maximum (mode) of the density of a random lognormal vector $\vec{x}$ is achieved at a point with coordinates $\exp(\vec{\mu} - \Sigma \times 1)$, where $\vec{\mu}$ is the vector of mathematical expectations of the logarithms of the components, Σ is the covariance matrix of the components' logarithms, and **1** is a vector consisting of ones. This result is easily obtained from the condition that all partial derivatives of the density function of the multivariate lognormal distribution are equal to zero:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt[2]{det\Sigma}}\frac{1}{\prod_{i=1}^{n} x_i} exp(-\frac{1}{2}(\Sigma^{-1}\overrightarrow{ln(x)}, \overrightarrow{ln(x)})), \quad (8)$$

where $\overrightarrow{ln(x)}$ is the centred random vector.

Thus, the formula for calculating *the conditional modal value* of a random vector $\vec{X}$ (vector of economic indicators of an enterprise) with a known annual revenue $X_0 = x$ takes the form:

$$MODE(\vec{\boldsymbol{X}}|X_0 = x) = \exp(\boldsymbol{E}(\vec{\boldsymbol{Y}}|Y_0 = y) - \boldsymbol{CV}(\vec{\boldsymbol{Y}}|Y_0 = y) \times \boldsymbol{1}). \quad (9)$$

We find *median values* using the formula:

$$MEDIAN(\vec{\boldsymbol{X}}|X_0 = x) = \exp(\boldsymbol{E}(\vec{\boldsymbol{Y}}|Y_0 = y)). \quad (10)$$

Considering that an artificial variable of a type $\bar{X} = -\bar{x} + 1{,}01$ introduced for the latter indicator, to produce the final result (as mode or median) reverse transformation of the form $ManCoef = 1.01 - x_8$ should hold for the manoeuvrability coefficient. The corresponding calculations were performed in the environment of the R statistical package. Tables 3 and 4 present the results of these calculations.

Table 3. Median values of indicators for different values of enterprises' revenue (sample size is 138 enterprises)

Sales revenue, billion rubles per year / Indicator	1	2	5	10	50	100	500	1000	2000	3000
Current ratio	1.5	1.55	1.61	1.66	1.78	1.84	1.97	2.03	2.09	2.13
Debt-to-equity ratio	1.05	0.99	0.91	0.86	0.74	0.69	0.6	0.56	0.52	0.51
Immobilization ratio	1.34	1.47	1.66	1.82	2.26	2.48	3.08	3.38	3.71	3.92
Return on equity	0.25	0.25	0.24	0.24	0.23	0.22	0.22	0.22	0.21	0.21
Receivables' turnover	62	66	71	75	86	91	103	109	116	120
Accounts payable turnover	135	129	122	117	105	101	91	87	84	81
Inventory turnover	26	24	22	21	18	17	14	14	13	12
Manoeuvrability coefficient	-0.32	-0.33	-0.34	-0.35	-0.37	-0.38	-0.40	-0.40	-0.41	-0.42

The median values of the current ratio and the immobilization ratio tend to increase depending on the values of the revenue of the oil and gas enterprises, [32]. So, the current ratio increases from 1.5 calculated for enterprises with revenue of 1 billion rubles, to 2.13 for enterprises with revenue of 3,000 billion rubles.

Table 4. Modal values of indicators for different values of the enterprises' revenue (sample size is 138 enterprises)

Sales revenue, billion rubles per year / Indicator	1	2	5	10	50	100	500	1000	2000	3000
Current ratio	3.99	4.11	4.28	4.41	4.73	4.88	5.23	5.39	5.56	5.66
Debt-to-equity ratio	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Immobilization ratio	0.75	0.83	0.93	1.03	1.27	1.40	1.73	1.90	2.09	2.21
Return on equity	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Receivables' turnover	11	11	12	13	15	15	18	19	20	20
Accounts payable turnover	21	20	19	18	17	16	14	14	13	13
Inventory turnover	2	2	2	2	2	1	1	1	1	1
Manoeuvrability coefficient	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95

For the modal values shown in Table 4, the dependence on the volume of revenue is less noticeable, i. e., indicators are more stable and can be used in assessing industry indicators for enterprises` financial analysis.

The difference between the median and modal results is explained by the right-sided asymmetry of the distribution.

Further, an analysis of differences between median and modal research results was carried out.

The importance of this difference is illustrated with an example. Since a 9-dimensional random vector, and a conditional 8-dimensional random vector were considered above, visualization of the results is impossible for obvious reasons. For clarity, consider a three-dimensional random vector (X_0, X_4, X_5) where X_0 is the sales revenue in billion rubles per year, X_4 is the return on equity, X_5 is the receivables turnover. Out of the original sample of 185 enterprises, after removing outliers for these components alone, 175 enterprises remain. For a random vector (X_0, X_4, X_5), a vector of average values of logarithms (0.72; -1.39; 4.33) and a covariance matrix for logarithmic components are obtained:

$$\begin{pmatrix} 9.18 & -0.20 & -0.23 \\ -0.20 & 1.38 & -0.049 \\ -0.23 & -0.049 & 1.95 \end{pmatrix}.$$

Consider the conditional modal and median values of indicators (components of a random vector) X_4 is the return on equity, X_5 is the receivables turnover at two significantly different revenue values, for $X_0 = 0.01$ billion rubles per year and for $X_0 = 3{,}000$ billion rubles per year (use formulas (9) and (10)).

In the first case, with $X_0 = 0.01$ billion rubles in a year, the conditional mode is equal to:

$$MODE\begin{pmatrix}X_4\\X_5\end{pmatrix} = \begin{pmatrix}0.075\\13\end{pmatrix},$$

the conditional median will be:

$$MEDIAN\begin{pmatrix}X_4\\X_5\end{pmatrix} = \begin{pmatrix}0.21\\63\end{pmatrix}.$$

In the second case, at $X_0 = 3{,}000$ billion rubles per year the conditional mode is equal to:

$$MODE\begin{pmatrix}X_4\\X_5\end{pmatrix} = \begin{pmatrix}0.057\\10\end{pmatrix},$$

the conditional median will be:

$$MEDIAN\begin{pmatrix}X_4\\X_5\end{pmatrix} = \begin{pmatrix}0.21\\63\end{pmatrix}.$$

Figure 4(a) and Figure 4(b) show in three-dimensional graphics the surfaces representing the density of joint normal distributions of the logarithms of the components X_4, X_5.

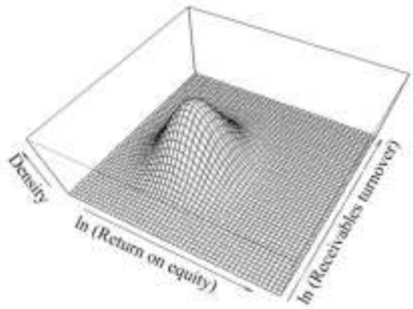

Fig. 4(a): Density surfaces of joint normal distributions of the logarithms of components X_4, X_5 for the case when $X_0 = 0.01$ billion rubles per year
Source: Developed by authors

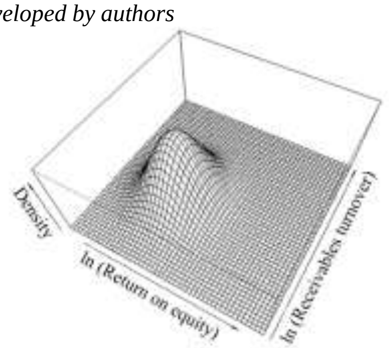

Fig. 4(b): Density surfaces of joint normal distributions of the logarithms of components X_4, X_5 for the case when $X_0 = 3{,}000$ billion rubles per year
Source: Developed by authors

Figure 5(a) and Figure 5(b) show surfaces representing the density of joint log-normal distributions of the components X_4, X_5.

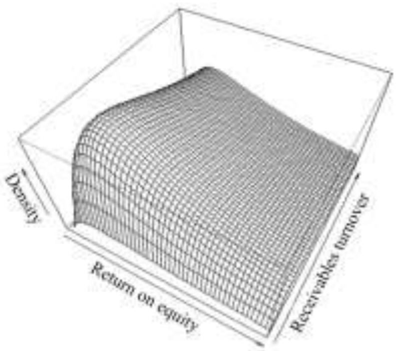

Fig. 5(a): Density surfaces of joint lognormal distributions of components X_4, X_5 for the case when $X_0 = 0.01$ billion rubles per year
Source: Developed by authors

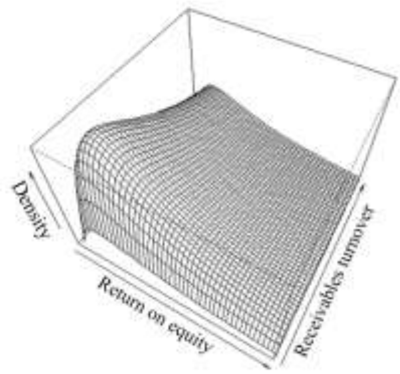

Fig. 5(b): Density surfaces of joint lognormal distributions of components X_4, X_5 for the case when $X_0 = 3{,}000$ billion rubles per year
Source: Developed by authors

In the transition from the coordinates in logarithms X_4, X_5 to the coordinates in X_4, X_5 there is a significant change in the surface density distribution. The point, which in the logarithms of the components X_4, X_5 corresponded to the median and coincided with the modal value (as in the case of the multivariate jointly normal distribution), with such a transformation will take a position on the corresponding level line, i. e., in the original coordinates X_4, X_5 there is a whole set of points that have the same probability as the median point obtained by formula (10). Could such a median point be used to assess industry performance? Many other equally probable and significantly different combinations of X_4 (Return on equity) and X_5 (Receivables' turnover) are on the corresponding level line. View of the level line indicates that such combinations of X_4 and X_5 (Figure 6) may have significant differences. That's why it makes no sense to talk about the typicality of any one of these combinations for the industry.

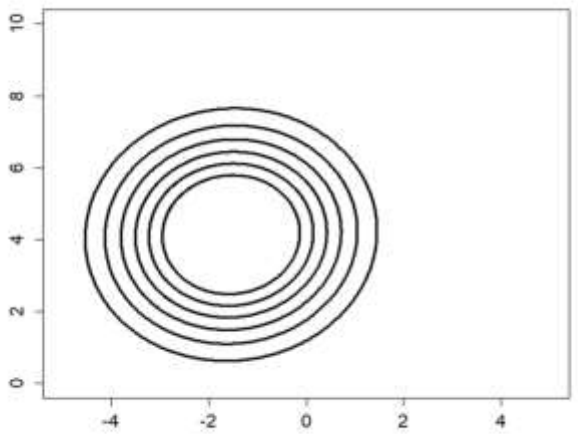

Fig. 6: Scatter plots (level lines of equal probability) for the normal distribution of the logarithms of the components X_4,X_5
Source: Developed by authors

All that has been said above also applies to estimates obtained not from the medians of coordinates, but the mathematical expectations of coordinates or any other point except for the mode of a two-dimensional distribution. The lognormal distribution component X_4, X_5 mode is unique. Its coordinates X_4, X_5 are the most likely value of the return on equity and the receivables' turnover indicators for a given value of the enterprise's revenue.

Moreover, this point is more typical than all the others because it has the highest probability that it is in line with the current understanding of the market value in the appraisal activity as a calculated amount of money corresponding to the most probable price. It seems that not only the market value but also many related concepts in the assessment would be logical to consider in terms of the highest probability. Cases of dimension more than three have the same logic, although they cannot be represented graphically.

It should be noted that the 8 indicators studied for the chosen gas and oil industry, when evaluated using the median formula (10) and indicated in Table 3, vary significantly depending on the sales revenue. On the contrary, the modal values shown in Table 4 change insignificantly when the volume of the sales revenue changes, and some do not change at all. This is because in the studied industry the coordinates of the modal point are not very sensitive to the value of the enterprise's revenue. Any other estimates, including median ones, are highly sensitive to the revenue value since they come to level surfaces similar to the lines shown in Figure 7(b).

A three-dimensional representation of the observed values of the random vector (X_0,X_4,X_5) for 175 gas and oil enterprises can serve as an additional illustration of the importance of studying modal values. Figure 7(a) and Figure 7(b) show the dispersion of empirical observations of indicators for 175 enterprises in the original coordinate values and in their logarithms, respectively. The cross marks the positions of the points of maximum density. The dispersion of 10,000 points generated in the R statistical package with the same parameters as the empirical observations is shown in Figure 8(a) and Figure 8(b).

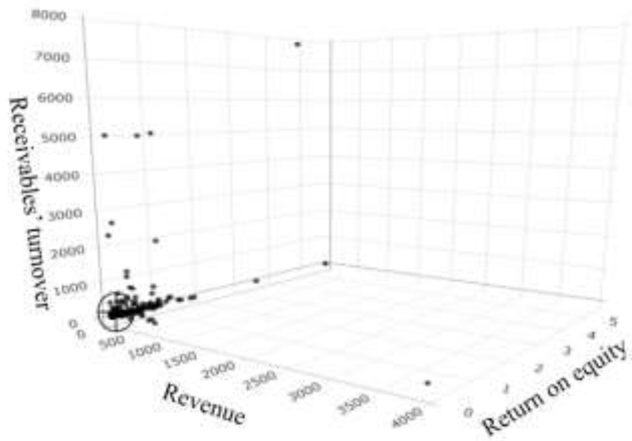

Fig. 7(a): Dispersion of empirical observations of indicators for 175 enterprises in the original coordinate values
Source: Developed by authors

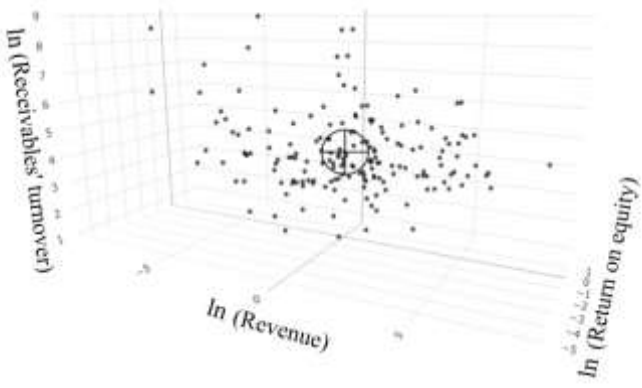

Fig. 7(b): Dispersion of empirical observations of indicators for 175 enterprises in their logarithms
Source: Developed by authors

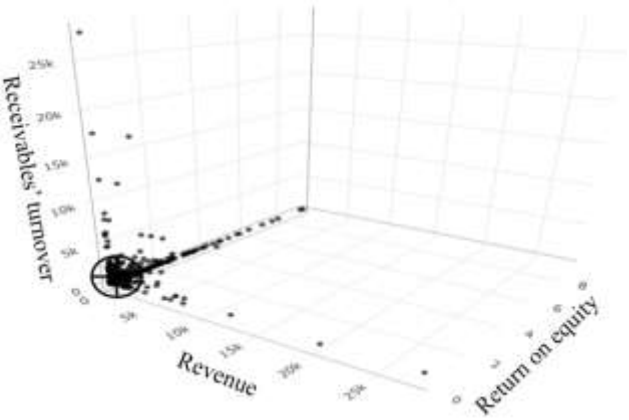

Fig. 8(a): Dispersion of 10,000 generated points with the same lognormal parameters as the model for the original sample in coordinates (X_0, X_4, X_5)
Source: Developed by authors

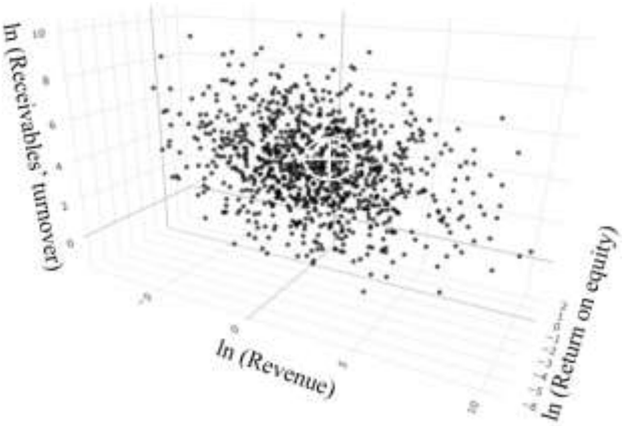

Fig. 8(b): Dispersion of 10,000 generated points with the same lognormal parameters as the model for the original sample in their logarithms
Source: Developed by authors

Thus, the choice of an ensemble of industry indicators that corresponds to the point of maximum density seems to be reasonable and consistent with the general content of appraisal activities, [33].

4 Discussion

Note that further research is carried out under the reasonable assumption that the random variables $X_0, \dots, X_8$ have a joint lognormal distribution, and the quantities $ln(X_0), \dots, ln(X_8)$ are the joint normal distribution.

In the paper, it is demonstrated that the orientation towards the generally established recommended values of the coverage indicators, financial leverage, immobilization, profitability, and turnover seems to be incorrect due to the specific features of the enterprises' activities. It is recommended to calculate the most probable industry average financial ratios by groups depending on the size of the enterprise (revenue or capitalization).

The paper focuses on the application of multidimensional statistical analysis, the use of a model of logarithmically normal distributions recommended for distributions with pronounced right-sided asymmetry. Financial information of oil and gas industry enterprises for 2016 was selected as statistical material.

To improve the proposed stages of the analysis of the definition of industry financial indicators, constant monitoring of the analyzed indicators is necessary. The authors plan to conduct similar studies on an annual basis. Possible directions of future research can be based on the derivation of the most probable industry financial coefficients by groups, depending on the scale of the enterprise (revenue or capitalization) for various sectors of the national economy.

5 Conclusion

To conclude, we note the following provisions of the research results.

1. In this research, the authors have developed the stages of determining industry average indicators for comparative financial analysis on the example of the oil and gas industry, but the proposed algorithm applies to all sectors of the economy.

2. The paper shows that the orientation towards the generally established recommended values of the current ratio, financial leverage ratio, immobilization ratio, return on equity, and turnover ratios seems to be incorrect due to the variety of specific features of the enterprise's activities in various industries.

3. We recommend calculating the most probable industry coefficients by groups depending on the size of the enterprise (revenue or capitalization), but the average and median indicators of industry coefficients derived by the 'Avdeev and Co' auditing firm on the TestFirm portal should be used with caution.

4. In the digitalization process, it is necessary to create a unified system of information and analytical support that would allow standardizing, optimizing, and unifying financial analysis procedures. To create such a system, it is required to accept uniform standards for electronic financial information and a calculation algorithm to ensure its automatic processing, as well as to ensure reliability (relevance) for comparative financial analysis.

The results of the research together form a complete solution to an actual scientific problem.

They also form a vector for further scientific research with the saturation of models and tools with additional and clarifying factors, with the consideration of additional digital technologies for the analysis of enterprises' financial and economic activities, taking into account industry and regional characteristics.

References:

[1] A. Althunibat, B. Alokush, R. Dawood, S. M. Z. A. Tarabieh, and I. Gil-Pechuan, "Modeling the factors that influence digital economy services acceptance," in *2021 International Conference on Information Technology, ICIT 2021 (14-15 July 2021) - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Ed., Amman, Jordan: IEEE, 2021, pp. 942–945. doi: 10.1109/ICIT52682.2021.9491686.

[2] S. Bag, A. K. Sahu, P. Kilbourn, N. Pisa, P. Dhamija, and A. K. Sahu, "Modeling barriers of digital manufacturing in a circular economy for enhancing sustainability," *Int. J. Product. Perform. Manag.*, vol. 71, no. 3, pp. 833–869, Feb. 2022, doi: 10.1108/IJPPM-12-2020-0637.

[3] G. Silkina, "From analogue to digital tools of business control: Succession and transformation," in *IOP Conference Series: Materials Science and Engineering*, St. Petersburg: IOP Publishing Ltd, 2019, p. 012018. doi: 10.1088/1757-899X/497/1/012018.

[4] M. A. Berawi, N. Suwartha, M. Asvial, R. Harwahyu, M. Suryanegara, E. A. Setiawan, I. Surjandari, T. Y. M. Zagloel, and I. J. Maknun, "Digital Innovation: Creating Competitive Advantages," *Int. J. Technol.*, vol. 11, no. 6, pp. 1076–1080, 2020, doi: 10.14716/ijtech.v11i6.4581.

[5] A. Dovgal and N. Parkhomenko, "Formation of the financial and economic security system at the enterprise," *Galic'kij Ekon. Visn.*, vol. 72, no. 5, pp. 78–84, 2021, doi: 10.33108/galicianvisnyk_tntu2021.05.078.

[6] S. Hilkevics and V. Semakina, "The classification and comparison of business ratios analysis methods," *Insights into Reg. Dev.*, vol. 1, no. 1, pp. 48–57, 2019, doi: 10.9770/ird.2019.1.1(4).

[7] Rahmatullah, Inanna, Sahade, Nurdiana, F. Azis, and Bahri, "Utilization of digital technology for management effectiveness micro small and medium enterprises," *Int. J. Sci. Technol. Res.*, vol. 9, no. 4, pp. 1357–1362, 2020.

[8] A. Borodin, I. Mityushina, M. Harputlu, N. Kiseleva, and A. Kulikov, "Factor Analysis of the Efficiency of Russian Oil and Gas Companies," *Int. J. Energy Econ. Policy*, vol. 13, no. 1, pp. 172–188, 2023, doi: 10.32479/ijeep.13763.

[9] B. K. Guru and I. S. Yadav, "Financial development and economic growth: panel evidence from BRICS," *J. Econ. Financ. Adm. Sci.*, vol. 24, no. 47, pp. 113–126, 2019, doi: 10.1108/JEFAS-12-2017-0125.

[10] L. Rodrigues and L. Rodrigues, "Economic-financial performance of the Brazilian sugarcane energy industry: An empirical evaluation using financial ratio, cluster and discriminant analysis," *Biomass and Bioenergy*, vol. 108, pp. 289–296, 2018, doi: 10.1016/j.biombioe.2017.11.013.

[11] A. I. Shinkevich, D. R. Baygildin, and E. L. Vodolazhskaya, "Management of a sustainable development of the oil and gas sector in the context of digitalization," *J. Environ. Treat. Tech.*, vol. 8, no. 2, pp. 639–645, 2020.

[12] S. P. Kyurdzhiev, A. A. Mambetova, and E. P. Peshkova, "An integral evaluation of the financial state of the regional enterprises," *Econ. Reg.*, no. 2, pp. 586–601, 2016, doi: 10.17059/2016-2-22.

[13] A. V. Kozlov, E. V. Rytova, S. S. Gutman, and I. M. Zaychenko, "The valuing of the indicator of a regional industrial development: The fuzzy logic approach," in *Proceedings of the 19th International Conference on Soft Computing and Measurements, SCM 2016*, Institute of Electrical and Electronics Engineers Inc., Ed., St. Petersburg, Russia: IEEE, 2016, pp. 545–547. doi: 10.1109/SCM.2016.7519842.

[14] E. Kiseleva, A. Efimov, and S. Sekar, "Predictive analytics as an instrument to prevent bankruptcy," in *Proceedings of the 33rd International Business Information Management Association Conference, IBIMA 2019: Education Excellence and Innovation Management through Vision 2020*, Granada, Spain: IBIMA, 2019, pp. 8783–8788. doi: 10.2139/ssrn.3677898.

[15] Y. A. Malyukov, A. O. Nedosekin, Z. I. Abdoulaeva, and A. V. Silakov, "The Fuzzy Model for Sectoral Resilience Analysis," *WSEAS Trans. Bus. Econ.*, vol. 20, pp. 2038–2047, 2023, doi:

10.37394/23207.2023.20.177.

[16] P. Adriaens and N. Ajami, "Infrastructure and the Digital Economy: Reinventing Our Role in the Design, Financing, and Governance of Essential Services for Society," *J. Environ. Eng.*, vol. 147, no. 5, 2021, doi: 10.1061/(asce)ee.1943-7870.0001866.

[17] M. Kotsupatriy, I. Ksonzhyk, S. Skrypnyk, I. Shepel, and S. Koval, "Use of international accounting and financial reporting standards in enterprise management," *Int. J. Manag.*, vol. 11, no. 5, pp. 788–796, 2020, doi: 10.34218/IJM.11.5.2020.071.

[18] D. Rodionov, A. Zaytsev, E. Konnikov, N. Dmitriev, and Y. Dubolazova, "Modeling changes in the enterprise information capital in the digital economy," *J. Open Innov. Technol. Mark. Complex.*, vol. 7, no. 3, p. 166, 2021, doi: 10.3390/joitmc7030166.

[19] U. Uwah, I. Aji, and E. Iniabasi, "Accounting Ratios and False Financial Statements Detection: Evidence From Nigerian Quoted Companies," *AKSU J. Adm. Corp. Gov.*, vol. 3, no. 1, pp. 61–74, 2023, doi: 10.61090/aksujacog.2023.105.

[20] Audit IT, "Ready Ratios. Online Financial Analysis Financial Reporting and Analysis Software", [Online]. https://www.readyratios.com/sec/ industry/ (Accessed Date: May 20, 2024).

[21] L. A. Chaldaeva, T. I. Chinaeva, and A. S. Bogopolskiy, "Analysis of financial and economic indicators, characterizing the activities of organizations in the oil and gas industry," *Stat. Econ.*, vol. 17, no. 1, pp. 69–78, 2020, doi: 10.21686/2500-3925-2020-1-69-78.

[22] F. Brunetti, D. T. Matt, A. Bonfanti, A. De Longhi, G. Pedrini, and G. Orzes, "Digital transformation challenges: strategies emerging from a multi-stakeholder approach," *TQM J.*, vol. 32, no. 4, pp. 697–724, Jul. 2020, doi: 10.1108/TQM-12-2019-0309.

[23] M. Abdel-Basset, W. Ding, R. Mohamed, and N. Metawa, "An integrated plithogenic MCDM approach for financial performance evaluation of manufacturing industries," *Risk Manag.*, vol. 22, no. 3, pp. 192–218, 2020, doi: 10.1057/s41283-020-00061-4.

[24] Scrin Ltd., "SKRIN POD / FT", [Online]. https://podft.skrin.ru/ (Accessed Date: May 20, 2024).

[25] Federal State Statistics Service, *Methodological recommendations for conducting an analysis of the financial and economic activities of organizations*. Moscow: Russian State Statistics Committee, 2002, [Online]. https://rosstat.gov.ru/bgd/free/b99_10/isswww.exe/stg/d040/7-2.htm (Accessed Date: May 20, 2024).

[26] V. V. Manuylenko, A. A. Mishchenko, O. B. Bigday, T. A. Sadovskaya, and T. S. Lisitskaya, "Evaluation method for efficiency of financial and innovative activities in commercial organizations based on stochastic modelling," *J. Appl. Econ. Sci.*, vol. 12, no. 4, pp. 1226–1246, 2017.

[27] J. M. Müller, O. Buliga, and K. I. Voigt, "The role of absorptive capacity and innovation strategy in the design of industry 4.0 business Models - A comparison between SMEs and large enterprises," *Eur. Manag. J.*, vol. 39, no. 3, pp. 333–343, 2021, doi: 10.1016/j.emj.2020.01.002.

[28] V. Schepinin and A. Bataev, "Digitalization of financial sphere: Challenger banks efficiency estimation," in *IOP Conference Series: Materials Science and Engineering. In Proceedings of International Scientific Conference "Digital Transformation on Manufacturing, Infrastructure and Service" 21–22 November 2018*, Saint-Petersburg, Russian Federation: IOP Publishing Ltd, 2019, p. 012051. doi: 10.1088/1757-899X/497/1/012051.

[29] The TestFirm, "Comparison of the company's financial condition with industry indicators and competitors. Gas and Oil Industry," Audit-it.ru. Accessed: Aug. 20, 2024, [Online]. https://www.testfirm.ru/otrasli/06/ (Accessed Date: May 20, 2024).

[30] O. V. Rusakov, O. I. Jaksumbaeva, A. A. Ivakina, and M. B. Laskin, "Pricing in the Real Estate Market as a Stochastic Limit. Log Normal Approximation," in *Proceedings - 2015 2nd International Conference on Mathematics and Computers in Sciences and in Industry, MCSI 2015*, Sliema, Malta: IEEE, 2016, pp. 235–239. doi: 10.1109/MCSI.2015.48.

[31] M. B. Laskin, "Multidimensional log-normal distribution in real estate appraisals," *Bus. Informatics*, vol. 14, no. 2, pp. 48–63, 2020, doi: 10.17323/2587-814X.2020.2.48.63.

[32] S. Guo, W. Ding, and T. Lanshina, "Global Governance and the Role of the G20 in the

Emerging Digital Economy," *Int. Organ. Res. J.*, vol. 12, no. 4, pp. 169–184, 2017, doi: 10.17323/1996-7845-2017-04-169.

[33] A. Damodaran, *Investment Valuation: Tools and Techniques for Determining the Value of any Asset, University Edition*. 2012.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Svetlana Pupentsova has implemented the research algorithm, collected data for analysis, and performed the initial processing of statistical data.
- Mikhail Laskin was responsible for developing the research methodology, implementing the computer code and auxiliary algorithms, and processing statistical data.
- Maria Livintsova conducted a literature review, collected data for analysis, prepared and edited the manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The research is partially funded by the Ministry of Science and Higher Education of the Russian Federation under the strategic academic leadership program 'Priority 2030' (Agreement 075-15-2021-1333 dated 30 September 2021).

Conflict of Interest

The authors have no conflicts of interest to declare.